Bashir Kasozi

IMPACT OF OPEN DATA ON EDUCATION
Using Open Datasets as an Open Educational Resource
IMPACT OF OPEN DATA ON EDUCATION
Using Open Datasets as an Open Educational Resource

Bashir Kasozi
Bachelor’s Thesis
Spring 2019
Degree Programme in Business
Information Technology
Oulu University of Applied Sciences
ABSTRACT

Oulu University of Applied Sciences
Degree Programme in Business Information Technology

Author: Bashir Kasozi
Title of Bachelor’s thesis: Impact of Open Data on Education: Using Open Datasets as an Open Educational Resource
Supervisor: Ilkka Mikkonen
Term and year of completion: Spring 2019 Number of pages: 58

The objective of this thesis work was to introduce the concept of Open Data and how its application in the education sector as an Open Educational Resource can have an impact on teaching, learning and research. Additionally, the work also discusses the relevant tools and skills needed to utilize Open Datasets for educational purposes, and the challenges involved in using Open Datasets for teaching and learning tools.

The theoretical work of this thesis is based on literature review examines the different formats, types and structures Open Data. The benefits of using Open Data for teaching and learning purposes, and the new innovations in Open Education. To understand the Open Data phenomena better, attention was also drawn to Data visualization and its significance in the field of education.

Finally, this thesis work shows that Open Data is a very important resource that has a very big impact on the way people teach and learn and can be used effectively to transform and revolutionize the education sector by making teaching, learning and research affordable by reducing costs, providing opportunities to people to access free education as well as enhancing skill development of learners and teachers. However, getting valuable information from large volumes of Open Data requires the right tools and skills needed to manipulate data. Similarly, understanding and interpreting the data requires both soft and hard or technical skills such as statistical, programming and communication

Keywords: Open Data, Open Education, Open learning, Data Visualization, Open Access.
CONTENTS

1 INTRODUCTION ............................................................................................................................................ 7
   1.1 Thesis structure .................................................................................................................................. 7
   1.2 Background and Motivation .............................................................................................................. 7
   1.3 Objectives .......................................................................................................................................... 8
   1.4 Research Questions .......................................................................................................................... 8

2 OPEN DATA .................................................................................................................................................. 10
   2.1 History of Open Data ....................................................................................................................... 10
   2.2 Data, Information and Knowledge ................................................................................................... 11
   2.3 Data .................................................................................................................................................. 11
   2.4 Information ...................................................................................................................................... 11
   2.5 Knowledge ...................................................................................................................................... 12
   2.6 Principles of Open Data ................................................................................................................... 14
   2.7 Benefits of Opening Data ................................................................................................................. 14
   2.8 Categories of Open Data ................................................................................................................ 16
   2.9 Sources of Open Data ...................................................................................................................... 16
   2.10 Formats of Open Data ..................................................................................................................... 19
   2.11 Linked Open Data .......................................................................................................................... 22
   2.12 Open Data Structures .................................................................................................................... 23

3 IMPACT OF OPEN DATA SETS ON EDUCATION AND LEARNING ......................................................... 24
   3.1 Using Open Data as Teaching and Learning Resource ..................................................................... 24
   3.2 New Trends in Open Education ...................................................................................................... 29
   3.3 Challenges and Bottlenecks to Open Education ............................................................................. 31
   3.4 Open Data Portals and Institutions .................................................................................................. 33
       3.4.1 European Data Portal ................................................................................................................ 33
       3.4.2 Avoindata.fi .............................................................................................................................. 34
       3.4.3 Helsinki Region InfoShare ....................................................................................................... 34
       3.4.4 Kaggle ...................................................................................................................................... 34
       3.4.5 World Bank Data Portal ......................................................................................................... 34
       3.4.6 6Aiika ....................................................................................................................................... 35
       3.4.7 Open Data Institute (ODI) ........................................................................................................ 35

4 OPEN DATA MANIPULATION TOOLS AND SKILLS ............................................................................ 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Apache Hadoop</td>
<td>36</td>
</tr>
<tr>
<td>4.2</td>
<td>Yahoo Query Language</td>
<td>37</td>
</tr>
<tr>
<td>4.3</td>
<td>Google charts</td>
<td>37</td>
</tr>
<tr>
<td>4.4</td>
<td>ArcGIS</td>
<td>37</td>
</tr>
<tr>
<td>4.5</td>
<td>Quadrigram</td>
<td>37</td>
</tr>
<tr>
<td>4.6</td>
<td>MongoDB</td>
<td>38</td>
</tr>
<tr>
<td>4.7</td>
<td>QGIS</td>
<td>38</td>
</tr>
<tr>
<td>4.8</td>
<td>Apache Storm</td>
<td>38</td>
</tr>
<tr>
<td>4.9</td>
<td>Skills required to work with Open Data</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>DATA VISUALIZATION IN OPEN DATA</td>
<td>42</td>
</tr>
<tr>
<td>5.1</td>
<td>What Is Data Visualization</td>
<td>42</td>
</tr>
<tr>
<td>5.2</td>
<td>Significance of Data visualization in Open Data and Education</td>
<td>43</td>
</tr>
<tr>
<td>5.3</td>
<td>Data Visualization Techniques</td>
<td>43</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Line Charts</td>
<td>44</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Bar Charts</td>
<td>44</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Pie &amp; Donut Charts</td>
<td>45</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Word Clouds</td>
<td>46</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Network Diagrams</td>
<td>47</td>
</tr>
<tr>
<td>5.3.6</td>
<td>Correlation Matrices</td>
<td>48</td>
</tr>
<tr>
<td>5.3.7</td>
<td>Geo-Mapping</td>
<td>49</td>
</tr>
<tr>
<td>6</td>
<td>CONCLUSIONS</td>
<td>51</td>
</tr>
<tr>
<td>7</td>
<td>REFERENCES</td>
<td>52</td>
</tr>
</tbody>
</table>
Abbreviations

CBE  Competence-Based Learning or Education
CSV  Comma Separated Values
EU   European Union
GIS  Geographic Information Systems
HTML Hypertext Markup Language
JPEG Joint Photographic Expert Group
JSON JavaScript Object Notation
KML  Keyhole Markup language
MOOCs Massive Open Online Courses
NASA National Aeronautics and Space Administration
OA   Open Access
OER  Open Educational Resource
OCW  Open Course Ware
ODF  Open Document Format
ODT  Open Digital Textbook
OER  Open Educational Resource
PDF  Pocket Document Format
RDF  Resource Description Framework
SHP  Shapefile Format
UN   United Nations
UNESCO United Nations Educational, Scientific and Cultural Organization
WMO  World Meteorological Organization
XML  Extensible Markup Language
1 INTRODUCTION

Open Data signifies datasets and information that is available and accessible without any restrictions (Open Data Tool Kit 2019, cited 02.01.2019). Open Data and technology are very crucial in dispensing information in our societies today that are data driven. Open Data is transforming education systems from traditional methods to a modern open learning platform. Stake holders in the education sector, parents, teachers, students and researchers can effectively use Open Data for educational purposes. This thesis work discusses how Open Data, can impact teaching and learning, in an education environment that is becoming predominantly web-based learning.

1.1 Thesis structure

This work is structured into five chapters. The introductory chapter gives information about the background and motivation, The main objective, the structure and the research questions for this thesis. The second chapter explains the origin of Open Data, the definition of Open Data and the evolution of Open Data. The third chapter explores the impact of Open Data in education and learning particularly its advantages and problems associated with using Open Data sets. The fourth chapter gives an insight of into Open Data manipulation tools and skills, and their importance in handling Open Data. The final chapter of this thesis work highlights data visualization, a core subject in the field of data presentation.

1.2 Background and Motivation

We are living in a digital era where data and information have become part of our daily lives. In an information-dependent society today, data is being used to solve common life issues, from scientific to social problems. Data is being used as a foundation for the creation of new knowledge, in addition, data is being utilized to create new theories, discoveries, and research. Unrestricted data can be used by society to study and understand their cultural, political, social or economic characteristics. Data has numerous uses where it can be practically applied. Despite its several uses; most crucial data remains blocked and restricted to open access to society and researchers.
It is very common that important data that would be used to accelerate development and solve problems has been subjected to legal and financial restrictions. Limitations to data sharing and access have led to a new concept of Open Data. Open Data has been defined as the type of data that can be freely used, re-used and distributed by anyone, subject to only, at most, to the requirement to attribute and share alike (Open Knowledge International 2019a, cited 04.01.2019). Open Data has been used by governments, individuals and companies to improve products and services offered to people, improve government efficiency and produce new innovations (Citizen's Guide to Open Data 2016, cited 04.01.2019)

1.3 Objectives

The major aim of this thesis to explore the impact of Open Data in education and learning environment. Education is an integral part of every modern society today. Knowledge formation and transfer can only be possible through education and learning, however, in many parts of the world, access to open educational material is still problematic despite new innovations in technology and communication. Many researchers, learners, and teachers can neither share nor access educational material because it is being restricted by copyright and financial laws. This prohibits the transfer of knowledge, it hinders innovation and creativity. It limits improvements in research and makes learning and teaching difficult and expensive. Open Data can remove all barriers to quality education. By making educational tools and resources free and open makes learning much easier, improves the effectiveness of teachers and making education more affordable.

1.4 Research Questions

This thesis work answers three key questions.

1. What is the impact of Open Data on Education?
2. What are the challenges of using Open Datasets for teaching and learning purposes?
3. What key tools and skills are needed to utilize Open Data in education?

These questions are answered by introducing and explaining the term “Open Data” and its purpose. Data manipulation tools, technologies and skills needed to handle Open Data are also introduced,
types and formats of Open Data are also explained. The challenges and barriers of using Open Data types are thoroughly discussed to show the impact of using Open Datasets.
Societies are being transformed by technology and data revolution. Many government agencies and private institutions are coming up with policies that can make this data readily available to the public without any restrictions. However, some of the important data still remains locked up away from the public domain. This section discusses the concept of Open Data, its history, categories, types and formats of Open Data. Similarly, importance of Open Data and its sources are also discussed in this chapter.

2.1 History of Open Data

The term “Open Data” is not new, several authors and journals have tried to explain the evolution and history of Open Data. Open Data awareness was first discussed by a famous an American sociologist researcher Robert King Merton in 1942 who emphasized that all data from scientific research should be made available to everyone (Paris Innovation Review 2013a, cited 05.01.2019). Merton further highlighted that for knowledge being passed on to the next generations, researchers should remove any type of restrictions from their work and stressed the benefits of sharing knowledge to the community.

In 1995, an American Scientific Agency authored a document emphasizing the importance of opening environmental and geophysical data. The agency stressed the point of sharing scientific information and data between different nations to understand the universe fully (Bode 2013, cited 05.01.2019).

However, the popularity of Open Data become more evident between 2003 and 2009 after the European Union and other several governments such as the USA, UK, Canada, and New Zealand declared that they will be disclosing their information to the public (Open Knowledge International 2019b, cited 05.01.2019). The European Union has been central in advocating and promoting the use of Open Data worldwide with more emphasis among its member states. In 2003, the European Union passed a directive on the re-use of public sector information also known as PSI directive among its member states (European Commission 2019a, cited 10.01.2019). The PSI Directive governs the re-use of public sector information, also known as Open Government Data. Member states were encouraged to make as much information available as possible. This directive came into effect on 31 December 2003.
2.2 Data, Information and Knowledge

The term Open Data contains two key words, “Open” and “Data” and to understand the meaning, we need to define each of them separately. The term Open can be used in different contexts. Openness can be used in respect to knowledge, in such a case, it becomes Open Knowledge. Similarly, openness can also be applied to other aspects for example in the case of software such as Open Source or free software. Open Source Software is defined as “any type of software that is free and available to anyone for use or modifications without any restrictions or payment for it” (Open Source Initiative 2018, cited 05.01.2019).

2.3 Data

Data has various definitions and several scholars have come up with different definitions of data. Data can be defined as a collection of texts, numbers or symbols without clear meaning. Data can mean anything for instance images, places, houses, cars etc. “Data is the basic raw material for the formation of information and knowledge cited (European Data Portal 2019a, cited 02.01.2019). Furthermore, data can also be defined as “the basic individual items of numeric or other information, garnered through observation; but in themselves, without context, they are devoid of information” (Burrell 2013, cited 02.01.2019 ). In addition, it can aslo be defined as “A series of disconnected facts and observations”. (Crnkovic 2013, cited 02.01.2019).

2.4 Information

Data that exists on its own is rendered useless and it only becomes information when it carries meaning. Therefore, Information is data that has meaning. Overall, information can be defined as “Data that has been processed or analyzed to make it readable and understandable” (Cambridge International 2017, cited 06.01.2019).

By accumulating and organizing data, we can form new information. Using the example of data above, one can build information about a place to stay by describing the kind of houses and cars in that place after data has been given a meaningful context.
2.5 Knowledge

Knowledge is a by-product of processing and understanding information. Knowledge is a result of interpreting the information you have and applying it where necessary to suit your basic needs or knowledge can be used to solve a certain problem. For information to be classified as knowledge, it must be put to practical use. In summary, knowledge is defined as “The general understanding and awareness garnered from accumulated information, tempered by experience, enabling new contexts to be envisaged” (Knowledge Map of Information Science 2013, cited 10.01.2019). The process of creating Knowledge from raw data can be summarized in the figure below;

FIGURE 1: Knowledge Creation Process from Data (Han & Pei 2011, 7-10)
To understand the concept of Open Data, we need to understand how raw data is converted into knowledge or open knowledge. According to opendefinition.org “Knowledge is open if anyone is free to access, use, modify, and share it-- subject, at most, to measures that preserve provenance and openness” (Open Knowledge International 2019b, cited 09.01.2019). The same definition of open knowledge is applied when defining Open Data. Therefore, from the above observations and deductions, several definitions can be drawn for Open Data. Open Knowledge International exclusively defines Open Data as follows;

“Open Data is data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike” (Open Knowledge International 2019c, cited 12.02.2019). Three major forms of openness have been defined,

1. Technological openness
2. Non-proprietary openness
3. Legal openness

**Technological Openness** means that data must be provided in a machine-readable format where users can access the data using freely available resources, software or hardware. Technological openness permits users of Open Data to access it via public servers without any restrictions such as passwords and firewalls. The main aim of technological openness is to provide data in a machine-readable format such as PDF, CSV, RDF, SPARQL, and any other machine-readable formats (World Bank 2019a, cited 14.01.2019).

**Non-proprietary openness** means that data is neither registered nor restricted by any form of trademarks or protection, but it is shared and reused freely since it is a public resource that must be used for a common good. Since most of the open Data is generated by government agencies, it is morally wrong to permit access to public data to a few selected individuals and restrictions are imposed on others from accessing the data (World Bank 2019a, cited 14.01.2019).

**Legal Openness** regarding Open Data means that the legal terms or the licenses that have been subjected to the data allow free use reuse and redistribution of the data. Legal openness gives users rights and freedom to use data in any appropriate way possible. It removes any form of restriction including commercial ones that allow users to use data for commercial purposes or specific purposes such as using Open Data for health, learning or research purposes (World Bank 2019a, cited 14.01.2019).
2.6 Principles of Open Data

The different forms of openness are further organized to form the eight principles of Open Data. These principles have been accepted by various International Open Data bodies, therefore, Open data must conform to these principles (Open Government Data 2014, cited 14.01.2019).

1. Data must be **complete**: All public data is made available and public data is exempted from any form of privacy, security or privilege restrictions.

2. Data must be **primary**: Data is issued as collected at the source, with the finest possible level of granularity, not in aggregate and modified formats.

3. Data must be **timely**: Data is made available in due time to preserve its value.

4. Data must be **accessible**: Data is available to the widest range of users for the widest range of purposes.

5. Data must be **machine-processable**: Data is reasonably structured to allow automatic processing.

6. Access must be **non-discriminatory**: Data is available to anyone with no registration requirements.

7. Data must be **non-proprietary**: Data are available in a format over which no entity has exclusive control.

8. Data must be **license-free**: Data is not subject to any copyright, patent, trademark or trade secret regulations. Reasonable privacy, security and privilege restrictions may allow as governed by other statutes.

The above principles govern the use, sharing, distribution and re-use of Open Data in regard to its quality, quantity, legal and technical aspects. These principles clearly demonstrate that Open Data can be efficiently utilized, this will be thoroughly discussed in the next chapter.

2.7 Benefits of Opening Data

Fundamentally, Open Data is a mixture of different interconnected things, it is therefore related to several applications and functions. Open Data is not only used in one discipline but rather in so many areas that are dependent on one another.

The value of data being left open will depend on the field where it is applied for instance, in governance, education, business, health or public sectors but Open Data alone doesn’t have any substantial significance. To emphasize the importance of Open Data, the founder of Open
Knowledge International Rufus Pollock stressed that “Openness of Data is essentially a means to an end and not an end in itself” (Pollock 2011, cited 17.01.2019).

The advantages of opening data are quite massive, and this work highlights various areas where openness of data is significant. Open Data fosters Transparency and Accountability. In well-managed societies and governments, the citizens have rights to access public government data and information. By opening data to the public, the citizens can know how their government is spending public resources and this helps to government officials to account for their spending and prevents wasteful expenditure and improves efficiency in procurement deals. A good example of transparency is The Finnish Government regularly updates the public about important public data in clear readable formats without paying for it and this information can be accessed from various websites such as municipal websites. The data includes environmental information and resources, weather and climate data, information from government ministries and agencies. Helsinki Metropolitan area has an Open Data portal, https://hri.fi/en_gb/ (Helsinki Region Infoshare 2019a, cited 16.01.2019). This Open Data portal updates the public about employment opportunities, business deals, procurement data, and tenders. In another example, according to David Eaves a leading activist for open governments, in Canada, it is believed that access to Open Data helped the public to expose one the country’s worst fraud in the history of Canada and helped to prevent the government losing over $3bn (Eaves 2010, cited 17.01.2019).

Open Data improve social services and the economy. Most of the information and knowledge that has been created is in most cases stored by the government or government agencies. By releasing such information to the public, governments can improve social services that are being provided to the citizens and business owners are able to connect and engage with their customers. Furthermore, entrepreneurs and investors can make use of available information about taxes, inflation, wages, laws, costs about the rent of a place to make decisions about new businesses and investment. In Finland, the six big cities of Finland (Helsinki, Espoo, Vantaa, Tampere, Turku, and Oulu) have a joint development project, https://6aika.fi/ that helps to tackle urban challenges and create jobs and businesses for the population. (6Aika 2019a, cited 16.02.2019.)

Open Data improves innovation and research. Developers and researchers can make use of available open government data to come up with new innovations and products. An example is Transport for London that has provided Open Data to the public and developers have made use of this information to build more than 800 transport applications (Transport For London 2019, cited...
16.02.2019 ). In Finland, information from Open Data is being used to come up with innovative applications in the environment, health, education, and other sectors. The use of Finnish digital geodata increased by 50 times in three months when the National Land Survey halted charging users for their services, and new applications were developed within a month. (maanmittauslaitos 2019, cited 16.02.2019)

2.8 Categories of Open Data

There are several types of Open Datasets but according to Paul Clarke. He categorized the different types of Open Data into four major categories (Clarke 2010, cited 16.02.2019).

1. **Historical Data.** This type of data includes events that happened in the past for example past forex rates, inflation rates in a country, the performance of an organization. Historical data can also include company past revenue earnings and stock prices, the performance of schools and rankings of universities.

2. **Planning Data.** This type of data includes information about projected or predicted events that will happen. This type of data includes information about new infrastructures such as buildings, roads, airports that are expected to be constructed in the future.

3. **Infrastructural Data.** Infrastructural Open Data type comprises of information about important services addresses of buildings, hospitals, police and so much more. In addition, information about maps, boundaries, company structures are also included.

4. **Operational Data.** This type of data gives real-time information and updates. Examples include transport schedules from buses, trains, and airports. Weather updates, accidents and any events that may be happening at a time.

2.9 Sources of Open Data

Open Data comes from various sources regardless of the type or category and eight key sources were identified that contribute to the large pool of Open Data. (Audsin & Vathana 2019, 1-4).

1. **Geo-Open Data.** This type of Open Data also known as Geographic or Geospatial data is used in the creation of maps and locations by using physical features such as mountains, rivers, lakes, buildings etc. Geodata is mainly used to show geographic locations and positions. Geodata is mainly used in displaying transport routes, for example, Google maps
and can also be used to show environmental data. Geodata is decoded using Geographic Information Systems (GIS) and can be downloaded in KML, shapefile and GeoJSON formats (Free GIS Data Sources 2019, cited 21.01.2019).

2. **Cultural Open Data.** This is a source of Open Data that is normally stored in historical places such as libraries, museums, and galleries. Examples include artifacts or regalia of ancient kingdoms. Od4h.org is an Open Data website that provides information about the history of cultural sites and their heritage. It also advocates for the discovery of abandoned cultural and heritage sites and connects ancient places to the digital world (Open Data 4 Cultural Heritage 2017, cited 21.01.2019).

3. **Science Open Data.** Open Data from this is basically centered about research and science or innovations and discoveries. This is one of the largest sources of Open Data. NASA earth observations provide information about the earth and the datasets include information about rainfall, solar insolation, global temperatures, and global warming, net radiation, sea surface temperatures and so much more. The image below shows the earth’s land cover classification mapping the year 2011. This information can be used by scientist to collect knowledge on how the earth is changing and what is causing the changes. By using these maps, scholars can study the trends of deforestation, urban growth, migrations, and desert expansion. This kind of information can be used by policymakers in respective countries to come up with solutions (NASA 2019, cited 17.02.2019)
4. **Financial Open Data.** This source of Open Data displays information about markets, business, investments, stock shares, bonds, inflation rates, forex rates. The information can be related to governments or private and public companies and organizations. An example is Yahoo Finance which gives information and data on markets and industries, currencies, commodities and forex rates. (Yahoo Finance 2019, cited 24.01.2019).

5. **Statistics Open Data.** Open Data from this source displays information from global and regional databases. The information can include population size, employment rate,
commodity trade statics, environmental statistics, financial statistics others. **UNdata** gives
global statics in different categories (Undata 2019a, cited 27.01.2019).

6. **Open Government Data.** This source of data is basically from government and its
agencies and this data should be free for anyone to access, use, modify or share it for any
purpose. Open government data includes data from research and statics, government-
reports, expenditure, environmental, population, crime rates, finance and markets,
geospatial data and data from other sources. (Open Knowledge International 2019d, cited
27.01.2019)

7. **Environmental Data.** This includes information about the environment for example
deforestation, deserts, pollution rates, natural disasters such as earthquakes, flooding,
waste disposal, fuel, and energy use. The UN data portal shows demographic statics of
countries and how they dispose of solid and liquid waste from household (Undata 2019b,
cited 27.01.2019)

8. **Weather.** The weather is a key source of Open Data because a lot of information is
extracted from it. The World Meteorological Organization (WMO) provides information
and data climate and weather changes. It also predicts and warns about disasters such as
cyclones, extreme heat conditions, droughts, and heavy rainfall. WMO data portal also
provides information about global warming and climate changes. (World Meteorological
Organzanzation 2019, ciiited 27.01.2019).

2.10 Formats of Open Data

The Open Data format is a method in which Open Data is organized in a pattern that is
understandable and readable to machines as well as human beings (European Data Portal 2019b,
cited 27.01.2019). Open Data is structured in different formats and users can use a format that is
appropriate and convenient for them to use.

1. **Comma-separated values (CSV).** This type of Open Data format is basically used to store
data in plain text formats. The data field are normally separated from the next by a comma.
CSV file formats are utilized when transferring large sets of data. CSV data formats can be
loaded onto excel and spreadsheet applications which can be downloaded by users. CSV
data formats require careful handling and accuracy since mistakes in recording cause
misinterpretation of the whole data (European Data Portal 2019b, cited 27.01.2019).
2. **Extensible Markup Language (XML).** This is an Open Data format that uses specific encoding rules. XML format is mostly used over the internet by software developers and can be used over excel applications and online application tools. XML allows exchange and transfer of data sets without changing the data structure (Open Knowledge International 2019e, cited 27.01.2019).

3. **JavaScript Object Notation (JSON).** This is a type of file format that is can be interpreted by humans and the data type is transported in pairs. JSON Open Data is basically easy to be interpreted by most programming languages and can be easily processed by computers. JSON Open Data format is widely used over the internet (Open Knowledge International 2019e, cited 27.01.2019).

4. **Resource Description Framework (RDF).** This Open Data format belongs to a family of international data standards that was created by the Wide Web Consortium (W3C) that allows sharing of data over the web. RDF uses web identifiers to locate data over the web, for example, HTTP URLs. RDF is used linking data from different sources over the web and is stored as XML and JSON (Open Knowledge International 2019e, cited 27.01.2019).

5. **Spreadsheets.** This is a type of data format that is used to store data in cells. Cells are pre-arranged in rows and columns and different cells can contain any type of data for example strings, date or text and numbers. an example is a data Cited from Microsoft excel (. _XLS), Oracle Discoverer Workbook (.DIS), Apache OpenOffice Calculation Spreadsheet (.SDC) and Google Drive Spreadsheet (. GSHEET) (Open Knowledge International 2019e, cited 27.01.2019).

6. **Text Document.** Open Data can be presented in document formats such as Microsoft Word documents (doc), Office Open XML (OOXML), Open Document Format (ODF) and Pocket Document Format (PDF). However, this type of format can only be used to store certain kind of data and it doesn’t consider keeping the data structure uniform and may be difficult for the computer to interpret the data if it is not logically entered, additionally, data cannot be entered automatically in text documents. It is therefore advisable to store text documents in templates because they are easier to transfer information in text document formats (Open Knowledge International 2019e, cited 27.01.2019).

7. **Plain Text documents(.txt).** This type of format is the simplest way of storing data and it is very easy for computers to read data in plain text format. All the data is stored as text files numeric values inclusive. Plain text documents do not require advanced applications and platforms to interpret or write them and they can be easily understood by human
beings. However, plain text data format does not have a standard way to format the data and it is not recommended to use it in large sets of data and lacks data integrity controls (Open Knowledge International 2019e, cited 27.01.2019).

8. **Proprietary formats.** Open Data can be represented in a proprietary format. This type of format is basically used by organizations, companies or individuals who prefer their data to be stored and shared in a systematic encoding structure. This type of format is owned and controlled by the company and they can make changes to the format. It is advisable not to share data in a proprietary format (Open Knowledge International 2019e, cited 27.01.2019).

9. **Scanned Images.** Open Data can be stored as images, but this type of method is not very convenient because it requires large amounts of memory to store images. Images can be stored in various formats and the most common file format used is Joint Photographic Expert Group (JPEG), Portable Network Graphics (PNG) and Tagged Image File Format (TIFF) (Open Knowledge International 2019e, cited 27.01.2019).

10. **Hypertext Markup Language (HTML).** Open Data can be stored and shared across various websites in form HTML format. HTML is the standard in which data and information can be stored over web pages. Data in this format can be easily downloaded and manipulated. There are various tools and applications that can be used to extract data from web pages for instance Yahoo Query Language (YQL) is a platform that can be used to query, filter, edit and join data from different websites to one place (Yahoo Developer Network 2019a, cited 20.01.2019).

11. **Databases.** Data can be stored and shared across databases and users can directly access the data by using query languages to extract the information they want. Databases are useful when storing large amounts of data and can be updated easily. In addition, they limit errors in data and data inconsistency. Databases also provide data privacy and advanced security policies for data (Open Knowledge International 2019e, cited 27.01.2019).

12. **ZIP.** This is a file format that uses archives to compress large amounts of data. They are normally used for storage of different files and directories that have been compressed or zipped. ZIP files can be used over different platforms, for example, Mac OS and Microsoft (Fileinfo 2019, cited 21.01.2019).

13. **Other Formats.** There are other formats of Open Data that can be used extensively, for example, Keyhole Markup language (KML) file format which is used in geographic annotations and visualization over the internet. Google Earth makes use of this data
format Shapefile Format (SHP) is another format of data that is used in Geospatial Vector Data for Geographic Information Systems (GIS). (Government Data Ireland 2019, cited 25.01.2019.)

2.11 Linked Open Data

Linked Data is a process of publishing structured data in a way that data is interconnected and can be accessed across different platforms on the web. Additionally, linked data can be defined as accessing data that is closely associated over the web, therefore, Linked Open Data is accessing linked data on the internet based on RDF and URIs standards of the web. The principle behind Linked Open Data is that data is more valuable and useful if it can be linked to other forms of data (W3C 2015a, cited 04.02.2019).

A five-star model has been used to set principles for linked Open Data. Table 1 below demonstrates the 5-star model of linked Open Data.

**TABLE 1. 5-star linked Open Data (W3C 2015b, cited 04.02.2019).**

<table>
<thead>
<tr>
<th>Number of Stars</th>
<th>Data Description</th>
<th>Example of Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>*</td>
<td>Data is available on the web in any format but with an open license</td>
<td>PDF</td>
</tr>
<tr>
<td>**</td>
<td>Data is available in a machine-readable format</td>
<td>Excel rather than PDFs</td>
</tr>
<tr>
<td>***</td>
<td>As in two stars and non-proprietary format</td>
<td>CSV instead of Excel</td>
</tr>
<tr>
<td>****</td>
<td>All the above plus one standard from W3C</td>
<td>RDF and SPARQL</td>
</tr>
<tr>
<td>*****</td>
<td>All the above plus linking your data to other people’s data to provide context</td>
<td>Related or interconnected data</td>
</tr>
</tbody>
</table>
2.12 Open Data Structures

Open Data appears in different file formats and can be presented depending on the type of data format. There are three main data structures.

1. **Tubular Structure.** This is the most available structure in which Open Data can be conveyed. This type of data structure organizes data into rows and columns. The tubular data structure is used in organizing linked data and non-related data which can be structured in CSV file format (European Data Portal 2019c, cited 27.01.2019).

2. **Hierarchical Data Structure:** This type of data structure is used in displaying associations between different data points. The hierarchical data structure is used organizing data in a tree-like structure of the hierarchically. An example is a company organization leadership structure of a family tree where you have grandparents, parents, children, grandchildren and so forth. JSON data format uses a hierarchical data structure (European Data Portal 2019c, cited 27.01.2019).

3. **Network Data Structure.** Network data structure enables links and associations between data in form direction. A clear example of the network data structure is a social network on Facebook or Instagram where someone has friends and friends have other friends. Internet or web is also an example of a network. One page can link you to different web pages (European Data Portal 2019c, cited 27.01.2019).
This chapter is divided into sub-chapters. The main chapter examines how Open Data can be applied in Education and learning environment. The main objective of this chapter is to analyze the value of Open Datasets and how they can be utilized in the education systems and their impact and benefits on teaching and learning. Furthermore, the challenges of using Open Data in teaching and learning environment is also discussed, new trends and innovations in open learning, important Open Data institutions and portals and finally, key skills that are required to manipulate Open Datasets.

3.1 Using Open Data as Teaching and Learning Resource

The value of Open Data has been comprehensibly discussed by several scholars, experts, articles and journals and the benefits of using Open Data have been mentioned. Whereas most of the attention has been centered around the economy, business and democratic governance, Open Data can be employed in so many other sectors where it can play a very important role. The main objective of this Thesis work is to discuss how Open Data can be used in the Education sector as an Open Educational Resource (OCR) particularly in the teaching and learning process. Large amounts of Open Datasets remain un-tapped, using Open Datasets can have a major impact on the education sector in several areas. The work in this chapter will discuss how Open Data can be utilized as an educational resource.

Open Data can create innovation and improves creativity in the teaching and learning process. By using Open Data, both teachers and students can improve their innovative skills by adapting new methods of teaching and learning respectively. A number of universities and schools have realized the importance of Open Data in fostering creative thinking and innovation in learners and teachers. Several instances of using Open Data to create new innovations exist, for example, the Irish Government is using competitions to come up with innovative ideas from both students and teachers on how Open Data can be used as an educational resource and the outstanding ideas are awarded cash prizes to develop their prototypes (Pandacan 2017, cited 10.02.2019). One of the winning ideas was “Our Raging Planet” an education simulation tool that allows students to use public data to predict possible natural disasters (Our Raging Planet 2018, cited 10.02.2019).
Open Data can be used to provide basic education for those who cannot access schools or educational resources such as textbooks. According to UN, more than 123 million young people lacked basic reading and writing skills, in addition, 57 million primary going children were out of school (UN Millennium Development Goals 2015, cited 13.02.2019). By utilizing Open Datasets, countries can meet the UN development goals to provide access to a meaningful and purposeful yet affordable education to their citizens.

Open Data is transforming the educational system from the traditional learning methods to 21st century learning and modern learning. Higher institutions of learning are producing students with relevant transversal skills and competencies, these are well-defined by UNESCO as being “critical and innovative thinking, interpersonal skills, intra-personal skills, global citizenship and media literacy” (UNESCO 2019a, cited 17.02.2019). These skills are very essential in today’s digital and information society. Advancing education systems are utilizing Open Data effectively in Open education to promote Competence-Based Learning or Education (CBL/E). CBE permits learners’ advancement and progress based on their skills and competencies. In this context, competence is the capacity to use personal knowledge, skills and social abilities in work and study situations (Competency Works, learning from the Cutting Edge 2019, cited 17.02.2019). Whereas traditional methods of learning are basically teacher centered CBL is a learner-centered approach that utilizes Open Education and Open Educational Resources to enable the development of students with the required transversal skills and competencies that can be used to solve societal needs.

Open Educational Resources (OER) are defined as “Instruction, Education and Research Materials in any medium, digital or otherwise, that exist in in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited limitations” (UNESCO 2019b, cited 17.02.2019). According to Jeremy Knox, OER and Massive Open Online Courses (MOOCs) are key themes in Open Education that are utilizing Open Data and technology to remove barriers in education and bring educational reforms (Knox 2013, 1-9). He further stresses that OER have been internationally recognized by bodies such as UNESCO and the European Union for their contribution towards open up Education. MOOCs began as simple course experiments in mathematics and engineering before making partnerships with major world universities (Forbes 2019b, cited 17.02.2019). In addition to Coursera, other successful players such as edX and Udacity are providing free high-quality Open Education.
A prominent promoter of “Open Content” David Wiley engineered a 4Rs framework which summarized the use of Open Data in Open Education environments (Opencontent 2007, cited 21.02.2019). “The four Rs” are;

1. **Reuse**: Use the work verbatim, just exactly as you found it.
2. **Revise**: Alter or transform the work so that it better meets your needs.
3. **Remix**: combine the verbatim or altered work with other works to better meet your needs.
4. **Redistribute**: Share the verbatim work, the reworked work, or the mixed work with others.

He stresses that understanding copyright permissions of Open Educational Resources is fundamental in unlocking and opening educational materials and this has an impact on open learning.

Open Education is defined as “resources, tools and practices that employ a framework of open sharing to improve educational access and effectiveness worldwide” (Open Education Consortium 2019, cited 19.02.2019). By using Open Data, stakeholders in the education sector can open up education by aiming at practices that promote the reuse of educational resources and fosters the sharing of these materials. In this way, both teachers and learners share the benefits of Open Education or Open Data.

Open Data is having an impact on the education systems and its usability and application is increasing every day. Advancements and improvements in digital communication technology as well as social media has led to a sharp increase in the use of Open Data in education. Open educational resources can be easily and freely shared across social media platforms. However, the impact of Open Data in education depends on the context in which they are used, applied, and executed. Similiarly the use of Open practices by learners and educators is complex, personal, and is also continually negotiated (Cronin 2017, 1-6).

Open Data is very vital to open research practices, it is an invaluable resource in research-based learning approaches. Research based-learning is defined as instruction and teaching activities that rely on scientific methods of inquiry. According to (Arzberger et al, 2016.) access to Open Data is essential in open scientific research. They recognize the need and value of Open Data in scientific data. They argue that Open Data and Open Access is the key to publishing valuable and quality scientific research and data. It promotes sharing of data and information, encourages discussions and opinions about research, encourages new research and lays foundation for improving theories and research methods.

Furthermore, Open Data is in promoting shared research data and information to improve on research findings. Several Open Data movements are facilitating researchers with relevant
knowledge, materials and advanced research tools. Researchers are openly sharing and collaborating with others globally (Borgmann & Rolando 2013, cited 19.02.2019) argue that data is the cornerstone of science. The ability to share, reuse, and combine data presents researchers to re-analyze their evidence and verify their results. They accentuate the importance of technology in research data distribution and collection.

By using research-based methods of instructions, educationalists can use common existing societal and world problems to develop skilled workers, competent and active students, who are self-aware and concerned citizens. Using Open Data in learning and teaching approaches has shown that learners can acquire several skills, competencies and attributes in the context of research based learning and competent-based learning. Many institutions are using Open Data as a skill development tool. Table 1 below demonstrates how Open Data can be used as a skill development tool at different levels of learning.

**TABLE 1. Data Manipulation Skills at Different Levels (Open Praxis 2015, cited 19.02.2019 ).**

<table>
<thead>
<tr>
<th>Skills/Level</th>
<th>Basic</th>
<th>Intermediate</th>
<th>Proficient</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Critical Thinking</strong></td>
<td>Students understand basic concepts of critical thinking</td>
<td>Students can use data to verify information from the media</td>
<td>Students can Analyze phenomena from their region using data and write reports critically analyzing solutions.</td>
<td>Students are able to develop and present complex, evidence-based arguments in key academic formats</td>
</tr>
<tr>
<td><strong>Data curation Skills</strong></td>
<td>Learners can arrange datasets in simple files.</td>
<td>learners can establish distinctive sources of datasets and organize them in databases</td>
<td>learners can use electronic mechanisms for data curation and share it with others</td>
<td>learners can develop databases and automate the process to organize and combine datasets, and insert metadata into the directories to ease access to the resources</td>
</tr>
<tr>
<td><strong>Data information management</strong></td>
<td>Students can classify datasets</td>
<td>Students can select datasets from various</td>
<td>Learners can extract, clean and study data from</td>
<td>Students clean and configure data in different</td>
</tr>
<tr>
<td>skills</td>
<td>from different sources</td>
<td>portals in diverse formats</td>
<td>different data sources building a single dataset</td>
<td>formats analyze it creating complex datasets.</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Data visualization skills</td>
<td>Students can design graphics and charts</td>
<td>Students can use online software to develop simple infographics</td>
<td>Students can use graphic design software to develop infographics</td>
<td>Students can use data visualization techniques to present their findings using complex statistical modelling</td>
</tr>
<tr>
<td>Data Mining Skills</td>
<td>Students can locate CSV files on the internet</td>
<td>Students can extract datasets from PDFs</td>
<td>Students can extract datasets from different sources</td>
<td>Students can use complex methods for developing datasets</td>
</tr>
<tr>
<td>Research skills</td>
<td>Students understand the scientific method and are familiar with the concepts of quantitative and qualitative methods</td>
<td>Students can structure them research and apply different techniques to obtain results</td>
<td>Students can replicate experiments and studies following research methods explained in the literature</td>
<td>Students can compare data and information from different data sources and research papers and replicate experiments and studies to produce new research findings</td>
</tr>
<tr>
<td>Statistical skills</td>
<td>Students can perform basic statistical operations such as averages, media and median</td>
<td>Students can perform statistical operations using clusters, standard deviations, significance, chi square, correlation or regression analysis</td>
<td>Students can use data modelling techniques for different statistical methods such as forecasting to predict future events</td>
<td>Students can write queries in order to perform complex statistical analysis functions and create models and complex graphs and visualizations</td>
</tr>
<tr>
<td>Data analysis skills</td>
<td>Students can analyze data using quantitative and qualitative methods</td>
<td>Students can analyze data using quantitative and qualitative methods</td>
<td>Students use proficiently software for data analysis which are relevant for their own disciplines</td>
<td>Students can present complex reports based upon data analysis in the form of research papers or posters</td>
</tr>
</tbody>
</table>
3.2 New Trends in Open Education

Education has become the main priority in both developed and developing nations. Many countries are formulating policies that can provide quality and sustainable education to their citizens. Education systems are being engineered towards providing solutions to current and future educational problems such as demographics problems, cultural and social differences, technological and economic problems. With the world changing every day, it is vital that learners are equipped with life-long skills and continuous learning to enable them to adopt to the changing world. For this reason, a number of trends and innovations are being formulated to enhance Open education and open learning.

More information and data are being released every day and more knowledge is being created. More information is becoming digital and openly accessible. There is need to take advantage of this freely accessible information and turn it into knowledge. Various institutes and movements for example OA, MOOCs, OCW, are advocating that information be open to the public and be used as an Open Educational Resource. These learning communities are providing educational solutions to large groups of people, allowing global exchange of knowledge and information. With more research and innovations in technology, more online courses are being created, this is part of future learning.

Globalization is setting new trends in so many sectors the education sector inclusive. Globalization is the growing interdependence of the world's economies, cultures and populations, brought about by cross-border trade in goods and services, technology, and flows of investment, people, and information (United Nations 2005, cited 23.02.2019). Therefore globalization will have an impact on the education sector in a number of ways. Whereas the traditional education systems were literally centered in individual countries, cross-border education is increasing every year and more universities are opening up their courses to learners outside their country boundaries. Collaborations between different universities is increasing especially in the field of research,
Science and technology. Globalization has removed educational boundaries and enabled migration of learners and teachers to different countries, an example is the Erasmus Programme in Europe.

Technology is changing modern education. Emerging trends and innovations in Open Learning systems are creating a shared learning environment where content is freely accessible and sharable to everyone. Billions of free open content and information online, enhanced web-learning infrastructure and knowledge-sharing culture are three factors that are opening up education to everyone (Curtis, 2009). Continuous development and advancement in technology will always have an impact on the education sector. Recent technological innovations are promoting openness and flexibility to Open Educational Resources.

The rise in digital content and hand gadgets such as tablets, iPads, smart mobile phones, e-book readers is having an impact on the way people teach and learn. Traditional textbooks are being replaced by E-textbooks commonly known as Open Digital Textbooks (ODTs) as Open Educational Materials. ODTs are more flexible than paper books for a number of reasons, first and foremost, material in ODT can be easily converted to audio files enabling teachers and students to listen anytime anywhere, secondly, ODT are comparatively cheaper than traditional text books, and, finally, ODTs are quite flexible, accessible, interactive and extensive (Allen & Seaman 2016, 17-32). In 2012, Ohio introduced an eTextbook pilot program for schools to use digital textbooks and electronic educational content in classrooms, and for professional development of the teachers as well as sharing content with other schools across the state (Ohio Department of Higher Education 2015, cited 19.02.109). The project identified the following benefits of using digital open content;

- Improved technology skills for students
- Access to updated curriculum
- Improved instructional materials
- Opportunity to differentiate and personalize learning
- Greater student engagement and learning
- Opportunity to be more innovative in classrooms and
- Improved skills in technology for teachers.

In summary, Open Digital Textbooks have helped in opening up education systems, connecting teachers with learners as well as fostering the growth of individual skills.
The emergence of blogging and social media sites such as Twitter, Facebook and Youtube is opening up education, creating opportunities for learners and making study material sharable and accessible. Communities and educational institutions can engage professionally via social media sites, people have the ability to voice their opinions and suggestions via blogs. Similarly, researchers can share their research data on social media sites especially Twitter is being used professionally by data scientist and researchers. Social media sites have become vital tools for communication and engagement of different experts. In summary, Social media is creating networking opportunities and promoting career development in the field of education,

3.3 Challenges and Bottlenecks to Open Education

Whereas Open Data is a vital resource in revolutionizing the education sector, several challenges exist on how to effectively use it as an Open Education Resource. We need to understand the dynamics of Open Data sets, and technology and skills should be available to identify relevant Open Datasets from raw data. First and foremost, Open Data is part of data as a whole, there are different players that are contributing to the general pool of Open Data. Therefore, it is important to note that Open Data co-exists in an eco-system with multiple components that are dependent on one another. Although data is a key player in this eco-system, other factors influence and have an impact on the type and quality of data released or opened up for educational purposes. Things like government policies, data skills, internet, licenses and technology can have an impact on Open Data. These factors influence on the usability of Open Data and subsequently the availability of Open Data resources.

The amount Open Data being released is growing every day. It is estimated that 2.5 quintillion bytes of data new data is released daily, and the amount of data storage will skyrocket to 40,000 exabytes in 2020. Forbes estimates show that Google processes 3.5 billion searches daily, 527,676 photos are uploaded on snapchat every minute and 4,146,600 videos are watched on YouTube every minute. In addition, Facebook which is the largest social media site with an average of 1.5 billion people active on the site every day and 300 million photos uploaded per day (Forbes 2018a, cited 19.02.2019). The large volumes of data being released require explicit data curation and handling skills which some learners and teachers do not possess to select quality and useful information
that is resourceful. Some of the resources lack quality and are not relevant to the current education trends.

The labor market and the economic environment in most countries is shaping open learning. Most countries are leaning towards knowledge-based economies, therefore low qualified workers are less likely to break into these labor markets. Knowledge-based economies require high skilled personnel who can perform complex tasks. This implies that people have to continuously learn and update their skills, similarly, Open education content need to change to meet the current needs of people who can compete for the global job market. The economic situation requires that employees continuously up-skill and re-skill in order to increase their job opportunities because there is disconnection between job requirements and people’s skill. The European Centre for the Development of Vocational Training highlights the need for lifelong learning especially for adult workers (CEDEFOP 2012, cited 19.02.2019).

The rise in Open Data volume also raises other issues such as sustainability and data preservation methods. Most of the Open Educational Resources are in digital formats which makes them more volatile. This information needs backing up and storage on secure servers. Most of the Open Educational Resources require internet to access the information, people with limited internet access may not be able to get the right information. This will create a data divide since those with access to internet will continue learning and acquiring new skills.

Open Educational Resources are mostly available in English language, this makes it difficult for people who do not understand the English language. English monopoly in Open Educational Resources is a limiting factor. English is the 3rd most spoken language in the world behind Chinese and Spanish (Babble 2018, cited 19.02.2019). English dominates the International stage because it is easy to speak and the influence of UK and USA in world affairs make it dominant over other languages. Therefore, English language continues to dominate the internet and most educational materials are published in English.

Open Education Resources do not consider enough people with disabilities. People with physical impairments such as blindness cannot see or read something on a computer, similarly, people with hearing problems may not be able to listen to audio recordings or files. However, recent innovations in technology are creating possibilities where all people can effectively use the web and internet resources without limitations. However, several challenges still exist in opening up doors to open teaching and learning environments. Some of these issues are technical, economic, social and legal limitations. Advancement in technology means that learners and teachers need to constantly train and equip themselves with modern digital skills of handling Open Datasets. Additionally,
problems like plagiarism and fake accreditation from some providers of open learning courses are also a constant threat to the quality of Open education, and, there is need for recognition of awards achieved from Open learning platforms since most of them are not recognized by some institutions and employers.

Finally, whereas so many hurdles are hindering the progress of open learning and teaching, we have to acknowledge that internet and Open Data have opened so many doors to open content and technology is changing the way we teach and learn. Open Educational Resources are saving learners lots of money by providing free and cheap education, it is engaging learners with teachers. The use of Open Data is likely to grow tremendously if all stakeholders in education can understand the value of Open Data and support the development of key skills and technology needed to utilize the Open Datasets that can be used as Open Educational Resources.

3.4 Open Data Portals and Institutions

This section highlights some examples of organizations and Open Data portals that are involved in the promotion and use of Open Data. They provide relevant information concerning Open Data use, tools and skills needed to utilize Open Data. They include both public and private organizations.

3.4.1 European Data Portal

This Open Data portal is used to access public data sets in the European Union agencies and institutions. It is also used in training researchers and data scientist in data management skills. The European Union Data portal, [https://www.europeandataportal.eu/en](https://www.europeandataportal.eu/en) has been fundamental in promoting Open Data strategy in the EU zone and the world at large. (European Data Portal 2019d, cited 17.02.2019.)
3.4.2 **Avoindata.fi**

This is an Open Data platform with over 1000 datasets from private individuals, business and public administration in Finland. It can be used for visualization and development of applications in Open Data. The platform, [https://www.avoindata.fi/en](https://www.avoindata.fi/en) also has an English platform for people who don’t understand Finnish language (Avoindata 2019, cited 15.02.2019).

3.4.3 **Helsinki Region Infoshare**

This is an Open Data platform for Helsinki Metropolitan area i.e., Helsinki, Espoo, Vantaa and Kauniainen. It has an Open Data portal, [https://hri.fi/en_gb/](https://hri.fi/en_gb/) that informs the public about employment opportunities, business deals, procurement data, tenders and governance matter in Finland (Helsinki Regional Infoshare 2019b, cited 15.02.2019).

3.4.4 **Kaggle**

Kaggle, [https://www.kaggle.com/](https://www.kaggle.com/) is an Open Data science platform used for publishing data in research. It provides datasets, tools that can be used in reporting and analysis of research reports. It is a platform for different groups of data scientists and machine learners. It allows exploration and publishing datasets and build data models that can be used freely. The platform also offers education and data science competitions (kaggle 2019, cited 15.02.2019).

3.4.5 **World Bank Data Portal**

3.4.6 6Aika

This an Open Data portal, https://6aika.fi/ provides information about research, jobs, health, environment and the economy in Finland’s six biggest cities. It is an open innovation platform and open participation relationship that helps in promoting Finnish smart city model, data openness and exploitation as well as service development (6aika 2019b, cited 15.02.2019).

3.4.7 Open Data Institute (ODI)

The ODI, https://theodi.org/ was co-founded in 2012 by the inventor of the web Sir Tim Berners-Lee and artificial intelligence expert Sir Nigel Shadbolt to show the value of Open Data, and to advocate for the innovative use of Open Data to affect positive change across the globe. ODI works with companies and governments to build an open, trustworthy data ecosystem (Open Data Institute 2019, cited 15.02.2019).
With the amount of global Open Data accumulating every second, robust and reliable tools that can handle huge amounts of Open Data are needed in the education sector. Several open and licensed tools are available and new ones are being created to handle Open Data. The advancement and innovations in the education systems for example, shifting from traditional classrooms to web-based learning, necessitates the need for techniques and tools that can assist in learning and conducting research.

Open Data handling tools may differ on the usability depending on the type of Open Data being handled. These tools are needed for preprocessing and analysis of the Open Data, additionally, technical skills are also required when working with Open Data. Open Data manipulation tools and skills are very vital in Open Education for visualizing and analyzing the data. The work of this chapter is to introduce the type of tools and skills required to work with Open Data.

4.1 Apache Hadoop

“The Apache™ Hadoop® project develops Open-Source Software for reliable, scalable, distributed computing” (Apache Software Foundation 2018a, cited 15.02.2019). Apache Hadoop was basically created due to the need to process big data. Hadoop is a brain child of the Apache Software Foundation which provides Open software and funds Open software projects for the public good. The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. (Apache Software Foundation 2018a, cited 15.02.2019). The Apache Hadoop framework is composed of mainly four major frameworks (Apache Software Foundation 2018a, cited 15.02.2019).

- Hadoop Common: This framework consists of libraries and utilities required by other modules of Hadoop
- Hadoop Distributed File System (HDFS): This is a file-system used for data storage on commodity machines and provides quick access to application data.
- Hadoop YARN: This is a resource management framework used mainly in cluster management and application scheduling.
• Hadoop MapReduce: This a programming platform used for processing large sets of data

4.2 Yahoo Query Language

Yahoo Query Language is platform that enables developers to query, filter and combine data across the web. YQL uses similar syntax like SQL to extract the data from the web in form of JSON or XML formats. A number of developers are using the platform to develop different applications. The API enables data to be collected from different web sources and can be converted into reusable Open Data on the web. (Yahoo Developer Network 2019b, cited 16.02.2019.)

4.3 Google charts

Google charts are powerful web and mobile applications that make use of JavaScript libraries to visualize data with interactive charting capabilities. The charts created are open source and can be customized according to one’s needs. Google charts are compatible across all platforms, android and iOS without the necessity of additional plugins. Additionally, they can also be used across all browsers including older versions of internet explorer (Google Developers 2019, cited 16.02.2019).

4.4 ArcGIS

ArcGIS is a data manipulation tool that is used in analysis and reporting of geographic data. ArcGIS can be used to connect people, locations and data by using interactive maps. It can be used as a desktop version or an online tool (Esri 2019a, cited 16.02.2019). ArcGIS provides tools for mapping and spatial reasoning, visualizing and analyzing the data. The data can be shared in form of maps, insights, apps or reports. It has the following key features used in data manipulations, spatial analytics, imagery and remote sensing, mapping and visualization-time GIS, 3D GIS, Data Collection and Management (Esri 2019b, cited 16.02.2019).

4.5 Quadrigram

Quadrigram is a drag and drop tool editor that is used in visualization of data in any format. This tool does not require any coding skills to manipulate the data. Data can be loaded into the editor
from cloud services such as google drive in form of XLS or CSV files and the visualized data can be shared in form of articles, analysis reports or presentations (Quadrigram 2019, cited 10.02.2019).

4.6 MongoDB

MongoDB is an open source cross-platform database program written in C++ and JavaScript that is used in manipulation of data using cloud services. MongoDB flexibility document model enables working with data anywhere anytime. It can be used to build apps, allows data migration from mainframes to public cloud. The platform allows users to have a full-time managed cloud database or serverless platform according to their choice, and a data visualization tool for their data (MongoDB 2019, cited 17.02.2019).

4.7 QGIS

QGIS is a free open source data manipulation tool used in Geographic Information Systems. QGIS can be used to create, edit, visualize and publish Geospatial information across different platforms. It can be used on windows, Mac, Linux and BSD (QGIS 2019, cited 17.02.2019).

4.8 Apache Storm

Apache storm is a free and open source distributed Realtime computation system that allows manipulation of large sets of data coming from various sources. Storm is easier to use and can be used with any programming language. It can be employed to handle Realtime data analytics, machine learning and computation processes (Apache Software Foundation 2018b, cited 20.02.2019).

4.9 Skills required to work with Open Data

The importance of Open Data in the education sector necessitates that people who work with Open Data have the necessary e-skills to manipulate the data. According to the European Open Data portal, to create value from Open Data, both soft skills and hard skills are essential. Hard skills are
teachable abilities that can be defined and measured, for example knowledge about a subject matter, mathematic and statistical knowledge as well as technical skills. Additionally, soft skills that are needed to work with Open Data include communication, storytelling, collaboration, curiosity, communication and creativity (European Data Portal 2019e, cited 20.02.2019).

The Type of e-skills or electronic skills needed to work with Open Data may vary depending on the type of data being handled and the person working with the data, for instance, the e-skills for a business professional are different from those of a programmer. According to European Data Portal, E-skills or electronic skills include those needed to make use of information and communication technologies (ICT) as well as those required to apply and develop them. The European e-skills forum identified three types of e-skills;

1. **ICT practitioner skills** are the capabilities required for researching, developing, designing, strategic planning, managing, producing, consulting, marketing, selling, integrating, installing, administering, maintaining, supporting and servicing ICT systems (European Commission 2019b, cited 10.02.2019).

2. **ICT user skills** are the capabilities required for the effective application of ICT systems and devices by the individual. ICT users apply systems as tools in support of their own work. User skills cover the use of common software tools and of specialized tools supporting business functions within industry. At the general level, they cover "digital literacy" (European Commission 2019b, cited 10.02.2019).

3. **E-business skills** correspond to the capabilities needed to exploit opportunities provided by ICT, notably the internet; to ensure more efficient and effective performance of different types of organizations; to explore possibilities for new ways of conducting business administrative and organizational processes; and or to establish new businesses (European Commission 2019b, cited 10.02.2019).

When dealing with data, a number of skills are required in addition to the technical skills that are mandatory.

- **Technical skills**: Are defined as the knowledge and abilities needed to accomplish mathematical, engineering, scientific or computer-related duties, as well as other specific tasks. Technical skills are essential in understanding how tools process the data, programming algorithms and machine language which are very crucial when handling complicated sets of data. Furthermore, technical skills are needed to develop standards and interoperability for Open Data and also in the development of Open Data platforms or databases (European Data Portal 2019e, cited 20.02.2019).
- **Mathematics and Statistical skills**: These are defined as knowledge of the science that deals with the collection, analysis, and interpretation of numerical data, often using probability theory. Statistical knowledge is fundamental in understanding how to analyze data, cleaning data and using common data tools such as Excel and SPSS. Statistical skills are used in creating metadata, using taxonomies and ontologies, data mapping and handling large sets of data (European Data Portal 2019e, cited 20.02.2019).

- **Analytical and Personality skills**: These can be explained as the ability to work with data, that is, to see patterns, trends and things of note and to draw meaningful conclusions from them. Analytical skills can assist in identifying the problem and finding a solution for that problem. In addition, the person working with data should have self-drive or motivation to work and should be able to communicate the results of his or her work (European Data Portal 2019e, cited 20.02.2019).

- **Legal skills**: A person working with open data should be well versed with copyright laws and any legal aspects relating to the use of Open Data. Some Open Data may be protected under certain circumstances.

- **Problem solving skills**: The process of working through details of a problem to reach a solution. Problem solving may include mathematical or systematic operations and can be a gauge of an individual's critical thinking skills.

- **Storytelling**: Storytelling is a method of explaining a series of events through narrative. It is used as a tool to illustrate an otherwise difficult concept or a point. A person dealing with Open Data should be able to explain the insights and outcomes drawn from datasets, make interpretation of the data as well as making written and oral reports about data (European Data Portal 2019e, cited 20.02.2019).

- **Collaboration**: To work with another person or group in order to achieve or do something. Effective method of transferring ‘know how’ among individuals. It is essential that a person dealing with Open Data can work with different people in working environments in a professional way (European Data Portal 2019e, cited 20.02.2019).

- **Communication**: Two-way process of reaching mutual understanding, in which participants not only exchange information, ideas, feelings but also create and share meaning. Handling Open Data requires engaging with other people and one should be
able to communicate with technical and nontechnical people about data (European Data Portal 2019e, cited 20.02.2019).

- **Creativity:** The ability to transcend traditional ideas, rules, patterns, relationships, or the like, and to create meaningful new ideas, forms, methods, and interpretations.

In conclusion, working with Open Data requires and involves a number of skills but the core skills for working with Open Data include Statistical, analytical and technical skills. It is difficult to possess all the skills needed to handle data, therefore, working in teams and extra training at school and work is needed to improve on data handling skills.
5 DATA VISUALIZATION IN OPEN DATA

In a data driven economy, more data and information are being open by government and private agencies. The number of datasets readily available in the public domain is quite massive. Large sets of data are being opened up in the education, health, finance, environment and other sectors. Whereas data is readily available especially in the education sector, it is important to note that a reasonable percentage of the population cannot use and benefit from this data because they lack the essential skills, knowledge and tools to analyze the data. Data Visualization can provide solutions to these problems by providing tools that can communicate and make sense out of large sets of Open Data.

5.1 What Is Data Visualization

Data visualization is the demonstration of information and ideas in graphical arrangement or graphical format. Data Visualization can be defined as a method or form of communication where data and information are conveyed to users in form of graphics or visual arrangement that can be easily interpreted and recognized.

Large amounts of Open Data are produced by different organizations and this vast amount of data is very difficult to be interpreted and analyzed very easily especially if this data is text format. Data visualization can assist in the extraction of the most useful information from vast amounts of Open Datasets to create associations through advanced data analysis visualization methods.

Data visualization is based on the principle that the human brain responds to visual information faster and better than any other form of stimuli, in fact, the human brain processes visual information 60,000 times faster than information in text format (Oracle Data Visualization 2019, cited 13.02.2019). The aim of Data visualization in Open Data is, therefore, to present information and data in very simple, clear and logical format by reducing cognitive effort while increasing information display in form of images, diagrams, simulations, maps, graphs, and other illustrations. Furthermore, society needs many qualified personal in the field of data visualization, therefore, data visualization is playing an important role in the education sector by providing several career possibilities such as data analysts, business intelligence analysts and data visualization engineers.
5.2 Significance of Data visualization in Open Data and Education

Data visualization is fundamental in interpreting and understanding Open Datasets that are used for educational purposes. There some alternative methods to interpret and analyze data for example descriptive statistical methods such as mean, median and range, however, these just describe the data. Data visualization provides a meaningful and a proper insight into Open Data which statistical methods cannot provide. As new information from Open Data keeps flowing and increasing, visualization tools can be used to manage and organize the new information.

Stakeholders in the education sector can organize Open Datasets faster and come up with better decisions on how to use the data. Data visualization helps in understanding and interpreting the large quantities of Open Data using visual representation of the data. The visual trends can be used in saving resources such as money and time that can be spent by using traditional methods of data analysis for example spreadsheets. In addition, data visuals can be shared and distributed over the web unlike spreadsheets. Likewise, data visualization can quickly identify data outliers that cause discrepancies in the data, these can be easily removed and eliminated from the data.

5.3 Data Visualization Techniques

Data visualization techniques refers to set of tools or applications that are used to represent or visualize data in a graphical format. The importance of data visualization techniques in Open Data is for people to have a better and clear understanding of the data. There are various tools and techniques used in data visualization, however some are not appropriate for use in Open Data sets, additionally, some data visualization techniques may be misleading with or without intention by manipulating data scales leading to wrong insights. For example, small differences can be manipulated to look big, this can lead to wrong decision making.
5.3.1 Line Charts

Line charts are mainly used to display association of one variable to another. Line charts are very functional when showing trends of events in a certain period or when comparing multiple items.

The line chart below compares the employment and unemployment rate in Finland from 2008-2018 in the months of November.

![Line Chart](Image)

**FIGURE 3. Employment Rate in Finland between 2008-2018**

*(StatisticsFinland 2019a, cited 20.02.2019.)*

5.3.2 Bar Charts

Bar charts are used in the comparison of quantitative groups of Open Data. Data is visualized either in a horizontal or vertical bar and the height of length of each bar representing the quantitative value of the data. Bars can have different colors to create visual variance.

The chart below shows Consumers’ expectations concerning their own and the Finnish economy.
5.3.3 **Pie & Donut Charts.**

Pie charts are a circle that has been divided into different sectors with varying sizes and each sector represents a certain percentage of the whole chart. Donut chart is like the pie chart, but it has space left in the center. These types of charts are not the best choice for data visualization especially if the data is in large quantities because it is very difficult to interpret the data, estimating the areas of the sectors and comparing many large visual angles. In addition, a comparison of sections, sectors or slices on the pie chart that is similar in size but not next to each other is complicated.

Pie and Donut charts are effective when visualizing information in limited components’ and percentages, figures and texts are included in the values to describe it. The Pie chart below shows total expenditure on pensions in Finland in 2017.

**FIGURE 4. Consumers’ Expectations of their Own and Finland’s economy**

5.3.4 Word Clouds

Open Data is being released in very large amounts, it is, therefore, challenging and difficult to visualize both semi-structured and un-structured Open Data. Some type of Open Data demands modern advanced techniques to visualize it. Word cloud visual is a modern method of data visualization where text data is represented by portraying the keywords in the text body. The size of the words is visualized depending on how much they appear in the text. Word clouds use visual analytics to make relationships and associations which are used in demonstrating the major topics within a text. In the picture below, a data visualization tool has been used to visualize some of the text,

https://www.jasondavies.com
5.3.5 Network Diagrams

A Network Diagram sometimes referred to as Network Map or Network Graph is a visualization technique that is used in the visualization of semi-structured and unstructured Open Data. Network diagrams are used in visualizing the connections and relationships between different actors on a network in form of nodes. The relationship or association within a network can be a business relationship, family, friendship or any other form of relationship. The nodes are shown as points and ties are expressed as lines on the network.

Network diagrams are utilized mostly by business and organizations to analyze customer behavior using social networking sites. Network diagrams are also used by law enforcement organizations to track connections between different suspected criminals. Figure 7 below shows the Airbus network map for Finnair & other airlines.
5.3.6 Correlation Matrices

A correlation matrix is visualized as a form of a table that is used to distinguish between different variables from large quantities of related Open Data. Correlation matrices are used in data visualization due to their ability to summarize and create associations of data from big datasets. Color codes are used to show the relationships between different variables, dark or strong colors indicate a very strong relationship between variables (sas 2019, cited 23.02.2019). Figure 8 below shows the correlation Heatmap of wine characteristics.
5.3.7 Geo-Mapping

Geographic Information Systems (GIS) are used to capture, analyze and visualize data. By using GIS, visual patterns can be drawn from data especially in Open Data where the interpretation of the data can be problematic and complex. GIS systems are being used in location Intelligence Systems (LI) to collect and organize geographic data of different places or areas, this data is in turn used by Business Intelligence to make decisions and create business opportunities Geographic data can be used by real estate developers, law enforcement agencies, environmental specialists, politicians, urban planners, and educations by visualizing map data to draw decisions.
FIGURE 9. GIS location of Oulu University Hospital (Google 2019, cited 02.03.2019).
Open Data is a vital resource and an important raw material for the education sector. This thesis work has demonstrated the impact of Open Data on the education sector particularly on the fields of teaching, learning and research. New innovations in Information and digital technology have increased the applicability of Open Data in the teaching and learning environments by enabling more openness and access to study and research material. Additionally, technology has removed certain boundaries and barriers to education, connecting learners, teachers and researchers together. Furthermore, Open Data is reducing costs in education, it is providing opportunities for people to access education from any part of the world and is promoting skill development of the learners.

However, using Open Datasets for teaching and learning purposes faces a number of challenges. Open Datasets exist in large volumes, therefore identifying useful and relevant information is quite a cumbersome process. It requires a combination of key factors such as data mining skills, right tools and technology to mine data which learners, teachers or researchers may not possess. Similarly, Open Access to Datasets may also be affected by government policies, legal, social and economic constraints.

Finally, benefiting from Open Data requires the right tools and skills to handle Open Datasets. The type of tools and technology needed to manipulate data will depend on the type of Open Data being handled. Tools are needed for preprocessing and analysis of the data and making insights into the processed data. It is important to note that getting value from Open Data requires both soft and hard skills. Hard skills or technical skills include statistical skills, mathematical and programming skills to mention a few while soft skills include proper communication and presentation skills.
7 REFERENCES

Clarke. (2010). There’s data, and there’s data. Cited 16.02.2019,
Competency Works – Learning from the Cutting Edge. (2019). Cited 17.02.2019,
https://www.competencyworks.org/
Crnkovic. (2013). What is the meaning of data, information, and knowledge? Cited 02.01.2019,
https://pdfs.semanticscholar.org/d04c/f02cf43bfb27245e2ec8b3c0b0fcf48837da.pdf
https://er.educause.edu/articles/2017/10/open-education-open-questions
https://www.esri.com/en-us/arcgis/about-arcgis/overview
https://livingatlas.arcgis.com/
http://www.europeaneskills.eu/
Han & Fei. (2011). Data mining: Concepts and techniques. (pp. 7-10).
https://hri.fi/fi/
Helsinki Regional Infoshare. (2019b). Open Data Service: Public information in the Helsinki Metropolitan Area for recovery. Cited 15.02.2019,
https://hri.fi/fi/
History of open data. Cited 21.01.2019,
http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data
Jason Davies. (2019). Cited 19.01.2019,
https://www.jasondavies.com/
kaggle. (2019). Datasets. Cited 15.02.2019,
https://www.kaggle.com/datasets
http://www.success.co.il/is/dik.html
MongoDB. (2019, 2 1). Why MongoDB. Cited 17.02.2019,
https://www.mongodb.com/
NASA. (2019). Land Cover Classification. Cited 17.02.2019,
https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MCD12C1_T1
https://od4ch.org/
https://theodi.org/
Open Data Tool Kit. (2019). Open Data in 60 Seconds. Cited 02.01.2019,
Open Education Consortium. (2019). What is Open Education? Cited 19.02.2019,
https://www.oecconsortium.org/about-oec/
https://opengovdata.io/2014/8-principles/
https://blog.okfn.org/2013/10/03/defining-open-data/
Open Knowledge International. (2019b). Introduction. Cited 05.01.2019,
http://opendatahandbook.org/guide/en/introduction/
Open Knowledge International. (2019b). Open Definition. Cited 09.01.2019,
https://opendefinition.org/od/2.1/en/
http://opendatahandbook.org/guide/en/what-is-open-data/
https://opendefinition.org/government/
Open Knowledge International. (2019e). File Formats. Cited 27.01.2019,
Open Source Initiative. (2018). Cited 05.01.2019,
https://opensource.org/
Oracle Data Visualization. (2019). Cited 13.02.2019,
http://ourragingplanet.co.uk/
http://parisinnovationreview.com/articles-en/a-brief-history-of-open-data
Pollock. (2011). Open Data: a means to an end, not an end in itself. Cited 17.01.2019,