Universiteit Leiden

ICT in Business

Data-driven Innovation in NGOs

Name: Aikaterini Deligianni

Student-no: s1462350

Date: 18/05/2016

1st supervisor: Dr. C. J. Stettina

2nd supervisor: Dr. A. Knobbe

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)

Leiden University

Niels Bohrweg 1

2333 CA Leiden

The Netherlands

Master Thesis

Data-driven Innovation in NGOs

Aikaterini Deligianni

In partial fulfillment of the requirements for the degree of

Master of Science (M.Sc.) of ICT in Business

Graduation: May 2016

Supervisor: Dr. C. J. Stettina

Second reader: Dr. A. Knobbe

Leiden Institute of Advanced Computer Science (LIACS)

Leiden University

Niels Bohrweg 1

2333 CA Leiden

The Netherlands

Acknowledgements

I would first like to thank my thesis supervisor Dr. Christoph J. Stettina, for his valuable guidance, ideas and recommendations during my master thesis. I would also like to acknowledge Dr. Arno Knobbe as the second reader of this thesis and to thank him for his comments.

I would like to say thank you to the Centre for Innovation at The Hague and especially to Dr. Ulrich Mans for introducing me to various NGOs in the Netherlands and for his valuable feedback on the interview questions.

I would like to express my gratitude to all the participants of the interviews for freeing up their time and sharing their experiences. Without their input, the research project could not have been successfully conducted.

Finally, I must express my very profound gratitude to my parents, my cousins Christos and Dimitris and to my friends, Kostas, Stephanos, Bill and Penny, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Aikaterini Deligianni

Executive Summary

Innovation already appears to be a vague, blurry and universally applicable concept. Whether it is an NGO, a governmental organization, or the UN, they feel compelled to use the term 'innovative approach' (Krause, 2013). This thesis identifies, evaluates and interprets the available literature relevant to the topic area of "data-driven innovation" and tries to connect it with the NGOs. A Systematic Literature Review has been conducted for that reason. Also, eight semi-structured interviews with various NGOs in the Netherlands have been conducted in order to identify the needs and challenges in practice. The definition of data driven innovation as concluded from the SLR and the interviews is also presented.

In short, the SLR analysis shows that either government or business sector implement the majority of the initiatives, with NGOs staying far behind. In relation to the industry areas concerned, two third of the data-driven articles in the inventory focus primarily on information, information technology, education, health, energy, public administration and transportation. As regards the paper classification categorization, more than fifty per cent of the articles in the inventory are philosophical papers, validation research and proposal of a solution. The empirical data from the interviews indicated the main challenges and needs from the NGOs' perspective. Data quality, technical and privacy issues were the main fields that cause NGOs to stay behind in data-innovation.

The conclusion of this study is to identify how NGOs define data-driven innovation and whether they use data to innovate. Our findings verify that most NGOs from the humanitarian sector are still at an entry level as regards data innovation. The majority uses data for monitoring and evaluation purposes but they do not innovate with data. Main issues that they are dealing with are data standardization, data capacity, innovation capacity and IT infrastructure. The positive outcome although, is that many have engaged in projects that will enhance information exchange in a standardized way.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

CDRs	Call Detail Records
DAG	Directed Acyclic Graphs
DANs	Data Archiving and Networked Services
EHRs	Electronic Health Records
EU	European Union
GB	Gigabyte
HICS	High Income Countries
IoT	Internet of Things
IT	Information Technology
KDD	Knowledge Discovery in Database
LMICs	Low-and-middle Income Countries
MDG	Millennium Development Goals
MNE	Monitoring Evaluation
NGOs	Non-Governmental Organizations

NN	Neural Networks
OECD	Organization for Economic Co-operation and Development
OGD	Open Government Data
OKF	Open Knowledge Foundation
OSCE	Organization for Security and Co-operation in Europe
OSNs	Online Social Networks
SLR	Systematic Literature Review
SVM	Support Vector Machine
UN	United Nations
UNICEF	United Nations International Children's Emergency Fund

1 INTRODUCTION

1.1 BACKGROUND

Innovation is usually expected to benefit the adopters of the innovation, the innovating firm and the society at large (Porter, 1990). Over the last few years, numerous examples of innovation in data analysis have emerged, creating new business models for Data-driven innovation. For example, businesses are developing ways for real-time weather information to be communicated to devices in the field that can advise farmers on pest activity, water supply, and inclement weather (Gray, 2013). While more data are generated today than ever before; this is a positive trend that will inevitably continue: 90 percent of the world's information generated through the history of mankind has been generated over the last two years while data generated per year is growing at a rate of 40 percent (IBM, 2013; Manyika et al., 2011).

In this research we will focus on the social value of data, but from the point of view of use and purpose rather than volume. Therefore we will talk about data driven-innovation instead of "Big Data". Innovative uses of data have been a key factor for developing new products and making more efficient decisions for quite a long time, and these activities have become more common and more efficient with the availability of modern computing. Crunching data, statistics, and trends in new ways has always helped change the way that entire sectors operate. Data alone do not possess inherent value; instead it is the processing of data in innovative ways that brings new economic and social benefits, and this value creates a virtuous circle to feed into more use of data-based decision-making and analysis. According to Hilbert (2013) "the crux of the 'Big Data' paradigm is actually not the increasingly large amount of data itself, but its analysis for intelligent decision-making". In other words, it is the use of data that really matters (Hemerly, 2013).

These data, emitted as a byproduct of technologies such as Call Detail Records (CDRs) from mobile phone network operators, as well as data generated through the use of social media (known as social data), are playing an important role in academic, development, and humanitarian research and practice due to its ability to provide dynamic data sources and methods of analysis (Taylor et al., 2014). Mobile phone calling records in particular can provide real-time information on mobility of people following natural disasters, as shown by the work of Flowminder after the 2010 Haiti earthquake (Bengtsson et al., 2011). There has been an exponential increase in the use of digital communication technologies in low-and-middle-income countries (LMICs[1]) from mobile phone usage to mobile and fixed broadband internet (ITU, 2013).

International development institutions and Non-governmental Organizations (NGOs) such as United Nations (UN) have become interested in using these types of data for helping solve social problems such as homelessness, human trafficking, hunger, poverty (United Nations, 2012). As Teegen (2003) mentioned: “One main characteristic of NGOs is that they have the resources to do what their individual members could not do alone”. Furthermore, multinational institutions such as Global Pulse initiate to organize the sharing of digital data from LMIC worldwide, and to operationalize the idea that “shared data constitutes a public good” (Kirkpatrick, 2011).

Many sectors benefit from data-driven innovation: healthcare (e.g., diagnosis and treatment), financial services (e.g., analyzing market trends and economic conditions), to name a few (Andrade et al., 2014). One example of innovative data use that has an economic value proposition is Google’s Flu Trends, which provides near real-time estimates of flu activity for a number of countries around the world (The Economist, 2013b).

1.2 RESEARCH OBJECTIVE

The aim of this research is to contribute to the identification of the innovation methods being used by different domains and compare the results with the NGOs’ process. Opportunities may arise for the development of new methods for NGOs. Through our research we will try to identify the state-of-the-art on scientific literature on Data-driven innovation methods and how NGOs concede innovation, whether they pioneer in unsearched areas and what kind of challenges may appear. To support this understanding this research aims to conduct a Systematic Literature Review (SLR) and semi-structured interviews with NGOs in the Netherlands. The overall research objective is to identify the challenges and needs from the NGOs’ perspective and of course to determine whether

[1] We use the World Bank’s definitions grouping countries, see: http://data.worldbank.org/about/country-classifications, where LMICS have income of $1,036-$12,616 and high income countries (HICS) above that threshold. Our particular focus is the low- and lower-middle-income countries, with an upper threshold of $4,085 per capita, which includes India and most of Africa.

NGOs can benefit from new possible data-driven innovation methods applied by other business sectors.

1.3 RESEARCH RELEVANCE

Data-driven intelligence has been used successfully in technical and business endeavors, but a very different situation prevails in the social arena. A chasm exists between the data-driven information and its use for solving social problems.

Social problems are called "wicked" problems. They are dynamic and complex due to the numerous stakeholders involved and the numerous feedback loops among inter-related components. A large variety of government agencies and nonprofit organizations are involved in tackling these problems, although there is limited cooperation and data sharing among them (Desouza & Smith, 2014).

1.3.1 THEORITICAL RELEVANCE

First of all, a Systematic Literature Review (SLR) will assist us to present a fair evaluation of data-driven innovation in various domains. By using this auditable methodology we will be able to identify the gap in literature, see the domains that data-driven innovation is already applied and be in position to evaluate the potential benefits for NGOs. One major advantage according to Kitchenham et al., (2009) is that SLR will provide information about the effects of some phenomenon (in our case is data-driven innovation) across a wide range of settings and empirical methods. If studies give consistent results, systematic reviews provide evidence that the phenomenon is robust and transferable. If the studies give inconsistent results, sources of variation can be studied. Moreover, the use for a SLR arises from the need to summarize all existing information about data-driven innovation in an unbiased manner. This will give us the opportunity to draw more general conclusion than is possible from individual studies.

1.3.2 PRACTICAL RELEVANCE

This research aims at identifying the challenges in practice. For that reason we will conduct interviews with various NGOs in the Netherlands to participate. That would give us a firm grasp of the issues regarding privacy and data protection. For example, Big Data deriving from sources such as mobile communications and internet use are extremely sensitive with regard to the protection of data subjects' information, particularly when those subjects live in places of limited statehood where enforceable regulation is often absent (Greenleaf, 2012). “The real challenge is not simply gaining more data, but gaining data which is relevant to country priorities and can gain traction in terms of supporting policy and lead more informedness and greater transparency” (Taylor & Schroeder, 2014). Furthermore, we will perform grounded theory to read from a corpus of data and label variables and their relationships. According to Strauss & Corbin (1990) a theory is a set of relationships that proposes a reasonable explanation of the phenomenon under study. By using Open Coding as part of our analysis we will be able to create more abstract categories and therefore be able to draw more general conclusions which will be applicable to NGOs.

1.4 RESEARCH QUESTIONS

The main research questions of this thesis are:

1. What is the state-of-the-art on scientific literature on data – driven innovation methods?
2. How can data – driven innovation methods help NGOs?

1.5 RESEARCH SCOPE

This Research Thesis will focus on data-driven innovation in NGOs. The twofold purpose of this study is from one hand identifying the gap in literature and on the other hand identifying the challenges in practice.

The potential solution we propose has a two-fold purpose. First of all, through our research we will try to identify, evaluate and interpret the available literature relevant to the topic area of “data-driven innovation” and try to connect it with the NGOs. A Systematic Literature Review will be conducted for that reason. Secondly, semi-structured interviews with various NGOs in the Netherlands will be conducted in order to identify the challenges in practice.

By conducting SLR, which is a rigorous method to analyze existing evidence, we will be able to identify the different domains and the various innovation methods that were used in. Moreover, we will introduce the potential of data-driven information and its actual use in helping solve social problems to the NGOs community. Last, the interviews will help us to identify the challenges and needs from the NGOs’ perspective and be able to discuss and propose our finding from the SLR. We would have the opportunity to verify the data-driven innovation methods that are used within the NGOs’ community.

1.6 THESIS OVERVIEW

Below in Table 1 a brief overview of this master thesis' outline is presented. The first chapter introduces the research and provides the objective, research questions and scope. After this a literature study will describe the core constructs of the study providing the scientific state-of-the-art context. After this the methodology section will state how the research process was executed scientifically leading into the result chapter discussing the empirical findings. In the next chapter these findings will be discussed and compared with the findings in literature. Finally, in the last chapter the conclusions and recommendations are presented.

Table 1 Thesis Chapter Outline

Chapter	Outline
1	The first chapter will provide an **overview** presenting the essential basics of the study by placing it into context of current top level scientific research. After this the overall objective, research relevance, research questions and scoping will be defined.
2	An elaborate scientific **literature review** of the core constructs of the study will be presented in this chapter describing the essence of Big Data, innovation, data-driven innovation and NGOs.
3	After the literature review the **methodology and design** will be presented. This chapter will elaborate on the scientific methods used to gather and analyze the data.
4	This chapter will present the **results** regarding the collected articles from SLR and coding of the interviews.
5	The fifth chapter, **Analysis and Discussion**, will answer the research questions by combining and analyzing the findings from literature with the empirical data gathered.
6	Finally, the last chapter will be designed to give **conclusions and recommendations** on how NGOs can benefit from data-driven innovation and presents possible areas for further scientific study.

2 THEORETICAL FRAMEWORK

Although the specific research regarding data-driven innovation in NGOs is lacking, the topics of Big Data, Innovation and NGOs have been studied extensively. In section 2.1 the definitions and main concepts of Big Data will be discussed, secondly in section 2.2 the concepts of Innovation and data-driven Innovation will be defined and discussed in detail, furthermore in section 2.3 the concept and principles of NGOs will be defined.

2.1 BIG DATA

Data are the lifeblood of decision-making. Without data, we cannot know how many people are born and at what age they die; how many men, women and children still live in poverty; how many children need educating; how many doctors to train or schools to build; how public money is being spent and to what effect. Big Data can offer insights on particular issues by making it possible to reveal more details of a given problem. Interactive tools for using this data can allow greater engagement with and understanding of a problem. In particular, the rapidly increasing use of digital communication via mobiles phones and Internet connectivity has created opportunities for activism using volunteered data in LMICs (Taylor et al., 2014). To put that in perspective, as Google CEO Eric Schmidt said at a 2010 Techonomy conference, the amount of data generated in two days is as much as all data generated in human history before 2003 (Kirkpatrick, 2010). However, too many countries still have poor data, data arrives too late and too many issues are still barely covered by existing data. Entire groups of people and key issues remain invisible.

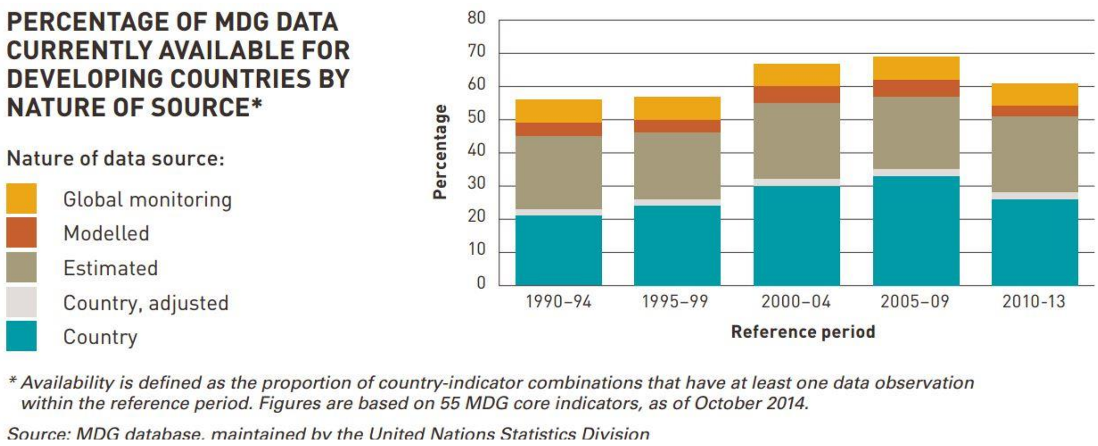

Figure 1Percentage of MDG data currently available for developing countries by nature of source*

The *Figure 1* above presents a summary snapshot of current data availability in the MDG databases (as of October 2014), covering 55 core indicators for 157 developing countries or areas. Overall, the picture is improving though still poor, so there is no five-year period when the availability of data is more than 70% of what is required.

2.1.1 THE DATA REVOLUTION

Data, or individual pieces of information, have been gathered and used throughout history. What's changed recently is that advances in digital technology have significantly increased our ability to collect, store and analyze data (Desouza & Smith, 2014). This enormous volume of information has been called Big Data, a term that is widely used (Hienz, 2014). For example, in 2012 Wal-Mart was generating more than 2.5 petabytes of data relating to more than 1 million customer transactions every hour (Open Data Center Alliance, 2012) and Facebook reported that it was processing 2.5 billion pieces of content (links, comments, etc.), 2.7 billion 'Like' actions and 300 million photo uploads per day (Desouza & Smith, 2014).

Because this is only the beginning of the Big Data era, the terms, definitions and ideas are still evolving. "Big Data is a step change in the scale and scope of the sources of materials (and tools for manipulating these sources) available in relation to a given object of interest" (Schroeder, 2014).Big Data "[…] refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered Big Data – i.e. we don't define Big Data in terms of being larger than a certain number of terabytes […]" (Manyika et al., 2011).The size of the datasets is only one condition for Big Data.

Laney (2001) at META Group introduced the so-called '3VS' framework in 2001 adding variety and velocity to volume, which has become the standard conceptual approach towards Big Data (Zikopoulos et al., 2012). According to Desouza and Smith (2014) there are multiple dimensions to Big Data, which are described in the set of seven "V"s that follows.

- *Volume:* considers the amount of data generated and collected.
- *Velocity:* refers to the speed at which data are analyzed.
- *Variety:* indicates the diversity of the types of data that are collected.
- *Viscosity:* measures the resistance to flow of data.
- *Variability:* measures the unpredictable rate of flow and types.
- *Veracity:* measures the biases, noise, abnormality and reliability in datasets.

➢ *Volatility:* indicates how long data are valid and should be stored.

Desouza and Smith (2014) also highlight that although all seven *"V"s* are increasing, they are not equal. As the volume of data increases along with the tendency to store multiple instances of the same data across varied devices, the science of information search and retrieval will have to advance.

Overall, the use of big data varies across sectors, where some sectors are poised for greater gains. *Figure 2* depicts the results of an analysis that McKinsey conducted in 2011 and illustrates differences among sectors in the use of big data (Manyika et al., 2011). The study divided the sectors into primarily 5 clusters. These include Cluster A: computer and electronic products; Cluster B: finance, insurance and government; Cluster C: construction, educational services, and arts and entertainment; Cluster D: manufacturing and wholesale trade; and Cluster E: retail, health care providers, accommodation, and food.

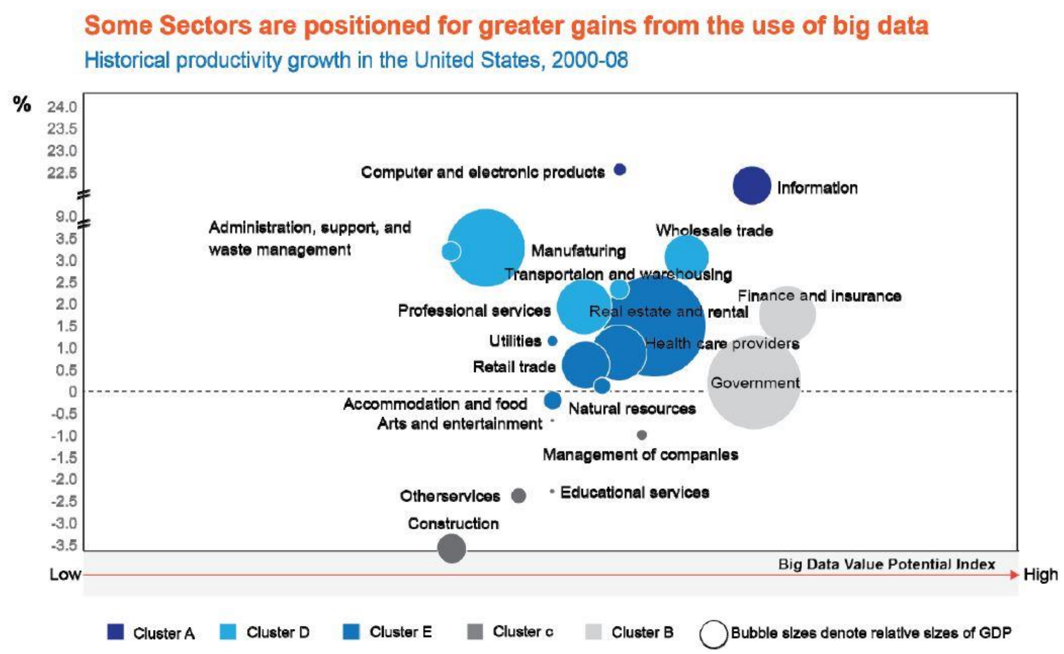

Figure 2Big Data Value across Sectors. Source: (Manyika et al., 2011)

According to Gartner, "data-driven innovation," will help to create 4.4 million information technology (IT) jobs globally by 2015, including 1.9 million in the United States (US) (Gartner, 2012). McKinsey's report indicates that big data has the potential to create massive saving and revenues in all sectors, i.e., create $300 billion in potential annual value to U.S. health care (more than double the total annual health care spending in Spain); €250 billion potential annual value to Europe's public sector administration (more than the gross domestic product [GDP] of Greece); and $600 billion in potential annual consumer surplus from using personal location data globally (Manyika et al., 2011). All in all, big data is considered to have a huge impact on all sectors, providing endless arrays of new opportunities for transforming decision-making; discovering new insights; optimizing businesses; and, innovating their industries.

2.1.2 TYPES OF DATA

The discussion of data-driven approaches should be distinguished between two main types of data. The first is the use of **public datasets** (**administrative** or **open data** and statistics about populations, economic indicators, education, etc.) that typically contain descriptive statistics, which are now used on a larger scale, used more intensively and linked. Open data is a movement of publishing digital data online in an open format bringing together a variety of societal actors ranging from individual hacktivists, NGOs and NPOs to businesses, governmental administrators and policy-makers (Hogge, 2010). The Open Knowledge Foundation (OKF), a leading NGO in promoting open data, gives the following, widely used definition: "A piece of content or data is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike" (Linksvayer, 2012). Open data, however, is not only about sharing data or making it accessible. More importantly, the term 'open data' connotes the publishing of data in ways that enable anybody to repurpose the data and to combine it with other datasets to create new, innovative online services (Marton, Avital, & Jensen, 2013).

Although, research in this area is still limited, there is some evidence that more open data and new methods of data collection and use, can save money and create economic, social and environmental value (Independent Expert Advisory Group, 2014). For example, the U-report social monitoring platform established by UNICEF in Uganda has more than 240,000 young people reporting on issues that affect their communities. Early reporting of an infectious disease in banana production contributed to halting the spread of the disease, which could have cost the country $360 million per year if left unchecked (Kumar, 2014).

The second main type of data is from **social media, sensors and mobile phones** that are typically new data sources. They are analyzed with novel methods such as sentiment analysis, location mapping or advanced social network analysis. The importance of these new types of data is increasing (Poel et al., 2015). Observed data such as CDRs from mobile phone network operators as well as data generated through the use of social media (known as “social data”) are playing an increasingly important role in academic, development and humanitarian research due to its ability to provide high-resolution, dynamic (in terms of time, space and coverage) data sources and methods of analysis. Researchers can combine data which have not previously been combined into higher-resolution datasets from which they can detect new correlations and surface new questions. Mobile phone calling records in particular can provide predictive capacity and real-time information on mobility of people following natural disasters, as shown by the work of Flowminder after the 2010 Haiti earthquake. Similar analyses may be possible to perform in crises, conflicts and elections, although multiple additional sensitivities would need to be addressed in these cases (Bengtsson et al., 2011).

The volume of data in the world is increasing exponentially: one estimate has it that 90% of the data in the world has been created in the last two years (IBM, 2013). As *Figure 3* below demonstrates, the volumes of both traditional sources of data (represented by the number of household surveys registered) and new sources (mobile subscriptions per 100 people) have been rising and openness is increasing (numbers of surveys placed online). People, economies and societies are adjusting to a world of faster, more networked and more comprehensive data. Data revolution is about the opportunity to improve the data that is essential for decision making, accountability and solving development challenges (Independent Expert Advisory Group, 2014).

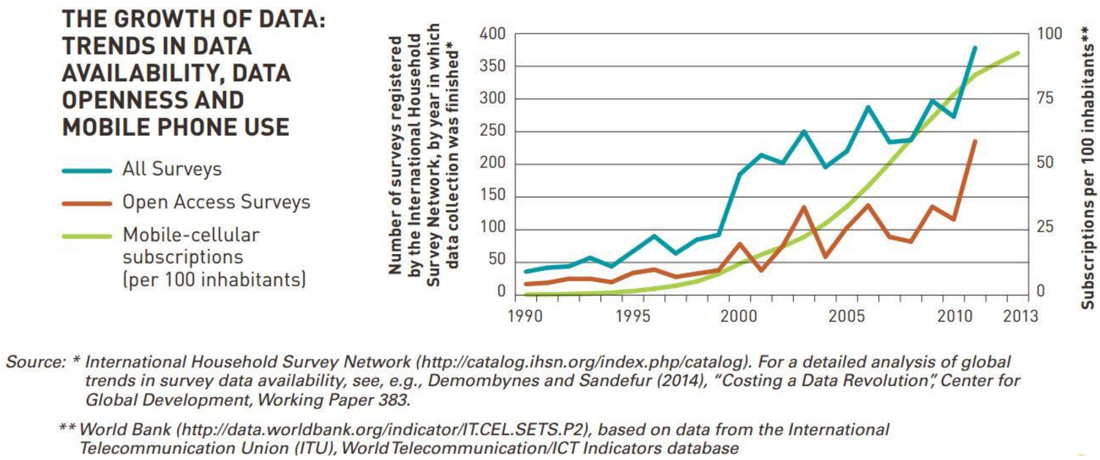

Figure 3The growth of data: Trends in data availability, data openness and mobile phone use

Big data is an intricate assemblage of techniques and infrastructures diffusing into the institutional fabric of society. Business and commerce are the obvious social domains in which Big Data will flourish (Davenport, Cohen, & Jakobson, 2005). Other domains, however, are expected to benefit as well (Manyika et al., 2011). Health care is expected to increase its effectiveness in terms of patient care and the development of new treatment regimens. On all levels of government, Big Data is expected to improve the governance of citizens, administration of services and cutting of costs. The natural sciences already gaze at the sub-atomic through the Large Hadron Collider at CERN and stargaze through telescopes, such as the Sloane Digital Sky Survey, each generating petabytes of data on a daily basis (Shiri, 2012). The social sciences hope for a new methodology based on behavioral data revealing what people are actually doing rather than what they say they are doing (Manovich 2011; Phillipe 2012). Whatever values, facts, truths or, generally, information one is looking for, it is supposedly there in the data waiting to be discovered (Marton, Avital, & Jensen, 2013).

Einav and Levin (2013) have surveyed the various ways in which economists have recently begun using datasets that were previously unavailable. Economists have used Big Data in many ways, such as gathering information about prices by scraping them from websites. Moreover, in terms of scale Eagle and Greene (2014) demonstrate that Big Data does not just apply to large populations: individual lives and interactions with others and with the environment, for example, can be measured in real time and using several types of measurement. They showed how data can scale up, to neighborhoods (for crime and traffic), to cities, then to nation-state and ultimately to the global population. The explosion in the production of Big Data, along with the development of new epistemologies, is leading many to argue that a data revolution is under way that has far-reaching consequences to how knowledge is produced, business conducted and governance enacted (Anderson,2008; Bollier,2010; Floridi,2012; Mayer-Schonberger & Cukier,2013).

2.1.3 PROBLEMS AND CHALLENGES FOR DATA

Boyd and Crawford (2012) were among the first to sketch the epistemological, ethical and social challenges of Big Data. As with any change, the data revolution comes with a range of new risks, posing questions and challenges concerning the access to and use of data. These risks must be addressed. For instance, Intel in 2012 conducted a survey asking IT Managers about their top Big Data challenges and *Figure 4* below (Intel, 2012) groups the critical issues in Big Data into three categories based on the commonality of the challenge.

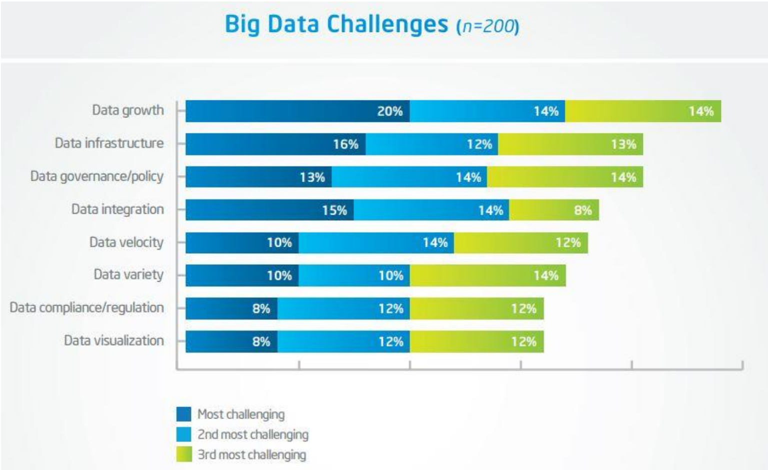

Figure 4Challenges in Big Data (Intel, 2012)

These challenges are also addressed in Cukier's and Mayer-Schoenberger (2013) influential book about big data. Fundamental elements of human rights have to be safeguarded: privacy, respect of minorities or data sovereignty requires us to balance the rights of individuals with the benefits of the collective. It raises the major ethical and legal issues of Big Data and especially the ethical issues of constraining freedom if certain behaviors (like crime) can be predicted and the legal issue of how an appeal about services offered on the basis of Big Data analysis (like credit scores or health and car insurance) is possible given that ordinary people and sometimes even the analytics experts themselves cannot understand how the analysis works.

Data are buried in administrative systems. Challenges in Big Data analysis include data inconsistency and incompleteness, scalability, timeliness, and security (Labrinidis & Jagadish, 2012; Kouzes et al., 2009). Prior to data analysis, data must be well constructed. However, considering the variety of datasets in Big Data, the efficient representation, access, and analysis of unstructured or semi structured data are still challenging. Understanding the method by which data can be preprocessed is important to improve data quality and the analysis results. Datasets are often very large at several GB or more, and they originate from heterogeneous sources. Hence, current real-world databases are highly susceptible to inconsistent, incomplete, and noisy data. Therefore, numerous data preprocessing techniques, including data cleaning, integration, transformation, and reduction, should be applied to remove noise and correct inconsistencies (Han, Kamber & Pei, 2006).

Data governance standards are lacking. Information is simultaneously increasing at an exponential rate, but information processing methods are improving relatively slowly. Currently, a limited number of tools are available to completely address the issues in Big Data analysis. The state-of-the-art techniques and technologies in many important Big Data applications (i.e., Hadoop, Hbase, and Cassandra) cannot solve the real problems of storage, searching, sharing, visualization, and real-time analysis ideally. Furthermore, the issues around privacy and consent, as well as ensuring the accuracy of Big Data analyses, are only starting to be tackled by researchers (Lane, Stodden, Bender, Nissenbaum, & (Eds.), 2014). Greenleaf (2012) provides an inventory of countries that have data privacy laws, though he points out that these laws are aspirational. Pasquale (2015) argues that “we should require data controllers to keep records of the original source of their data, noting how it was collected, purchased or bartered”.

As more data become available in disaggregated forms and data-silos become more integrated, privacy issues are increasingly a concern about what data is collected and how it is used. Furthermore, human rights cut across many issues related to the data revolution. These rights include but are not limited to the right to be counted, the right to an identity, the right to privacy and to ownership of personal data (Khan, et al., 2014).

Data are often unreliable. The abundance of data provides great opportunities to researchers trying to understand and solve social problems, but unfortunately much of the data is unreliable. Simply having a lot of data does not necessarily mean that the data are representative. As Taylor et al., (2014) are mentioning “in relation to using mobile data for research and activism, it is first important to find out whether the phone users in the data set are representative of the population.” Often mobile data is qualitatively incomplete. It offers a way to observe population displacement, but can only provide an indication that people are moving and not why they are moving or what they think, feel or need in a crisis situation. The data are only available from places where the mobile network is functional, so there may be gaps and biases that the researcher cannot evaluate (Graham, 2012). There is also an issue of data source bias. Where the data comes from only one operator or is produced by a particular social group, this may skew the research findings toward a particular demographic (Cowls, 2014).

Data can cause unintended consequences. Another ethical issue is that the relations between the data sources used (search engines, Wikipedia, mobile phone records) and the phenomena under investigation skew the knowledge towards those about whom the data are captured. This issue may arise in highly developed countries (e.g. urban vs. rural, rich vs. poor, high vs. low use of internet and mobile apps) as well as in emerging economies (Schroeder & Taylor, 2015).

Much of the new data is collected passively, from the 'digital footprints' people leave behind, from sensor-enabled objects or is inferred via algorithms. As more is known about the people and environment, there is a greater risk that the data could be used to harm, rather than to help. People could be harmed in material ways, if the huge amount that can be known about people's movements, their likes and dislikes, their social interactions and relationships is used with malicious intent, such as hacking into bank accounts or discriminating in access to services. People and societies can be harmed in less material, but nonetheless real ways if individuals are embarrassed or suffer social isolation as a result of information becoming public (Eagle & Greene, 2014).

2.2 INNOVATION

Innovation is defined as the activity of people and organizations to change themselves and the environment (Pianna, 2003). The scope of innovation varies from products and processes to organizations or even societies. The advent of the Internet has expanded the scope and volume of information. "While traditional market research techniques have been proven valuable in obtaining customer requirements, the need for innovation prompts for a more efficient and comprehensive process of information acquisition and analysis. Those who will make a better use of information will gain an advantage."(Kusiak & Tang, 2006). Nevertheless, Galanakis (2006) proposed a much broader definition of innovation: "the creation of new products, processes, knowledge or services by using new or existing scientific or technological knowledge, which provides a degree of novelty either to the developer, the industrial sector, the nation or the world, to succeed in the market place."

The study of innovation – the development of new knowledge artifacts – is of interest to engineering, business, social and behavioral sciences and spans across sociology, history, philosophy, economics, psychology and political science (Troyer, 2005). Researchers from different areas have examined how innovation is formed, progressed and is disseminated (Parthasarthy & Hammond 2002; Gopalakrishnan & Damanpour 1997 ; Klepper 1996). According to Krause (2013), "Innovation has become the keyword that should not be missing in any strategy, project proposal, concept or annual plan. Whether it is an NGO, a governmental organization, or the UN, they feel compelled to use the term 'innovative approach'. Innovation appears to multiply, to be almost universally applicable – everything and anything can be innovative."

"The innovation trend is still relatively new in the development world. Innovation had already grown to be a key phenomenon in the private sector years ago and has now caught the development industry. Yet, it lacks a precise and overreaching definition. Some focus on new products or processes, others on meeting ultimate needs, and yet others on adapting to the context" (Krause, 2013). Consider the following:

- OECD and Eurostat define innovation as 'the implementation of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, workplace organization or external relations' (OECD/ Eurostat, 2008).
- UNICEF grasps innovation as 'improving the lives of children and their families through new or better products, services and systems' (UNICEF, 2013).
- Oxford University's Humanitarian Innovation Project understands innovation 'not as novelty or invention but as the adaptation of products or processes to a particular context. It is based on the recognition that there may be alternative, untapped solutions and solution-holders "out there" that can provide new and better ways to approach the different sectors that comprise humanitarianism – water, sanitation, nutrition, communications, livelihoods, shelter, and health, for example. Furthermore, it is based on the recognition that sometimes private actors – including refugees themselves and businesses at the local, national and global levels – may offer creative and sustainable alternatives to state-led humanitarian dependency' (Betts, Bloom & Omata, 2012).

"Owing to the increased usage of the term innovation in the context of development aid, there is a distinct need for agreeing on a definition. Innovation already appears to be a vague, blurry and universally applicable concept. It could be commendable to refer to technological and social innovation to highlight the fundamental differences. In spite of the different approaches to apply innovation, the process is mainly driven by organizations from the Global North and implemented in the Global South. In addition, it seems that rather well-established and well-funded agencies have the means to engage in technological innovation while smaller ones rather employ social or soft approaches with little necessary costs."(Krause, 2013)

2.2.1 TYPES OF INNOVATION

Innovation research has distinguished between innovation types because they have different characteristics and their adoptions are not affected identically by environmental and organizational factors (Jansen et al., 2006; Kimberly & Evanisko, 1981; Light, 1998). Prior research also suggests that the process of generation of different innovation types at the industry level, and their adoption at the organizational level, is not similar (Tornatzky & Fleischer, 1990).

Innovation researchers have introduced many conceptual typologies of innovation. For instance, Zaltman et al., (1973) identified approximately 20 innovation types grouped in terms of the state of the organization, and the focus and outcome of innovation. The variety of innovation types notwithstanding, the best known and most widely studied typology of innovation is the distinction between product and process innovations (Kotabe & Murray, 1990; Light, 1998). Another widely recognized but less researched typology is the distinction between technological (also called 'technical') and administrative (also called 'organizational' and 'management') innovations (Birkinshaw et al., 2008; Kimberly & Evanisko, 1981;).

Edquist and colleagues (Edquist et al., 2001; Meeus & Edquist, 2006) compare these two established typologies and offered a taxonomy that distinguishes between two types of product innovations ('in goods' and 'in services') and two types of process innovations ('technological' and 'organizational'). Hamel (2006) distinguished between two types of process innovation that resemble Meeus and Edquist's distinction: innovations in operational processes (such as customer services, logistics, and procurement) and innovations in management processes (such as strategic planning, project management, and employee assessment). From Meeus and Edquist's (2006) innovation types, we analyze three that are applicable to service organizations: service innovations, technological process innovations, and administrative process innovations.

Service innovations. Barras (1986) defines a product as a good or service offered to the customer or client. Innovation research has not generally distinguished between product and service innovations; that is, services offered by organizations in the service sector are conceptualized to be similar to products introduced by organizations in the manufacturing sector (Sirilli & Evangelista, 1998). This view has been prevalent because product and service innovations have external focus, are primarily market driven, and their introduction results in differentiation of the organization's output for its customers or clients (Damanpour & Gopalakrishnan, 2001). Hence, like product innovations, the drivers of service innovations are mainly clients' demand for new services and executives' desire to create new services for existing markets or find new market niches for exiting services (Matthews & Shulman, 2005). Given the focus on meeting client needs in the service sector, the nature of service innovation is best understood through its relationship to service user.

Process innovations. Contrary to product or service innovations, process innovations have an internal focus and aim to increase efficiency and effectiveness of the internal organizational processes to facilitate the production and delivery of goods or services to the customers (Boer & During, 2001). The new processes can be associated with the 'technological core' or the 'technical system' of the organization (technological process innovations) or to the 'administrative core' or the 'social system' of the organization (administrative process innovations) (Meeus & Edquist, 2006).

Technological process innovations are new elements introduced into an organization's production system or service operation for producing its products or rendering its services to the clients (Damanpour & Gopalakrishnan, 2001). The drivers of these innovations are primarily reduction in delivery time, increase in operational flexibility, and lowering of production costs (Boer & During, 2001). Technological process innovations, therefore, modify the organization's operating processes and systems (Meeus & Edquist, 2006). In service organizations, these innovations are primarily innovations associated with information technology (Uchupalanan, 2000).

Administrative process innovations are new approaches and practices to motivate and reward organizational members, devise strategy and structure of tasks and units, and modify the organization's management processes (Birkinshaw et al., 2008; Light, 1998). Whereas technological innovations are directly related to the primary work activity of the organization and mainly produce changes in its operating systems, administrative innovations are indirectly related to the organization's basic work activity and mainly affect its management systems (Damanpour & Evan, 1984). Administrative process innovations pertain to changes in the organization's structure and processes, administrative systems, knowledge used in performing the work of management, and managerial skills that enable an organization to function and succeed by using its resources effectively.

The three classical types of innovation nature are incremental, radical, and disruptive. Incremental innovations are by far the most common form of medical technology innovations and build on existing knowledge (Ali, 1994; Ettlie et al., 1984). Radical innovations produce fundamental changes in products, services, or processes (Hall & Martin, 2005; Lettls et al., 2006). Disruptive innovations change the very basis of practice through breakthrough and transformational change (Hwang & Christensen, 2007).

Notwithstanding the lack of a harmonized understanding of innovation or innovative development projects the OSCE and Eurostat outline four types of innovation: product, process, marketing and organizational innovation (OECD/Eurostat, 2008). Product and process innovations are closely related to the concept of technological developments. A *product innovation* is the introduction of a good service that is new or significantly improved regarding its characteristics or intended uses, including significant improvements in technical specifications, components and materials, incorporated software, user friendliness or other functional characteristics (OECD Oslo Manual, 2005). Product innovations can utilize new knowledge or technologies, or can be based on new uses or combinations of existing knowledge or technologies. The term product covers both goods and services. Product innovation is a difficult process driven by advancing technologies, changing customer needs, shortening product life cycles and increasing global competition. For success, it must involve strong interaction within the firm and further between the firm and its customers and suppliers (Ben-Akiva & Gershenfeld, 1998).

A *process innovation* is the implementation of a new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software. Process innovations can be intended to decrease unit costs of production or delivery, to increase quality, or to produce or deliver new or significantly improved products (OECD Oslo Manual, 2005). Fagerberg et al., (2004) stressed that while the introduction of new products is commonly assumed to have a clear, positive effect on the growth of income and employment, process innovation, due to its cost-cutting nature, can have a more hazy effect.

A *marketing innovation* is the implementation of a new marketing method involving significant changes in product design or packaging, product placement, product promotion or pricing (OECD Oslo Manual, 2005). Marketing innovations target at addressing customer needs better, opening up new markets, or newly positioning a firm's product on the market with the intention of increasing firm's sales. Marketing innovations are strongly related to pricing strategies, product package design properties, product placement and promotion activities along the lines of four P's of marketing (Kotler, 1991).

Finally, an *organizational innovation* is the implementation of a new organizational method in the firm's business practices, workplace organization or external relations. Organizational innovations have a tendency to increase firm performance by reducing administrative and transaction costs, improving work- place satisfaction (and thus labor productivity), gaining access to non-tradable assets (such as non-codified external knowledge) or reducing costs of supplies (OECD Oslo Manual, 2005). Examples would be the introduction of practices for codifying knowledge by establishing databases of best practices, lessons learnt and other knowledge, so that they are more easily accessible to others. Thus, organizational innovations are strongly related with all the administrative efforts of renewing the organizational routines, procedures, mechanisms, systems, etc. to promote teamwork, information sharing, coordination, collaboration, learning and innovativeness.

2.2.2 DATA – DRIVEN INNOVATION

Data-driven science is more open to using a hybrid combination of abductive, inductive and deductive approaches to advance the understanding of a phenomenon. It differs from the traditional, experimental deductive design in that it seeks to generate hypotheses and insights 'born from the data' rather than 'born from the theory' (Kelling et al., 2009). Data-driven innovation is a new concept that may meet the needs of practitioners. Its underlying belief is that "new knowledge or valuable innovative ideas are embedded somewhere in the data." Since the collected data may document the environmental variables, the decision process and the limiting constraints, valuable insights are usually hidden in it. It is estimated that more than 75% of the new design initiatives use the previous design knowledge (Iyer et al., 2005). Past knowledge could be improved and applied in other areas (Kusiak & Tang, 2006).

Without data, decisions are guesses; with it, decisions are targeted, strategic, and informed. These lead to better business, better government, and better solutions. Data-driven innovation, as Dr. Joseph Kennedy describes in Chapter 2 in the Future of Data-driven Innovation (Hienz, 2014), has enormous economic value, with Big Data product and service sales exceeding $18 billion in 2013, expected to reach $50 billion by 2017 (Kelly, 2014). This value comes in the form of: new goods and services; optimized production processes and supply chains; targeted marketing; improved organizational management; faster research and development; and much more.

What is important about data is not their volume, but how they may contribute to innovation and therefore be used to create value. Data alone do not possess inherent value; instead it is the processing of data in innovative ways that brings new economic and social benefits, and this value creates a virtuous circle to feed into more use of data-based decision making and analysis (Hilbert, 2013). In other words, it is the use of data that really matters (Hemerly, 2013). For instance, The Royal Netherlands Meteorological Institute has found a way to generate extremely accurate rainfall information using nothing more than existing data from cell-tower installations (The Economist, 2013a). Nevertheless, the ability to innovate, like many other skills, should be taught. Different individuals can come up with different innovative ideas using the same data as we can see in *Figure 5* (Kusiak & Tang, 2006).

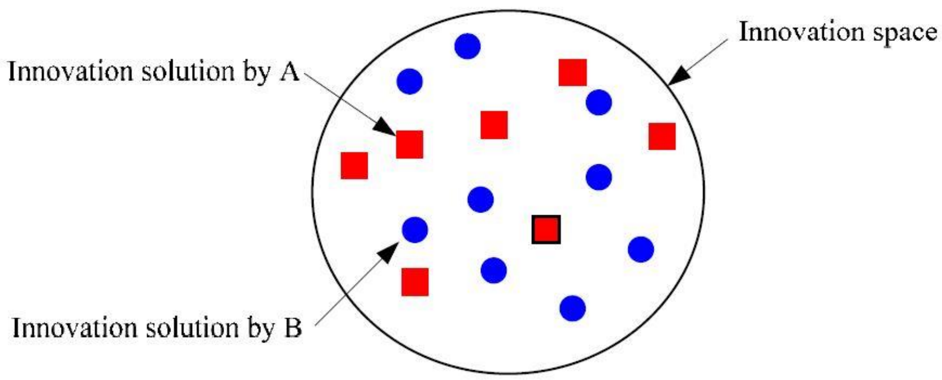

Figure 5Innovation Space (Kusiak & Tang, 2006)

"The data-driven innovation by no means automatically produces innovation nor is it the only way to generate innovation. However, it is a pragmatic and systematic approach toward innovation. Assisted by a human it iteratively synthesizes heterogeneous data. The interpretation of information acquired by analyzing data is the key to successful innovation endeavor. It is important to note that the process of data-driven innovation is dynamic and it is scenario dependent. Direct inferences and conclusions made from other so-called "best practices" running the danger of observation of bias."(Kusiak & Tang, 2006).

Furthermore, data analysis is critical in data-driven innovation. Traditional statistical methods are hardly applicable. Data mining and evolutionary computation have created new possibilities for discovery of interesting patterns, trends and associations from the data (Kusiak & Tang, 2006). A brief overview follows:

(1) Data mining

Data mining, an integral part in the process of knowledge discovery in database (KDD), is the process of generating useful information from raw data (Tan et al., 2005). Many successful applications in business, biosciences and engineering have been reported (Ganguly & Gupta, 2004). It is an interdisciplinary field incorporating machine learning, artificial intelligence, and statistics. It is largely based on supervised learning and unsupervised learning. Algorithms such as neural networks (NN), decision tree, support vector machine (SVM), and k-means clustering are widely applied. Readers interested in more details may refer to Tan et al., (2005) and Larose (2005).

(2) Evolutionary computation

Evolutionary computation covers the study of the foundations and applications of heuristics algorithms based on the principles of natural evolution. Examples of techniques include genetic algorithms and ant colony optimization (De Jong, 2006). The link between innovation and evolutionary computation has not been built yet, however, there appear to be a natural match between the two.

In principle, it is understood that the collection of information from digital transactions and interactions is something that is unstoppable. Whether we like it or not, the digital trail we leave behind in the e-world is amazingly large (Al-Khouri, 2014). O'Harrow (2006) indicates that although the emergence of a data-driven surveillance society has provided the conveniences of access to information and services (such as cell phones, discount cards, and electronic toll passes), it also has created new approaches to watching us more closely than ever before. He also points to the fact that as companies customarily collect billions of details about nearly every connected individual, the world will reach a state where people will lose control of their privacy and identities at any moment. *Figure 6* depicts an illustrative diagram of the evolving possibilities of capturing a data trail of individuals in the digital world.

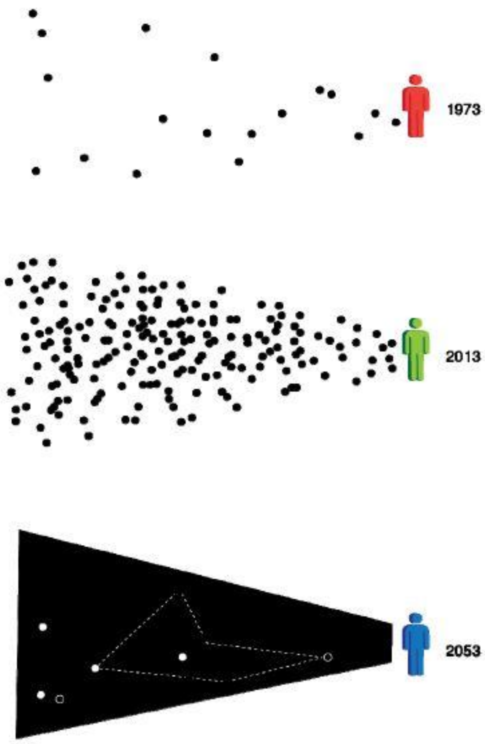

Figure 6Digital Behavior and Data Trail in Big Data (Al-Khouri, 2014)

2.2.2.1 THE BENEFITS OF DATA – DRIVEN INNOVATION

The benefits of smart data use are being realized by companies large and small, and by society more broadly. For example, a start-up called FlightCaster (www.flightcaster.com) uses data analytics to predict flight delays before airlines alert customers. Individuals can participate in hackathon events and visualization competitions. These public events facilitate innovation and data literacy, breaking down the boundaries of organizational research and development. However, with more opportunities for innovation comes more competition, which in turn drives new products and services, improvements to existing offerings, and novel approaches to old problems. Two broad categories of data-driven innovation that have already shown measurable, positive returns are decision making and efficiency improvement (Hemerly, 2013).

Every industry and domain wants to make use of data and find new ways to make decisions, improve processes, and come up with methods and technologies that solve problems—in short, to use data to drive innovation. Semantic analysis of posts on Twitter can tell us about pop culture trends or the response to a particular crisis. Data are the ultimate renewable resource. But data are meaningless until compared to other data, visualized in context or analyzed for significance. The number of people with basic data science skills needed to transform data into meaningful action or insight—increases exponentially. As Edd Dumbill wrote in a January *Forbes* piece, "In the broader business consciousness, at the end of 2012, 'big data' really means 'smart use of data'". Long after the phrase "big data" has run its course, data will continue to catalyze innovation. Solutions for problems will come from rigorous analysis and new ways of interpreting data. The products, services, and processes enabled by data and developed to support smart uses of them can be thought of as data-driven innovation. These kinds of data-driven innovations are the future of "big data," but ultimately, public policy will be a critical factor in determining their place in society (Hemerly, 2013).

As Less Andrade et al., (2014) mentioned many sectors benefit from data-driven innovation: healthcare (e.g., diagnosis and treatment), financial services (e.g., analyzing market trends and economic conditions), and transportation and public administration (e.g., metrics on what citizens want and where economic development is headed), to name a few. In one example, a philanthropic research center stores and analyzes the cancer genome and the sequences and mutations of more than 10,000 cancer cases to understand the complexity of the disease (Burke, 2012). Moreover, the UN is working with governments around the world to understand global trends related to hunger, poverty, disease, and job loss (United Nations, 2012). However, data are neither a good nor a service and so they escape traditional economic analysis. This highlights the complication of discussing data: although the value often creates an economic reward, such measurements are not easy to make. One example of innovative data use that has a difficult-to-quantify economic value proposition is Google's Flu Trends, which provides near real-time estimates of flu activity for a number of countries around the world. Flu Trends provides its analysis based on aggregated search queries (The Economist, 2013b). This example illustrates how the openness and accessibility of data are extremely crucial for innovation.

Studies suggest that there is a direct connection between data-driven decision-making in business and improved firm performance. Firms that adopt data-driven decision-making have an output and productivity that is 5 percent to 6 percent higher than would be expected, given their other investments and their information technology (IT) usage (Brynjolfsson et al., 2011). Another study has shown that the use of Internet computing tools can also help firms reach decisions more efficiently, across a broad range of industries, as they allow firms of all sizes to leverage data-driven analysis without needing to make huge investments in their IT infrastructure (Cacciola & Gibbons, 2012). For most companies, data-driven decision making improves output and productivity, which leads to increased revenue. Data for decision making means using both real-time and historical data to inform decisions in the present. Today's computational capabilities enable data scientists to put large sets of historical data to use. In combination with our other decision-making skills and tools, data sight improves a problem's granularity, depth, and time horizon in extraordinary ways (Hemerly, 2013).

By providing a way to check assumptions, detect problems, clarify choices, prioritize resources, and identify solutions, data-driven policymaking injects data-based rationality into the policymaking process, all of which could also create economic benefits (Esty & Rushing, 2007). According to OECD, by fully exploiting public data, governments in the EU could reduce administrative costs by 15 percent to 20 percent, creating the equivalent of €150 billion to €300 billion (Manyika et al., 2011).

2.3 NON-GOVERNMENTAL OREGANIZATIONS (NGOs)

The definition of NGOs is evolving to what an NGO is rather than what an NGO is not (Luxmore & Hull, 2011). Teegen et al., (2003) list several characteristics that delineate NGOs, including that NGOs are 'civil society counterparts of MNEs and governments'. NGOs are non-profit organizations that operate by providing services and advocating change through organizing, mobilizing resources and disseminating information (Doh & Teegen, 2003; Spar & La Mure, 2003). Lambell et al., (2008) define NGOs as including any actor who is not part of the market or governmental sector. NGOs are growing more numerous, their number has risen from roughly 1,000 in 1914 to hundreds of thousands today (Leverty, 2009). Thus, the role of NGOs in shaping government regulations is important (Phillips, 2006). NGOs derive their financial support from both public and private sources. A few will accept no public sector money, while others get between 60%-70% of their income from donor governments (U.S. Agency for International Development, 2000).

How NGOs are organized and governed affects their work. The largest NGOs comprise multiple national affiliates under various forms of confederation. Some are purely fundraising and recruitment vehicles, while others operate independent programmes. The majority of the NGOs has expanded its programme portfolios from initial emergency aid deliveries to long-term, anti-poverty activities throughout the developing world. NGOs are controlling a larger share of humanitarian resources than ever before (Macrae et al., 2002). Calculating the relative importance of emergency relief as against rehabilitative or developmental work is, however, difficult because most NGOs' budgets do not differentiate between 'emergency' and 'development'. In general, very few NGOs doing humanitarian work bill themselves as exclusively relief organizations. The vast majority maintain both relief and development programmes and long-term missions.

2.3.1 THE THREE TRADITIONS

Today's NGOs have evolved from one of three main historical strands: the *religious*, the '*Dunantist*' and the '*Wilsonian*'. The religious humanitarian tradition is the oldest of the three, and is predicated on the basic tenets of compassion and charitable service. Although religious humanitarianism has its antecedent in missionary work in the European colonial empires, most religious humanitarian agencies do not proselytize in any direct way, though many may combine religious values with social goals. Of the Christian faiths, Catholicism provides some of the largest and most visible aid organizations; CRS, Caritas and CAFOD, for instance, are all Catholic organizations.

The second strand could be labeled the 'Dunantist' tradition, for Henri Dunant, whose horrified reaction to the aftermath of the Battle of Solferino in 1859 launched the Red Cross as a humanitarian movement based on the protection of civilians in war. Some of today's largest humanitarian NGOs have their roots in this tradition such as Save the Children UK, Oxfam and MSF.

The third strand, the 'Wilsonian' tradition, characterises most US NGOs. It stems from US President Woodrow Wilson's ambition of projecting US values and influence as a force for good in the world. Wilsonian NGOs tend to have a practical, operational bent, and practitioners move back and forth between NGOs and government (Stoddard, 2003).

Dutch NGOs may also be included in the Wilsonian realm. The Dutch government gives generously to humanitarian causes and is politically liberal, so that its NGO recipients have few qualms about accepting large government donations (O'Malley & Dijkzeul, 2002).

3 RESEARCH METHODOLOGY

The following chapter describes the overall research methodology and design used in this research. The following sections explain the overall research approach taken, the structure of the literature review, the research protocol for the gathering and analysis of the data and the identification of new possible methods to assist NGOs.

3.1 RESEARCH APPROACH

The two-fold purpose of this study is from one hand identifying the gap in literature and on the other hand identifying the challenges in practice. First of all, through our research we will try to identify, evaluate and interpret the available literature relevant to the topic area of "data-driven innovation" and try to connect it with the NGOs. A Systematic Literature Review will be conducted for that reason. Secondly, semi-structured interviews with various NGOs in the Netherlands will be conducted in order to identify the challenges in practice. A qualitative research approach, grounded theory, is taken for answering the research questions. In *Figure 7*, a representation of the general research approach is shown.

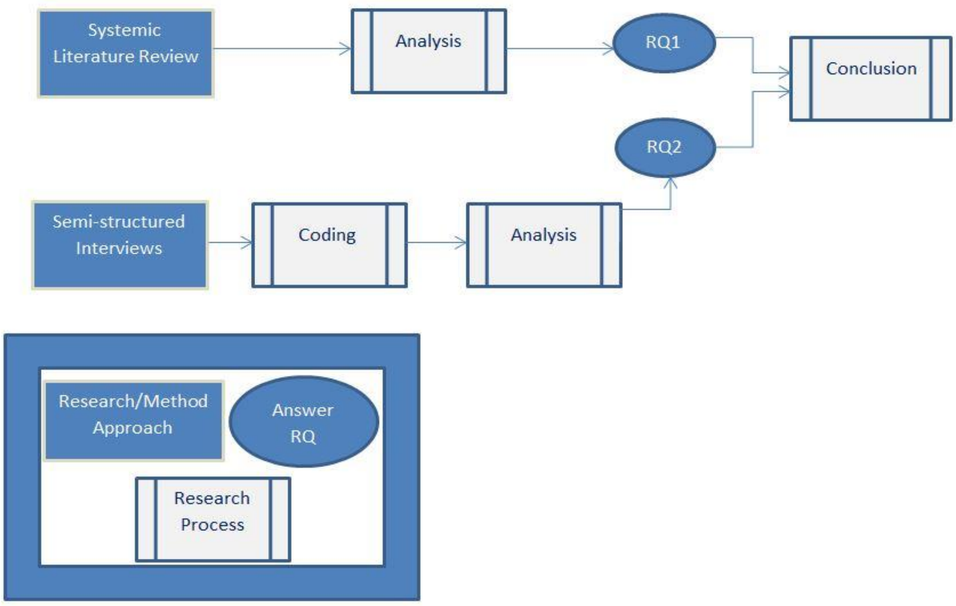

Figure 7General Research Approach

3.2 SYSTEMATIC LITERATURE REVIEW (SLR)

In order to answer the first question a systematic literature review was conducted. The SLR began with a set of keywords that were used to search for relevant literature. The following keywords were used: “data-driven innovation”. It is a rigorous method to identify and evaluate all the available research relevant to a particular topic area. A systematic review aims to present a fair evaluation of a research topic by using a trustworthy and auditable methodology. The method was originally introduced to analyze evidence in medical trials and since then has been applied in many fields including Software engineering (Kitchenham et al., 2009) and decision support systems (Garg et al., 2005). Moreover, the method helped us to identify gaps in current research in order to suggest areas for further investigation and improvement.

Specifically, the stages associated with planning the review are:

- *Identification of the need for a review* - The need for a systematic review arises from the requirement to summarize all existing information about data-driven innovation in a thorough and unbiased manner.
- *Development of a review protocol* - A review protocol specifies the methods that will be used to undertake a specific systematic review. A pre-defined protocol is necessary to reduce the possibility of research bias.

Once the protocol has been agreed, the review will take place as we can see from *Figure 8* below.

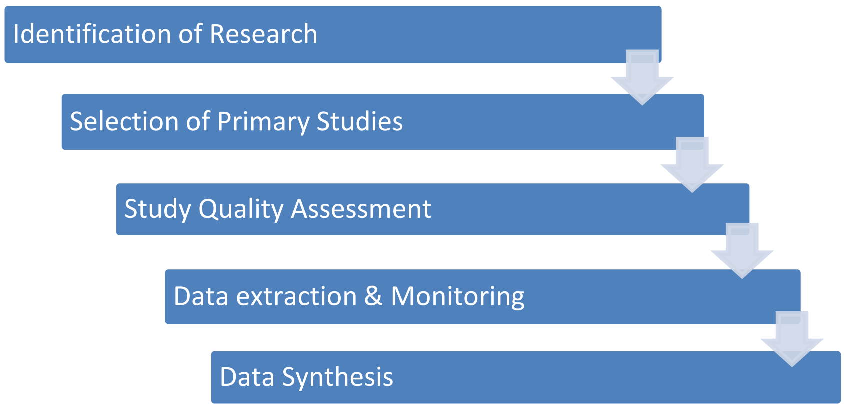

Figure 8Stages of Review

The stages associated with conducting the review are (Kitchenham et al., 2009):

1. *Identification of research* - The aim of a systematic review is to find as many primary studies relating to the research question as possible using an unbiased search strategy.
2. *Selection of primary studies* - Once the potentially relevant primary studies have been obtained, they need to be assessed for their actual relevance.
3. *Study quality assessment*:
    - To provide still more detailed inclusion/exclusion criteria.
    - To investigate whether quality differences provide an explanation for differences in study results.
    - To guide the interpretation of findings and determine the strength of inferences.
    - To guide recommendations for further research.
4. *Data extraction & monitoring* - The objective of this stage is to design data extraction forms to accurately record the information we will obtain from the primary studies.
5. *Data synthesis* - Data synthesis involves collating and summarizing the results of the included primary studies. Using statistical techniques to obtain a quantitative synthesis is referred to as *meta-analysis*.

3.3 SEMI – STRUCTURED INTERVIEWS

Based on the state-of-the-art scientific literature results, the next step was to define a qualitative questionnaire on the challenges and needs of NGOs. The scope of the questionnaire included various NGOs in the Netherlands. The outcomes are valuable in order to investigate the challenges regarding data and also to identify ways that data-driven innovation methods could help NGOs.

The questions of Appendix II formed the basic structure of the interviews. The interviewees were asked upfront whether they gave permission for recording of the interview. Recordings made it possible to accurately transcribe the interviews. Through semi-structured interview a consistent line of inquiry is pursued while avoiding being too rigid.

3.4 CODING

To apply grounded theory, the method of Corbin and Strauss is used. Grounded theory seeks not only to uncover relevant conditions but also to determine how the actors under investigation actively respond to those conditions, and to the consequences of their actions (Strauss & Corbin, 1990).

The data found in the interviews are analyzed to find potential indicators of phenomena, which are then given potential conceptual labels. Different phenomena with the same conceptual term accumulate as the basic units for theory. These concepts in the grounded theory approach become more numerous and more abstract as the analysis continues (Straus & Corbin, 1990). What is important for grounded theory is to group concepts that are found to refer to the same phenomenon into categories. Corbin and Strauss call these categories the “cornerstones” of a developing theory because they provide the means by which a theory is integrated.

To go into more detail about the coding, Corbin and Strauss divide the coding process into three stages.

1 **Open coding:** here events/actions/interactions etcetera are compared against others for similarities and differences. These phenomena are conceptually labeled. The next step in this stage is to categorize these concepts into categories and subcategories.

2 **Axial coding:** In the next phase these categories are interconnected. These relationships are tested against the data. Also the categories are further developed. Subcategories are related to categories base on conditions, context, strategies and consequences, this is done to make the conceptual linkage more specific.

3 **Selective coding:** is the process by which all categories are unified around a central core category, which represents the central phenomenon of the study. Also, categories which need further explanation are filled-in with descriptive detail.

The data analysis followed the following steps:

1 Recordings are transcribed.

2 Transcriptions are made.

3 Each transcript is coded based on Corbin and Strauss method. This means the first step is to begin with open coding. The text is conceptually labeled and first efforts are made to identify categories.

4 The codes are categorized based on frequency of coding in combination with concepts found in the Systematic Literature Review and concepts discovered during the analysis phase.

5 Each coded transcription is analyzed once more and first attempts are made to relate codes to each other on paper.

6 Categories which need further explanation are filled-in with descriptive detail.

4 RESULTS

In this chapter the results of the data collection are described. The results are divided in two parts. Firstly, we will introduce the outcomes from the Systematic Literature Review (SLR) and secondly the results from the performed semi-structured interviews will be presented.

First of all Section 4.1 describes a general overview of the SLR. Secondly section 4.2 contains the findings from SLR. 4.3 describe the results which were extracted from the interviews.

4.1 SYSTEMATIC LITERATURE REVIEW

Systematic Literature Review is a process for reviewing the literature using a comprehensive preplanned strategy to locate existing literature, evaluate the contribution, analyze and synthesize the findings and report the evidence to allow conclusions to be reached about what is known and, also, what is not known (Denyer & Tranfield, 2009). As Kitchenham (2004) mentioned, a Systematic Literature Review is a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest. Based on that process, we were able to answer the *RQ1: "What is the state-of-the-art on scientific literature on data-driven innovation methods?"*. Our objectives were to identify all available research relevant to the topic of data-driven innovation and also to spot any possible gap in the literature review and thus space for innovation.

4.1.1 SEARCH STRATEGY

The SLR began with a set of keywords that were used to search for relevant literature. The following keywords were used: "data-driven innovation". Our initial search in Google Scholar provided us with 304 articles, books references, etc. We came up with a strategy to remove all duplications and articles beyond the scope of this study. Using the flow diagram (*Figure 9*) allows the number of studies reviewed in stages one and two in the SLR process to be reported clearly.

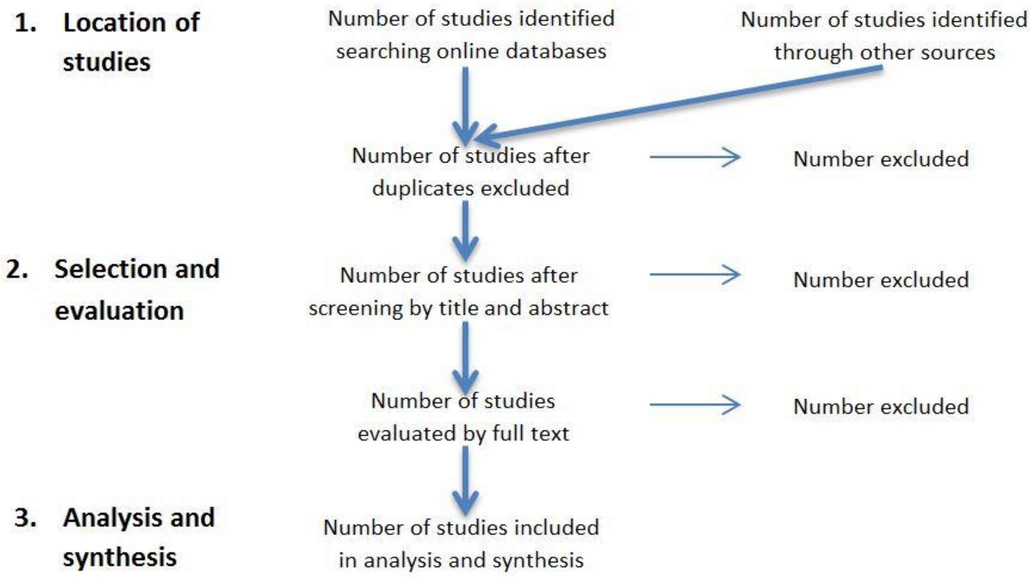

Figure 9Flow diagram for reporting Systematic Literature Review study location, selection and evaluation (Moher et al., 2009)

4.1.2 SELECTION CRITERIA

We chose only English-language articles, published in peer-reviewed journals. After removing duplicates and articles beyond the scope of this study, we ended up with 187 articles that met our criteria. We developed the analysis and synthesis (*Figure 9*) based on the findings.

4.1.3 KEY POINTS USED FOR ANALYSIS AND SYNTHESIS

Our analysis was based on 187 articles that stemmed from our search for relevant research with the keywords “data-driven innovation”. The analysis and synthesis phase included the categorization of those 187 articles into several sectors. We basically created an Excel table that included all those articles and categorized them accordingly (APPENDIX I). Below follows the categories/columns of the table and their description.

- *Title*, the title of the article.
- *Abstract*, the abstract of the article.
- *Link*, the URL link of the article.
- *Business/Government/NGO or Non-Profit*, types of authorities responsible for data that appeared in the article.

- *Business Sector*, primary policy areas as defined by the Harvard Business School (HBS Working Knowledge).
- *Business Sector Description*, a more detailed description of the policy areas including information such as subject, geographical etc.
- *Paper Classification*, categorization of the articles according to evaluation criteria defined by Wieringa et al., (2005).
- *Research Method*, the process used to collect information and data for the article. The methodology may include case study, survey, field experiment, small example, mathematical analysis, a new conceptual framework, author's opinion etc.
- *Type of Innovation*, based on what drives the innovation as described in the article we applied the conceptual typologies of innovation (product, service, process, data-driven etc.).

4.2 SYSTEMATIC LITERATURE REVIEW FINDINGS

Based upon the literature (Kitchenham, 2004) and the findings from the SLR (APPENDIX I) after the collection of the necessary and relevant data (research articles), we were able to identify the available research to the topic of "data-driven innovation". In the following sectors we will show and discuss the frequency and proportion of occurrences of various categories.

The inventory of relevant articles contains 187 articles. In short, the inventory analysis shows that either government or business sector implement the majority of the initiatives with NGOs staying far behind. In relation to the industry areas concerned, two third of the data-driven articles in the inventory focus primarily on information, information technology, education, health, energy, public administration and transportation. The majority of the articles were describing a data-driven innovation method, however, most articles were introducing more than two types of innovation. The distinction between the innovation types was difficult and the percentages for each category do not differentiate a lot. As regards the paper classification categorization, more than fifty per cent of the articles in the inventory are philosophical papers, validation research and proposal of a solution. The significant number of articles in these categories indicates the novelty of the data-driven innovation domain.

4.2.1 TYPES OF AUTHORITIES RESPONSIBLE FOR DATA

Government bodies and business sector implement the majority of data-driven innovation initiatives in the inventory (*Figure 10*). It should be noted that the inventory was collected with a clear focus on English-written research articles related to “data-driven innovation”. Examples of usage for the government bodies are: the Data Archiving and Networked Services (DANS) which promotes sustained access to digital research data in the Netherlands. Researchers can deposit their data through the online archiving system EASY. Via the portal NARCIS the research data are shown in context, namely in relation to e-publications, and other research information. The data-driven directed acyclic graphs (DAG) representation used in India showed that energy consumption causes carbon emissions and economic growth while there is a bidirectional causality between carbon emissions and economic growth. The importance of electronic health records (EHRs) in the healthcare system is leading to great expectations for all stakeholders. This article discussed the role of data data-driven innovation in the reform of the U.S. healthcare system by focusing on providing real-time data to its increasingly consumer-minded customers and embracing technology to engage stakeholders in health promotion and disease prevention. This means that the healthcare industry's use of data-driven innovation needs to be based on modernized data systems, increased access to data, and implementation of easy-to-use technology that will allow stakeholders to participate in achieving desired health outcomes and cost reductions.

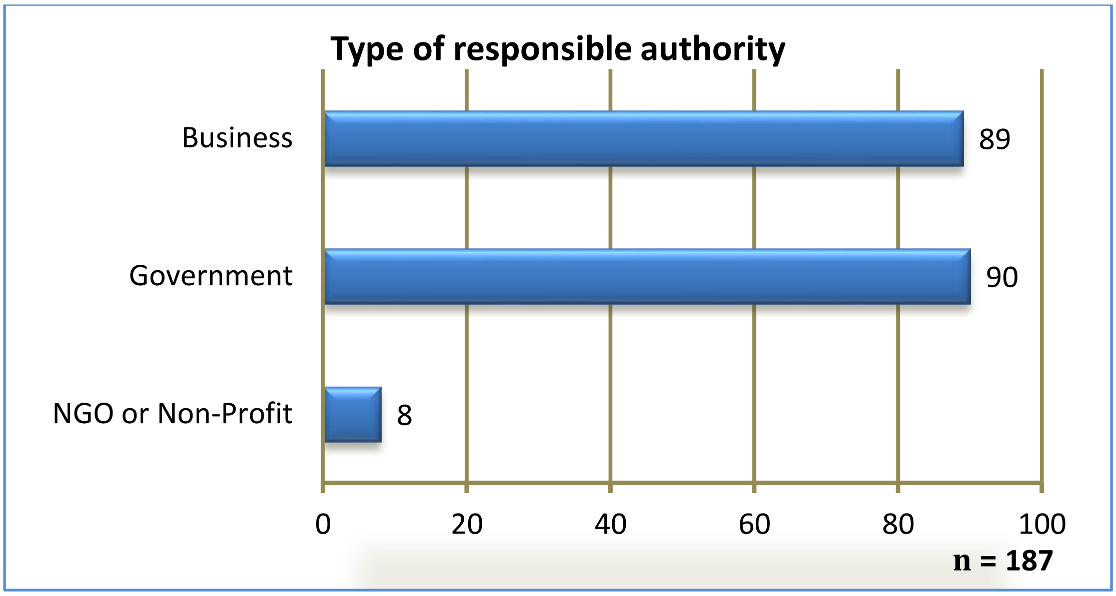

Figure 10Type of responsible authority

On the other hand, multiple are the examples of data-driven innovation method that we encountered in the business sector. The data-driven learning can be used to increase online sales yet the interpretative skills possessed by humans are still needed. By reading the behavior of individuals and building the e-commerce offerings based on real-time adaptation to their individual-level responses, e-commerce applications are able to adapt dynamically to individual situations without any relationship to any understanding about how customers are theoretically supposed to behave, or how other customers have behaved on aggregate. Another example is the one with Casey Family Services, the direct services agency of the Annie E. Casey Foundation, a multi-service child welfare agency that provides foster care, permanency planning, and family preservation and support services. The agency has incorporated its work in benchmarking and data-driven dash boarding processes into its model of "learning while doing." The need for data-driven change management makes the learning while doing approach particularly well suited for organization leaders and model developers in the human services. One of the most significant developments in the online environment over the last few years has been the rise of social media. More and more individuals are making use of Online Social Networks (OSNs) to stay in touch with family and friends. As the number of actors engaging with OSNs and OSN data increases, so does the risk for potential privacy infringements. The data collected by apps may also be used to facilitate behavioral targeting. Various entities collect and/or analyze data relating to the browsing activities of OSN users.

NGOs are also implementing data-driven innovation initiatives. Relevant NGOs include proponents of open government, who analyze available data in order to check on their governments' policies and processes in countries such as Kenya, Nigeria, Liberia or Pakistan. For instance, reported school attendance and enrolment in Kenya. Trends in maternal mortality in Nigeria and mining Indonesian Tweets to understand food price crises. Other articles focus on the very local level of the city of Rotterdam and Almere (the city, the neighborhood, the street level). The main focus is on what is happening in the field of data-driven innovation in relation to liveability in the city.

4.2.2 INDUSTRY AREAS

Figure 11 provides an overview of the primary industry areas, using a categorization of business sectors as defined by the Harvard Business School (HBS Working Knowledge). The inventory includes data-driven initiatives in a broad range of industry areas and illustrates the potential of data-driven innovation approaches for all of them.

A substantial number of initiatives concerns information and information technology being applied in different fields e.g., government transparency, citizen engagement, economic growth, business creation. An example is the Open Government Data strategy and portal of the City of Vienna, which offers the release of public data and manages to increase transparency of public administration and processes, as well as, reduces the administrative efforts. Not to mention, the high-tech firms that collect and apply user involvement and data throughout the whole innovation process, with the aim of developing new products and features that are appreciated by their users.

If one aggregates all industry areas by the 187 initiatives in the inventory, information and information technology as well as education and health are the four industry areas mentioned most often. A second observation is that the majority of the initiatives are relevant for two or more industry areas. This indicates that data-driven innovation initiatives cut across several industry areas, e.g. public administration, service and information technology.

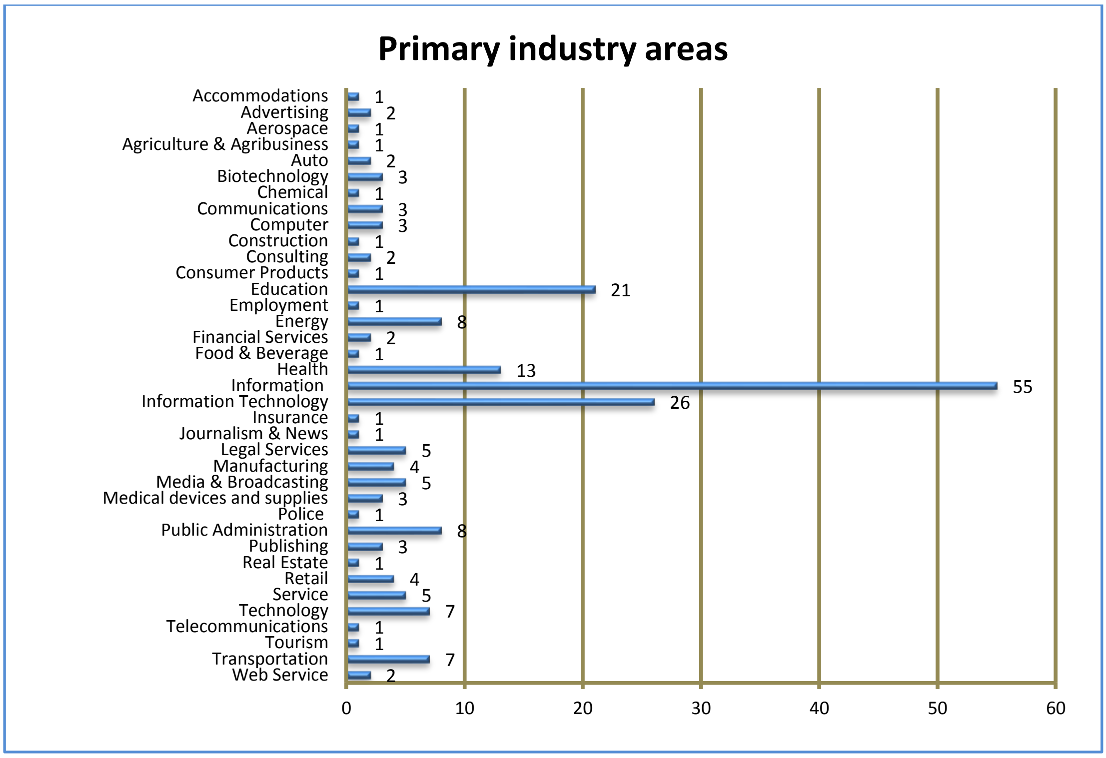

Figure 11Primary industry areas

As regards the papers written by NGOs, we can clearly observe that the majority of them is focusing on the information area. Problems and issues regarding data policies, data openness and data accuracy are prevailing. Nearly all initiatives are relevant for two or more policy areas e.g., information, health, education. An example is the maternal mortality in developing countries, due to the lack of adequate data and functioning systems for births and deaths registration.

4.2.3 TYPE OF INNOVATION

Figure 12 shows the type of innovation as described in the articles from the SLR. As it is illustrated data-driven, policy/business model and social innovation rank in the first three places. Product/service and system are following. It is worth mentioning that the majority of the articles refer to more than one type of innovation. The distinction of what drives innovation in many cases is difficult to make due to the overlapping of the innovation drivers.

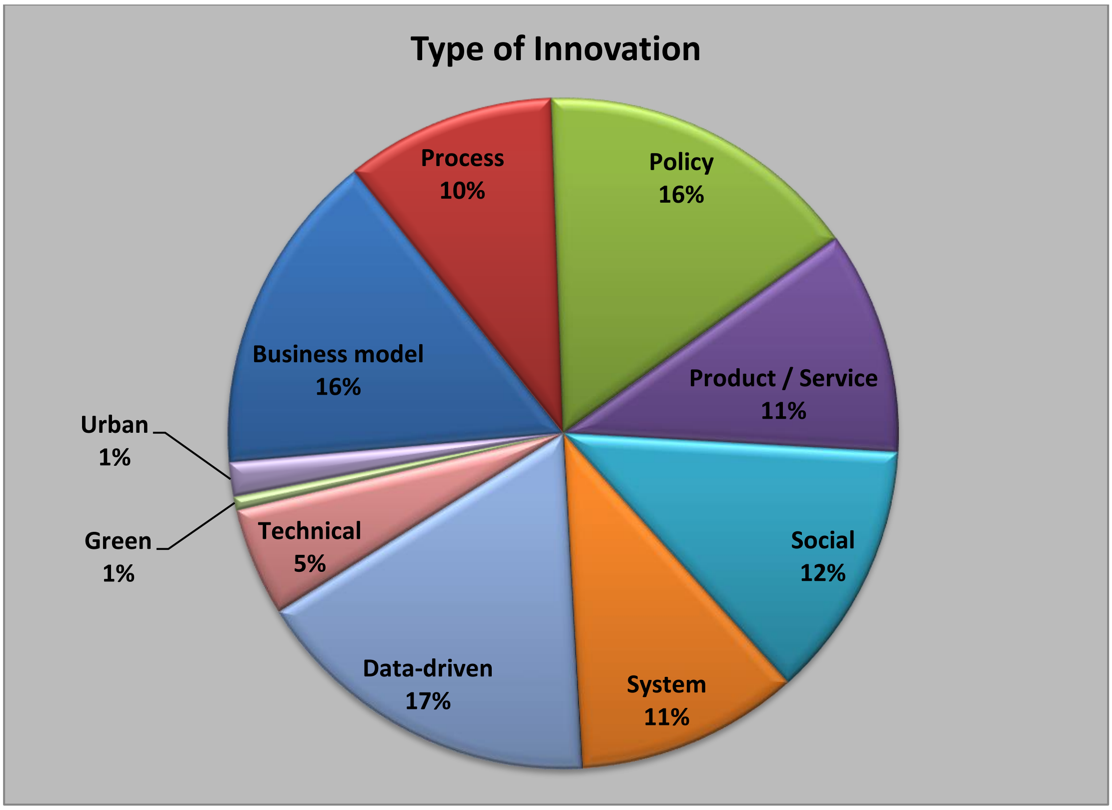

Figure 12Type of innovation

Many online companies have adopted business models that rely on personal data as a key input. One common business model involves two-sided markets, where companies offer consumers free technologies, services and products with the aim of acquiring more valuable data from these consumers to assist advertisers to target the right audience. Other companies undertake data-driven strategies to obtain and sustain competitive advantages over their rivals. Furthermore, the advertising-supported media businesses, such as Facebook, Google, Twitter and many other online services in which user data are important, are not just technology companies- they are media companies. The innovation drives in such companies are so diverse.

Another example would be the policy-makers that need to integrate open data use into social strategies to encourage open data use. The ability to use open data partly depends on the availability of open data technologies. The rise of the so-called "Internet of Things" and wearable technology drives innovation and as with other new and highly disruptive digital technologies, they challenge existing social, economic and policy norms.

Many are the example of governments across the world that change their policies based on data-driven initiatives. Iran seeks to develop an Internet free of Western influences or domestic dissent. The Australian government places restrictions on health data leaving the country. South Korea requires mapping data to be stored domestically. Vietnam insists on a local copy of all Vietnamese data. The nations of the world are erecting zones for data, undermining the possibility of global services. Data localization requirements threaten the major new advances in information technology — not only cloud computing, but also the promise of big data and the Internet of Things.

4.2.4 PAPER CLASSIFICATION

Figure 13 provides an overview of the paper classification, using a categorization of the articles according to evaluation criteria defined by Wieringa et al., (2005). The inventory includes all five categories: evaluation research, proposal of a solution, validation research, philosophical papers, opinion papers and personal experience papers. All in all, what we can extract from this information is that the data-driven innovation field is novel and most articles try to introduce the data-driven innovation method either by experiments, new ways of looking at things or even by empirical studies. Below we will demonstrate some examples that fit in each categorization.

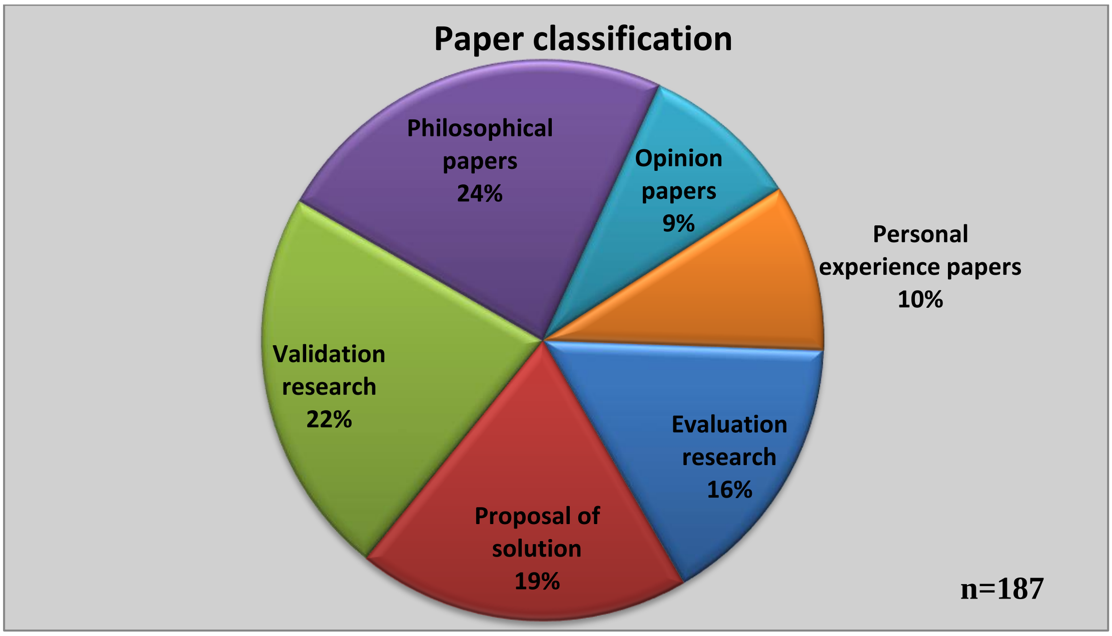

Figure 13Paper classification

Philosophical papers. As we can observe the majority of the papers is philosophical. These papers describe a new conceptual framework, implying a new way of looking at things can fit in this category. Jess Hemerly (2013) proposes a framework for policy discussions that examines three main areas of policy interest: privacy and security, ownership and transfer, and infrastructure and data civics. She discusses how regulation could preclude some of the economic and societal benefits of data-driven innovation. A new way of looking at Open Government Data (OGD) from an academic literature perspective is also being discussed. The authors developed a conceptual model portraying how data as a resource can be transformed to value. They conclude that openness is complemented with resource governance, capabilities in society and technical connectivity, use of OGD will stimulate the generation of economic and social value through four different mechanisms: Efficiency, Innovation, Transparency and Participation. Another article tries to define innovation opportunities in medicine by identifying ongoing and future trends. These trends include personalized medicine (i.e., genetically focused diagnosis and treatment), computerized decision support, quality-centric economics, patient empowerment, telemedicine, social networking and gaming, and medical nanotechnology. It is also mentioned as a source of future innovation, the large group of clinical users, which have the potential to become a valuable source of medical imaging and IT innovation. Furthermore, another paper is introducing a conceptual landmark and a discussion-opener for linking dynamic real-time data-based learning and selling by comparing testing data- vs. theory-driven approaches in e-selling. The results indicate that data-driven learning can be used to increase online sales yet the interpretative skills possessed by humans are still needed.

Validation research. As regards the validation research paper' main characteristic is that the proposed solution has not yet been implemented in practice. Possible research methods are experiments, mathematical analysis, simulation, etc. The "learning while doing" approach was implemented in Human Services as an experiment. The approach relies on data and on adjusting the implementation plan and the underlying ways of doing business. In the article it was described one agency's successful large-scale administrative, program, and practice change that resulted in better outcomes, stronger infrastructure as a learning organization, and a set of lessons with implications for change management in child welfare and other human services organizations. Moving on to another example, the BrightBrainer integrative cognitive rehabilitation system was tested in practice by conducting a clinical feasibility with nursing home-bound dementia patients. Interactive serious games were designed to improve basic and complex attention (concentration, short term memory, dual tasking), memory recall, executive functioning and emotional well-being of the patients. Statistical analysis is also being used to investigate the dynamics of the radiation belt Phase Space Density (PSD) and its correlation with the plasmapause location.

Proposal of a solution. In the third place we see the proposal of a solution paper where solution techniques are investigated. A proof-of-concept may be offered by means of a small example or a sound argument. Arguments have been raised for the correlation of innovation and engineering or service domains. For that reason a requirements-driven approach to innovation is proposed. The proposed approach is implemented as an example to a Living Innovation Laboratory. Furthermore, another article explores the smart city concept and proposes a strategy development model for the implementation of IoT (Internet of Things) systems in a smart city example. A novel approach to modeling the performance of individual wind turbines as well as their collection (a wind farm) is also being discussed in one of the articles and the proposed solution can be scaled down to optimize residential wind turbines.

Evaluation research. In this category we see the investigation of a problem in RE practice or an implementation of a RE technique in an empirical way such as by case study, survey, etc. The Intervention Design and Development model posited by Rothman and others was applied to a staff development initiative which provided evidence that an interstitial system composed of academics and agency personnel can work together in improving the quality of practice while simultaneously meeting the separate needs of each constituency. Another article focused on how private companies can use open data in order to innovate. The research investigates how certain resources and moderating variables can influence the open data use by the companies. The research used four research methods in order to collect the necessary data: Literature review, open data scenario, interviews, and survey. While companies make use of these data for marketing purposes and governments also discuss their application to social systems, individuals not only find them difficult to access, but also do not have any means of benefiting from their own data, even though they generate the data themselves. So the authors of this article focused on the results of a questionnaire survey on handling buying information that was conducted to investigate the resistance that users feel to allowing companies to use the data.

Personal experience papers. In these papers the emphasis is on what and not on why and most importantly it is based on author’s personal experience. For example, the Laboratory of Computational Physiology at the Massachusetts Institute of Technology organized the

Critical Data Conference. The article contained a list of lessons learned by the authors from their experience. The experience of the meeting at Cornell's School of Hotel Innovation was described on another article. A group of service researchers and practitioners gathered to examine the latest concepts in service, with a goal of sharing innovative ideas and processes, and expanding a culture of innovation in the hospitality industry. Based on their experiences, the authors of another article suggest that opportunities for partnerships between academics and community are enhanced when faculty members are already engaged in community-based activities.

Opinion papers. These papers contain the author's opinion about what is wrong or good about something. In an article the author mentions the compelling benefits of IoT to consumers, companies, and on national level. On the other hand, he mentions the privacy and security challenges associated with the Internet of Things and how a one-size-fits-all regulation will limit innovative opportunities. In another case, we see the author's opinion about the recent evolution of the British Library's collection development strategy and policies. He discusses how we are changing what we collect in core areas such as commercial publishing, new web-based media and heritage collections and he explains how these changes are shaping our thinking about the future of services.

4.3 SEMI-STRUCTURED INTERVIEWS

Table 2 presents an overview of the characteristics of the interviewees. Due to privacy reasons and ethical considerations we anonymized our data and will identify the described cases by interviewees with the letters A-H. A total of 8 interviews have been conducted, transcribed and analyzed. The eight interviewees included programme and policy officers, digital strategists, data and business development analysts, founders. All the interviewees are currently working for NGOs in the Netherlands. The participants that have been interviewed have experience ranging between six months and six years, with an average of 2.87 years. All the interviewees have done a comprehensive amount of projects within the field of data management. All interviews lasted between 20 and 50 minutes and were conducted by one interviewer and transcripts of the interviews were made. The interviews allowed us to gain an overview of the data-related challenges that NGOs are facing and of course to determine whether NGOs can benefit from new possible data-driven innovation methods applied by other business sectors.

Table 2 Characteristics of Interviewees

Interview	Years of experience	Interviewees and their roles
A	0.5	GIS and IM support
B	2.5	Business Development
C	5	Head of Data Innovation team
D	2	Digital Strategist
E	6	Founder and Director
F	3	Policy Officer
G	1	Program Officer
H	3	Senior Associate

4.3.1 DATA SOURCES

The inventory of the transcriptions (Appendix IV) indicates that a variety of data sources are used as information by the NGOs (*Figure 14*). Open data/Public datasets were mentioned by all the interviewees and especially they were mentioned in the transcripts in 27 instances. NGOs receive even in a daily basis governmental data, community data and exchanged data. All this kind of data is open data publicly available to use and republish. This main data source along with all the others indicates new trends and societal challenges and help NGOs to benefit their programs and to decide whether they would like to be involved or not.

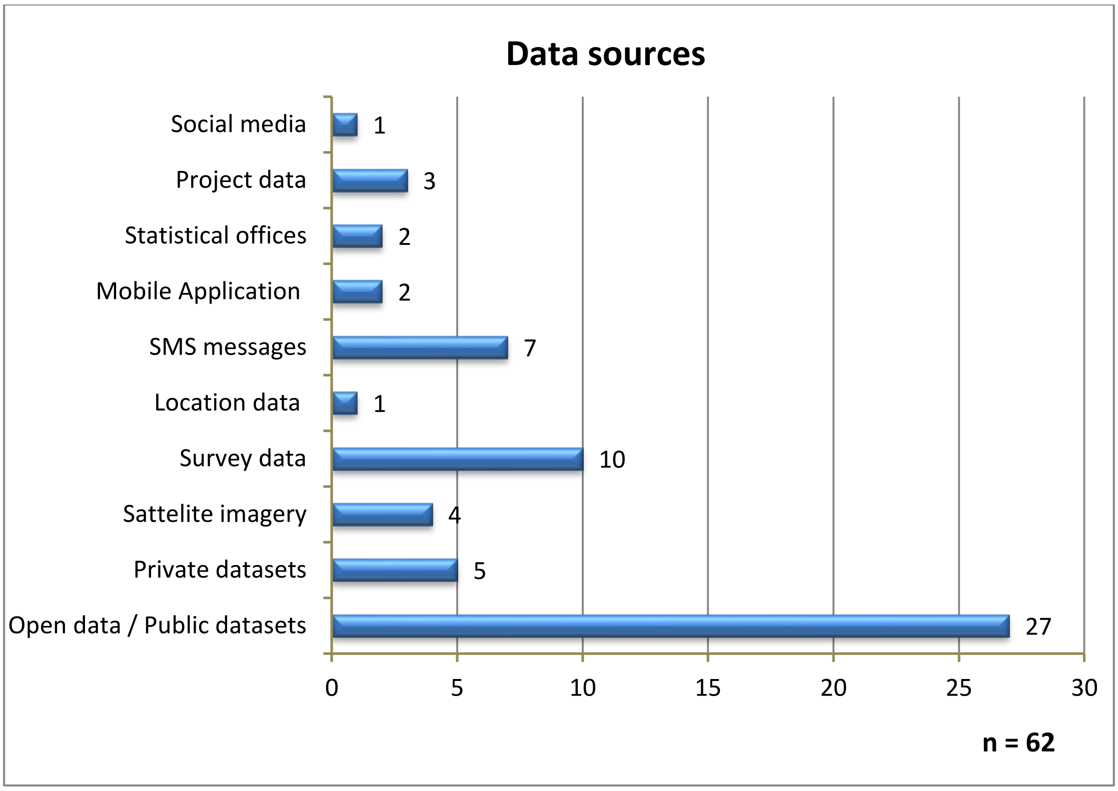

Figure 14 Data sources

Survey data and SMS data continue to play an important role. 62.5% of the interviewees mentioned Survey data as a key a data source. They mostly collect this kind of data from field surveys by people in ground. Survey data may include interview or observation of people in their natural environments to learn their everyday problems or emergency situation. Two interviewees mentioned SMS data as a main data source that they based on for policy making and get a better understanding of the actors on a local situation. This type of data was generated by citizens using SMS and they were definitely not suitable for Open data format.

Satellite imagery data (obtained from data brokers such as Microsoft Bing and Dataglobal). The main issue with this kind of data is that the resolution may not be high enough. Social media data (such as Facebook, Twitter) can provide different insights in trends mostly. Three interviewees mentioned project data that includes financial and project related information. This type of data is at the beginning private because it entails private organizational information such as financial statements and resource allocation but at the end these NGOs make this information publicly available to ensure transparency, accountability and flexibility. This whole process was initiated by the Dutch Government and the project is called IATI (International Aid Transparency Initiative). The IATI platform allows people to choose specific sets of aids activities to view in depth. IATI brings together donor and recipient countries, civil society organizations, and other experts in aid information who are committed to working together to increase the transparency and openness of aid. IATI is based on a standard format and framework for publishing data on development cooperation activities, intended to be used by all organizations in development, including government donors, private sector organizations, and national and international NGOs. Organizations implement IATI by publishing their aid information in IATI's agreed electronic format (XML) – usually on their website – before linking it to the IATI Registry. Other data sources used include statistical offices, mobile applications (such as WhatsApp) and location data. Especially for location data, the interviewee emphasized the need for trends rather than the exact numbers due to privacy issues.

4.3.2 DATA FORMATS, DATA INTEROPERABILITY AND DATA LINKING

Across the 8 interviewees, three data formats were mentioned more than 5 times:

- Microsoft Excel,
- CSV (Comma Separated Values standard file format),
- XML (Extensible Markup Language)

Other data formats mentioned are:

- SQL data format and programming language (Structured Query Language),
- GIS database format (Geographic Information Systems),
- JSON (JavaScript Object Notation),
- PDF (Portable Document Format),
- Microsoft Word but also plain text (that allows text mining)

In terms of different levels of data openness, as developed by Berners-Lee (2006), the inventory provides a positive picture. Still there are opportunities to move from data formats such as PDF, Excel and Word (with limitations in terms of data processing and linking) to formats such as CSV and XML that can be processed with different types of software and databases. CSV text-based files are containing nothing but data values with commas separating them and they can be read by any computer system. All interviewees mentioned Excel format as one the main dataset format they are working with. A couple of interviewees mentioned that ideally they would like to work only with PDF formats and especially with narrative data. On the other hand, some others working with maps indicated that PDF maps are not suitable because they cannot show the changes as quickly as happening on the ground. We can clearly understand that the interviewees are dealing with many and different formats of data as part of their daily work. Many of them encountered issues with linking and analyzing all these different types of formats and made it clear that a standardized type of format will save them a lot of time and effort.

Based on the transcriptions we can see that combining and linking different data sources has become one of the main characteristics of innovative data-driven approach for the NGOs. The majority of them are already involved in projects that enhance data interoperability, standardization and data linking. Others are considering this kind of approach for the upcoming future.

Challenges in data interoperability and data linking are related to data quality, privacy issues, IT affinity and expertise of the people handling the data, governmental decisions on data exchange, different institutional regulations and so on. Many are the factors that can affect interoperability and linking of data. Initiatives such as the IATI project mentioned by three interviewees showed that the NGOs involved are willing to adopt this standardized way of handling data to such an extent that they will replace the existing data formats in order to be IATI compliant.

4.3.3 DATA ANALYSIS TOOLS

In this category, we found a variety of approaches, all the way from simple desktop base analysis such as Excel or programming languages for statistical computing such as R and Python or machine learning algorithms and data mining tools (*Figure 15*).

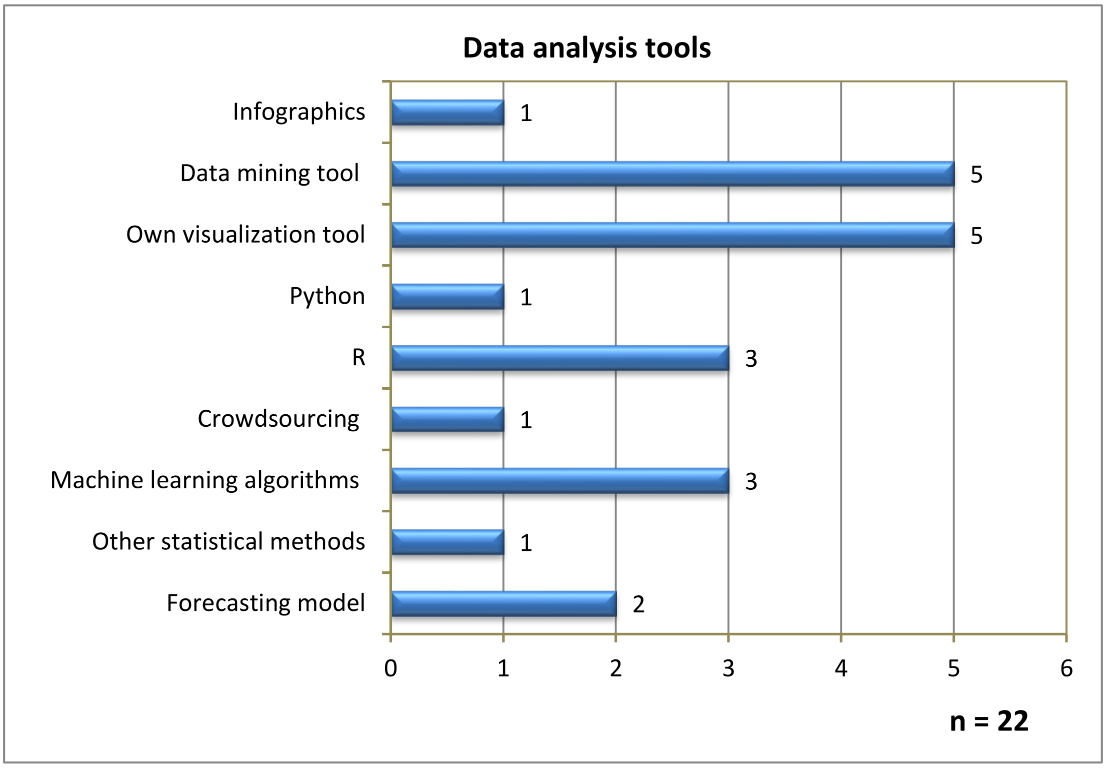

Figure 15 Data analysis

Five interviewees mentioned that within the NGOs they have developed their own tool for data analysis and visualization. They all feel comfortable using their tools and they are much more satisfied than they were using other outsourced software. Three interviewees mentioned 5 instances in the transcripts that they are using data mining tools (such as Tableau, Qlik Sense and Microsoft Business Intelligence, Power BI). According to one interviewee the data analysis tool is not the problem. All these tools are very powerful; the availability of the data is a much bigger problem.

In terms of data analytic approaches we can identify simple descriptive statistics, trend analysis, advanced statistical analysis and predictive modeling. One interviewee mentioned the urgent need for a standardized and unified way to do data analysis. We can clearly see from the interviews that the majority of NGOs are using more than one ways to analyze their data. That can be related to the variety of data sources or the small need to adopt a unified

IT architecture. Much of the data is sourced from Open data sets and statistical offices. These are updated anything from real-time to annual or less.

4.3.4 VISUALIZATION TOOLS

Some of the more common visualization tools that were mentioned during the interviews are for displaying information geographically, for instance: specialist GIS tools including Mapbox and Carto DB or more general purpose tools such as Google Maps. Since many of these initiatives are bounded by political geography, the choice to include mapping information spatially makes sense: for example, calculating the distance from a person's house to the nearest school or health care facility. There are also visualizations that have to do with visualizing the patterns in a wide variety of data: for example using public datasets and machine learning algorithms to develop a model that will inform of the number of refugees that are arriving in the next day to each country.

Interestingly, two interviewees mentioned that they do not currently have any analysis or visualization tool. However, the majority of the interview participants are using interactive visualization internally to decide where to deliver their aid. At this point, 6 interviewees mentioned the urgent need to upgrade the way they visualize the data and make data visualizations openly available in order to better communicate what they are doing. As one interviewee said "How can we make things interactive? How can we make things up-to-date with what other people are presenting?" It is likely that they all have data in bar and graph charts. The skills needed to operate Excel are relatively low, but for other programs some training or self –taught skills may be required.

One interviewee mentioned that data visualization can help NGOs to evaluate their projects before and after. A good data visualization tool can help a lot of NGOs with their Monitoring and Evaluation approach based on data. Of course, according to the same interviewee, NGOs should embrace young people with affinity to technology; people that would have broad horizons and an innovative spirit.

It can be noted that although many examples of innovative visualizations exist in the academic literature (Börner & Scharnhorst, 2009), in NGOs there is perhaps more of a need to make visualizations accessible on the web or usable, which may produce innovation more of in terms of best practice rather than advanced techniques.

4.3.5 PERCEIVED CHALLENGES IN PRACTICE

Data governance covers aspects such as data quality, data management, metadata management, access rights, decision rights, accountability and data policies (Eckartz et al., 2014). Data governance literature shows that barriers to data sharing differ when considering open data, which is a form of data sharing by public organizations with private organizations, and data sharing in between private organizations (Weber et al., 2009).

From a data governance perspective, the most dominant open data barriers are found to focus on data quality. Data quality is specified in more detail by Batini et al. (2006). Domain specific metadata describing the data origin, the data production date, data provenance, and for which applications the data can be used is of crucial importance. Data quality aspects that should be considered with respect to the entire data set are: accessibility, data format, semantics, conciseness, completeness, believability and reputation (Knight & Burn, 2005). Data quality aspects that should be considered with respect to data elements are: validity, completeness, consistency, uniqueness, timeliness, accuracy and preciseness (Nousak & Phelps, 2002). *Table 3* summarizes the biggest data quality challenges as they appeared in the transcripts (Appendix IV).

Table 3 Data quality challenges

Data Quality Challenges	Definition
Accessibility	Data is not easily accessible or understandable
Accuracy	Data is incorrect or out of date
Completeness	Data is missing or is unusable
Conformity	Data is stored in a non-standard format
Consistency	Data values give conflicting information
Duplicates	Data records or attributes are repeated
Integrity	Data is missing or not referenced
Timeliness	Information recorded and made available to systems is not as rapidly as is required

Other technical barriers to re-using open data are the publishing of data in a format that is not machine readable, the lack of an Application Programming Interface (API), difficulties to processing data sets, the lack of a linking or combining functionality, and difficulties in configuring data transformation (Zuiderwijk et al., 2014; Janssen & Zuiderwijk, 2012; Barry & Bannister, 2013).

Besides technical barriers, Zuiderwijk et al. (2014), Janssen et al. (2012), and Barry et al. (2013) analyze potential barriers to publishing open data according to various perspectives: political, social, economic, institutional, operational, and legal. Political barriers include a lack of support, a lack of attention and a lack of knowledge about open data. Among the social barriers are a lack of interaction with users, difficulty to measure impact, cultural differences and risks and liability with respect to providing low data quality. The lack of business models is a main economic barrier to open data. Institutional barriers include a lack of standards, a lack of an open data policy, an inability to handle user requirements and a lack of guidelines. Data fragmentation, a lack of services, a lack of metadata, changing or a lack of clear semantics, and a lack of information on data quality are among the operational barriers. And the legal barriers include licensing, policy differences, lack of (detailed) policy. Eckartz et al. (2014) identify five main categories of barriers to data sharing: technical, data quality, ownership, privacy and economic.

After transcribing and coding all interviews we identified 41 exclusive thematic codes and 6 sub-codes related to Data Governance and Open Data barriers that NGOs are struggling with. We organized the identified challenges into eight categories as presented in Figure 16: Data quality, Technical, Institutional, Social, Privacy, Economic, Operational and Ownership. The coded themes were mentioned in the transcripts in 174 instances of which 49 were related to Data quality, 29% were related to conformity, 25% were related to accessibility, 14% were related to accuracy, 10% were related to integrity, 8% were related to consistency, 6% were related to completeness and 4% were related to duplicates and timeliness.

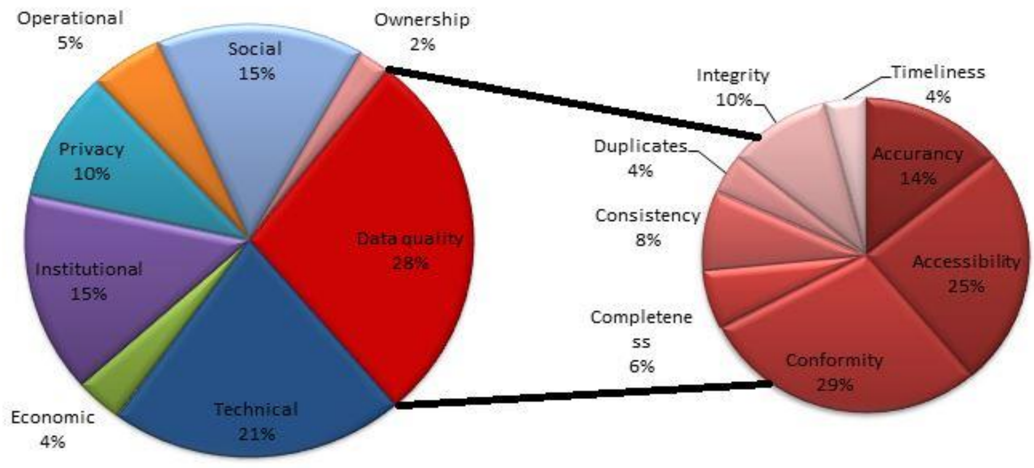

Figure 16 Open Data Barriers

The technical barriers were mostly related to the publishing of data in a format that is not machine-readable. As one interviewee mentioned it takes too long to “clean” the data and prepare it for analysis. The lack of an API causes difficulties to processing different data sets. Also, quite many interviewees mentioned problems with configuring data transformation and especially the lack of a linking or combining functionality.

Regarding the institutional challenges, the majority of them were related to the lack of knowledge and capacity to innovate. As one interviewee mentioned innovation can be a buzzword to people who are not familiar with technology. Those people and therefore organizations might want to innovate but they do not know how or where. A big challenge for NGOs is the fact that they are humanitarian aid-driven and that restraints them from setting long term goals. So, the change in IT infrastructure for becoming more innovative takes longer than the actual goal to help communities around the world.

The social challenges were mostly related to NGOs’ aid communication. NGOs are based on donors (government, people, organizations, etc.) to continue working and delivering their aid. Therefore, losing the trust of their donors would be a major disadvantage. Moreover, many interviewees mentioned the lack of people working with data in NGOs. The majority of the interviewees were part of small teams (5-6 people) handling large scale and diverse data sets coming for countries all around the world even in a daily basis.

All interviewees raised their concerns for privacy challenges. Issues regarding data disclosure and different countries' regulations applied each time were mostly mentioned. NGOs are handling super sensitive data including information for criminalized people in their countries, sex workers, LGBT communities. As two interviewees mentioned they are dealing with ethical issues when it comes to publishing their own data because at some cases they might do more harm than good.

We found also challenges related to small budget allocation for data-driven innovation projects. As one interviewee mentioned it is hard for them to justify spending a significant portion of funds or staff on possibilities that might not work out, especially when the core purpose of their organization is to help people in crisis. NGOs rely heavily on donors and without them they could not do anything. People donate with the expectation of providing critical support, not for a development project. Ownership challenges were also mentioned, specifically for governmental permission and censorship of data in some cases. Furthermore, we found challenges related to lack of information on data quality and projects communication. As two interviewees mentioned there is a lot of misunderstanding of what innovation is and it is one thing to be innovative and another to do something with innovation.

4.3.6 COMPETITION AMONG NGOs

As our interview guide has been tuned to the collection of data related challenges in practice by NGOs, the question of competition for identifying and implementing innovation emerged directly from the interviews. The findings reported here have been mentioned by the interviewees and identified during the coding process. More than half of the interviewees acknowledged healthy competition for identifying or implementing data-driven innovation solutions. As one interviewee mentioned the donors are setting up a competitive system by calling for proposals. NGOs have to compete in order to receive the funds to be able to implement their projects. As a side effect of this competition some NGOs might overlook the real needs of the people and fall in the trap where they would push a specific technology in order to solve communities' problems. Whilst, they should focus on people's needs and then find the technology that would assist solving their problems. As three interviewees mentioned the healthy competition has less to do with not allowing others to do the same, on the contrary it has more to do with finding the drivers to improve your field. Four interviewees have raised the matter of collaboration between NGOs. The main reasons for collaboration are data exchange, increase humanitarian aid capacity, transparency and flexibility.

4.3.7 NGOs AND BENCHMARK ANALYSIS

After transcribing and coding all interviews we identified five industry sectors that NGOs are following to better manage their data in innovative ways (*Figure 17*).

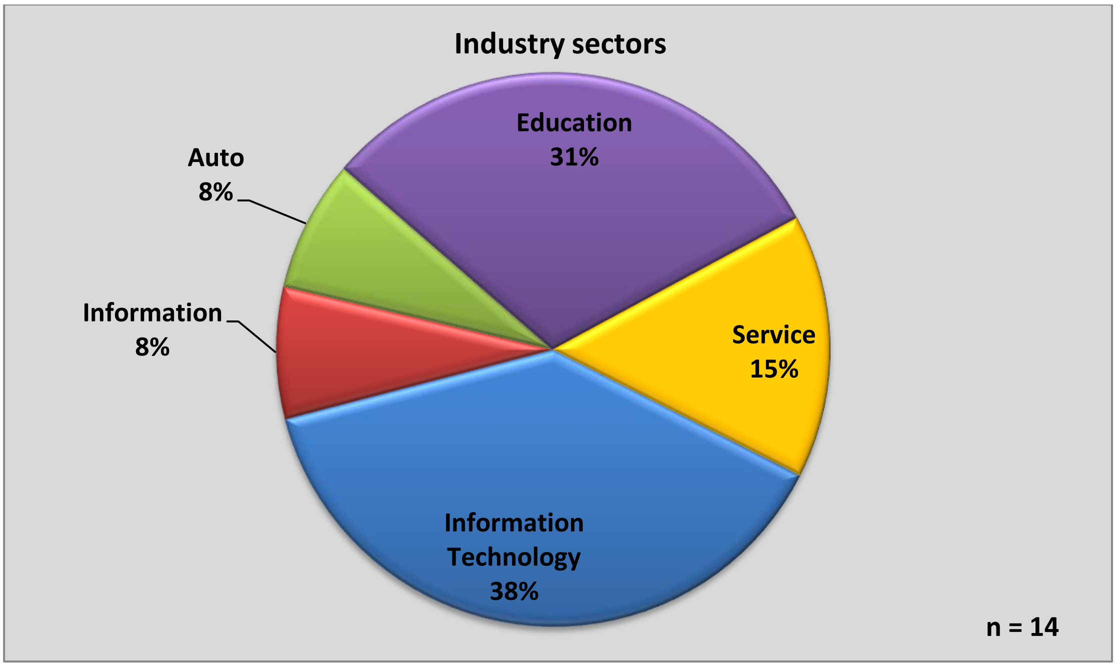

Figure 17 NGOs and benchmark analysis

38% of the interviewees mentioned that they checked companies from the Information Technology sector in order to find new tools and technology that would help them improve their data analysis, data collection and data integration process. One interviewee mentioned that they are mostly looking for new technologies that do not require a lot of expensive hardware and that they prefer freely available software applications. Two other interviewees mentioned that they check how other organizations are using their data and therefore evaluate whether to develop their own tool or not. Three interviewees mentioned that they have established cooperation with either IT, or Service or Education business sectors in order to use their experience to drive data innovation. 8% mentioned that they are looking for innovative ways in information management. One interviewee mentioned Auto industry as good example to drive innovation. According to this interviewee, only big organizations can lead in innovations that would bring great changes. Therefore, NGOs are not capable of such big innovative changes.

5 ANALYSIS AND DISCUSSION

In this section the gained insights regarding the Systematic Literature Review and the semi-structured interviews will be discussed.

In section 5.1 the definition of data-driven innovation as concluded from the SLR and the interviews is presented. 5.2 present the challenges and solutions for the NGOs regarding data use.

5.1 DEFINITION OF DATA-DRIVEN INNOVATION

After the SLR and the interviews' transcriptions we tried to define what data-driven innovation is. We will present the data-driven definitions as they concluded from various research papers (SLR) and from the interviewees' opinions. The *Tables 4 and 5* below are based on Kettunen (2009) and Laanti et al., (2013) method. A third column is added, explaining where the emphasis is in the corresponding definitions. After the tables our proposal for the definition is following.

5.1.1 DATA-DRIVEN INNOVATION DEFINITION FROM SLR

Table 4 Definitions of data-driven innovation from SLR [adapted from Kettunen (2009) and Laanti et al., (2013)]

Source	**Definition of data-driven innovation**	**Emphasis on the corresponding definition**
Iyer et al., (2005)	New knowledge or valuable innovative ideas are embedded somewhere in the data.	Knowledge, new ways of data use.
Kusiak & Tang (2006)	Interpretation of information by analyzing data. A dynamic process.	Data analysis, knowledge, dynamic process.
Kitchin (2014)	The use of guided knowledge discovery techniques to identify potential questions worthy for further examination.	Knowledge, methods, technology, question and problem identification.
Zuiderwijk et al., (2014)	Increase efficiency and effectiveness by better decisions.	Better decisions, improve efficiency and improve effectiveness.
Krause (2013)	Directly connected to invention and technological progress.	Technological progress, invention, change.

Hemerly (2013)	Finding new ways to manage data and improve processes in order to make decisions and derive insights to problems. Using both real-time and historical data to inform decisions in the present. Today's computational capabilities enable better solutions.	Better data management, improved processes, decision making, new insights, problem solving, present decisions, technological capabilities, better solutions.
Eckartz et al., (2014)	A potential for the development of innovative services that have economic value and address societal challenges.	Innovative services, economic value, societal challenges.
Al-Khouri (2014)	New opportunities for transforming decision-making, discovering new insights, optimizing businesses and creating new technologies.	Change, decision making, new insights, improve processes, improves businesses, new technologies.

Based on Table 4 we formulated the data-driven innovation definition below:

"Data-driven innovation is about finding new ways to use data to *make decisions*, improve processes and create methods and technologies in order to *solve problems*."

Making decisions

For most companies, data-driven decision making improves output and productivity, which leads to increased revenue (Caroleo et al., 2015). Google experts in data-driven innovation state that "data for decision making means using both real-time and historical data to inform decisions in the present" (Hemerly, 2013). In general, the Data-Driven Decision Making (D3M) is considered in the literature as a "style" of decision making that heavily involves data in the decision process to achieve stakeholder goals. A deep review about D3M reveals that its domains of application include business, medicine, education and government (Marsh et al., 2006; McAfee & Brynjolfsson, 2012; Sackett, 1997).

Solving problems

The concept of using data for more informed D3M relies on the bigger "container" of data-driven innovation (U.S. Chamber of Commerce Foundation, 2014), which accounts both for making decisions and improving efficiency. Literature unanimously agrees in considering data of great value for decision making as a catalyst of innovation and scholars often tried to quantify such values, starting from the evidence of a positive relationship between data, economic growth and societal challenges. A fundamental aspect of data-driven innovation is the ability to find new insights by analyzing existing data and combining them with other data. Hybrid decision making helps us put insights from data into action and illustrates how "data sight" helps us see things we could not see before. In combination with our other decision-making skills and tools, data sight improves a problem's granularity, depth, and time horizon in extraordinary ways (Hemerly, 2013).

5.1.2 DATA-DRIVEN INNOVATION FROM INTERVIEWS

Table 5 Definitions of data-driven innovation from interviews [adapted from Kettunen (2009) and Laanti et al., (2013)]

Source	**Definition of data-driven innovation**	**Emphasis on the corresponding definition**
Interviewee A	A creative solution, something new that solves a problem. A process to come up with a solution.	Problem solution, change, new insight.
Interviewee B	Make decisions based on better data.	Decision making, better data.
Interviewee C	Right information in order to make decisions on how changes can happen. Decisions based on aid delivery.	Decision making, changes happen, aid delivery.
Interviewee D	Better decision making on aid delivery and work communication.	Decision making, aid delivery, work communication.
Interviewee E	Different data source collection in order to show change and new insights of progress.	Change, new insights, progress.
Interviewee F	Better decision making. Better aid delivery.	Decision making, aid delivery.
Interviewee G	Better decisions for change aiming to aid –delivery	Decision making, change, aid delivery, flexibility.

	adaption. Flexibility is important.	
Interviewee H	Processes and tools improvements for aid-delivery. New insights and trends.	Aid delivery, change, technological progress.

Based on Table 5 we formulated the data-driven innovation definition from the NGOs' perspective below.

"Data-driven innovation is about making decisions based on data in order *change to happen* on how to *deliver the aid*."

Change happening

Data holds an important role for NGOs in informing the work that they do. By making use of data that is around the world, NGOs can get a more deal-time idea of what is going on and be able to adapt and shift their programs based on the data that they are gathering. For NGOs flexibility is the key for better policy making.

Aid delivery

NGOs are gathering data from a lot of different sources around the world. All this kind of information assists NGOs to see where they have to be working, in which regions and what kind of interventions they need to carry out. They are using data to get more insights and indications of where to focus and to deliver their aid. NGOs rely heavily on their donors and therefore they are using data to enhance their projects' transparency and accountability and thus be able to better communicate their work.

We can see that there is a slight difference in both definitions. The interviewees form NGOs gave their perception of data-driven innovation from a humanitarian-aid perspective. The data they are handling is based on human's sensitive information and so the concern of using this data, with a critical mind and ethical responsibility in order to make decisions, has been raised.

5.2 CHALLENGES AND PROPOSED SOLUTIONS OF DATA USE FOR NGOs

Technology has been and will continue to be a fundamental driver of the data revolution. To harness the benefits of new technology, large and continuing investments in innovation are required at all levels, but especially in those institutions which are currently lagging behind. As one interviewee mentioned “It is always good to try to improve your field of work. Why would you want to stay behind? In the commercial field they are already so much further than NGOs are.” It is crucial for NGOs to identify critical research gaps, such as the relationships between data, incentives and behavior. One good example would be to engage research centres, innovators and governments in the development of publicly available data analytics tools and algorithms to better capture and evaluate the trends affecting sustainable development.

5.2.1 LACK OF DATA LITERACY

A data-driven economy will demand highly skilled workers well versed in data analytics, interoperability, information science, metadata and data visualization (Hemerly, 2013). If NGOs want to prosper and continue to be part of this new era, they should be no exception. “Sometimes NGOs do not understand what new media or data-driven innovation can offer. I think is important for NGOs –who are getting funded by governments to embrace young people with affinity to technology. It is one thing to be innovative and another thing to do something with innovation” said one interviewee.

The nonprofit DataKind (http://datakind.org) has had success by putting public data creators, researchers, and developers in the same room together to use data. Hackathons, visualization competitions and crowdsourced data predictions are also engaging the public by making data analysis and interpretation a game (Hemerly, 2013). It is surprising that only one out of eight interviewees’ teams was composed of people with different qualities. Some of them were data analysts, some others were working with visualization or programming. According to this interviewee this was their method to work with different expertise on data problems. On the contrary, in another NGO - where they did not have a structured way of gathering, using and analyzing data, they had a great problem in proving the work they were doing because they were missing data.

5.2.2 DATA QUALITY AND TECHNICAL CHALLENGES AND SOLUTIONS

NGOs must often handle multiple data even when handling local phenomenon. When the scale becomes larger (such as with the Refugee crisis project), the challenge is to find the technical capacity to scale up its representation in response. Data quality and technical issues are the first to appear. As one interviewee said "I am spending to long cleaning my data. I am collecting data from twenty different countries who are all reporting what they are doing in slightly different measurement and way- this is the way organizations work. That could change if everyone is reporting data in the exact same way". The technical format of a data set may be a constraint to open the data (Hofman & Bastiaansen, 2013). If the data is unstructured it may be difficult to convert it into a machine-readable format relevant to a data user. The size of the data set, the existence of a semantic model, and identifiers are other technical issues that need to be considered. For this group of issues, many interventions are possible. Examples include: offering the data in a structured format; reusing existing vocabularies and ontologies; publish the data according to existing data standards.

The metadata (attached information which can be used to help to understand the structure, features and limitations of a data set) is often missing or incomplete in the case of big data, given that it tends to come from multiple sources (Taylor et al., 2014). One example is to explore whether the data set can be extended with other data to improve completeness. In any case metadata describing the data quality should be added (Miller et al., 2008; Barry & Bannister, 2013). To allow for re-use of data by others, as much context information as possible about data should be provided. Furthermore, social interaction with the data should be supported: data is often most used and most easy to interpret when a community can be built around the data platform where the data is published. Data visualization can play an important role in this (Eckartz, 2014). Thus the inaccuracy of data used can lead to using data without fully understanding it, or discounting it, which may itself lead to bias.

The majority of the interviewees were handling publicly available data while two interviewees were mostly dealing with SMS generated data. None of the participants mentioned data generated through the use of social media (known as social data). However, one interviewee mentioned the benefits of social data (Facebook, Twitter) and the fact that they provide different insights in trends mostly. "I think that is a real opportunity for NGOs to work with this kind of data. This is something that commercial organizations have been working with for 10 years already. I think is time for NGOs to catch up on that.”, the interviewee said. Indeed, social data are playing an increasingly important role in academic, development and humanitarian research and practice due to its ability to provide high-resolution, dynamic (in terms of time, space and coverage) data sources and methods of analysis (Taylor et al., 2014). The use of new digital technologies such as mobile phones and internet-based search, communications and transactions mean that in the fields of economic and development we are seeing an emerging process. On the other hand, mobile phone data can provide predictive capacity and real-time information, as one interviewee described for a project in Uganda that was focusing on public engagement and civil society. According to this project several radio stations were set up in Uganda and citizens had the opportunity to express their opinions through opinion polls for their real-life problems. The extracted results were delivered to the politicians to take actions.

From the SLR we conducted we found organizations that are looking to the observation of social media as another powerful and insightful data source for gauging problems and issues as identified by ordinary people. The MDGs (Millennium Development Goals) constitute a roadmap to halve poverty worldwide focusing on poverty, health and environment. Global Pulse, a project of the United Nations, is working to supplement this process with big data and visual analytics, showing how people around the world are debating the topics. By searching approximately 500 million new posts on Twitter every day for 25,000 keywords relevant to 16 global development topics, the project’s dashboard shows which different countries talk the most about given topic (Taylor et al., 2014).

An alternative data source that we found from the SLR is observed data such as Call Detail Records (CDRs) from mobile phone network operators. In relation to using mobile data for research and activism, it is important to find out whether the phone users in the data set are representative of the population. There are several approaches to deal with this problem, using survey data in combination with the CDR. Once the level of correlation between data set and population is known, the mobile data set can then be used as s proxy for population in other models, such as of transport or disease vectors (Eagle & Greene, 2014). Another option might be to use multiple sources to validate a CDR: researchers can layer different data sets (some of which may be open data) and use modeling techniques to estimate how accurate each data set is. It may be possible to add survey questions on top of population data to sharpen mobile data’s accuracy around a specific question. Various firms provide a service in surveying mobile users, one being Jana (www.jana.com), where users sign up and are paid to answer questions. This type of CDR analysis is done for population movements after natural disasters (e.g., in Haiti) and to model the spread of a disease (e.g., cholera; Bengtsson et al., 2011). One possible next evolution is for researchers to combine big data signals (such as real-time mass movement of populations) with crowd-sourced data (e.g., SMS-based surveys) to ask for direct feedback on crisis response from affected communities (Taylor et al., 2014).

5.2.3 INFORMATION EXCHANGE IN NGOs

Big data on human activities stems predominantly from digital communications tools such as Facebook, Twitter and other platforms), and therefore a large part of its utility for objectives of social change is also related to its origins in communication and information exchange. Ways that digital data can be used to facilitate information exchange currently focus around the use of particular platforms to connect citizens and service users with organizations who are representing their needs or conducting interventions. The use of SMS is becoming especially prevalent in developing countries as a mechanism to invite feedback on service delivery, mainly by civil society organizations to support their advocacy strategies, and platforms developed by humanitarian-focused organizations such as Ushahidi focus on SMS in particular as a tool to invite reporting on specific issues. Beyond this, increasingly interactive platforms are also emerging, designed to promote a continuing process of dialogue and feedback between information providers and service users.

Grameen Foundation's work is one example we found in SLR about the way large datasets can be merged and linked around a base layer of data stemming from mobile phones. Grameen Bank's work on microfinance for the poor and unbanked is the largest such enterprise in Asia, and now extends into much of Africa. The related Grameen Foundation has set up an AppLab in Kampala to develop tools which can make use of the information stemming from its clients' microfinance transactions and related information flows. The AppLab processes and analyses the large datasets stemming from two mobile-based information-sharing applications: 'CKW Search' for information on livestock, crops, and weather; and 'Pulse messaging' – a mobile app which generates popup survey questions and then syncs clients' questions and answers over the web. Through these apps Grameen Foundation is able to check clients' farms' performance and evaluate its agricultural program. Using these linked datasets and ArcGIS or QuantumGIS (geographical information systems software) to do in-depth analysis, the AppLab can track the quantity and type of requests by region and thus identify local problems. The survey app, Pulse, has also led to data-driven discoveries such as the realization that while those involved in active farming are generally women, men are travelling to markets and selling the products. One interviewee also mentioned a similar project that facilitates information exchange by the use of SMS as a means to invite feedback on a specific crop in Uganda. The interviewee explained that they wanted to see if people are growing or eating this kind of nutritious plant. The ultimate goal was to promote that crop to be grown in different places so that more people can get more vitamin A in their diets.

5.2.3.1 CHALLENGES FOR INFORMATION EXCHANGE

Outcomes from data

Data quality and interpretation can be problematic, for two reasons. First, because sometimes inputs are bad – survey respondents may answer questions incorrectly. Second, the data is generated under locally specific conditions and this can make the data difficult to analyze remotely without local understanding. As one interviewee explained in some circumstances they collect geographical data from local people that they actually draw information on PDF maps. For a number of interviewees data are being collected from different organizations (such as regional or local government, or where crowdsourcing or volunteering is used), the main inclusion issue is that it is not clear how the completeness or the updating of the data is taking place, so that uneven or out-of-date coverage may emerge as main problem over longer term in making data sets less useful or usable.

Data storage and processing

The AppLab for example has collected four years of data, comprising a dataset that is now too large for conventional analytical tools such as Excel. Collaborating with analytics specialists (in their case, a technology company, which cooperates with them on a pro bono basis) raises questions of who are appropriate 'data handlers', for which new processes and standards have to be developed. Storage and processing issues may arise with sensitive data, particularly where it is backed up or stored with a third party. This problem is set to grow, given that the cloud is often the cheapest place for organizations to store and manage their data, but is also subject to jurisdictional issues if things go wrong and data ownership is challenged (Dannatt, 2012).

5.2.3.2 NEW TRENDS FOR INFORMATION EXCHANGE

Data sharing

Four interviewees mentioned that they are involved in data sharing initiatives between other NGOs and/or the government. The majority of them mentioned the importance of finding ways to better communicate what they are doing as NGOs and to strengthen the trust and faith of their donors. One initiative for data sharing according to our interviewees is the IATI (International Aid Transparency Initiative) reporting process. The Dutch ministry of foreign affairs enforced a new way for NGOs to report their projects and financial information in a transparent and open way. Accountability towards the government but also towards society is a big deal for NGOs.

Capacity for NGOs is a big deal as well. They should focus on building upon existing structures and publicize collaborative learning environments, for example consortia of universities or other NGOs. As one interviewee mentioned they do collaborate with another NGO with bigger capacity for humanitarian aid. According to Taylor et al., (2014) the key is to create programs and environments that can bring together data science capacity with area-specific knowledge. NGOs should encourage open tool development. Another good example is the UNOCHA initiative (United Nations Office for the Coordination of Humanitarian Affairs). They aim to create a standardized dataset for all humanitarian actors to use. UNOCHA will host all the different admin areas of the world in their servers. In that way, whenever a NGO is making visualization it can be freely available online for other NGOs to process. Two interviewees mentioned their consortia with a Dutch University for establishing a new platform that would include IATI data, government data, Open data and programmatic data.

Crowdsourcing

A prominent example of an organization using such an approach is Chequeado, an Argentine non-profit independent media outlet that specializes on fact-checking. It is the first initiative of this kind in all of Latin America and one of the four fact-checking organizations --from over 40 worldwide -- conducting some form of crowdsourcing. In 2014, Chequeado incorporated to this event the use of DatoCHQ, a public database and crowdsourcing platform through which its followers were able to participate in the live fact-checks by sending data via twitter, such as references, sources, facts, articles and questions. DatoCHQ was created as a response to the experience of 2013, where over 40,000 individuals followed the event and participated spontaneously sending information using Twitter. Chequeado is continuing to improve this tool and developing other crowdsourcing gadgets, such as its mobile app DatoDuro to receive and send relevant data in real time about ongoing debates, political events and speeches (http://datochq.chequeado.com/). There are two other fact-checking initiatives in Chile and Costa Rica, and several throughout the world including in South Africa, Egypt, Ukraine, Italy, in the European Union, among others. In the case of Chequeado, data is used in fact-checks, so something concrete and visible to users is produced with the information shared. The data is both processed and fed back to the user who submitted it. Data is made available to the public through DatoCHQ, which as well as being a tool for crowdsourcing operates as a public database. Users can share data or they can use the database as a source (university students writing papers, or journalists who sometimes struggle to access data). This is possible because Chequeado verifies all the data it makes available.

Another example of an organization which leverages large amounts of data to increase transparency and accountability is Ushahidi, a nonprofit technology development company which develops free and open source software for information collection, aggregation and visualization, providing users with guidance towards planning around deploying their software through toolkits. They provide an API (Application Programming Interface – a way to access and collect data from websites) with CSV (comma-separated values, a common format) and XML (Extensible Markup Language, a newer format for transporting and storing data) download options, and recently released a firehose (crisis.net) that moves data into a similar format for purposes of activism and advocacy.

5.2.4 DATA PROTECTION AND PRIVACY

As more data becomes available in disaggregated forms and data-silos become more integrated, privacy issues are increasingly a concern about what data is collected and how it is used. Clear international norms and robust national policy and legal frameworks need to be developed that will regulate data mining, transfer and dissemination. They should enable citizens to better understand and control their own data and protect data producers from demands of the governments, while still allowing for rich innovation in re-use of data for the public good. There is a case to be made for building a data commons for private/public data, and for setting up new and more appropriate ethical guidelines to deal with big data, since aggregating, linking and merging data present new kinds of privacy risk. In particular, organizations advocating for opening datasets must admit the limitations of anonymization, which is currently being ascribed more power to protect data subjects than it merits in the era of big data.

NGOs in particular are handling super sensitive personal data revealing racial or ethnic origin, political opinions, religious or other beliefs and data concerning health or sexual life among others. As all interviewees mentioned the data they use and republish should be anonymized and extra precautious meters should be taken under consideration for the data they reveal. Safeguarding the personal data of individuals, particularly in testing conditions, such as armed conflicts and other humanitarian emergencies, is an essential aspect of protecting people's lives, their physical and mental integrity, and their dignity – which makes it a matter of fundamental importance for the NGOs. Among the challenges the interviewees mentioned regarding privacy and data protection are data disclosure issues and different countries' regulations that affect the way NGOs are collecting and revealing data. For example, three interviewees mentioned that they had to follow specific exclusion policies in order to protect their data and the people that might be criminalized in their countries. There are many examples with governments that enforced laws and make it punishable for local NGOs to accept money from a foreign country. "The data fetishism is something to watch out because not everything should be published. We work with LGBT, sex workers, criminalized people and a lot of organizations that are doing great work. However, they do so hidden in the shadows of society. So, if you publish their names and activities you can actually do more harm than good. We are dealing with ethical issues when it comes to publishing our own data.", said one interviewee. Possible responses to these challenges according to Taylor et al., (2014) would be the use of templates evaluating data sources, for example, Rebecca McKinnon's Ranking Digital Rights project (rankingdigitalrights.org). Also, templates for user agreements for NGOs to standardize behaviors should be established. Data security is all about ethics, storage/infrastructure and regulation dimensions.

5.2.4.1 ETHICS AND LIMITATIONS

''Critics have raised concerns regarding the potential for re-identification of aggregated or anonymized data, even in the context of open data. The question becomes one of balance: Does the value of the innovation outweigh any risk of re-identification, and what harms could result to individuals? As we move to data-driven everything, we will see a widening gap between skilled data professionals who interpret and analyze data and the general public, who are unable to critically analyze claims and figures. Data can easily be misused and figures misconstrued—this is not new. Models have their vulnerabilities, and algorithms rely on assumptions. However, as the amount of data feeding the various models increases, patterns will begin to appear where they do not actually exist, just by virtue of mathematics. We must approach models with the knowledge that they are not perfect, and that machines cannot capture all variables. This ultimately means that professionals will not only be required to analyze data but also to check models, which will create an important space for social scientists to bring a critical eye, adding another layer of skill needed in cross-domain teams. Success will depend largely on the contextual relevance of models, access to data, ethical checks and balances, and the regulatory environment in which they operate.'' (Hemerly, 2013).

It surprisingly that only two interviewees touched the topic of data interpretation with a critical mind. As they both mentioned their organization is mostly humanitarian-aid driven and that brings great importance in the interpretation of data. "For example, if you have a particular goal that you want to support and you are using the numbers in such a way that it supports your goal. Is data neutral? One of the dangers of data-driven innovation is that people start trusting the numbers to such an extent that they stop being critical. Within the NGOs we should watch out of not being data obsessed", one of the interviewees explained. In that case a policy framework should focus on reasonable principles and best practices that are consistent with the rapid pace of technological evolution. Leadership within the teams handling data in the NGOs about how to think critically about existing standards that may not be adequate or contextual, should be taken also into consideration. Data-driven innovation has amazing potential for improving economies and societies, and a positive policy landscape can make sure this potential is realized in a responsible way.

Furthermore, three out of eight interviewees raised the limitation of institutional knowledge. Within the NGOs there is so much experience and so many people that have worked in this field for twenty years. There is knowledge that cannot be captured in numbers. "Our data is based on humans and that field is so dynamic- it changes every day", explained one interviewee. It is crucial for NGOs to act and do things right and ethically responsible.

6 CONCLUSION AND RECOMMENDATIONS

This research thesis focused on data-driven innovation in Non-governmental organizations (NGOs). The two-fold purpose of the study was from one hand identifying the gap in literature and on the other hand identifying the challenges in practice from the NGOs' perspective. The problem is the lack of literature in the field of data-driven innovation in NGOs. To address the question regarding the literature gap we conducted a Systematic Literature Review (SLR) for 187 articles. We identified quite many articles about data-driven innovation applied in fields like Information Technology, Education, Healthcare, Automobile Industry, etc. In order to identify the challenges and needs that NGO's are facing in practice, eight semi-structured interviews with people working for NGOs dealing with data (e.g. GIS analysts) were conducted. Approximately 250 minutes of interviews were recorded and over 12.000 words were transcribed. After the SLR and the interviews' transcriptions we tried to define what data-driven innovation is. We presented the data-driven definition as it concluded from various research papers (SLR) and from the interviewees' opinions.

The SLR findings verify the fact that very few research articles have been written for data-driven innovation in NGOs. The majority of them were about Information Technology, Health, Education, Energy and Public Administration sectors. We also categorized the articles according to the described type of innovation each time. Data-driven, business model, policy, product/service were among the highest percentages. According to Wieringa et al., (2005) we classified the research papers into six categories. Among first were validation research papers, philosophical papers and proposal of solution papers.

Traditional data analysis has focused on data manipulation rather than knowledge exploration from the data. Much valuable information is embedded in the data and potential benefits can be paramount for NGOs if the information is extracted and utilized. Constraints, trends and implications of the existing process are often in the data and any innovative idea or solution needs to be synchronized with these parameters. The situation in NGOs is somewhat more complex given the diversity of their philosophical underpinnings.

Following the perception of the interviewees, most NGOs are lacking a strategy and direction for the development of data-driven innovation initiatives. The majority of them have project monitoring, evaluation and they work with data but they do not really innovate with the data due to lack of data innovation capacity. The findings from the interviews indicated that the three top data sources for NGOs are Open data, survey data and SMS data. The most common data formats are Microsoft Excel, CSV and XML, although they are dealing with a vast amount of different data formats. NGOs' main challenges with regards to data use are Data Quality (such as conformity, accessibility and accuracy), Technical (such as lack of an API and difficulties in processing data sets) and Privacy (such as data sensitivity, data disclosure, ethical and critical view of data).

Based on the findings and discussion, the conclusion of this study is to identify how NGOs define data-driven innovation and whether they use data to innovate. We came to the conclusion that most NGOs from the humanitarian sector are still at an entry level as regards data innovation. The majority uses data for monitoring and evaluation purposes but they do not innovate with data. Main issues that they are dealing with are data standardization, data capacity, innovation capacity and IT infrastructure. The positive outcome although, is that many have engaged in projects that will enhance information exchange in a standardized way.

6.1 VALITIDY CONCERNS

Considering the fact that even though the interviews took approximately half an hour each, the semi-structured form of the interviews combined with anonymity of the participants obviously provides room for frank conversations.

Regarding external validity, the consideration should be made that the interviewees were all working with data within NGOs. Meaning while they did have excellent knowledge regarding the issues within the discussed NGO and all have a brought range of experience, they could possibly be slightly biased regarding the success of their projects.

Considering internal validity, the discussed cases could have been investigated more vigorously by including multiple viewpoints and more interviewees, while this does require more time and resources. However the current approach allowed this research, given the time constraints into account, to consider multiple cases and enables it to generalize the findings as a result.

6.2 RECOMMENDATIONS FOR PRACTICE

NGOs should adopt a data-driven innovation approach towards the way they are handling their data because "new knowledge or valuable innovative ideas are embedded somewhere in the data" (Iyer et al., 2005) and that will eventually help them to better communicate their work, enhance the trust of their donors and become more transparent, flexible and therefore accountable. Another topic that came along and interviewees expressed great interest is new trends in information exchange. Data sharing initiatives for collaboration between NGOs could assist in that direction. Furthermore, NGOs can make use of alternative data sources and technologies that would not require a lot of expensive hardware such as social media (e.g., Facebook, Twitter), sensors and mobile phones (e.g., SMS, WhatsApp). These data sources have stimulated development of methods such as sentiment analysis, location mapping or advanced social network analysis. A broad range of skills is also needed. For example, skills related to data collection, technical and programming skills and interpretation of data. Young people with affinity to IT would be an asset to NGOs as well as collaborating with outside experts (such as universities and corporations doing pro-bono projects).

6.3 RECOMMENDATIONS FOR FUTURE RESEARCH

Future research could focus on further evaluating the data-driven innovation methods in NGOs. It would be interesting to gain deeper knowledge on critical and qualitative GIS, radical statistics and mixed-method approaches. As we enter the age of Big Data, it is clear that we need critical reflection and research about Big Data as well as studies using data-driven innovation in NGOs. An urgent need for wider critical reflection on the epistemological implications of Big Data and data analytics in NGOs, a task that has barely begun despite the speed of change in the data landscape.

7 REFERENCES

U.S. Agency for International Development. (2000). *Voluntary Foreign Aid Programs.* Retrieved 2016, from http://pdf.usaid.gov/pdf_docs/Pnach955.pdf

Ali, A. (1994). Pioneering versus incremental innovation: review and research propositions. *J Prod Innovat Manag*(11), pp. 46-61.

Al-Khouri, A. M. (2014, September). Privacy in the Age of Big Data: Exploring the Role of Modern Identity Management Systems. *World Journal of Social Science, 1*(1).

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. pp. 7-16.

Andrade, P. L., Hemerly, J., Recalde, G., & Ryan, P. (2014, April 24). From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential? *The Global Information Technology Report 2014: Rewards and Risks of Big Data*(Chapter 1.8).

Barras, R. (1986). Towards a theory of innovation in services. *Research policy, 15*(4), pp. 161-173.

Barry, E., & Bannister, F. (2013). Barriers to open data release: a view from the top. *In: 2013 EGPA Annuaul Conference*.

Batini, C., & Scannapieco, M. (2006). Data Quality: concepts. *Springer, Heidelberg*.

Ben-Akiva, M., & Gershenfeld, S. (1998). Multi-featured products and services: analysing pricing and bundling strategies. *Journal of Forecasting, 17*(3-4), pp. 175-196.

Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & von Schreeb, J. (2011, August 30). Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti. *PLoS Medicine, 8*(8).

Berners-Lee, T. (2006). Linked data: Design issues. *https://www.w3.org/DesignIssues/LinkedData.html*.

Betts, A., Bloom, L., & Omata, N. (2012). Hummanitarian Innovation and Refugge Protection. *Working Paper Series*(85).

Birkinshaw, J., Hamel, G., & Mol, M. J. (2008). Management innovation. *Academy of management Review, 33*(4), pp. 825-845.

Boer, H., & During, W. E. (2001). Innovation, what innovation? A comparison between product, process and organisational innovation. *International Journal of Technology Management, 22*(1-3), pp. 83-107.

Bollier, D. (2010). The Promise and Peril of Big Data.

Börner, K., & Scharnhors, A. (2009). Visual Conceptualizations and Models of Science. *Special Issue on the Science of Science: Conceptualizations and Models of Science,Journal of Informetrics, 3*(3).

Boyd, D., & Crawford, K. (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society, 15*(5), pp. 662-679.

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? *Available at SSRN* .

Burke, K. (2012). Perspective on density functional theory. *The Journal of chemical physics, 136*(15).

Cacciola, S., & Gibbons, R. (2012, November 28). Coase Meets the Cloud: How and When Can Outsourcing IT Improve Organizational Performance? *White Paper*.

Caroleo, B., Tossato, A., & Osella, M. (2015). Making Sense of Governmental Activities Over Social Media: A Data-Driven Approach. *Springer Link*, pp. 34-45.

Cowls, J. (2014, September 25-26). The Crowd in the Cloud? Three Challenges for Gauging Public Opinion Online. *Paper presented at IPP2014: Crowdsourcing for Politics and Policy, The Internet Policy and Politics Conference*.

Cukier, K. &., & Mayer-Schoenberger, V. (2013). Rise of Big Data: How it's Changing the Way We Think about the World. *The. Foreign Aff., 92*(28).

Damanpour, F., & Evan, W. M. (1984). Organizational innovation and performance: the problem of" organizational lag". *Administrative science quarterly*, pp. 392-409.

Damanpour, F., & Gopalakrishnan, S. (2001). The dynamics of the adoption of product and process innovations in organizations. *Journal of Management studies, 38*(1), pp. 45-65.

Dannatt, G. (2012). *E-INTERNATIONAL RELATIONS STUDENTS.* Retrieved May 2016, from http://www.e-ir.info/2012/09/14/how-cloud-computing-complicates-the-jurisdiction-of-state-law/

Davenport, T. H., Cohen, D., & Jakobson, A. (2005). Competing on analytics. *Babson Executive Education*.

De Jong, K. (2006). Evolutionary Computation: A Unified Approach. *MIT Press*.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review.

Desouza, K. C., & Smith, K. (2014, Summer). *Stanford Social Innovation Review.* Retrieved September 7, 2015, from http://ssir.org/articles/entry/big_data_for_social_innovation

Doh, J. P., & Teegen, H. (. (2003). *Globalization and NGOs: Transforming business, government, and society.* Greenwood Publishing Group.

Dumbill, E. (2012). *Big Data, Big Hype: Big Deal.* Retrieved April 2016, from Forbes: http://www.forbes.com/sites/edddumbill/2012/12/31/big-data-big-hype-big-deal/#63d3869f6cc0

Eagle, N., & Greene , K. (2014). Reality Mining: Using Big Data to Engineer a Better World. *Cambridge: MIT Press*.

Eckartz, S. M., Hofman, W., & Van Veenstra, A. (2014). A decision model for data sharing. *Springer Berlin Heidelberg*, pp. 253-264.

Edquist, C., Hommen, L., & McKelvey, M. D. (2001). *Innovation and employment: Process versus product innovation.* Edward Elgar Publishing.

Einav, L., & Levin, J. (2013). The data revolution and economic analysis. *National Bureau of Economic Research*.

Esty, D., & Rushing, R. (2007). The promise of data-driven policymaking. *Issues in Science and Technology, 23*(4), p. 67.

Ettlie, J., Bridges , W., & O'Keefe , R. (1984). Organization strategy and structural differences for radical versus incremental innovation. *Manage Sci*(30), pp. 682–695.

Fagerberg, J., Mowery, D., & Nelson, R. (2004). Handbook of Innovation.

Floridi, L. (2012). *Philosophy & Technology*, pp. 1-3.

Galanakis, K. (2006). Innovation process. Make sense using systems thinking. *Technovation, 26*(11), pp. 1222-1232.

Ganguly, A. R., & Gupta, A. (2004). Data mining tehchnologies and decision supprot systems for business and scientific applications, Encyclopeida of Data Warehousing and Mining. *Idea Group Publishing*.

Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., et al. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA, 293*(10), pp. 1223-1238.

Gartner. (2011). *Big Data Creates Big Jobs: 4.4 Million IT Jobs Globally to Support Big Data By 2015.* Retrieved April 2016, from http://www.gartner.com/newsroom/id/2207915

Google Scholar. (n.d.). *Google Scholar*. Retrieved 2015, from https://scholar.google.com/

Gopalakrishnan, S., & Damanpour, F. (1997). A review of innovation research in economics, sociology and technology management. *Omega, 25*(1), pp. 15-28.

Graham, M. (2012). What Can Twitter Tell Us About Hurricane Sandy Flooding? Visualised. The Guardian.

Gray, J. (2013, August 7). Semios: Big Data and Digitizing the Farm. *Canadian Startup News*.

Greenleaf, G. (2012). The influence of European data privacy standards outside Europe: Implications for globalization of Convention 108. *International Data Privacy Law*.

Hall, J., & Martin , M. (2005). Disruptive technologies, stakeholders and the innovation value-added chain: a framework for evaluation radical technology development. *R&D Manage*(35), pp. 273–284.

Hamel, G. (2006). The why, what, and how of management innovation. *Harvard business review, 84*(2), p. 72.

Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques. *Elsevier*.

Harvard Business School. (n.d.). *Working Knowledge. The thinking lthat eads*. Retrieved October 2015, from http://hbswk.hbs.edu/industries

Hemerly, J. (2013, June). Public Policy Considerations for Data-driven Innovation. *Computer (IEEE Computer Society), 46*(6), pp. 25-31.

Hienz, J. (2014, October 7). *www.uschamberfoundation.org.* Retrieved November 2015, from U.S. Chamber of Commerce Foundation : https://www.uschamberfoundation.org/defining-data-movement

Hilbert, M. (2013, January 15). Big Data for Development: From Innovation to Knowledge Societies.

Hofman, W., & Bastiaansen, H. (2013). A global IT infrastructure improving container security by data completion. *In: ECITL, Zaragoza, Spain*.

Hogge, B. (2010). *Transparency and Accountability Initiative.* Retrieved from http://www. soros. org/initiatives/information/focus/communication/articles_publications/publications /open-data-study-20

Hwang, J., & Christensen , C. (2007). Disruptive innovations in healthcare delivery: a framework for business-model innovation. *Health Affair*, pp. 1329–1335.

IBM. (2013). *IBM*. Retrieved October 2015, from www.ibm.gom/big-data

Independent Expert Advisory Group. (2014, November). *DATA REVOLUTION .* Retrieved September 2015, from http://www.undatarevolution.org/report/

Intel . (2012, August). *Intel.* Retrieved February 2016, from http://www.intel.nl/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf

ITU . (2013). *Mobile overtakes fixed: Implications for policy and regulation.* Retrieved May 2016, from http://www.itu.int/osg/spu/ni/mobileovertakes/Resources/Mobileovertakes_Paper.pdf

Iyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., & Ramani, K. (2005). Three-dimensional shape searching: state-of-the-art review and future trends. *Computer-Aided Design, 37*(5), pp. 509-530.

Jansen, J. J., Van Den Bosch, F., & Volberda, H. (2006). Exploratory innovation, exploitative innovation, and performance: Effects of organizational antecedents and environmental moderators. *Management science, 52*(11), pp. 1661-1674.

Janssen, M., & Zuiderwijk, A. (2012). Open data and transformational government. *In: TGov Conference*.

Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., et al. (2009). Data-intensive science: a new paradigm for biodiversity studies. *BioScience, 59*(7), pp. 613-620.

Kelly, J. (2014, February). Big data vendor revenue and market forecast. *Wikibon Article*.

Kettunen, P. (2009). Agile software development in large-scale new product development organization: team level perspective. *Doctoral Dissertation*.

Khan, N., Yaqoob, I., Hashem, I., Inayat, Z., Mahmoud Ali, W., Alam, M., et al. (2014). Big Data: Survey, Technologies, Opportunities, and Challenges. *The Science World Journal*.

Kimberly, J. R., & Evanisko, M. J. (1981). Organizational innovation: The influence of individual, organizational, and contextual factors on hospital adoption of technological and administrative innovations. *Academy of management journal, 24*(4), pp. 689-713.

Kirkpatrick, M. (2010, August 4). Google, Privacy and the New Explosion of Data . *Techonomy*.

Kirkpatrick, R. (2011). Data philanthropy: Public and private sector data sharing for global resilience. *UN Global Pulse*.

Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele University, 33*(2004), pp. 1-26.

Kitchenham, B., Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009, January). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology, 51*(1), pp. 7-15.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1).

Klepper, S. (1996). Entry, exit, growth, and innovation over the product life-cycle. *American Economic Review, 86*(3), pp. 562-583.

Knight, S., & Burn, J. (2005). Developing a framework for assessing information quality on theWorld Wide Web. *Informing Science*, pp. 159-172.

Kotabe, M., & Murray, J. Y. (1990). Linking product and process innovations and modes of international sourcing in global competition: A case of foreign multinational firms. *Journal of International Business Studies*, pp. 383-408.

Kotler, P. (1991). Philip Kotler explores the new marketing paradigm. *Marketing Science Institute Review, 1*(4/5), pp. 1-5.

Kouzes, R. T., Anderson, G. A., Elbert, S. T., Gorton, I., & Gracio, D. K. (2009). The changing paradigm of data-intensive computing. *Compute, 1*, pp. 26-34.

Krause, U. (2013). Innovayion: The new Big Push or the Post-Development alternative? *Development, 56*(2), pp. 223-226.

Kumar, R. (2014). *The World Bank.* Retrieved February 2016, from http://blogs.worldbank.org/youthink/how-youth-saved-bananas-uganda

Kusiak, A., & Tang, C.-Y. (2006, June). The Data-Inspired Innovation Model. *Proceedings of the 36th International Computers and Industrial Engineering Conference, C&IE*, pp. 1-8.

Laanti, M., Similä, J., & Abrahamsson, P. (2013). Definitions of Agile Software Development and Agility. *In Systems, Software and Services Process Improvement*, pp. 247-258.

Labrinidis, A. &., & Jagadish, H. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowmen, 5*(12), pp. 2032-2033.

Lambell, R., Ramia, G., Nyland, C., & Michelotti, M. (2008). NGOs and international business research: Progress, prospects and problems. *International Journal of Management Reviews, 10*(1), pp. 75-92.

Lane, J., Stodden, V., Bender, S., Nissenbaum, H., & (Eds.). (2014). *Privacy, big data, and the public good: Frameworks for engagement.* Cambridge University Press.

Laney, D. (2001, February 6). 3D data management: Controlling data volume, velocity and variety. *Application Delivery Strategies, 6*(70).

Larose, D. T. (2006). Data mining methods & models. *John Wiley & Sons*.

Less Andrade, P., Hemerly, J., Recalde, G., & Ryan, P. S. (2014). From Big Data to Big Social and Economic Opportunities: Which Policies Will Lead to Leveraging Data-Driven Innovation's Potential? *The Global Information Technology Report*.

Lettl C., Herstatt , C., & Gemuenden , H. (2006). Learning from users for radical innovation. *Int J Technology Management*(33), pp. 25–45.

Leverty, S. (2009). Non-Governmental Organizations, the UN, and APA. *APA Office of International Affairs* .

Light, P. C. (1998). Sustaining innovation: Creating nonprofit and government organizations that innovate naturally. *Jossey-Bass*.

Linksvayer, M. (2012, December 17). *Open Definition.* Retrieved January 2016, from http://opendefinition.org/2012/12/17/open-definition-forges-ahead-get-involved/

Luxmore, S. R., & Hull, C. E. (2011). Innovation and NGOs A framework of interaction. . *The International Journal of Entrepreneurship and Innovation, 12*(1), pp. 17-27.

Lyer, N., Jayanti, S., Lou, K., Kalyanaraman, Y., & Karthik, R. (2005). Shaped-based searching for product lifecycle applications. *Computer-Aided Design, 37*(13), pp. 1435-1446.

Macrae, J., Collinson, S., Buchanan-Smith, M., Reindorp, N., Schmidt, A., Mowjee, T., et al. (2002). *Uncertain power: the changing role of official donors in humanitarian action.* Overseas development institute (ODI). Humanitarian policy group (HPG).

Manovich, L. (2011). Trending: The promises and challenges of big social data. Retrieved from http://www.manovich.net/DOCS/Manovich_trending_paper.pdf

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011, May). Big Data: The Next Frontier for Innovation, Competition and Productivity. *McKinsey Global Institute Report*.

Marsh, J., Pane, J., & Hamilton, L. (2006). Making Sense of Data-Driven Decision Making in Education. *Rand, Santa Monica*.

Marton, A., Avital, M., & Jensen, T. (2013). Reframing Open Big Data. *ECIS*, p. 146.

Matthews, J., & Shulman, A. D. (2005). Competitive advantage in public-sector organizations: explaining the public good/sustainable competitive advantage paradox. *Journal of Business Research, 58*(2), pp. 232-240.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Bus, 90*(10), pp. 60-66.

Meeus, M., & Edquist, C. (2006). Introduction to Part I. Product and Process Innovation. Innovation, Science and Institutional Change. A Research Handbook. pp. 23-37.

Miller, P., Styles, R., & Heath, T. (2008). Open data commons, a license for open data. *In: LODW 2008, Beijing* .

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*(4), pp. 264-269.

Nousak, P., & Phelps, R. (2002, January). *A scorecard approach to improving data quality.* Retrieved May 2016, from http://www2.sas.com/proceedings/sugi27/p158-27.pdf

O'Malley, S., & Dijkzeul, D. (2002). A Typology of International Humanitarian Organizations. In International Expert Conference 'Security and Humanitarian Action: Who is Winning.

OECD Oslo Manual. (2005). Guidelines for Collecting and Interpreting Innovation Data.

OECD/Eurostat . (2008). *Oslo Manual.* Retrieved 2015, from The Measurement of Scientific and Technological Activities, Guidelines for Collecting and Interpreting Innovation Data : http://epp,eurostat,ec,europa,eu/cache/ITY_PUBLIC/ OSLO/EN/OSLO-EN,PDF

O'Harrow, R. J. (2006). No Place to Hide. *New York: Free Press*.

Open Data Center Alliance. (2012). *www.opendatacenteralliance.org.* Retrieved February 2016, from https://www.opendatacenteralliance.org//docs/Big_Data_Consumer_Guide_Rev1.0.pdf

Parthasarthy, R., & Hammond , J. (2002). Product innovation input and outcome: moderating effects of the innovation process. *Journal of Engineering and Technology Management, 19*(1), pp. 75-91.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press.

Phillipe, O. (2012). Reassembling the data - How to understand opportunities using an interdisciplinary approach. *Internet, Politics, Policy*.

Phillips, R. (2006). The role of nonprofit advocacy organizations in Australian democracy and policy governance. *International Journal of Voluntary and Nonprofit Organizations, 17*(1), pp. 57-73.

Piana, V. (2003). Retrieved 2016, from Innovation: http://www.economicswebinstitute.org

Poel, M., Treperman, J., Schroeder, R., Rubinstein, M., Meyer, E., Mahieu, B., et al. (2015, July 17). *data4policy.* (T. Group, Oxford Internet Institute, & Centre for European Policy Studies, Eds.) Retrieved January 2016, from http://www.data4policy.eu/#!sota/cbiv

Porter, M. (1990). Competitive Advantage of Nations. *Free Press, New York.*

Sackett, D. (1997). Evidence-based medicine. *Semin. Perinatol*, pp. 3-5.

Schroeder, R. (2014, December 17). Big Data and the brave new world of social media research. *Big Data & Society*.

Schroeder, R., & Taylor, R. (2015). Big data and Wikipedia research: social science knowledge across disciplinary divides. *Information, Communication & Society, 18*(9), pp. 1039-1056.

Shiri, A. (2012). Typology and analysis of big data: An information science prospective. *Internet, Politics, Policy*.

Sirilli, G., & Evangelista, R. (1998). Technological innovation in services and manufacturing: results from Italian surveys. *Research policy, 27*(9), pp. 881-899.

Spar, D. L., & La Mure, L. T. (2003). The power of activism: Assessing the impact of NGOs on global business. *California Management Review, 45*(3), pp. 78-101.

Stoddard, A. (2003). Chapter 3-Humanitarian NGOs: Challenges and Trends. *Humanitarian Policy Action Group Report*, p. 14.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research* (Vol. 15). Newbury Park, CA: Sage.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduciton to Data Mining. *Pearson Eduction*.

Taylor, L., & Schroeder, R. (2015). Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal, 80*, pp. 503-518.

Taylor, L., Cowls, J., Schroeder, R., & Meyer, E. (2014, December). Big Data and Positive Change in the Developing World. *Policy & Internet, 6*(4), pp. 418-444.

Teegen, H. (2003). International NGOs as global institutions: using social capital to impact multinational enterprises and governments. *Journal of International Management, 9*, pp. 271-285.

The Economist. (2013a, February 9). Retrieved 2016, from Counting Raindrops: Available at http://www.economist.com/news/science-and-technology/21571384-how-use-mobile-phone-networks-weatherforecasting-counting-raindrops.

The Economist. (2013b, January 25). Have Germs, Will Travel. *The Economist*.

Tornatzky, L. G., Fleischer, M., & Chakrabarti, A. K. (1990). *Processes of technological innovation.* Lexington Books.

Troyer, L. (2005). Personal communication. *Department of Sociology, The University of Iowa*.

U.S. Chamber of Commerce Foundation. (2014). The future of data-driven innovation.

Uchupalanan, K. (2000). Competition and IT-based innovation in banking services. *International Journal of Innovation Management, 4*(4), pp. 455-489.

UNICEF . (2013). *Innovation.* Retrieved 2015, from http://www.unicefinnovation.org/frequently-asked-questions#sthash

United Nations. (2012). *World Population 2012*. Retrieved 2016, from http://www.un.org/en/development/desa/population/publications/pdf/trends/WPP2012_Wallchart.pdf

Veryzer, R. (2005). The roles of marketing and industrial design in discontinuous new product development. *Journal of Product Innovation Management, 22*(1), pp. 22-41.

Weber, K., Otto, B., & Osterle , H. (2009). One size does not fit all - a contigency approach to data governance. *Journal of Data and Information Quality, 1*(1).

Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2005, November). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Original article*.

Zaltman, G., Duncan, R., & Holbek, J. (1973). *Innovations and organizations* (Vol. 1973). New York: Wiley.

Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2012). Understanding big data. *New York et al: McGraw-Hill, 5*(8).

Zuiderwijk, A., Helbig, N., Gil-Garcia, J., & Janssen, M. (2014). Guest Editors' Introduction. Innovation through open data: a review of the state-of-the-art and an emerging research agenda. *Journal of Theoretical and Applied Electronic Commerce Research , 9*(2).

8 APPENDIX

8.1 APPENDIX I – SYSTEMATIC LITERATURE REVIEW TABLE

This part of the appendix is delivered separately.

8.2 APPENDIX II – INTERVIEW QUESTIONS

This interview guide is part of a master thesis project at Leiden University and the Centre for Innovation about data-driven innovation in NGOs. The goal of the study is how NGOs work with data-driven innovation.

The study will involve interviewing people working for non-profit and non-governmental organizations and are familiar with data-driven innovation. The interview will take approximately 30 minutes of your time and will focus on collection of information. It will be a non-judgmental procedure and you are free to skip any question you do not feel comfortable answering. We will treat the information you share with us carefully and keep you anonymous. We will ensure that no references to your name will be traceable. Any extracts from what you say that are quoted in the thesis will be entirely anonymous.

Should you have any questions or comments regarding this questionnaire, you are welcome to contact me by Phone: +31 626914501, Skype: katia_deligianni, or E-mail: katiadlgn@gmail.com

Interviewee information

Please fill out the information below.

Name:

Job position:

E-mail:

Introduction

Personal introduction of the interviewee and the researcher, research background (data-driven innovation), what would researcher like to learn (data-driven innovation process, key people, what triggers innovation, etc.). Researcher asks to record the interview and emphasizes on the anonymity of the data collection. None of the statements should be linkable to the participants, and if a participant will be quoted she/he will be asked in advance.

General questions

1. What is your role and responsibility within the NGO?

2. How long are you in your current position and within the organization itself?

3. What is your current assignment?

First Phase – Definition data-driven innovation

4. How would you define “data-driven innovation”?

5. Are you, and, how are you using data-driven innovation methods?

6. Are there in your organization any showcases / pilot projects for data-driven innovation you could tell us about? (Looking for anecdotal evidence)

Second Phase – Concrete Project Examples

Now I would like to ask you to bring in mind a recent data-driven innovation project that you were involved into during your work for the NGO. The questions that will follow will be related to this project.

7. What kind of data did you use? (e.g., Public datasets /administrative Open data, Social media, Sensors and mobile phones)

8. If you used Open data, what issues did you encounter? (e.g., Heterogeneity, Quality, Registration needed)

9. Where did you collect your data? (e.g., Administrative data, Statistical offices, Survey data)

10. In what type of format were the data you used? (e.g., Microsoft Excel, CSV, SQL, GIS, PDF, HTML)

11. What type of data analysis tools did you use? Would you prefer something else?

12. What were the agreements or guidelines regarding data protection and privacy that you had to apply?

13. Please indicate who /what were the key elements for your project? (e.g., key people with experience, the initial idea, resource allocation, the urgency of the project, the estimated time of completion)

Third phase – NGOs and data-driven innovation

14. Do you know if other NGOs are using data-driven innovation models? If yes, how?

15. Is there any kind of competition between NGOs for identifying or implementing innovation?

16. Do you benchmark with other different industry sectors to improve your ways to innovate? Which sectors according to you are good examples?

17. Which are the biggest challenges in data-driven innovation for a non-governmental organization?

Feedback

18. Do you have any feedback on this interview?

19. Can I contact you for any follow-up questions on your answers?

8.3 APPENDIX III – RAW INTERVIEW TRANSCRIPTIONS

This part of the appendix is delivered separately.

8.4 APPENDIX IV – CODEBOOK

Codes used in results and discussion section.

Code	*Sub-category*	Sources	References	Description
Data Governance Aspects				
Data sources		**8**	**62**	The sources for the data collection.
	Open Data / Public datasets	8	27	Data files prepared by investigators or data suppliers with the intent of making them available for public use and republish.
	Private datasets	5	5	Data files prepared by investigators or data suppliers with the intent of not being publicly accessible.
	Sattelite imagery	3	4	Consists of images of Earth collected by satellites. Imaging satellites are operated by governments and businesses around the world.
	Survey data	5	10	The collection of information outside a laboratory. May include interview or observation of people in their natural environments to learn their everyday problems or emergency situation.
	Location data	1	1	Data with specfic location information
	SMS messages	2	7	Data collected by SMS messages send by the participants.
	Mobile Application	2	2	Data collected by mobile phone application send by the participants.
	Statistical offices	2	2	Statistical offices that provide Open data portals.
	Project data	3	3	
	Social media	1	1	
Data formats		**8**	**35**	The format of the used data.
	Microsoft Excel	8	9	
	Google spread sheets	1	1	
	CSV	5	9	
	SQL	2	2	

	JSON	2	2	
	Java script libraries	1	2	
	XML	3	5	
	PDF	2	2	
	Microsoft Word	2	2	
API		3	4	Centralized Application Programming Interface
	Standardization	3	4	
	Transparent	3	4	
	Easily accesible	3	4	
Data visualization		**6**	**9**	Ways of data visualization.
	Front-end dashboards	2	2	
	GIS maps	3	4	
	Interactive maps	2	2	
	Unified visualization	1	1	
Data analytics		**6**	**6**	Data analysis tools.
	Forecasting model	1	2	
	Other statistical methods	1	1	
	Machine learning algorithms	1	3	
	Crowdsourcing	1	1	
	R	1	3	
	Python	1	1	
	Own visualization tool	5	5	The NGO has developed its own visualization tool.
	Data mining tool	3	5	
	Infographics	1	1	
Better management		2	3	
Set up networks for better data collection		3	3	Collaboration with other organizations for implementation of data-driven solutions.
	Add new applications	1	3	New applications for data collection.
Variety of expertise		2	2	Team qualification composition of people working with data.
A methodology and its platform		5	8	A methodology for data-driven innovation.
Monitoring and Evaluation approach		3	5	NGOs use this approach to evaluate the effectiveness of their work.
	Evidence-based approach	3	7	

Same admin areas		1	1	Same admin areas and same data sets.
Quality management		3	4	
Exchange information		3	5	
	Between countries	2	2	
	Between organizations	2	2	
	Between platforms	1	1	
Response to challenges		3	4	Data collected for better response to challenges.
	Faster and more effective way	1	2	Increased response time and effectiveness.
	Reduce costs of humanitarian purposes	1	1	
Public engagement	*Community involvement*	1	7	Public enagagement for participation.
	Politicians involvement	1	2	
Policy making		4	8	Data collected for policy making.
	Make decisions	5	12	Make new decisions based on new data.
	Change policies	4	7	Indication for change needed in policies.
Privacy				
Sensitive information		5	8	Data contains sensitive information.
Show trends not exact numbers		1	4	
Anonymity		3	5	Protect the anonymity of the participants.
Exusion policy		3	6	A policy created for the exclusion of some fields in the data set.
Project's key elements		8	8	Key factors that initiated the project
	People with experience	2	2	People with experience in the topic.
	Human resource allocation	1	2	
	Initial idea	6	6	The initial idea triggered the project.
	The urgency of the project	2	2	
	Resource allocation	3	3	Budget related.

	Compatability with government regulations	3	5	
Data Governance & Open Data Barriers				
Data quality challenges		8	55	NGOs challenges related to Data quality barriers
Accurancy		5	7	Data is incorrect or out of date
Accessibility		4	12	Data is not easily accessible or understandable
Conformity		4	14	Data is stored in a non-standard format
Completeness		2	3	Data is missing or is unusable
Consistency		4	4	Data values give conflicting information
Duplicates		2	2	Data records or attributes are repeated
Integrity		4	5	Data is missing or is not referenced
Timeliness		2	2	Information recorded and made available to systems is not as rapidly as is required
Technical challenges		8	37	NGOs challenges related to Technical barriers
Configuring data transformation		5	7	Difficulties in configuring data transformation
Not machine readable format		2	2	The publishing of data in a format that is not machine readable
Lack of an API		3	10	Lack of an Application Programming Interface (API)
Processing of Data sets		5	7	Difficulties to processing Data sets
Linking or combining functionality		3	6	The lack of a linking or combining funtionality
Low resolution for satellite data		2	2	Difficulties with satellite data resolution.
Capacity		2	2	
	Fields had word limitation	1	1	
Economic challenges		4	6	NGOs challenges related to Economic barriers
Small budget		4	6	Small project budget for Information Technology advancements
Institutional challenges		4	26	NGOs challenges related to Institutional barriers
Lack of knowledge		4	6	Lack of knowledge about new

				Information systems
Lack of support		2	3	Lack of support to employees to learn new technologies
Lack of standards		2	4	Lack of standard way in handling data
Not able to set long term goals		4	6	Hard to establish long-term goals due to the nature on NGOs work.
Lack of guideliness		1	1	
Lack of capacity to innovate		4	6	
Privacy challenges		8	17	NGOs challenges related to Privacy barriers
Not enough location data		1	1	Lack of complete location information due to privacy concerns
Data disclosure issue		4	9	Disclosure of data to specific group of organizations only.
Nature of data		2	2	Raw data nature.
Different countries' regulations		3	5	Different laws are applicable in different countries.
Operational challenges		1	9	NGOs challenges related to Operational barriers
Services				Lack of services
Lack of actions		1	1	Difficulty to put in action initial projects.
Lack of clear semantics		2	2	
Lack of information on data quality		3	4	
Project communication		1	2	
Social challenges		4	26	NGOs challenges related to Social barriers
Aid communication		2	3	Lack of an efficient way to communicate their aims/goals
	Trust of the donors	2	3	Difficulty in gaining donors' trust.
Lack of interaction		1	1	Lack of interaction with employees, especially for a beginner
Lack of people		3	5	Few people on team
	Data expertise	2	2	Lack of pepole with data innovation capacity.
	Generation gap	1	3	Lack of young people with affinity to IT.
Measure impact of innovation		4	4	Difficulty to measure the impact of inoovation for the organization
Interpretation of data		2	5	

Ownership challenges		1	4	NGOs challenges related to Ownership barriers
Registration is needed	*Governmental permission*	1	2	Governmet should grand permission.
	Censorship	1	2	Censorship from governments.
Data-driven Innovation Definition		**8**	**8**	Terminology used to define data-driven innovation
Specific type of innovation		1	1	
Creative solution		1	1	
	A process which leads to a solution	2	2	
	Better tools to address challenges	1	1	
	Something new that solves a problem	1	1	
	Drive change	2	4	Ways to improve changes that need to happen.
	Enhance our projects	4	4	
	Collection of different data sets	2	2	
	New insights and trends	3	8	Data provides new insights and trends to start working with.
Structured information		1	1	
Quantitative method		2	2	
Qualitative method		1	1	
Data collection		3	4	
	Technology can help	2	3	Better data collection.
Decisions based on data		4	7	Data as information to make decisions.
NGOs and Data-driven Innovation		**8**	**8**	
Aware of others NGOs actions		8	10	Awareness of NGOs data-driven innovation methods.
Competition between NGOs		5	7	
	Personal level	1	1	At employees level for ideas that have not been implemented.
	Healthy competition between NGOs for data-driven innovation methods	3	3	

	First comes the innovation for technology and then the humanitarian aid	1	1	NGOs first found the technology and then try to implement it for humanitarian purposes.
	Compete for funding	3	3	Donors set a competitive system for funding NGOs.
Collaboration with other NGOs	*Bigger capacity for humanitarian aid*	5	5	NGOs collaborate in order to innovate in hummanitarian aid.
Benchmark analysis		**7**	**17**	Improve ways to innovate from other industry sectors.
Information Technology		3	5	
	Text analytics Startups	1	1	
	Software applications	1	1	
	Companies dealing with Geospatial data	1	1	
Information		1	1	
Aerospace		1	1	
Auto		1	1	
Education		2	4	
Service		2	2	