Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Technologies

Diploma Thesis

Open Data for Ethiopia

Author: Leul Aberra Challa

Supervisor: Ing. Miloš Ulman, Ph.D.

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

B.Sc. LEUL CHALLA

Informatics

Thesis title

Open data for Ethiopia

Objectives of thesis

The main goal of the thesis is to determine how open government data can facilitate the efforts of meeting the development goals of Ethiopia by studying the impact of data openness on the community.

Partial goals of the thesis are such as:
- to make a review of the current state of Ethiopia in terms of public access to government data,
- to analyse available data sets and to evaluate their level of openness,
- to analyse and design a new open data application for agricultural sector in Ethiopia.

Methodology

The methodology used for the thesis is based on the study and analysis of open data initiatives in the world and particular case studies in Africa. In the practical part, the analysis and design of open data application will be done by using various software engineering methods. Recommendations on how to achieve an open government in Ethiopia will be stated based on the practices of the successful case studies.

The proposed extent of the thesis

60 – 80 pages

Keywords

Open data, Open Government, Data Portal, Government data, Right to information, Data for development

---

Recommended information sources

Attard, Judie et al. ”A Systematic Review Of Open Government Data Initiatives”. Government Information Quarterly 32.4 (2015): 399-418. Web.

Goldstein, Brett, Lauren Dyson, and Abhi Nemani. Beyond Transparency. San Francisco, Calif.: Code for America Press, 2013. Print.

Opendatahandbook.org,. ”The Open Data Handbook”. N.p., 2016. Web. 20 Feb. 2016.

Tauberer, Joshua. ”Open Government Data: The Book”. Opengovdata.io. N.p., 2016. Web. 20 Feb. 2016.

Yannoukakou, Aikaterini, and Iliana Araka. ”Access To Government Information: Right To Information And Open Government Data Synergy”. Procedia – Social and Behavioral Sciences 147 (2014): 332-340. Web.

Expected date of thesis defence

2017/18 WS – FEM (February 2018)

The Diploma Thesis Supervisor

Ing. Miloš Ulman, Ph.D.

Supervising department

Department of Information Technologies

Electronic approval: 18. 10. 2016

Ing. Jiří Vaněk, Ph.D.

Head of department

Electronic approval: 24. 10. 2016

Ing. Martin Pelikán, Ph.D.

Dean

Prague on 28. 11. 2017

Declaration

I declare that I have worked on my diploma thesis titled "Open Data for Ethiopia" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the diploma thesis, I declare that the thesis does not break copyrights of any third person.

In Prague on 29.11.2017 _____

Acknowledgement

I would like to thank Ing. Milos Ulman, PhD for his continuous advice and support during my work on this thesis.

Otevřená data pro Etiopii

Souhrn

Vlády jsou zodpovědné za sběr, uspořádání a šíření údajů (dat) týkající se různých aspektů jejich země. Údaje mohou být zpřístupněny veřejně nebo na základě oficiální žádosti. V současnosti je stále větším trendem vlád a institucí tyto údaje zpřístupňovat. Zpřístupnění vládních údajů není pouze „módní záležitost“, ale velmi důležitý zdroj ve vytváření nových obchodních hodnot. Tato práce analyzuje význam zpřístupnění údajů a jejich důsledky pro rozvoj. Diskutuje o pohybu těchto údajů, o tom jak zpřístupnění dat podnítit, dále o poptávce (po zpřístupnění údajů) a zapojení občanů. V rámci dopadu zpřístupnění údajů na rozvoj byly prozkoumány případové studie z Afriky. Studováno bylo zpřístupnění údajů ze strany etiopské vlády, etiopského portálu s těmito údaji, jeho přínosy i nedostatky. Práce poskytuje analýzu zpřístupněných dat v současném kontextu země zahrnující právo, technologii, ekonomiku, společnost i organizace. Dále obsahuje návrh a implementaci webové aplikace se zveřejněnými daty na zabezpečení potravin (asi jak aplikaci využít v problému potravinové bezpečnosti). Implementace také ukazuje, jak může být nová hodnota přidána do zveřejněných datových souborů za pomoci volně dostupných softwarových nástrojů.

Klíčová slova: Otevřená data, Otevřená vláda, Datový portál, Vládní data, Právo na informace, Data pro vývoj

Open Data for Ethiopia

Summary

Governments are responsible for collecting, organising and disseminating data on different aspects of their country. This data can be accessed openly by the public or through an official inquiry. There is a recent trend of more and more governments and institutions opening up their data to the public. Therefore, open government data is not just a buzz word anymore but an essential resource for creating new business values. This thesis analyses the importance of open data and its implications for development. It discusses the open data movement, how to start an open data initiative, its demand and civic engagement. Case studies from Africa have been examined to converse the impact of open data on development. The recent open data initiative by the Ethiopian government, the Ethiopian open data portal, its merits and shortcoming were studied. This thesis also provides an analysis of open data in the current context of the country including law, technology, economy, society and organisations. Furthermore, it includes the design and implementation of an open data-driven web application on food security. The implementation also shows how a new value can be added to openly accessible data sets using freely available software tools.

Keywords: Open data, Open Government, Data Portal, Government data, Right to information, Data for development

Table of content

List of Figures

1 Introduction

Open data in recent years has been getting a tremendous attention by governments, public institutions, and civil societies altogether. Initiatives in Europe and the United States are good examples of this movement. Open data portals like data.gov.uk, data.gov and data.gov.sg which resulted from such movements provide government information to citizens and interested parties. (Attard, Orlandi, ScerriAuer 2015)

Open data is defined as "data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike". It has three primary attributes: availability and access, reuse and redistribution and universal participation. (Open Knowledge Foundation 2016)

The immense amount of data produced by governments and public institutions to perform their operations on health, education, taxation, population and so forth, is stored in government body in electronic or printed archives. It should be a priority for governments to give access to this information. It would benefit different groups of people and organisations, including the government, in creating value from the information. Several areas can be mentioned including transparency, improved private products and services, innovation, improved efficiency and effectiveness of government services and new knowledge from combined data sources and patterns in big data (The Open Data Handbook 2016).

In 2011 the African development bank group started a platform called open data for Africa as an initiative to boost access to quality data to manage and monitor the development results of African countries (Open Data For Africa 2016). On a similar account, for the African continent as a whole, a non-governmental initiative, OpenAFRICA, was started in 2013 by a public technology hub called Code for Africa (*openAFRICA* 2016). Another good example from eastern Africa would be the Kenyan open data initiative, which is the first government initiative in Africa. This initiative started in 2011 (Kenya Open Data 2016"). The efficiency and the effectiveness of the above initiative will be discussed in this thesis.

As of the moment of writing this thesis, there were no substantial and impactful initiatives done by the Ethiopian government to make government data accessible to the public for free with easily reusable formats. This thesis will try to assess the open data readiness of the

Ethiopian government, identify critical values of opening up government data to the Ethiopian public, the resources it will require and finally design a simple web application for visualizing data which are pulled from available resources.

2 Objectives and Methodology

2.1 Objectives

The main goal of the thesis is to determine how open government data can facilitate the efforts of meeting the development goals of Ethiopia by studying the impact of data openness on the community.

Partial goals of the thesis are such as:

- To make a review of the current state of Ethiopia in terms of public access to government data,
- To analyse available data sets and to evaluate their level of openness,
- To analyse and design a new open data application for agricultural sector in Ethiopia.

2.2 Methodology

The methodology used for the thesis is based on the study and analysis of open data initiatives in the world and particular case studies in Africa. Recommendations on how to achieve an open government in Ethiopia will be stated, based on the practices of the successful case studies. And finally, an open data application will be implemented in sector of agricultural goods using tools for data visualization.

3 Literature Review

3.1 Open data movement

Three different groups can be noted when discussing the open data movement: open government advocates, open source software and open scholarly data advocates, and open innovation entrepreneurs. The first group is focused on freedom of information and money in politics, the second ones have open access advocacy as a significant part of the movement, and the latter group can also include government agencies and staffs looking to the public for expertise. (Open Government Data: The Book 2016)

The right to information plays a significant role in opening government data to the public, as stated by the organization for the right to information. Forty countries have the right to government-held information protected by their constitution. Kenya was one of the latest countries to adopt this to their newly ratified constitution in 2010. (Open Government Data: The Book 2016)

3.1.1 How is it defined?

Opendefinition.org summarises the definition of open regarding data and content as *"Open means **anyone** can **freely access, use, modify, and share** for **any purpose** (subject, at most, to requirements that preserve provenance and openness)."*

Data openness has two dimensions. Firstly, it should be legally open, meaning it must be placed in the public domain or in terms of minimum restriction. Secondly, it must be technically open meaning it must be in an electronic format which is machine-readable and preferably non-proprietary, which makes it accessible by freely available software tools. It should be published on a public server without any restrictions. (*A Primer on Machine Readability for Online Documents and Data - Data.gov* 2017)

3.1.2 Open data licences

Open data license is used by governments and organisations to clearly define the conditions on how any user may use their data. The most recognised and used licences are discussed below.

The Creative Commons (CC) Licences are the most recognised standard licences to provide access to data. They allow freely copying, reusing, distributing and sometimes modifying a creator's work without permission from the rights holder every single time of access. The baseline rights of this licence include Attribution [BY] – given credit to the works creator, Non-commercial [NC] – using the work for non-commercial purpose, No-Derivatives [ND] – no adaptations of the work may be created or no merger may be done with other works, and ShareAlike [SA] – if adaptations of the work are created then they should be publicly available under the same licence terms of the original CC work. This baseline rights can be used in conjugation with each other like, BY, BY-ND, BY- NC, BY-NC-ND, BY-NC-ND, BY-NC-SA, BY-SA.

CC Zero (CCO) is created to facilitate the release of content, data, datasets and databases to the public domain, while the copyright owner waives all right to be identified as the creator. This license is irrevocable once it is issued it cannot be changed. It can be used to prevent attribution stacking, but it would also mean that we cannot recognise the source and authority of the data.

Open Data Commons provides three solutions for data, datasets and databases. Open Data Commons Attribution Licence (ODC-BY), Open Data Commons Open Database Licence (ODC-OSbL) and Public Domain Dedication Licence (PDDL), which are compatible with Creative Commons CCBY, CC BY SA and CC Zero respectively. Irrevocability and third-party rights should be considered when using these licences.

The Open Government Licence (OGL) has been recently launched by the National Archives UK to facilitate the reuse of Government and other public organisations information. It makes data available in machine and human readable form similarly with the creative commons licences (Korn and Oppenheim, 2011).

3.2 Starting an Open data initiative

As it is often the case with stating any initiative, open data also involves some expenditure of resources and effort. The benefits of open data are comparably higher than the efforts invested in the initiatives. We must also keep in mind that data must be transformed into a usable form in order to realise its full potential as any raw commercial commodity. Individuals, organisations and even governments can innovate and collaborate in new ways when government data are made accessible and reusable.

Before opening data, certain principles must be considered. The G8 summit in 2013 has defined principles to be followed by nations that want to open their data to the public. Open data by default, quality and quantity, usable by all, releasing data for improved governance and releasing data for innovation are the five basic principles (*European Data Portal 2016*).

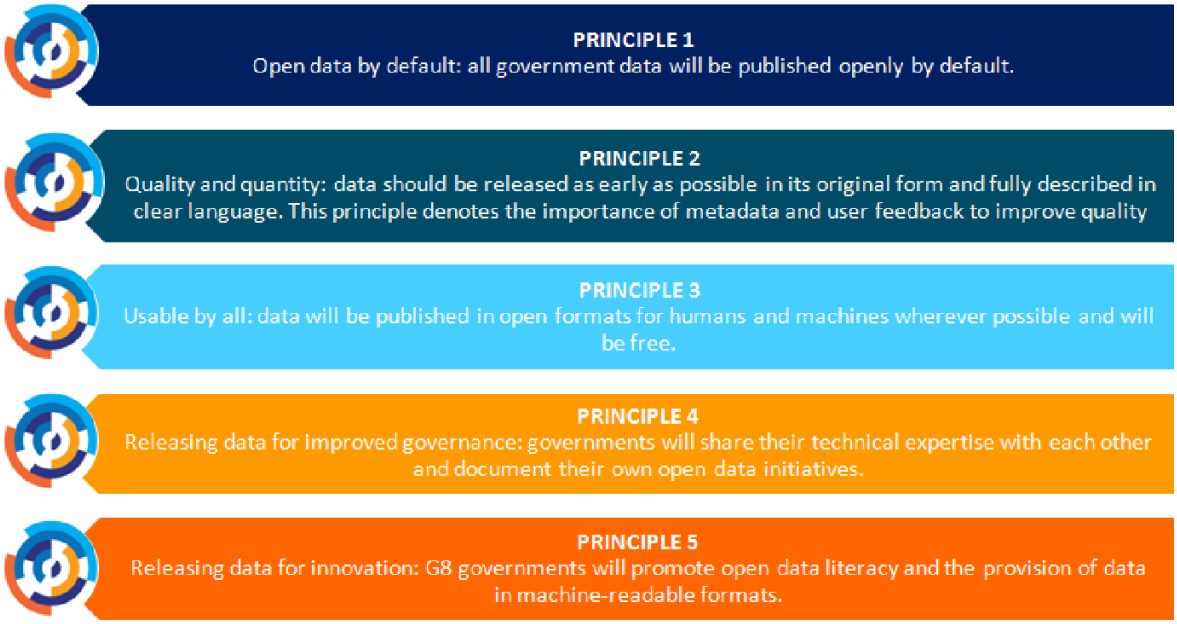

Figure 1: Open data charter principles [Source: (European data portal 2016)]

3.2.1 Building an open data strategy

After taking these principles into consideration, an open data strategy should be developed to clearly define the key goals and to set the ambition. This is best tackled using the following five steps. Setting the ambition, creating the strategy, drafting open data policy and overcoming barriers and assessing critical success factors.

Setting the ambition: Defining what needs to be achieved is the first step to build the open data strategy. The as-is and the to-be situations, as well as the measurable goals, must be defined.

Creating a strategy: this should cover the topics of how to achieve the to-be goals. This step should at least cover the list of data sets within the organisation, assessment of the data status, alignment with legal aspects, technical and budgetary implications, a time plan and key performance indicators to have a measurable progress.

Drafting an open data policy: the policy must cover definitions and benefits the scope of the outcome, the legal aspects the type of data and its quality and establishing or assigning a body which will be responsible in implementing the data policy.

Overcoming barriers to publishing open data: there are a variety of excuses given by government not to publish data as open data. Usually not to get a negative reputation and fear of the data might not be sufficiently accurate. However, almost all these barriers have reasonable remedies.

Assessing critical success factors: the three dimensions which the success of an open data initiative measured are the quality of open data publication, the use of open data and documenting the impacts and benefits emerging from it. The quality includes the accuracy, completeness, timeliness and consistency. The legislation, regulation and licences are critical success factors for both publishing and re-use of data (“*European Data Portal, 2016”*).

3.2.2 Technology Options

Data catalogue will be discussed as a technology option to provide open data to the public. A list of datasets available in an open data initiative is called a data catalogue. Metadata, clear license information, access to the database and searching capability are essential elements of a catalogue. Another option is a platform; it includes the data catalogue with other information and services which are part of the open data ecosystem. This platform will include technical support, online forum, and a knowledge base. (Herzog 2014)

3.2.2.1 Common characteristics of data catalogues

The data catalogues must provide access to publicly available data. It should either store the data in the system and provide necessary visualisation options or give a link to the complete dataset and provide its metadata. It must provide users to access tools to search, explore and engage with the datasets.

Developers must have an entry point to access the data catalogue and its contents directly to provide similar services as the catalogue. This entry can be achieved by application programming interface (API). Integration and customisation make the catalogue flexible to be embedded with other websites to add other layouts and custom features though modules ("Herzog, 2014").

3.2.2.2 Software delivery model

One consideration to create an open data catalogue will be the cost. Open source software, although they require management costs to host, maintain, update and training, they can be acquired for free. If the organization has the technical proficiency in the area of open source programming languages and an excellent infrastructure to host reliable and fast catalogue, self-managed open source catalogue might be a good option worth considering. Software as a service products (SaaS) are proprietary products from a vendor providing the software hosting services, which comes with recurring monthly or annual fee. SaaS products provide fast and reliable IT infrastructure but the flexibility to customize them might be limited to the vendor's capability and willingness. When considering the SaaS product the procurement policies of government organizations might make it difficult to acquire them. (Herzog, 2014)

3.2.2.3 Common open data platforms

Out of the available platforms in use for open data publishing. Some of the leading ones are discussed below.

CKAN is an open source data catalogue which is supported by the Open Knowledge Foundation. It is installed on a Linux server locally or on a cloud-based configuration. Examples include initiatives from Africa open data, Edo (Nigeria), Brazil and United States.

Junar is a Software as a Service open data platform which is managed by Junar's infrastructure. A complete data catalogue or an API is provided. Examples include initiatives from Chile, Costa Rica and Peru.

Open Government Platform (OGPL) is also an open-source data catalogue, developed jointly by the Indian and the U.S. governments. Examples include initiatives from Ghana and India.

Socrata is a SaaS platform which provides an API, catalogue and data manipulation tools. It also allows users to create their own visualizations and save it for other users to use it. It also provides an open source API which helps organizations which want to move from the cloud-based model to self-managed open source model. Examples include initiatives from Chicago (U.S.), Kenya and United Nations Development Program (UNDP). (Herzog 2014)

3.2.3 Deployment scheme

Tim Berners-Lee, one of the inventors of the web and initiator of open data, has developed a 5-star deployment scheme for open data. This helps in determining the level of openness of an open data initiative.

One star: making data available online under an open license in any kind of format (e.g. pdf, images). This allows users to be able to view, print, store, enter the data to any system, to change it and share it with anyone. This format is easy to publish.

Two stars: making your data available in a structured format (e.g. excel sheet). It allows the user to amylase, perform visualizations and combine data using a proprietary software on top of what a one-star deployment has to offer. This is also easy to publish.

Three stars: making your data available in a non-proprietary open format (e.g., comma-separated values(CSV)). Manipulating this kind of format doesn't need any proprietary software, and users are able to perform everything from two-star deployment. The publisher would need additional tools to convert the data from a proprietary format, but it is still easy to publish.

Four stars: making data available through a universal resource identifier(URI). Users can link to it, bookmark it and reuse parts of the data. Understanding this part of format takes more effort than a tabular or a tree format. Publishers need to invest some time arranging data and assigning URIs.

Five stars: linking data to other sources so it can provide context. Users can discover more linked data while consuming the data provided. Publishers need to invest resources in linking their data with others. Publishers will also get the same benefits as users. (Berners-Lee 2017)

3.3 Demand and Engagement

Full benefits of data are accomplished by engaging civil hackers, which are programmers and developers. The improvement of quality of life using data-driven apps and contents is achieved through developers who contribute actively to the civil society. Therefore, open data initiatives must put a primary focus on engaging stakeholders to empower citizens to use open data and maximize its value.

3.3.1 Civil Hacking

A creative approach and often a technological approach to solving a common problem is called civic hacking. It involves the use of government data to make governments more accountable. Civic hackers can be data scientists, designers, entrepreneurs, programmers, government employees and anyone who wants to solve problems. Hacking does not mean making a computer crime as is known by its ordinary meaning; it means modifying the original purpose of something to solve a problem. (Tauberer 2014)

Civil hacking using government data has an enormous contribution to the national economy, by helping consumers more informed and making governments more efficient.

3.4 Open Data in Africa

Under this topic, some of the currently existing open data platforms and the initiatives which started them are discussed. The focus is on Africa since the thesis is particularly written for a country in eastern Africa.

3.4.1 Open data for Africa

Open Data for Africa is an initiative by the African development bank is designed to provide the user with access to a wide range of data on Africa managed by the African development bank and other regional and international partners.

It also facilitates the collection, analysis and sharing of data on emerging and crucial development topics such as food security, gender equality and climate change. Users can visualize socio-economic indicators over a period of time, utilise presentation-ready graphics or create their own, performs comprehensive analysis at regional and country levels, blog and share their views with others by creating an informed community of users.

The open data for Africa platform also responds to a number of important global and regional initiatives to scale up the availability of quality data on Africa and to foster evidence-based decision making, public accountability and good governance. The initiative forms part of the worldwide effort to strengthen statistical capacity in different aspects of development. (*Home - African Information Highway Portal* 2017)

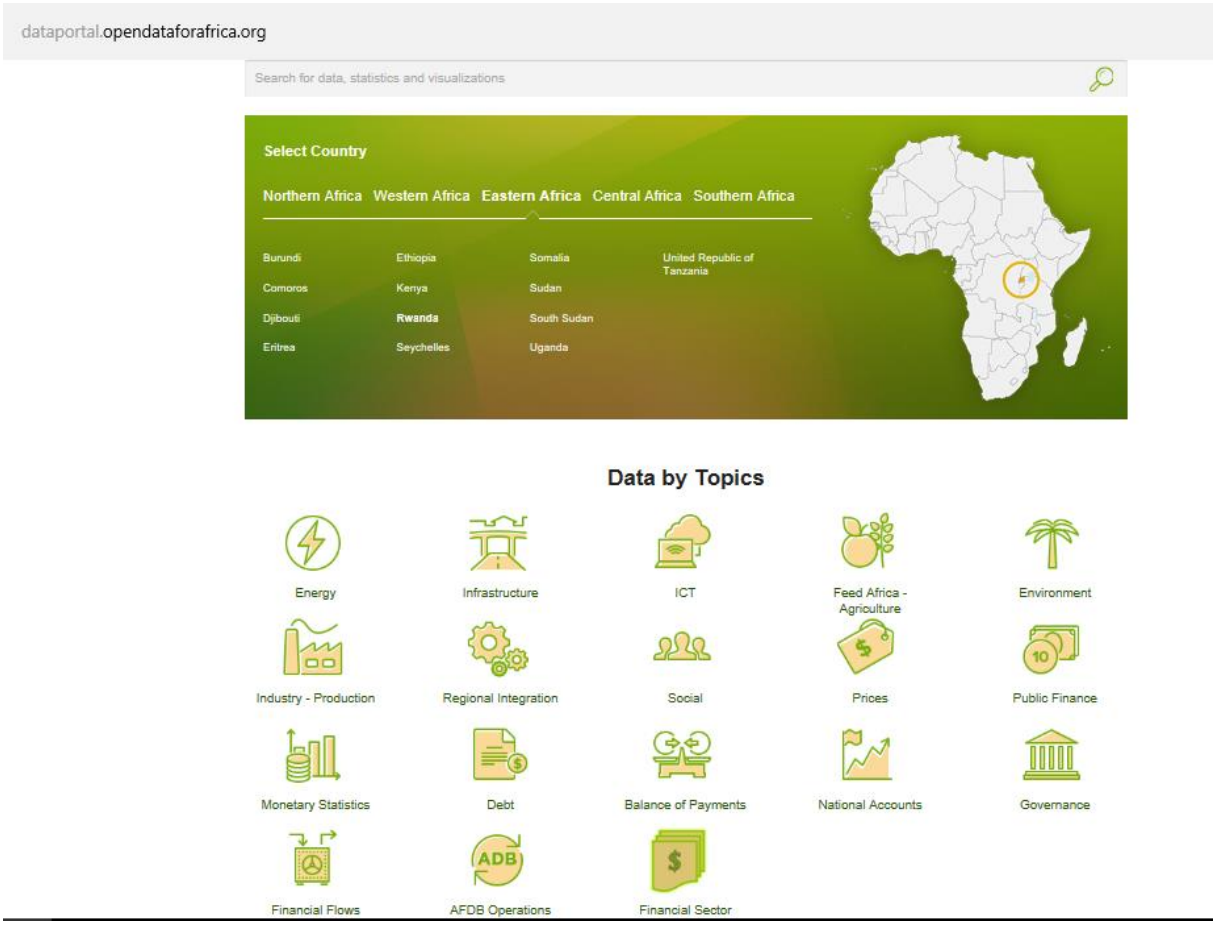

Figure 2: Open data for Africa [Source: (Home - African Information Highway Portal 2017)]

3.4.2 OpenAfrica

OpenAfrica is the most extensive independent data repository for the African continent. It is a non-governmental grassroots initiative started and maintained by Code for Africa as a public service. It provides resources to citizens, governmental and non-governmental agencies and civil societies as a free platform.

This platform running with the support of the World Bank and Google contains over 2500 datasets on the African continent. Code for Africa maintains this portal with the help of its affiliated labs in Cameroon, Ethiopia, Ghana, Kenya, Nigeria, Sierra Leone, South Africa, Tanzania, and Uganda. It currently runs over 100 projects in creating data-driven applications using open data to promote citizen engagement with government and enhance data-driven journalism with a user-centric approach. (*openAFRICA* 2017)

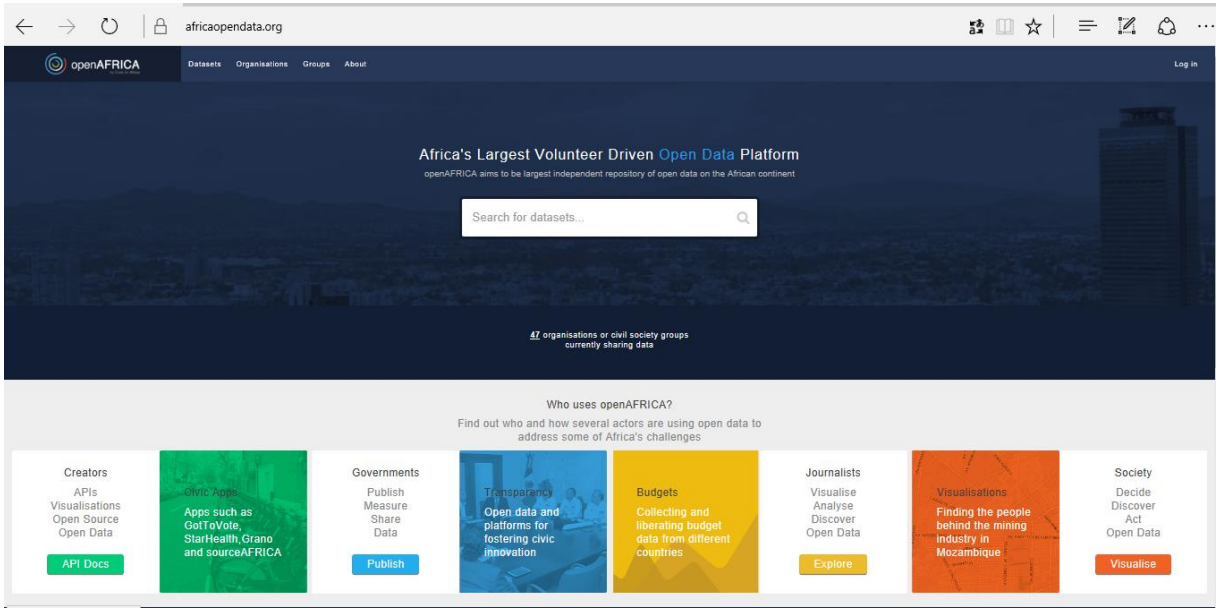

Figure 3: Open Africa [Source: (openAFRICA 2017)]

3.4.3 Kenya Open Data Initiative

In July 2011, the Kenyan open data Initiative (KODI) was launched with the purpose of making key government data freely available to the public through a single online portal. The first datasets released included national and regional expenditure, information on vital public services such as agriculture, education and health together with the 2009 census. With the primary objective of propagating better governance, Kenya was the pioneer in Africa in launching such kind of initiative**.**

Data is provided through this portal in accordance with the principles and definitions of open data stated by the open knowledge foundation. The data outlines timeliness, accessibility, ability to manipulate, data completeness, non-proprietary and license free. The portal provides different views of the datasets and open for manipulation by users. The main formats used in this portal are spreadsheets, tables, charts and digital maps. Filtering exporting visualising, and embedding data into other website is allowed for users.

The open data catalogue currently, at the time of writing this thesis, houses 990 datasets. These datasets can be filtered with category, view types and tags. The website of this initiative uses Socrata as an open data platform. (Kenya Open Data 2017)

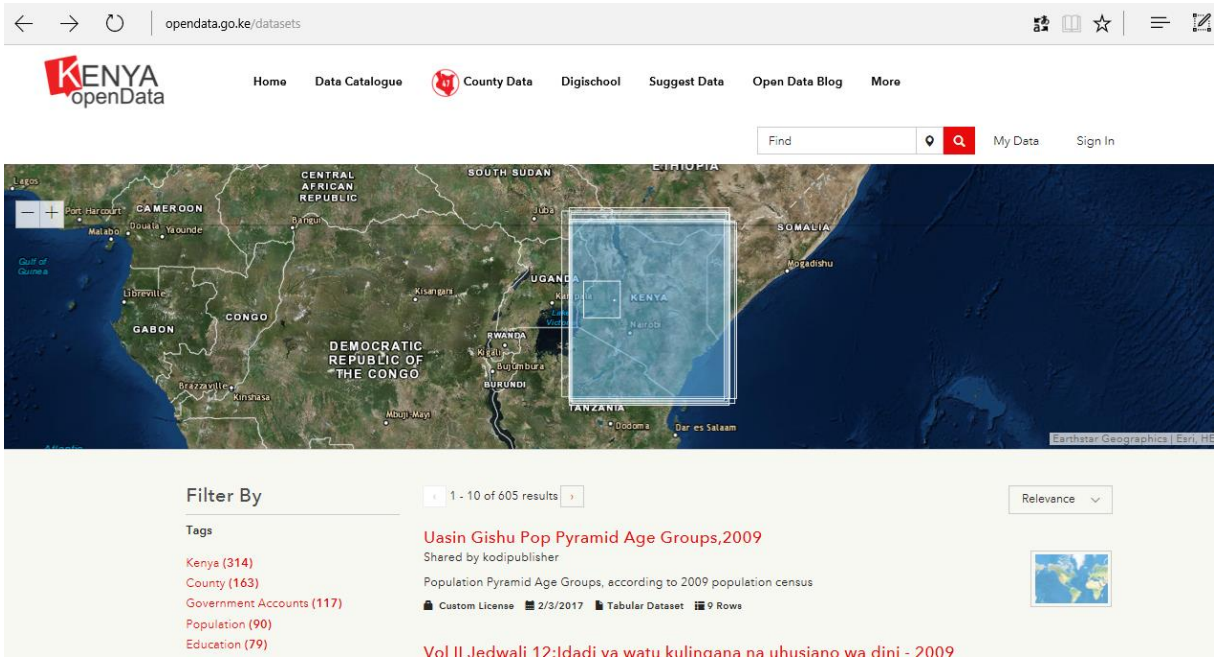

Figure 4: Kenyan open data portal [Source: (Kenya Open Data 2017)]

3.4.4 The code4kenya initiative

The expected utilization of data from the Kenyan open data initiative was not satisfactorily achieved therefore another outreach was needed in Kenya. The code4kenya was launched in November 2012 to test a model that could potentially increase the uptake of government data by the community. It started by creating technology-based applications, services and platforms to accelerate the awareness and ability of the public to make sense of the data provided and to promote an active engagement with the community around critical public issues.

The initiative was funded by mainly the World Bank, the African media initiative and other partners include the Kenyan ICT authority, Strathmore University, iHUB and the open institute.

The open data platform was used to build relevant tools and applications that increased data consumption collaboratively. The critical issue addressed on this platform were health, education, election and country administration which are areas affecting every citizen and can be tackled by the use of open data tools.

The objectives of code4kenya were specific, first was to create demand for government data and facilitate the engagement of citizens with government and second was to develop capacity data journalism within institutions working with communities. The applications built in code4kenya initiative were released in early 2013 to the public.

3.4.4.1 Applications Provided by code4kenya

- **Star Health**

    - Citizens get localised answers about everyday health challenges through the web and SMS. This health tool helps the community to get a better healthcare by providing data from a reputable official source and verified by the crowd. Which helps to alleviate overpriced medicine and to inform where the nearest medical specialist is available. (the Star 2017)

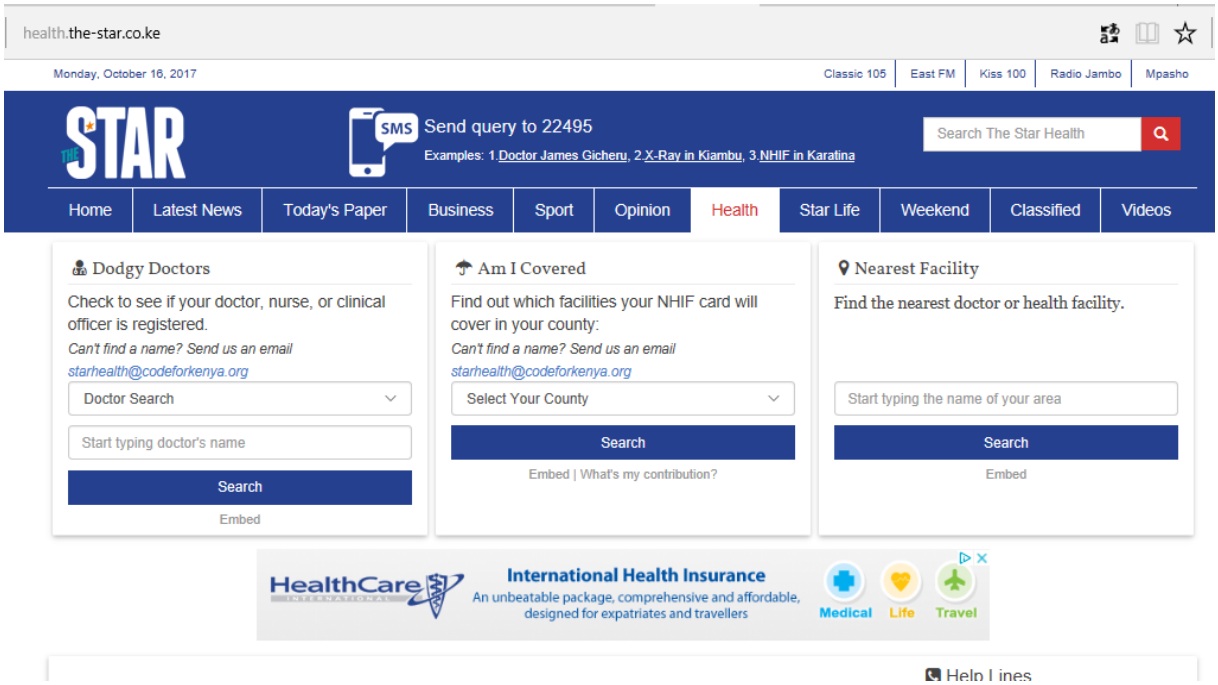

Figure 5: Star Health [Source: (the Star 2017)]

- **Go to Vote**

    - Built as a data journalism project "GotToVote!" helps citizens to understand how significant national events like elections affect their personal life as well as their communities. This project is also adopted by the Code4Ghana initiative to make voter registration a smooth process by avoiding the need to go to a

registration centre which might present, at times, a complicated bureaucracy. (*GotToVote Kenya* 2017)

Figure 6: Go to Vote [Source: (GotToVote Kenya 2017)]

- **HURUmap**
    - HURUmap aims to facilitate data backed journalism. Civic activists and journalists get an easy 'plug and play' tool for embedding interactive data visualizations into their stories using "HURUmap". This project promises that the data is quality-checked from a reputable official source including the government. (*HURUmap Kenya: Making Census Data Easy to Use* 2017)

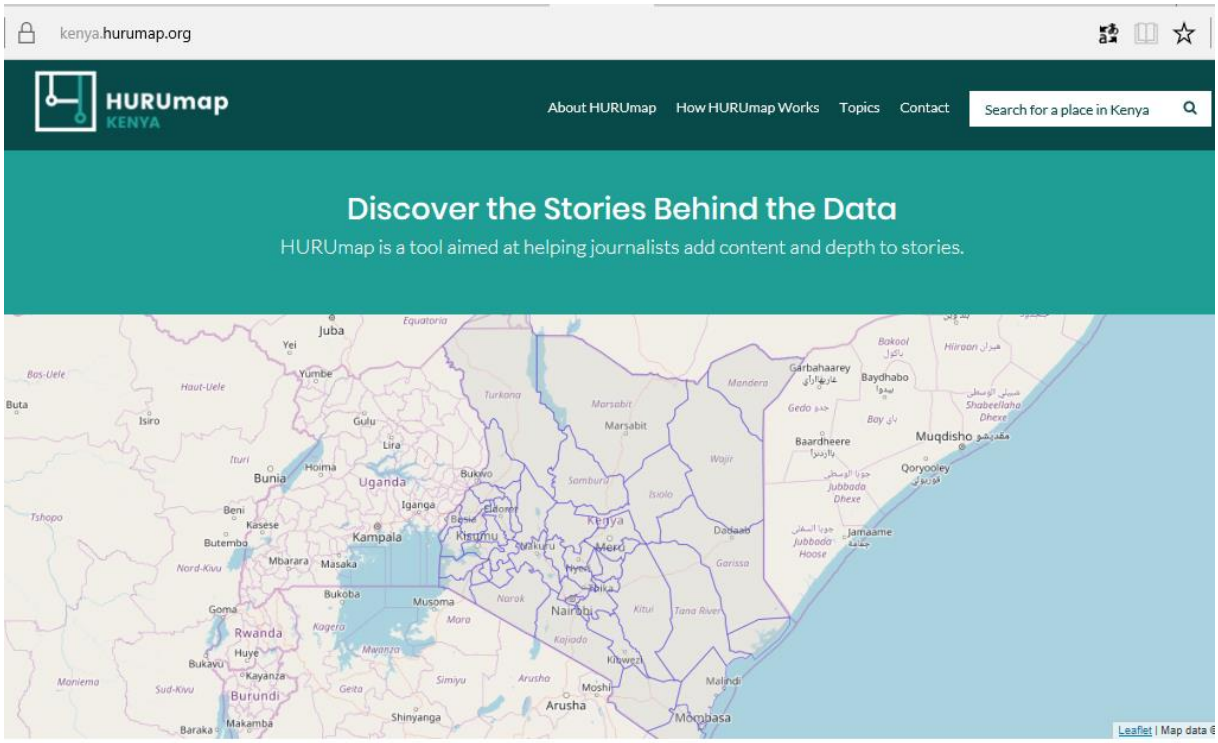

Figure 7: HURUmap Kenya [Source: (HURUmap Kenya: Making Census Data Easy to Use 2017)]

- **Pesa Check**
    - Policymakers and the public often get incorrect information because of wrong or confusing reports by public leaders and journalists. PesaCheck provides a data-driven fact-checking tool which tackles this misinformation. Misinformation is tackled by analysing the data behind the reports published. (PesaCheck 2017)

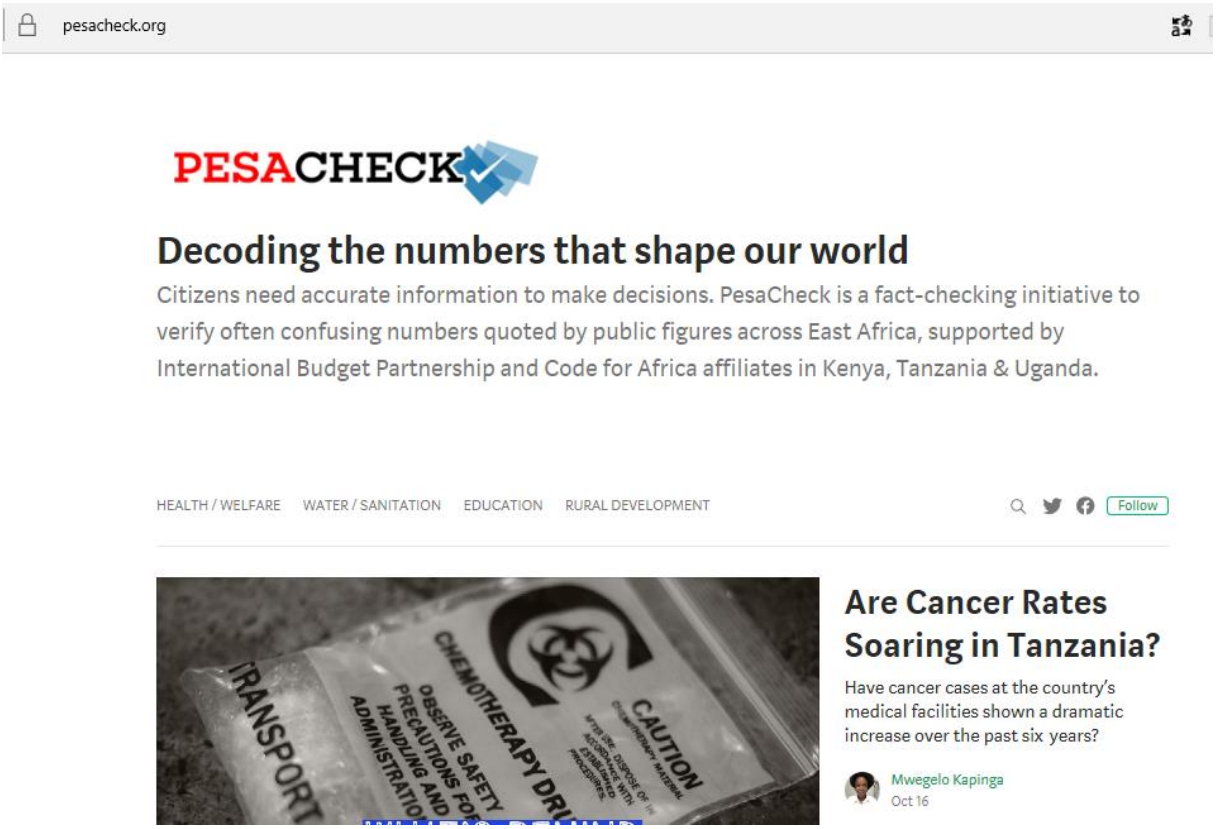

Figure 8: Pesa Check [Source: (PesaCheck 2017)]

3.4.5 Open data Nigeria

The Nigerian open data access platform was launched as an initiative to improve solutions, from research and development and civic engagement, on issues regarding the Nigerian people. This improvement aims to achieve a smart and sustainable environment. This open data platform was launched by the ministry of communication technology of Nigeria in January 2014. (Coetzee, 2017)

Their mission statement states: "This portal was borne out of the passion for the need of the development of local Content applications by innovators, increased problems focused on solving local problems, promoting civic engagement and transparency in the government" (E. 2017)

The diverse data sets available range from health, education and agriculture to budget and tourism.

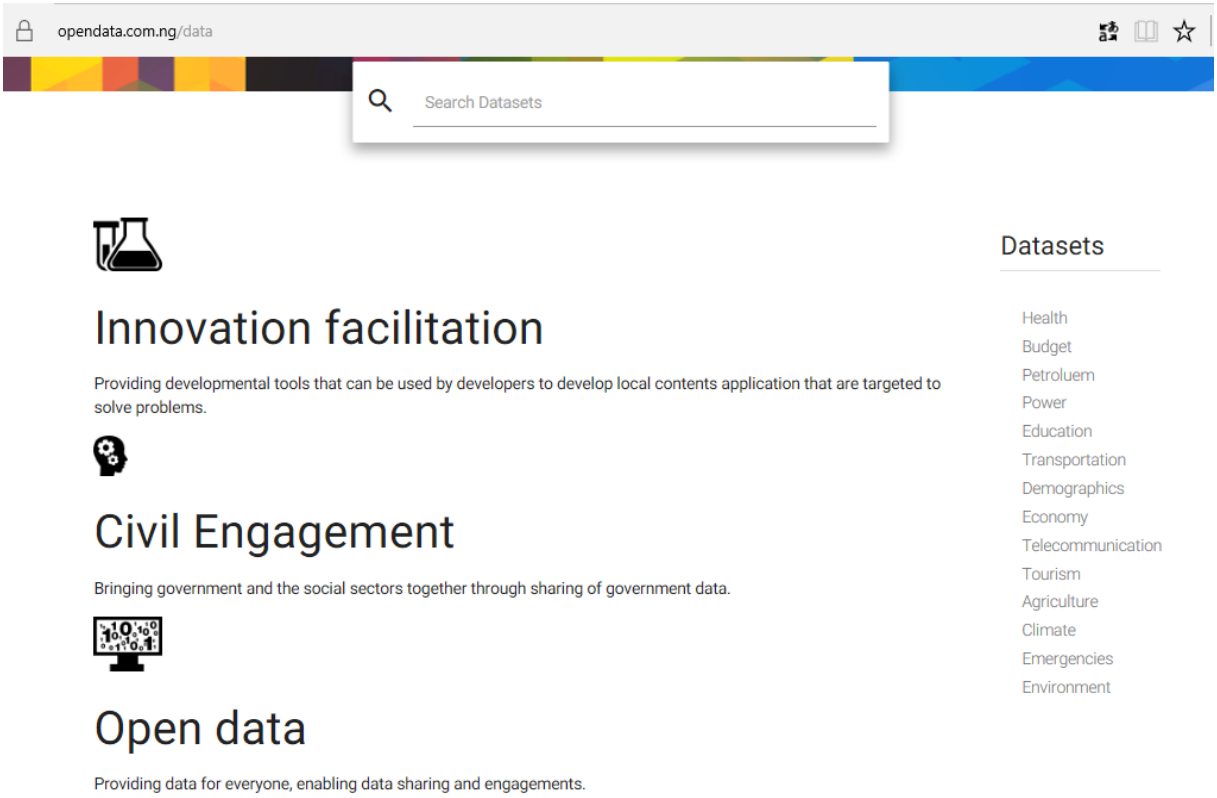

Figure 9: Nigerian Open data portal [source: (E 207)]

3.4.6 Open data Ghana

The Ghanaian open data initiative was started with the vision to coordinate citizens, developers, academic institutions, media organisations, civil societies and government. to make this coordination have an impact on the development of Ghana an interaction through open data platform was necessary.

The open data portal was launched by the communication ministry of Ghana, in 2011 by opening up data related to budget and procurement systems, agriculture, health and education. (*Ghana Open Data Initiative | Ghana Open Data Initiative* 2017)

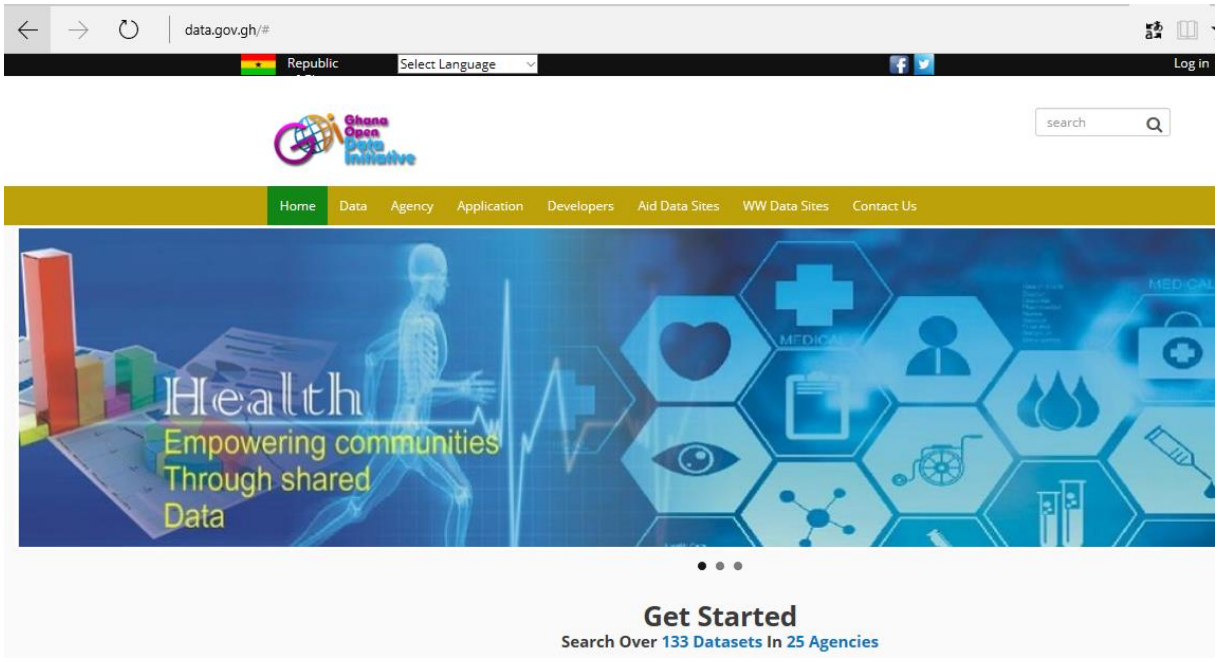

Figure 10: Ghanaian open data portal [Source:(Ghana Open Data Initiative | Ghana Open Data Initiative 2017)]

3.4.6.1 Applications from Code for Ghana

A notable application from the projects of code for Ghana an affiliation of code for Africa is “Where my money dey”. It helps the community understand where and when the mining industries in Ghana are returning some percentage of their revenues back to the local communities as part of a corporate social responsibility duties.

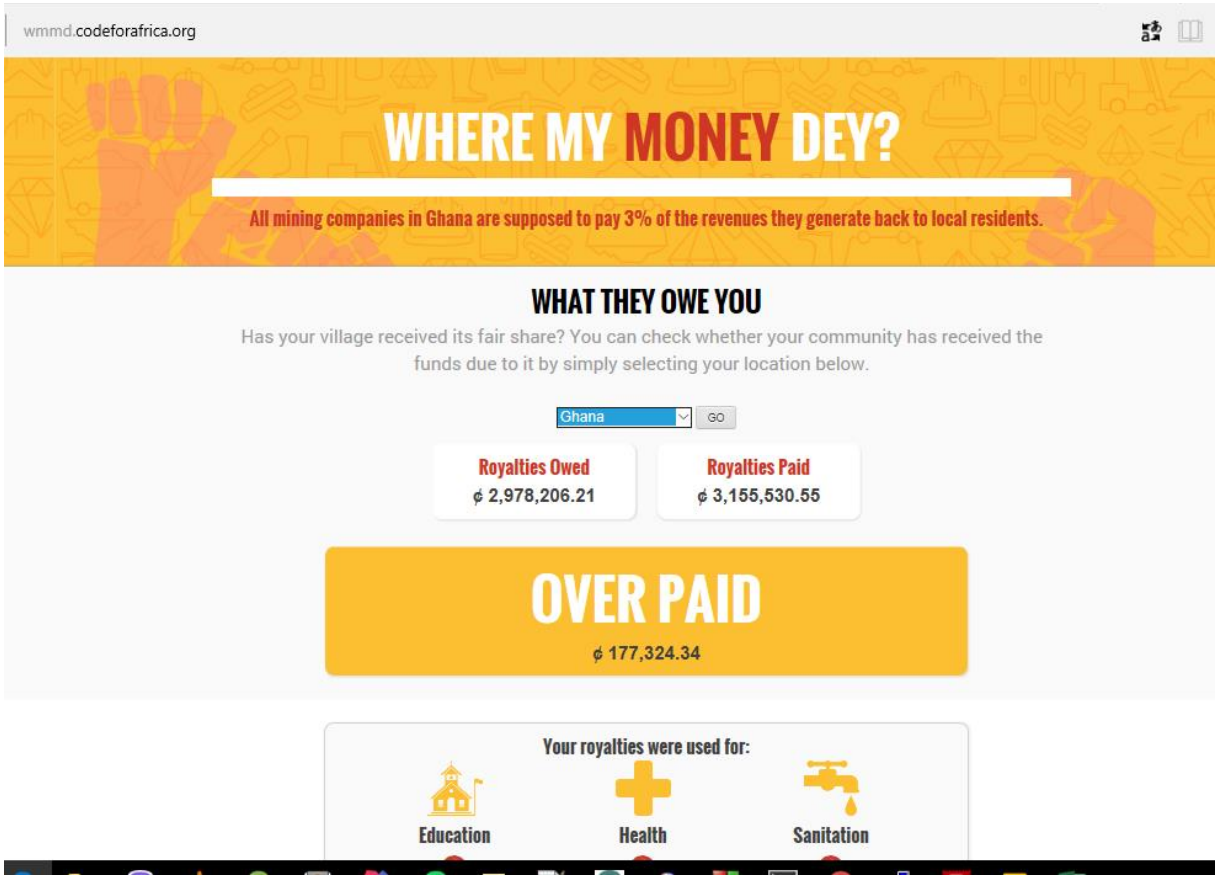

Figure 11: Where my money dey? Source: (Where My Money Dey? 2017)]

3.5 Open Data in Ethiopia

An assessment conducted by the world bank on open data readiness in March 2014 the government of Ethiopia has a strong technology and policy mandate for creating a national open data initiative and integrating the principles of open data into government systems and processes. Following this readiness assessment, the government engaged in the implementation of open data initiatives.

In a study done by open data barometer, Ethiopia is included amongst the countries in the category having a capacity constraint, as a result of limited government, civil society or private sector capacity weakness in digital data collection and management and a significant limit on affordable, widespread internet access. The research also shows that the impact of open data in Ethiopia is at a zero level. (Davies 2017)

According to this study, a number of measuring qualities were analysed. These qualities include impact on social, economic, and political; readiness of citizens and entrepreneurs; and data availability on innovation, accountability and social life. To put these into perspective, the following visualisation helps see how three countries, Ethiopia, Kenya and Nigeria, categorised in the above mentioned "capacity constrained" category fit in the scale.

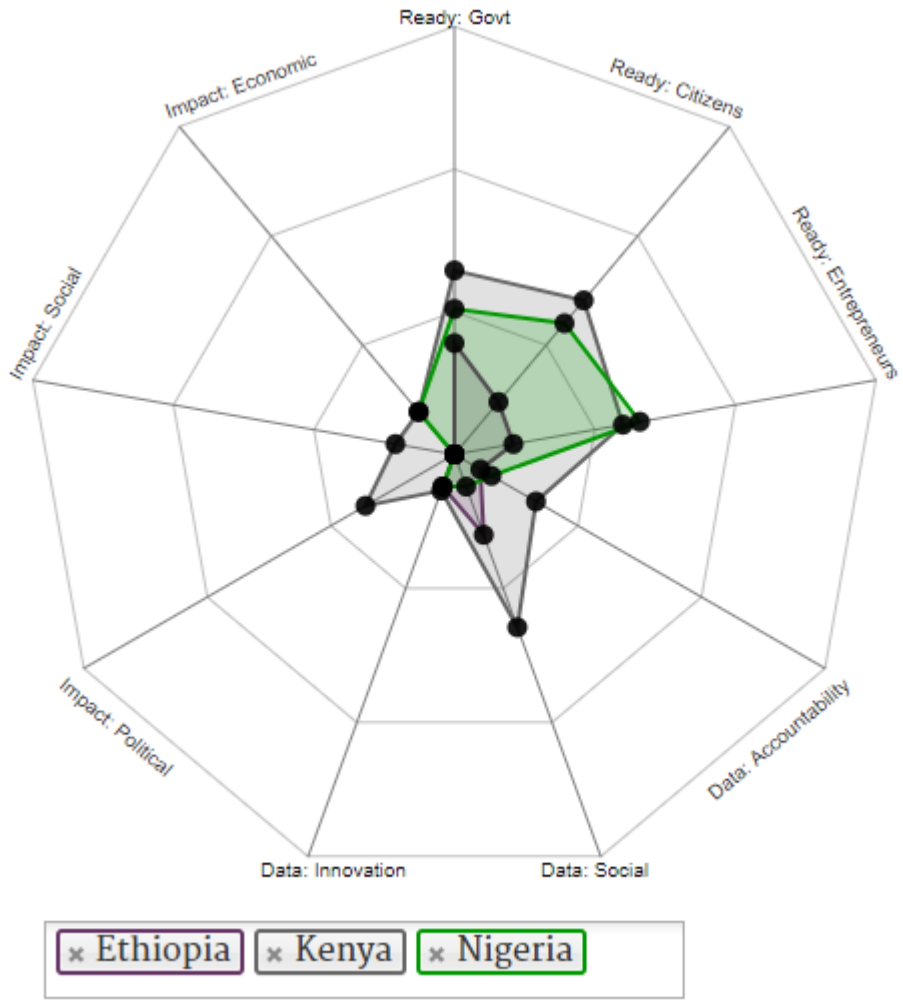

Figure 12: Open data barometer for Ethiopia, Kenya and Nigeria [source: (Davies 2017)]

In the above visualisation, the external nodes are for Kenya, intermediate nodes for Nigeria and the innermost for Ethiopia. The comparison of Ethiopia to a country that is in the category of "emerging and advancing", the Czech Republic, and a country in the category of "high capacity", the United Kingdom, is illustrated in the following visualisation. The external nodes being the United Kingdom, the intermediate nodes for the Czech Republic and the innermost nodes for Ethiopia.

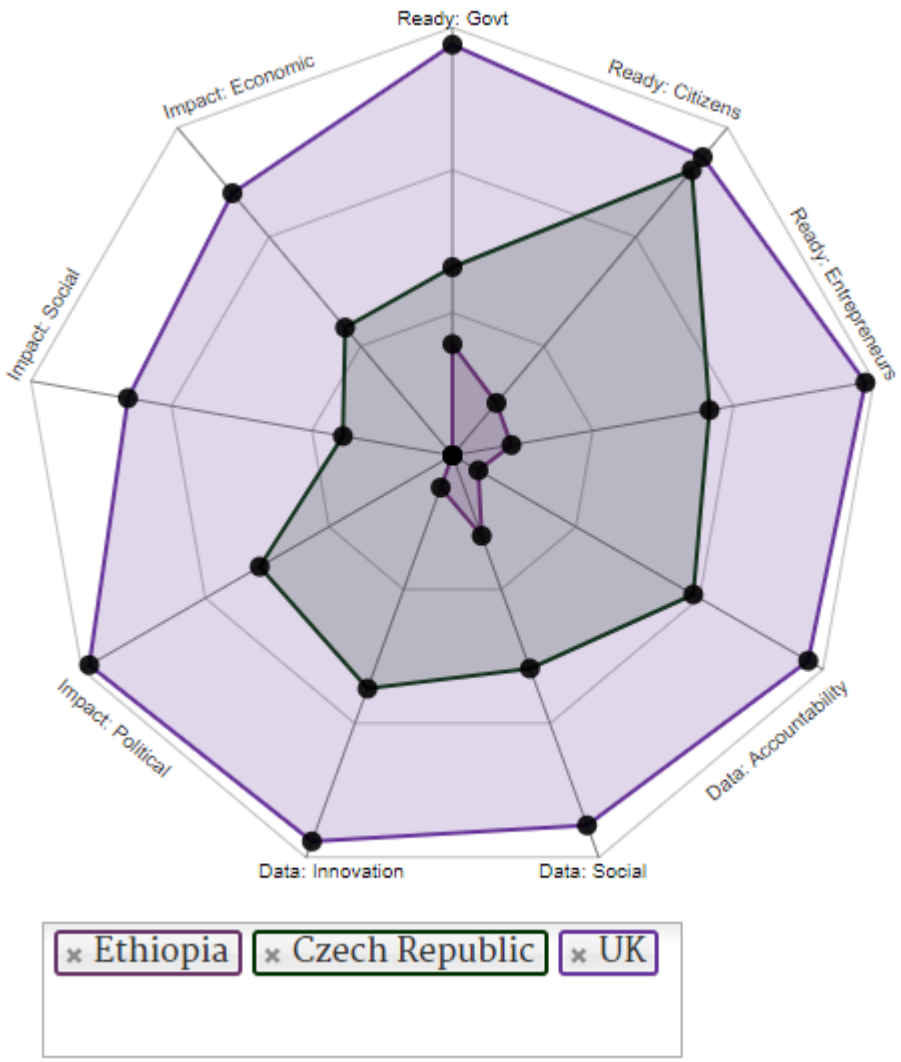

Figure 13: Open Data Barometer Ethiopia, Czech Republic and United Kingdom [Source: (Davies 2017)]

3.5.1 Data Sources

3.5.1.1 Central Statistical Agency of Ethiopia

Data collection in Ethiopia is mainly done by government institutions together with non-governmental and international organisations like the United Nations and the African Union. The primary government office to collect, organise and generate statistical data related to the socio-economic condition of the country is the Central Statistical Agency of Ethiopia (CSA).

CSA conducts, produces, disseminates and administers data generated from surveys and censuses in Ethiopia. Their websites provides dynamic consumer price database, the Ethiopian National Data Archive(ENADA), an indicator database called EthioInfo together with functions and responsibilities of the organisation.[1]

Under the data repository of this website price indices for consumer goods and agricultural products is presented along with reports on producer price index on manufacturing.

Data is only presented in a pdf format which is exported from an excel sheet. Most of the data is not up to date, except for the country and regional consumer price index which are made available monthly. Such presentation of data makes it unfriendly to users and will not have any engagement from most part of the community.

3.5.1.2 World Bank

World bank's open data portal is another source of developmental data from Ethiopia. Data provided on this portal ranges from GDP (gross domestic product) to health and education.[2]

3.5.1.3 Food Security Portal

Facilitated by the International food policy research institute, the Food security portal aims to provide improved food security for the world's poor and increased resilience of global food systems against food and financial crises. The project brings together international, regional, and country-level data, news, and research aimed at meeting countries' immediate food security needs, as well as building long-term global food security. Data about Ethiopia can be accessed by selecting a data set from the available list.[3]

[1] http://www.csa.gov.et
[2] http://data.worldbank.org/country/ethiopia
[3] http://www.foodsecurityportal.org/ethiopia

3.5.1.4 African Development Bank

Open data for Africa is sponsored by the African Development Bank for collects development related data from African countries and makes it available for everyone as an open data platform. Data for Ethiopia can be found on their microsite.[4]

3.5.1.5 FAO

The food and agricultural organisation of the united nation provides a web-based information system for food and agricultural statistics, CountryStat, at regional, national and subnational levels. It aims at facilitating informed policymaking and monitoring the goal to eradicate extreme poverty and hunger by supporting data analysis and evidence-based decision making. Data for Ethiopia in this area is provided on their website under FAOStat.[5]

3.5.2 Ethiopian government open data portal

An initiative to have an openly available data by the Ethiopian Government is an open data portal (https://www.data.gov.et), which is started with the aim to bring available public data to one searchable website to help citizens understand how government works and how policies are made. Thus, making it easy for people to make decisions and suggestion about government policies based on detailed information. The technology used to build this portal is the CKAN open source application.

In the Ethiopian government's open data portal, there are only seventeen datasets found. These datasets include economic sector with fourteen datasets, Trade sector, financial sector and agricultural sector with the last three consisting of one data set each.[6]

The datasets presented here are in Excel sheet format. In the data policy, information about linking to the website is stated. Linking to the website is for free but should be done in an appropriate context which can be as a service to people to find official information about

[4] http://ethiopia.opendataforafrica.org
[5] http://www.fao.org/faostat/en/#country/238)
[6] https://www.data.gov.et/

the Ethiopians government. Moreover, all data provided must not contain information related to national security and other data protected by legal precedent. (*Welcome - Ethiopian Government Open Data Portal* 2017)

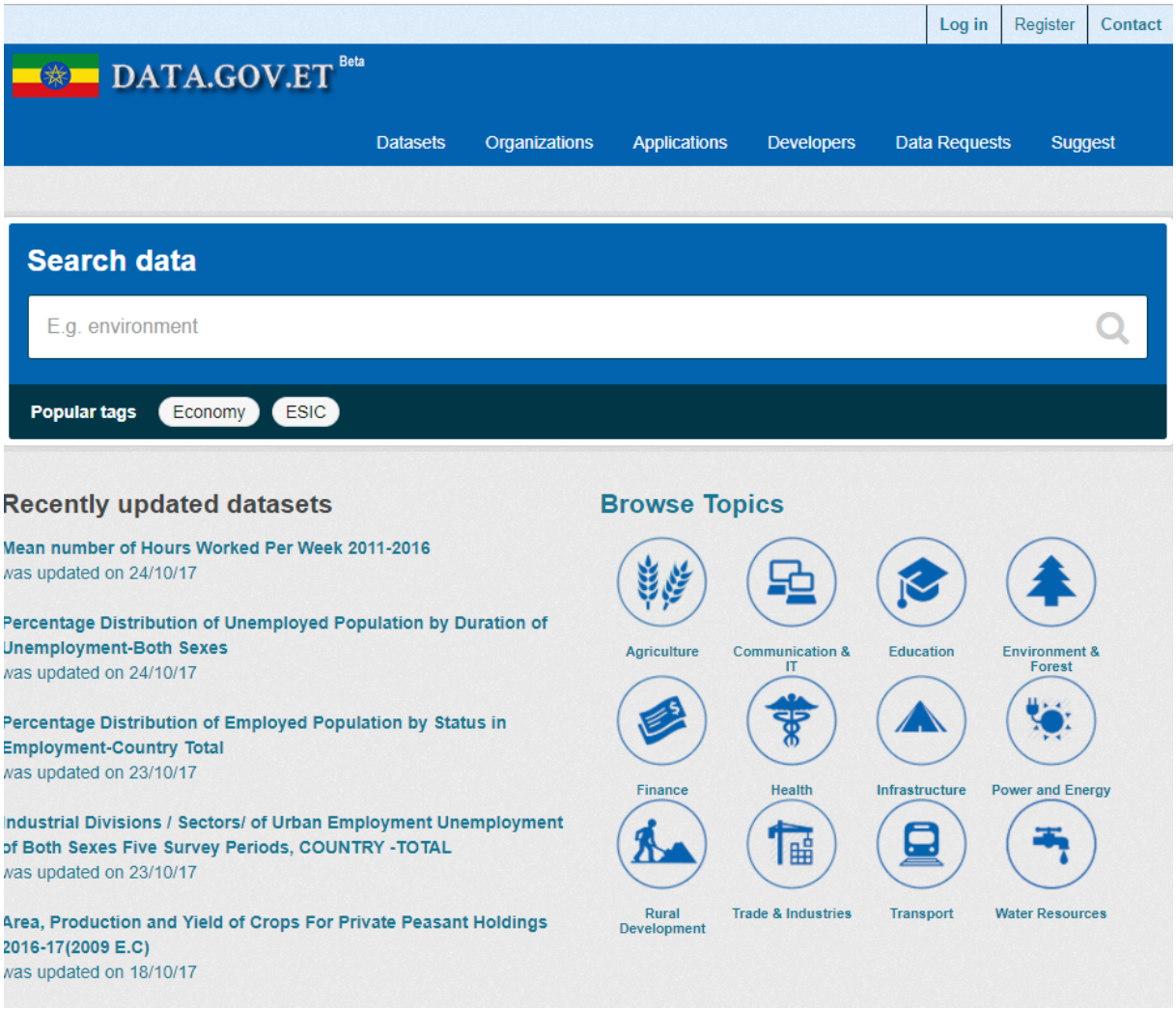

Figure 14: Ethiopian open data portal [source: (Welcome - Ethiopian Government Open Data Portal 2017)]

3.5.2.1 Application provided

At the writing of this thesis, there were no applications provided by via this portal.

3.5.3 Contextual analysis

International donors fund most of these open data programs in developing regions, creating an avenue for governments to receive additional resources. Open data initiatives have

been viewed as great PR for most governments, but those that do it for this reason usually have low-quality data, as their intentions do not align with the essence of what open data is and should do (*Schwegmann 2013*).

3.5.3.1 Legal context

Freedom of the Mass Media and Access to Information Proclamation (Proclamation No. 590/2008 of 2008) has been passed on 4th of December 2008. stating "Cognizant of the necessity of preserving and consolidating past achievements and positive practices pertaining to freedom of expression while removing structural and institutional impediments that hinder the independent operation of the mass media and the free exchange of information and ideas." (Refugees 2017)

However, in reality, the Ethiopian government has been known for cracking down on multiple privately-owned media outlets so much as to say it has a monopoly over mass media broadcasting to the citizens. Numerous accounts can be cited from the reports of "Reporters without borders". (*Ethiopian government's witchhunt against privately-owned media | Reporters without borders* 2017)

3.5.3.2 Technical Context

The ICT sector in Ethiopia is one of the worst in the world. As discussed in previous chapter 3.5, Ethiopia is categorised as one of the capacity constrained countries with a shallow internet penetration compared to the rest of the world. The mobile cellular prescription is 50.513% of the total population. Based on a report from the World Bank, in 2016 only 15.4% of the population has access to the internet and the international internet bandwidth is 2.424 bits per second per user. (*World Development Indicators | The World Bank* 2017)

Ethiopia has recently launched an open data portal which is hosted on CKAN platform. This portal has been discussed above in chapter 3.5.2.

3.5.3.3 Economic Context

Ethiopia has seen a steady double-digit growth for the last few years, making it one the fastest developing countries in the world. In a report about ICT sector development in Ethiopia,

the Ministry of communications of Ethiopia states that there is a substantial increase in ICT development. From a mere 1.2 million subscribers of in 2007 to 23.7 million subscribers in 2013. As the second populous country in the sub-Saharan Africa Ethiopia has still much work to do when it comes to ICT infrastructure development. Ethiopia ranked 169th out of 175 countries in the ICT development index of 2016 with a 1.5 IDI, which is an increase of .22 IDI value from 2015 rank of 172. (*ITU | 2017 Global ICT Development Index* 2017)

3.5.3.4 Social Context

Ethnically diverse Ethiopia has a population of over 100 million comprised of over 80 different ethnic groups. The world bank statistics on poverty in 2011, about 35.5% of the population live under the international poverty line of US$ 1.90 per day per capita. This poverty was more noticeable with individuals who are less educated, more remote and more engaged in agriculture. Since this statistic, Ethiopia has been hit by a series of climate shocks including the worst drought in 50 years which happened in 2015/16.

These natural disasters coupled with corruption and inequality makes the country face significant social problems including poverty, food shortage and insecurity. As a result, compared to healthcare, schools and critical infrastructure, improved digital access can seem like a luxury to many Ethiopians. The need for delivery of efficient services plays a vital role in promoting development and citizen empowerment. Therefore, the community would be highly benefited with the adoption of open data.

3.5.3.5 Organizational Context

Economic growth is the best way to reduce poverty. Tufts University study found that digitization is one of the biggest drivers of a nation's economic success. The report argues that careful policy-setting mostly achieves economic growth. Moreover, the government, as a policymaker, is the best driver of development.

In March 2015, the Ethiopian government with the help of the World bank group launched an open data initiative to provide data for agriculture and socio-economic wellbeing. As part of a global move by WB Group to further support open data initiatives, the central

statistical agency of Ethiopia started opening up data for the public in collaboration with the Living Standards Measurement Study-Integrated Surveys on Agriculture team of the WB Group. (*FBC - Ethiopia launches open data on agriculture, socioeconomic wellbeing* 2017)

4 Practical Part

This chapter describes the development of an open data-driven application in relation to food security in Ethiopia. Two data sources were mainly used for this purpose. A significant portion of the data has been provided by the food security portal from the international food policy research institute. The input of food security statistics from the Food and Agricultural organisation of the United Nations was further combined for the completion of the practical implementation.

4.1 Open data-driven application on food security

Food security is defined as having adequate access to safe, sufficient and nutritious food at all times to maintain a healthy life. Its three principal elements are availability, access and utilisation of food supplies.

The repetitive headline of intense drought having millions on the brink of starvation is becoming a common thing for the country. Seasonal drought with millions depending on rainfall to cultivate agricultural land is a major setback to the country's sustainable development goal of “zero hunger” which is also affecting the country’s agriculture-led economy. This agricultural led economy is heavily dependent on the regional and worldwide export of raw agricultural materials and goods, such as coffee beans and a wide range of flowers.

The primary motivation behind the choice of working on a web application based on the topic of food security in Ethiopia and the creation of the following use cases is the current ongoing severe drought caused by El Nino. El Nino is a climate cycle which creates a heavy weather impact throughout the globe. One of the known effects of this disastrous natural phenomena is the drought. It began affecting Ethiopia starting from 2015 and, two years forward, the situation has not seen much of an improvement. During the second half of 2017, the consequences are numerous for many unfortunate people. It could be mainly highlighted by the fact that 10.5 million people do not have regular access to drinking water. Another 8.5 million people required emergency food supply and more than 3.6 million mothers, and children required supplementary feeding in the profoundly affected regions of the country.

4.1.1 Use Case

Every Ethiopian understands the issue of food insecurity which the country faces every year. The part of the population who is not directly affected since it does not reside in the drought-covered areas, acquire the information regarding the situation from the local media and reports published by the government. However, getting informed may not always be enough. In order to grasp a thorough understanding of the published reports, one should be able to analyse the measured data behind the information provided.

Use Case1: Ethiopia is one of the biggest food aid recipients in Africa. One can ask how this is helping alleviate population undernourishment. Is the country anywhere close to meeting the sustainable development goals?

On this scenario, the visualisation will be done using data sets of "receipts of food aid (cereals)", "cereals import dependency ratio" and "population undernourished (%)". Additionally, relative comparisons with two more countries in Africa are provided. The two countries which were selected for this purpose are the Republic of Kenya and the Democratic Republic of Congo. The reason to join all the selected three countries is that they are considered as countries which need improvement on food security in the region of sub-Saharan Africa. Ethiopia and the Democratic Republic of Congo are also mainly considered to have a severe risk of food insecurity. Kenya has managed to reach a better score on the food security index than Ethiopia. Meanwhile, the Democratic Republic of the Congo has the lowest score among this group of three but also measures as one of the lowest scores in the world. On the global scale, out of 113 countries surveyed Kenya has ranked on the 86^{th} position, Ethiopia on the 99^{th} and the Democratic Republic of the Congo on the 112^{th} place.

Receipt of food aid is considered the overall amounts of cereals: mainly wheat, barley, maize, rye, oats, millet, sorghum, rice, buckwheat, canary seed, fonio, quinoa, triticale, wheat flour. The cereal component of blended foods is added too. The combined supply is finally assigned as food aid and transported from all donors to a recipient country. Cereals import dependency ratio (CIDR) allows us to understand how much of the domestically available cereal supply is imported and how much comes from the country's own production.

The minimum requirement for dietary energy is usually between 1750 and 2030 kilocalories per person per day. Population undernourished is the percentage of the country's population which has a food intake less than this minimum dietary energy.

Use Case 2: The use of technological innovation in agriculture has seen its advantages in many countries. Farmers' use of drought and disease resistant crops and drip irrigation to increase agricultural yield can be mentioned. Both measures have had a great impact in many countries' efforts of becoming self-sufficient in their food supply for their population. Agricultural yield is defined as the output of a crop per unit area of cultivated land. The same positive output of agricultural innovation has also been recorded in Ethiopia. For this use case, there are two component measurements to be visualised: the improvement of calorie supply from agricultural products and the reduction of food deficit in Ethiopia.

Calorie supply per capita is measured in kilocalories per capita per person. The standard calculation is done by dividing the available food supply by the number of the population of a country. This figure gives, as a result, the amount of food available for human consumption. However, the actual consumption by a person varies significantly among the population. Therefore, the average supply available for human consumption is chosen to be used as a more appropriate measure.

The depth of food deficit shows, while everything else is held constant, how many calories are needed to raise a person from an undernourished status. This deficit is calculated from the difference between required average dietary energy and the average consumption of dietary energy of the undernourished population. The obtained result is finally multiplied by the number of the undernourished people. Thus, the final number will provide us with an estimate of the food deficit of the whole country.

Use case diagram

The main end users of the web-based application will be food aid donors and government policymakers. The website can also be expected to be used by researchers and the public as a source of credible information.

The use cases described above will take into consideration three major actors: the user, the administrator and the data provider.

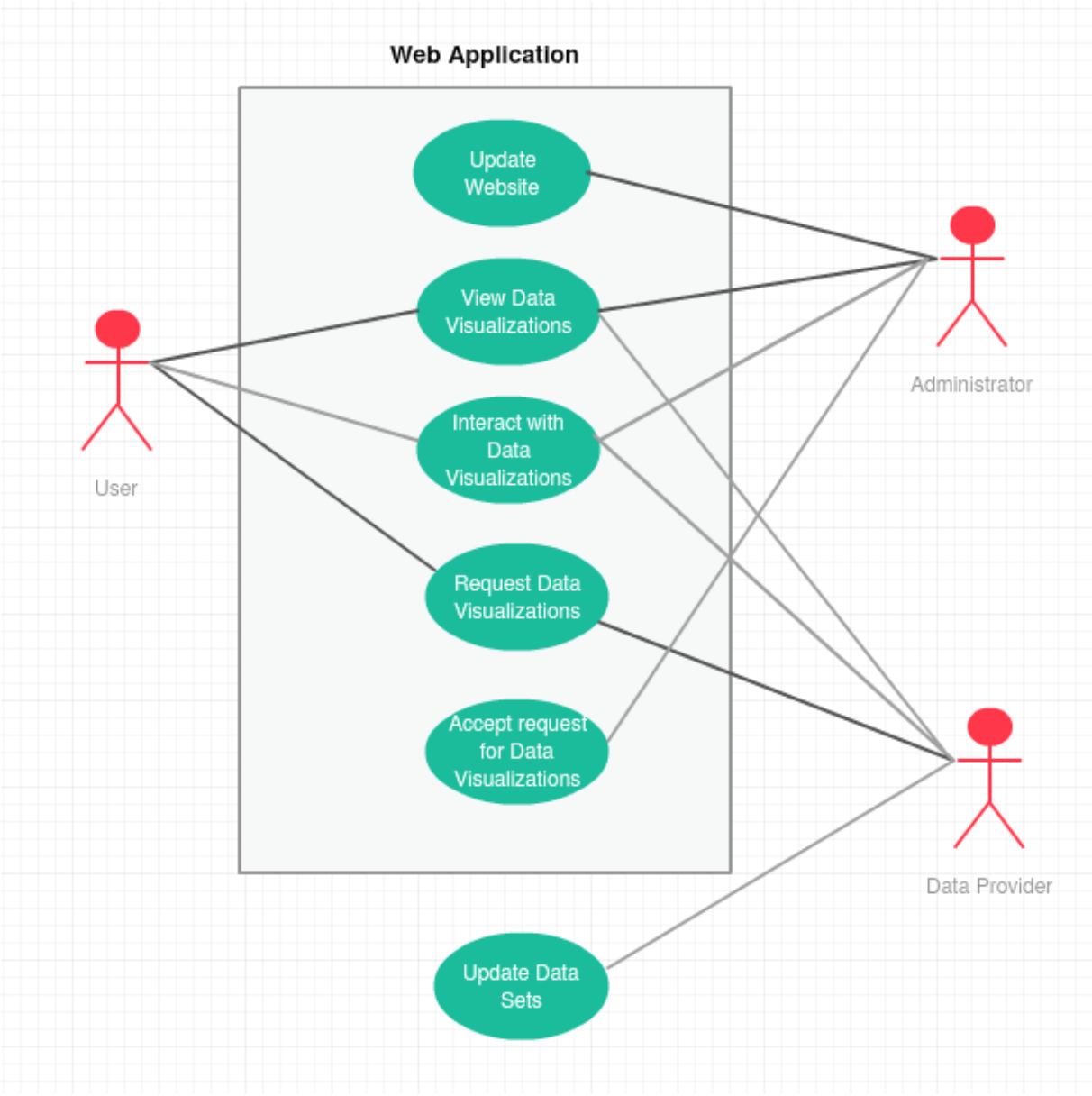

Figure 15: Use case diagram [Source: own]

The **user** of the web application will be able to perform the following actions:

1. Access data visualisations.

Through the web browser of an internet-connected device, the users can access the web application, and within few clicks, they will obtain instant clear, intuitive and comprehensive visualisations of the available open data in the website.

2. Interact with data visualisations.

The web application will allow the user to change few selection criteria for the data of interest to analyse, compare and understand.

3. Request for additional visualizations.

In case the user will not be able to find the exact information and the visualisation for the particular data needed, the website offers the option of contacting directly with the website administrator by requesting the needed information they expected from an open data portal.

The **administrator** of the web application shares with users the first two actions. Additionally, the admin has rights to perform:

1. Update and maintenance of the website.

The admin is responsible that the website content is with up-to-date and accurate information. Regular tests and maintenance improvements will be performed to ensure the full technical functionality of the web application.

2. Receive requests for data visualization.

The admin will handle the requests for additional data visualization by the users. Through assessing the available information and the newly proposed visualisation, the admin will be able to deliver the complete request.

The **data provider** is expected to be able to interact with the website in the same way as the user. The main difference is the ability of the data provider to have rights to update the data set with most recent measurement and indicators.

4.1.2 Wireframes

As part of the design process, the first step taken was to design the web pages. This design was done with the help of wireframes to materialise how the end user will view and interact with the website.

The first page is the “landing page”. This page is to give users simple information about what the project is about and to provide additional reliable sources of information on where data about Ethiopia and food security in Ethiopia can be found including the actual data sources which were used for the implementation of this thesis.

Figure 16: Wireframe for “landing” page [Source: own]

The second page is designed for the output of the two use cases which are mentioned in the previous subchapter. The page will contain a description of the use case and the corresponding data visualisations. The data visualisations will also include comparison with the two countries, i.e. Kenya and the Democratic Republic of the Congo.

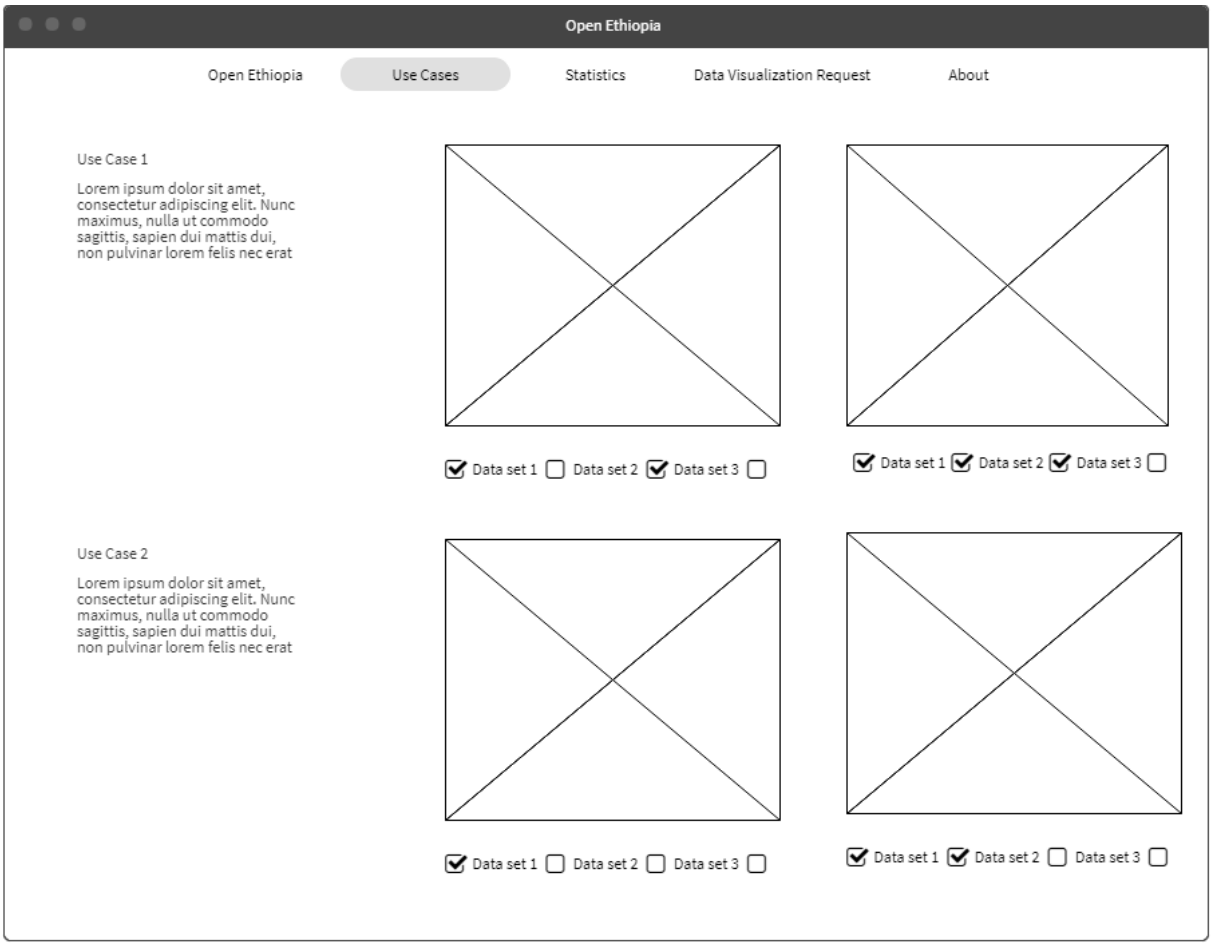

Figure 17: Wireframe for "Use cases" page [Source: own]

The third page served by the website is the "statistics" page. This page will contain data visualisation of different indicators for achieving food security. This statistic will also include a comparison of Ethiopia with different special groups. These groups are categorised as least developed, landlocked and developing, low income and lower-middle income countries. These special groups were created by the food and agricultural organisation of the United Nations.

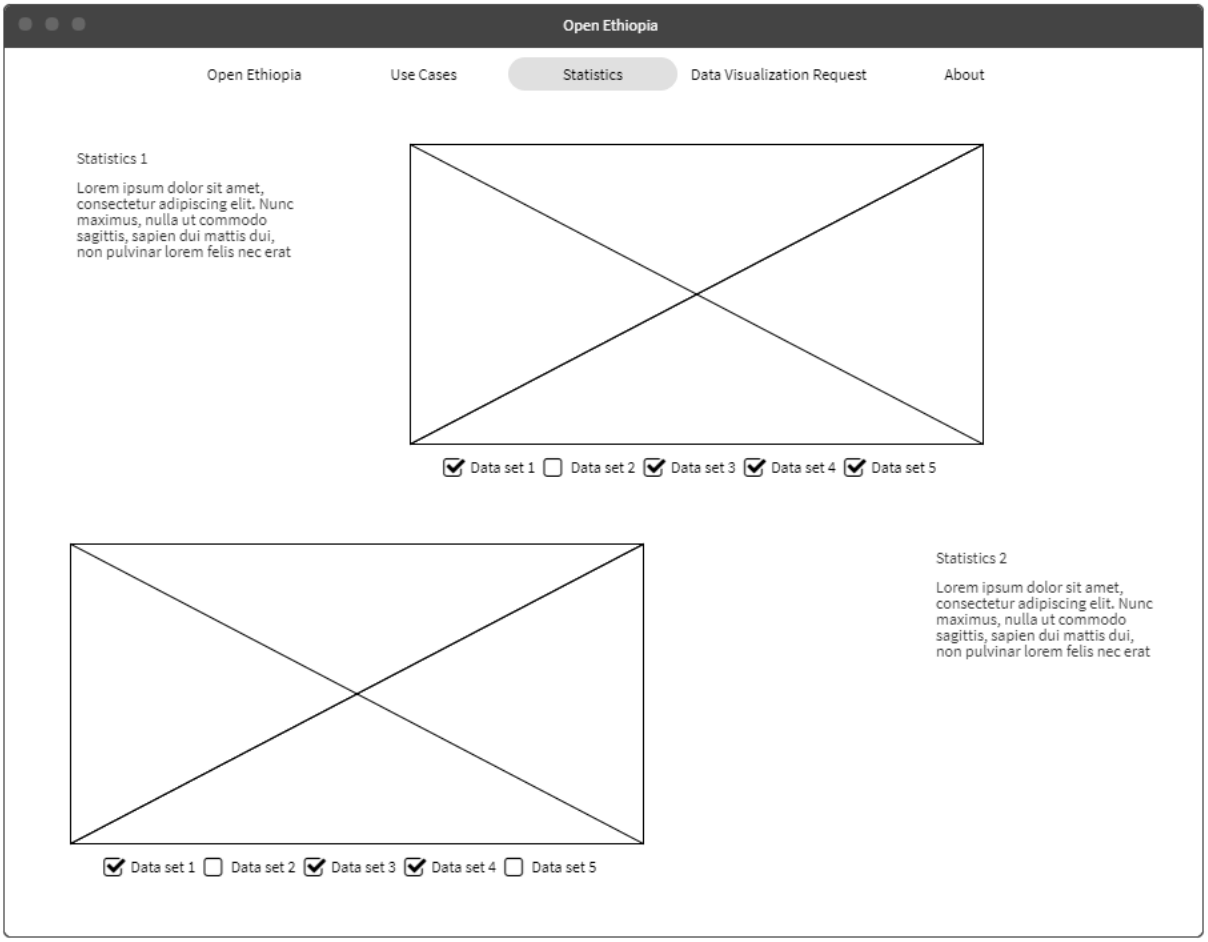

Figure 18: Wireframe for "Statistics" page [Source: own]

The fourth page served on the website is the "data visualisation request". To ensure the continuous delivery of useful data visualisation, the website should be able to accept user request. This page will provide a simple information entry form to collect user's contact information and description of their request. The primary goal of the project is to provide data visualisation for food security area. Therefore, this page will accept user requests related to the subject.

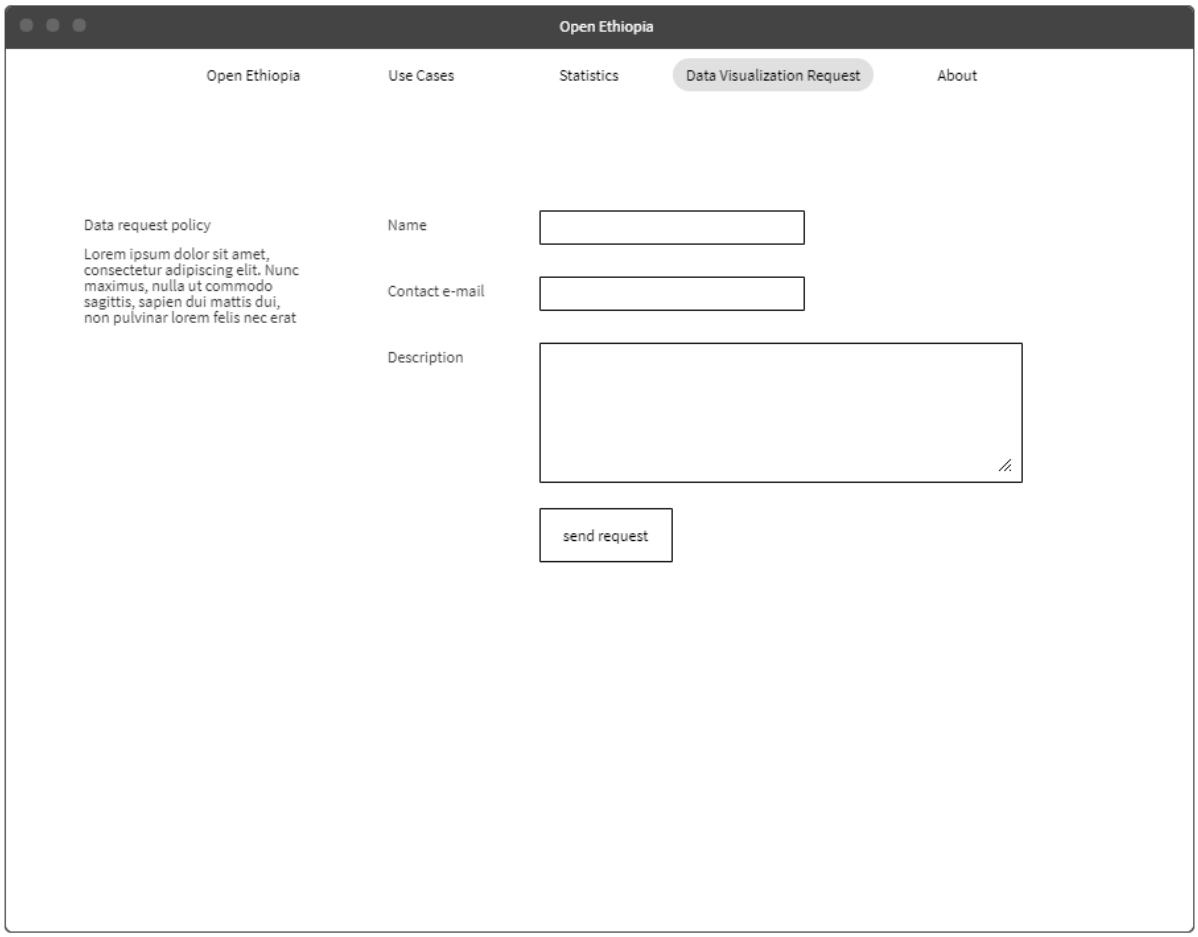

Figure 19: Wireframe for "Data visualisation request" page [Source: own]

The last page is the "about" page. The page will include general information about the objective of the project, the methodology used, future works and information about collaboration.

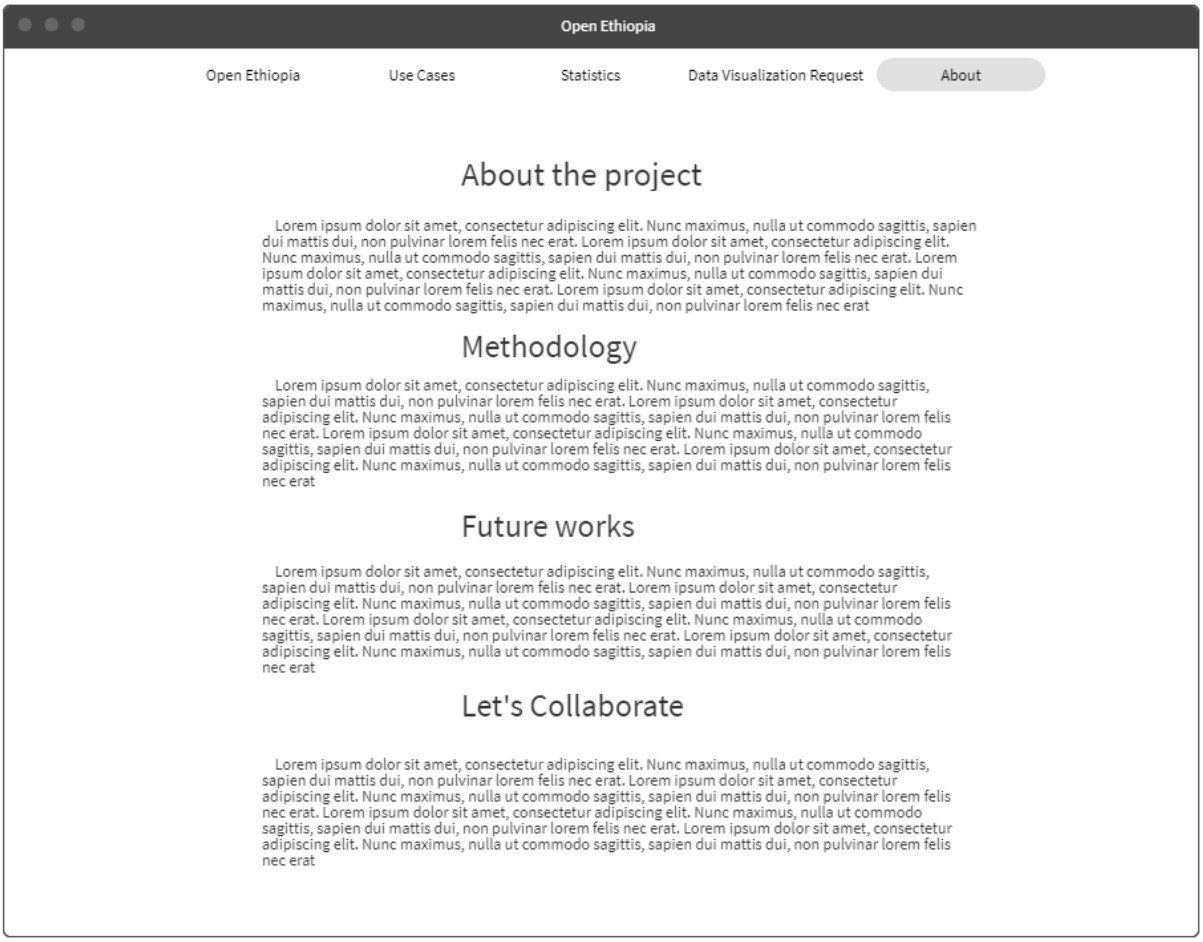

Figure 20: Wireframe for "About" page [Source: own]

4.2 Data sources and formats

The data sources used for the implementation of the practical part are from the International food policy research institute(IFPRI)[7] and the food and agricultural organisation of the United Nations(FAO)[8].

FAO provides a dataset on food security indicators for all countries in the world. These countries are divided into regions and subregions including special groups such as Least developed countries, Landlocked developing countries, Small island developing States, Low-income economies, Lower-middle-income economies and Low-income food-deficit countries. This dataset is provided as an Excel sheet which was revised in December 2016.

[7] http://www.foodsecurityportal.org/api/countries
[8] http://www.fao.org/economic/ess/ess-fs/ess-fadata/en/#.Wh092LpFxYc/

IFPRI provides data dashboard which can be accessed as an HTML table, can be downloaded as a CSV file or can be accessed through a data API by using a query builder. The data API also returns a CSV format for a requested query.

4.3 Implementation

Implementation of this design was done using conventional website development techniques and approach. The first step was choosing technologies which would be best suited for the goal of the design. The goal of the design was presenting data as charts or visualisations. This data presentation should help a user without an advanced statistical knowledge understand the meaning behind the data.

4.3.1 Technology chosen

For this implementation, the technologies used were HTML, Bootstrap, PHP, MySQL, HighCharts, JavaScript, jQuery and Nginx server. The reasons for the selection of each of these technologies are described below.

Bootstrap provides an excellent documentation which includes demonstrations and examples on how to use the stylings offered for many elements required to design web applications and websites. This detailed documentation makes it one of the most sought-after CSS frameworks amongst web designers. Bootstrap's get started guide makes it easy for someone new to web design avoids the hassle of writing a CSS script. The grid system and an extensive list of other components made it preferable to be used for the implementation of this practical part.

JQuery, a JavaScript library, was used to handle the parsing of data and also the presentation of the data visualisation. JQuery takes many lines of JavaScript code and packs them into methods which can be called in just a line. The methods provided makes it easy to use JavaScript on a website than writing a pure JavaScript code. JavaScript is a lightweight, interpreted programming language which allows client-side scripts to interact with the user and help the website serve dynamic pages.

The data visualisation was done by using HighCharts JavaScript library which offers modules for creating charts and maps. Highcharts provides professional looking 2D/3D charts and maps which gives an appealing look for data presentation. It works out of the box for many use cases but can also be customised based on the user's demand. It was chosen for this project because it is free and open source, it boasts of providing one of the fastest renderings of charts, having a responsive design, and being touch optimised.

PHP is a server-side scripting language which is used to create websites dynamic and allows for users to interact with web pages. It is free, efficient and widely used by the web development community compared to other server-side scripting languages.

Nginx web server was chosen to host the website. Nginx is a lightweight server which is becoming more popular for the ability to handle a large number of connections at a time. In comparison to other web servers which can handle many requests per second, Nginx shows a steady performance when it comes to concurrent connections.

Agile method of development was used throughout the implementation of the practical part of this thesis, in which an iterative process was used with an incremental approach. From a single JavaScript and HTML file which was used to implement the basics of the web application, i.e. creating a data visualisation from raw data, the project was incremented one web page at a time after testing the result and seeking consultation.

4.3.2 Server-side implementation

The first step done was acquiring a virtual server. Development was done using a virtual cloud server with a 512MB of memory and a 20 GB of Disk Space. The server was acquired from digital oceans, a cloud infrastructure provider. This provider was selected for its pricing and easy to set up servers, or as the provider calls it “droplets”.

As briefly mentioned in the chosen technologies, Nginx is used to server the web application. This set up was done as part of implementing the LEMP software stack on a server droplet. The LEMP stack is found to be an efficient way to host a dynamic webpage and web application by using open source software. This software stack includes Linux operating system

(Ubuntu 16.04.3 LTS), a Nginx web server, a MySQL database to hold data and PHP (version 7.0) for the processing of dynamic requests.

The following figure shows a code snippet which was used to configure the server side to support PHP post-processing and to assign the servers IP address to the domain name acquired.

```
server {
        listen 80;
        listen [::]:80;
        server_name openethiopia.info;

        root /var/www/openethiopia.com/html;
        index index.html;

        location / {
                try_files $uri $uri/ =404;
                if ($request_method ~* "(GET|POST)") {
                      add_header 'Access-Control-Allow-Origin' '*';
                 }
       }
        location ~ \.php$ {
        include snippets/fastcgi-php.conf;
        fastcgi_pass unix:/run/php/php7.0-fpm.sock;
        }

        location ~ /\.ht {
        deny all;
        }
}
```

Figure 21: Nginx script for configuring the web server [Source:own]

A MySQL database was created to hold data request entries. Through PHP post method, data is saved to a simple table of four columns. This table will hold user's name, contact information and description about the data request or a general comment about the project.

The table has the following structure:

- Datavisrequest
    - ID: integer, autoincrement, serves as a primary key
    - Name: Varchar
    - Email: Varchar

- Description: Varchar

The following SQL query was used to create this table:

```
mysql> CREATE TABLE IF NOT EXISTS datavisrequest (id INT NOT NULL PRIMARY
KEY AUTO_INCREMENT, name VARCHAR(255) NOT NULL, email VARCHAR(255) NOT
NULL, description VARCHAR(255) NOT NULL);
```

Figure 22: SQL query for creating table

The following code snippet shows a PHP script which accepts a form from the web page, implements a form validation to avoid SQL injection attack on the database and finally pushes the data on the form to the MySQL table created to hold this data.

```
<?php

require 'dbconnect.php';
$conn    = Connect();
$name    = $conn->real_escape_string($_POST['name']);
$email   = $conn->real_escape_string($_POST['email']);
$descr   = $conn->real_escape_string($_POST['description']);
$query   = "INSERT into datavisrequest (name,email,descr) VALUES('" . $name
. "','" . $email . "','" . $descr . "')";
$success = $conn->query($query);
if (!$success) {
    die("Couldn't enter data: ".$conn->error);
}
$conn->close();
header("Location: {$_SERVER['HTTP_REFERER']}");
Exit;

?>
```

Figure 23: PHP script for inserting data into table [Source: own]

The above PHP code snippet calls another PHP file which includes security credentials to access the database. After the connection with the database is established the table will be populated with the data passed by the user.

```php
<?php
function Connect()
{
 $dbhost = "localhost";
 $dbuser = "root";
 $dbpass = "**********";
 $dbname = "requests";

 $conn = new mysqli($dbhost, $dbuser, $dbpass, $dbname) or
die($conn->connect_e$

 return $conn;
}

?>
```

Figure 24: PHP script to connect to database [Source: own]

4.3.3 Client-side implementation

The client-side processing of the data sets and to be able to use HighCharts library for presenting the data visualisation JavaScript was used.

For the data sets available from the data API of the IFPRI, the URL links for the specified data sets were used. These URL links were generated by the query builder on their data dashboard. This URL link returns a comma-separated value (CSV) when called by a jQuery get method.

The following is an example of a CSV data fetched by an online tool which supports cross-origin request. It can be observed that the returned output is a comma separated values in a string format. The first line gives us the categories, in this case, they are the years when the data was collected. Moreover, the rest of the lines are the values of the series. The first position of the array holds the name of the series.[9]

[9] https://robwu.nl/cors-anywhere.html

```
GET https://cors-
anywhere.herokuapp.com/http://www.foodsecurityportal.org/api/countries/fao-receipts-of-
food/ethiopia.csv
200 OK

"Country","1961","1962","1963","1964","1965","1966","1967","1968","1969","1970","1971","19
72","1973","1974","1975","1976","1977","1978","1979","1980","1981","1982","1983","1984","1
985","1986","1987","1988","1989","1990","1991","1992","1993","1994","1995","1996","1997","
1998","1999","2000","2001","2002","2003","2004","2005","2006","2007","2008","2009","2010",
"2011","2012","2013","2014"
"Ethiopia","","","","","","","","","","","","","","","","","","","","","","","","","","","","","
","","","","","","1455920","1396230","1989700","3164180","2986500","2344300","2832070","26
82940","3298330","2825555","2743880","2906314","3911869","4029630","3336795","3776440","38
97163","4986125","6069413","6158318","6491540","7234955"
```

Figure 25: Results received from IFPRI data API [Source: (Demo of CORS Anywhere 2017)]

4.3.3.1 Processing of data from API

The following code snippet shows how the data was fetched from the API using the get method of jQuery and how JavaScript was used to parse through the lines of CSV data. This parsed data is then divided into categories, the name of the series and values. This data is at the end is fed as coordinates for the chart to be rendered to a specific division id on the corresponding HTML file by using HighCharts library.

```
$(document).ready(function() {
	var options = {
		chart: {renderTo: 'usecase10',type: 'line'},
		title: {text: 'Receipts of Food'},
		xAxis: {categories: []},
		yAxis: {title: {text: 'Metric Tons'}},
		series: []
	};

$.get('https://cors-anywhere.herokuapp.com/http://www.foodsecurityportal.org/api/cou
ntries/fao-receipts-of-food/ethiopia.csv', function(data) {

$.get('https://cors-anywhere.herokuapp.com/http://www.foodsecurityportal.org/api/cou
ntries/fao-receipts-of-food/congo.csv', function(data2) {

$.get('https://cors-anywhere.herokuapp.com/http://www.foodsecurityportal.org/api/cou
ntries/fao-receipts-of-food/kenya.csv', function(data3) {
			var lines  = data.split('\n');
			var lines2 = data2.split('\n');
			var lines3 = data3.split('\n');
			$.each(lines, function(lineNo, line) {
				var items = line.split(',');
				if (lineNo == 0) {
					$.each(items, function(itemNo, item) {
						item = item.replace(/^"(.*)"$/, '$1');
						if (itemNo > 0)
options.xAxis.categories.push(item);
					});
				}
				else {
					var series = { data: []};
					$.each(items, function(itemNo, item) {
							item = item.replace(/^"(.*)"$/,
'$1');
							if (itemNo == 0) {  series.name
= item;}
							else {
series.data.push(parseFloat(item));}
					});
				options.series.push(series);
				}
			});
			/*
			...
			...
			*/
		var receipts_of_food_chart = new Highcharts.Chart(options);
		    });
	  });
	});

});
```

Figure 26: Processing of data from a data API [Source: own]

For the other data visualisations which are not using this data API provided by IFPRI, an alternative approach was necessary.

4.3.3.2 Processing of data from an Excel sheet

The second data source which was used was FAO, and FAO only provides its dataset on food security indicators in an excel sheet format. Since this data set is updated only once a year, hard-coding the data in the JavaScript files was found to be a practical approach.

The following code snippet shows how this second approach was achieved using JavaScript and HighCharts framework. Dictionaries were used to hold the data for this approach. The dictionaries hold the name of the series and an array for the rest of the values. These dictionaries are then used as coordinates values to render the charts using HighCharts library.

```
$(function () {
	Highcharts.chart('usecase12', {
		chart: {type: 'line'},
		title: {text: false},
		xAxis: {
			categories:
['1990-92','1991-93','1992-94','1993-95','1994-96','1995-97','1996-98','1997-99','1998-00',

'1999-01','2000-02','2001-03','2002-04','2003-05','2004-06','2005-07','2006-08','2007-09','2
008-10','2009-11','2010-12',
                                        '2011-13','2012-14*','2013-15*','2014-16*']
		},
		yAxis: {min: 0,title: {text: 'Cereal import dependency ratio'},
				stackLabels: {enabled: true,style: {fontWeight: 'bold',
                                                                color:
(Highcharts.theme && Highcharts.theme.textColor) || 'green'}}
		},
		legend: {
			align: 'right',
			x: 0,
			verticalAlign: 'top',
			y: -10,
			floating: true,
			backgroundColor: (Highcharts.theme && Highcharts.theme.background2) ||
'white',
			borderColor: '#CCC',
			borderWidth: 1,
			shadow: false
		},
		tooltip: {
			headerFormat: '<b>{point.x}</b><br/>',
			pointFormat: '{series.name}: {point.y}'
		},
		plotOptions: {column: {stacking: 'normal',
                                dataLabels: {enabled: false,color: (Highcharts.theme
&& Highcharts.theme.dataLabelsColor) || 'white'}}
		},
		series:[{name: 'Ethiopia',data:
[0,7.9,17.3,11.1,8.9,4.6,4.4,5.7,9.9,10.8,10.7,12.1,10.6,9.9,6.0,6.0,7.3,9.9,11.3,10.7,,,,,,
,]},
                        {name: 'Kenya',data:
[8.2,8.8,13.7,12.0,12.1,18.4,24.4,26.9,24.7,25.4,24.8,20.8,19.5,20.6,21.7,22.5,24.3,34.9,34.
3,36.4,,,,,,,]},
                        {name: 'Congo',data:
[95.3,95.5,94.8,93.3,91.4,91.4,92.3,94.3,95.5,95.2,95.2,94.5,92.9,91.9,91.9,89.9,89.9,90.0,8
9.4,90.0,90.1,92.9,,,,,,,]}]
	});
});
```

Figure 27: Processing of data obtained from Excel sheet [Source: own]

4.3.3.3 Displaying data visualisations

Bootstrap card element was used to align the data visualisations created as shown in the following code snippet. Because of the support a responsive design by Bootstrap and HighCharts, the visualisations on the website will adjust to any screen size.

```
<div class="card">
<div class="card-block" id="usecase10" ></div></div></div>
<div class="col-md-6 col-lg-6">
<div class="card">
<div class="card-block" id="usecase12" ></div></div></div>
<div class="col-md-6 col-lg-6">
<div class="card">
<div class="card-block" id="usecase11" ></div></div></div>
```

Figure 28: HTML script to display data visualizations created [Source: own]

4.3.3.4 Accepting user requests

Accepting of requests was done using a simple HTML form. After the user fills in all the required fields and clicks send a request, the form data is sent to a PHP file for processing.

```
<form id="request-form" method="post" action="pushtodb.php" role="form">
<div class="row">
	<div class="col-md-6">
		<div class="form-group">
			<label for="form_name">Name *</label>
			<input id="form_name" type="text" name="name"
				class="form-control" placeholder="name *" required="required"
				data-error="Name is required.">
			<div class="help-block with-errors"></div>
		</div>
	</div>
</div>
<div class="row">
	<div class="col-md-6">
		<div class="form-group">
			<label for="form_email">Email *</label>
			<input id="form_email" type="email" name="email"
			   class="form-control" placeholder="contact email *"
required="required"
					data-error="Valid email is required.">
			<div class="help-block with-errors"></div>
		</div>
	</div>
</div>
<div class="row">
	<div class="col-md-12">
		<div class="form-group">
			<label for="form_message">Description *</label>
			<textarea id="form_message" name="descr"
				class="form-control" placeholder="description of your request
here*" rows="4" required="required"
				data-error="Please,leave a description.">
			<div class="help-block with-errors"></div>
		</div>
	</div>
	<div class="col-md-12">
		<input type="submit" class="btn btn-success btn-send" value="Send Request">
	</div>
</div>
</form>
```

Figure 29: HTML form to accept user request [Source: own]

5 Results and Discussion

This chapter discusses the approaches used for completing the practical part and the results obtained. The implementation of the practical part used two different approaches, which were used for the purpose of presenting the data visualisation. It is possible to achieve the same results through the implementation of both a data API or hard-coding the raw data directly in the web application.

In the case of the first approach, there is the distinct advantage of maintenance. Once the data provider updates the data sets with new entries, the new information will be automatically reflected. Thus, there will not be any need for manipulating the source code of the web application

On the second approach, it is less convenient when there is any change in the data provider side. In this case, it is necessary that the web administrator should be aware of such changes in advance in order to be able to maintain the source code of the application.

Considering the fact that a majority of the data sets in the discussion are of a time-series nature which is updated annually, or at most monthly, it would not be a problematic part. Generally speaking, the modification expected in the source would be adding one more entry to the data series affected.

A more crucial consideration that needs to be handled in both approaches is the interoperability between systems. The absence of an established way to ensure system and data interoperability according to open standards means that the value of open data will only be partial and leading to discrepancies with the source data.

When using the data API for pulling the datasets from the source, it was discovered that the target server does not support the cross-origin resource sharing (CORS). The lack of such feature means that any web browser will reject any request to the target server because the web server hosting the application is using a different domain. The reason behind this situation is a security mechanism. It forces web browsers to block cross-origin HTTP requests which are

generated from scripts, in the case of this implementation from the JavaScript code. (*Cross-Origin Resource Sharing (CORS)* 2017)

However, there are three possible solutions to the security-mechanism problem that is worth discussing. One proposed solution is modifying the target server by adding to it the code for a CORS header. The second possibility is through the setup and configuration of a proxy server. It would serve to proxy the requests coming from the client through a web browser. On the other hand, when the server acquires the data it will include a CORS header, which will then be allowed by a web browser to present the requested data. The third viable option is to make use of another API which will perform the second option of implementing a proxy server.

Because the author of this thesis does not have direct access to the target server, and to avoid setting up a proxy server, the third option was chosen as the most suitable to be used. CORS-anywhere is an API which enables cross-origin requests to any server. This approach is made by simply appending the uniform resource locator (URL) of the API at the beginning of the URL of the target server when using the get method of jQuery.

5.1 Results of use cases

Although a detailed statistical analysis of the output of the data visualisations is beyond the scope of this thesis, a preliminary analysis was done.

The first use case was to combine different data sets to create visualisations for analysing how food aid and import of cereals affect reducing the number of population undernourished. We can observe that the cereal import dependency is more or less constant throughout the last decade(Fig: 31), but there is a massive rise in food aid in the past few years(Fig: 30). The weather changes and extreme drought the country is facing, even in the current time of writing this thesis, can be accounted for this increase of food aid from foreign countries. We can also see that the percentage of the population which is undernourished has been decreasing gradually throughout the decade(Fig: 32). There can be many factors for this decline in addition to the increase in received food aid. One of which is the distribution of stored food supply by the government's relief agency to the areas which are highly affected by drought.

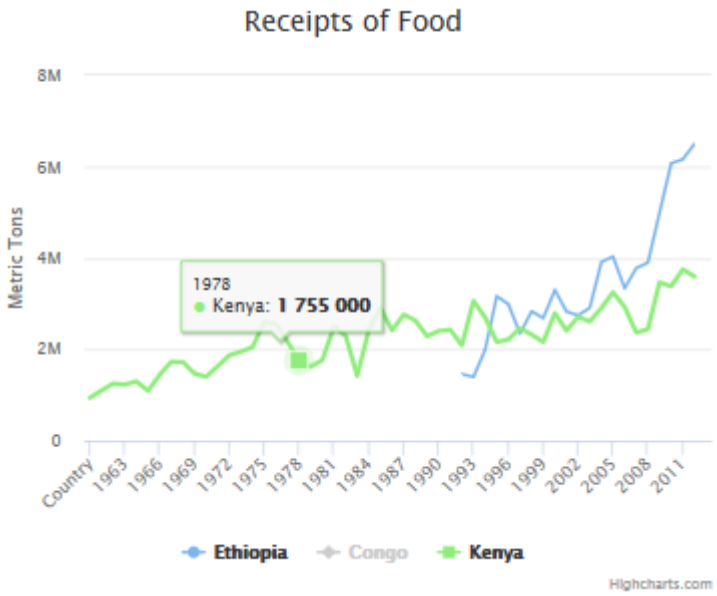

Figure 30: Receipts of food aid by Ethiopia and Kenya [Source: adapted from (Receipts of Food Aid (cereals) | Food Security Portal 2017)]

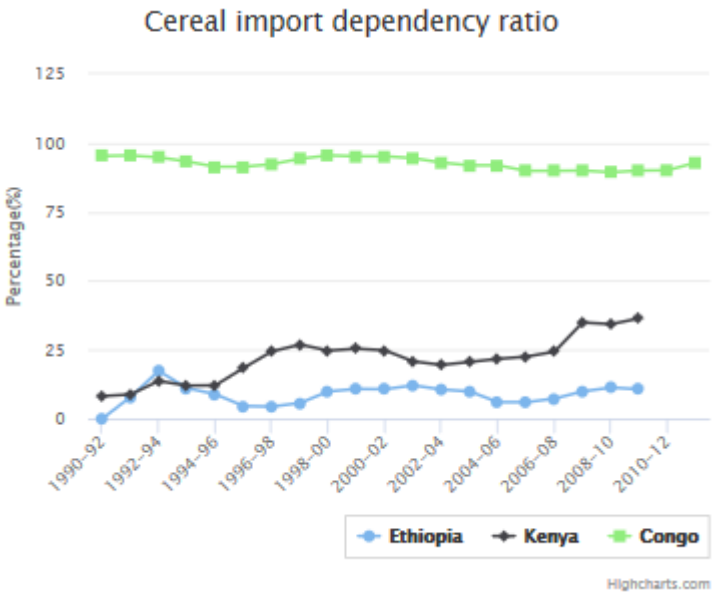

Figure 31: Cereal import dependency ratio for Ethiopia, Kenya and Congo [Source: adapted from (FAOSTAT 2017)]

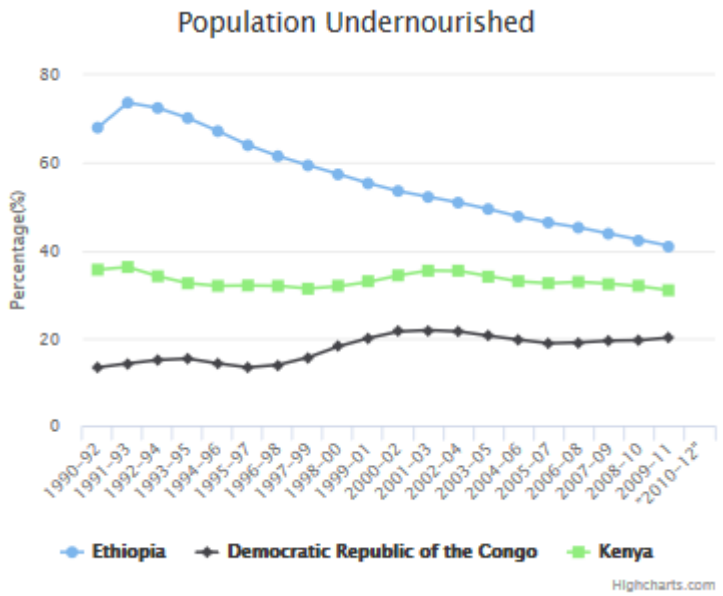

Figure 32: Percentage of population undernourished in Ethiopia, Congo and Kenya [Source: adapted from (Calorie Supply Per Capita, Crops Equivalent | Food Security Portal 2017)]

In the second use case, the calorie supply from plant origin and the depth of food deficit is visualised. The visualisation shows an increasing trend throughout the decade for calorie supply and a decreasing trend for the depth of food deficit. This data presentation shows that there is an increasing yield in the agricultural sector(Fig: 33) which, amongst other factors like calorie supply from animal origin, is reducing the depth of food deficit (Fig: 34)

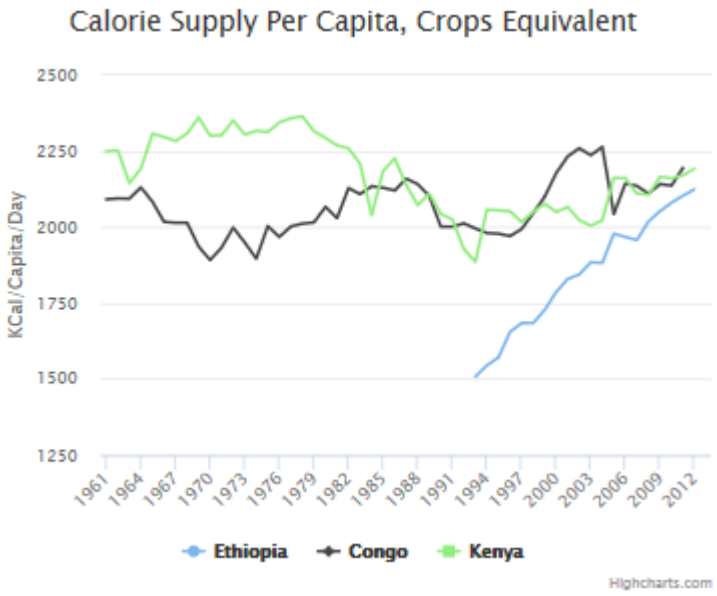

Figure 33: Calorie supply per capita, crops equivalent for Ethiopia, Congo and Kenya [Source: adapted from (Countries fao-calorie-supply- | Food Security Portal 2017)]

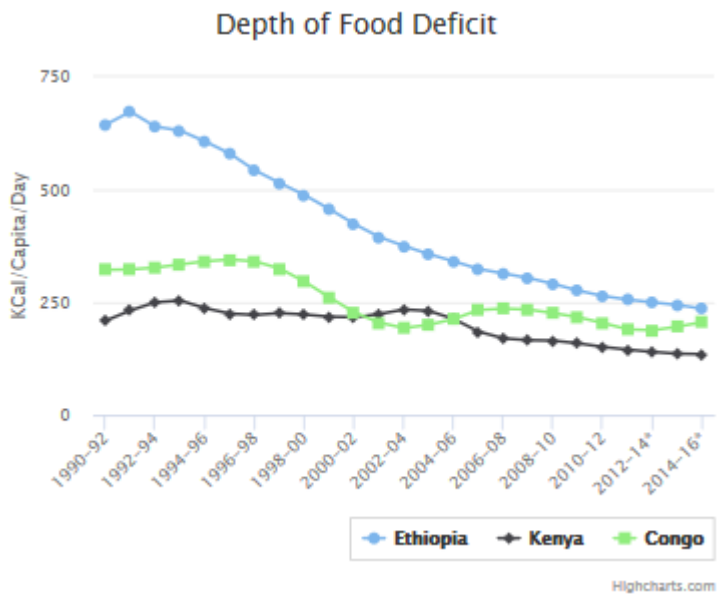

Figure 34: Depth of Food Deficit in Ethiopia, Kenya and Congo [Source: adapted from (FAOSTAT 2017)]

As viewed in the results of the data visualisation, some of the datasets are not up to date. The main reason behind this is the gap between data collections by FAO. In the datasets like percentage of population undernourished or depth of food deficit, and so forth, to reduce the impact of possible errors, FAO has used a three-year average. In some cases, FAO also states that it has used linear interpolation of data to make up for missing entries in their data series.

Open data in the areas of agriculture and nutrition has many merits. In the Ethiopian context, two significant benefits can be mentioned. Driving effective policy and better decision making, and enabling innovation in agriculture.

The main constraint of effective policymaking is not having easy access to a quality data and unbiased information. The cost of accessing data can be divided into direct and indirect cost. The time and resources which are required to find who is responsible for disseminating the data and requesting for access can be considered as indirect costs. In addition to this indirect cost, there are licence and subscriptions fees, i.e. direct costs which will lessen the interest of any potential user who wants to be informed or a developer who wants to create additional

value and product to the available data. Therefore, a widespread availability of data at low cost for access will drive a better decision making and an efficient policymaking.

Innovation in agriculture is already undergoing in Ethiopia. The mass use of drought and disease resistant crops for agriculture is one example. The author firmly believes that innovation should not stop here. Agricultural research should be available for the masses; this will push people to innovate by finding the gaps which are not covered by currently proposed solutions.

As mentioned in the contextual analysis there is a vast digital divide between people living in urban and rural areas. Access to the internet for the majority of parties involved in agriculture is virtually unavailable. Half the population of Ethiopia has a mobile subscription which makes them reachable with short message service (SMS). However, sending data visualization thorough SMS is virtually impossible. A mere one sixth of the population has access to data and broadband internet. Out of which the target user group of the application provided by this thesis will be limited to policymakers in disaster relief organisations in the country and external food aid donors.

The practical implementation part of this thesis was aimed at adding a new value to the data sources used. The data sources as mentioned were the IFPRI and FAO which is one of the leading, and mainly responsible, of collecting data on food and nutrition on a global scale. It collects data from governmental and non-governmental agencies all across the world, combines these data with its own resources and openly provides datasets and statistical reports for everyone. IFPRI, on the other hand, takes this data provided by FAO and adds addition values as making the data searchable, filterable and downloadable as well as provide a data API which helps integrate applications from other developers.

Development using open data resources, by default, is assumed to introduce new value. This thesis provide data visualisations to selected datasets which are not provided by the two data sources mentioned.

5.2 Load test

A load test was performed to see if there will be any degradation of the web application if there is a request from many users at the same time. The load test was performed on the website using a free online application against 5000 requests within one minute and was found to have a reasonable average response time of 96 milliseconds, and no errors were reported. [10]

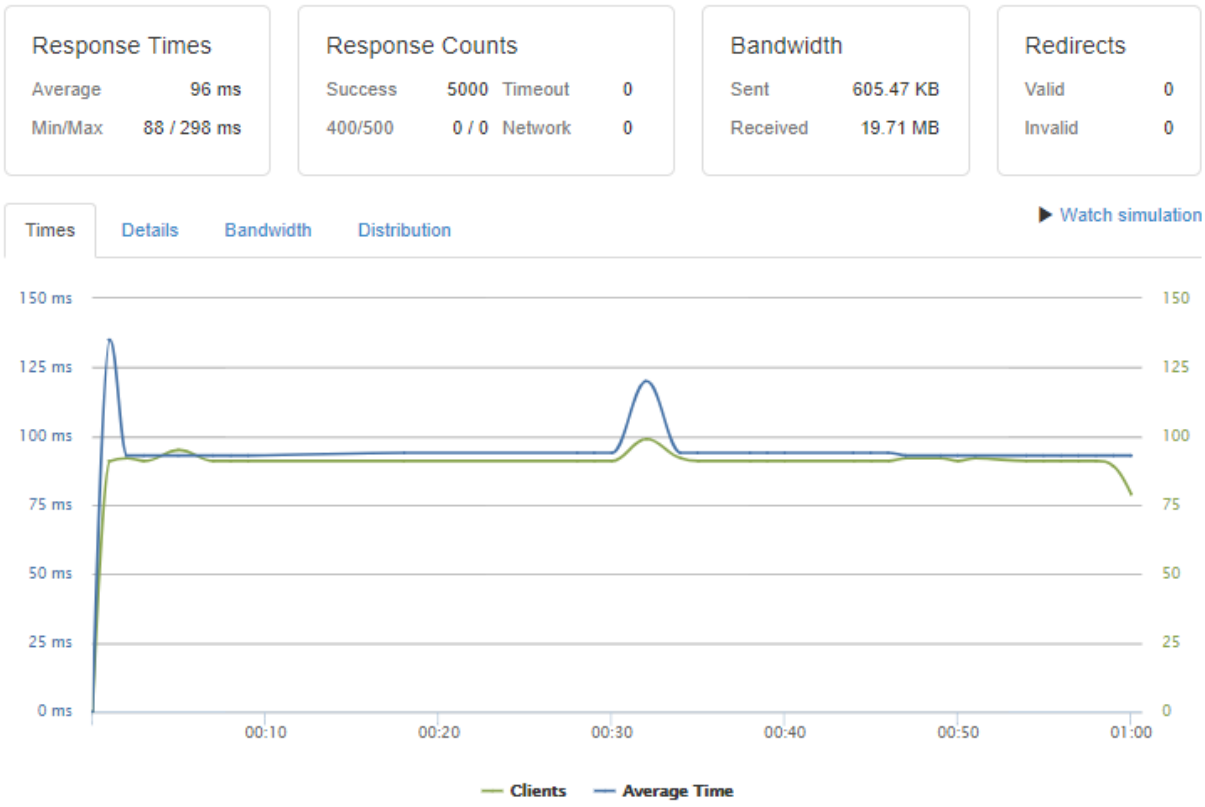

Figure 35: Result from load test on openethiopia.info [Source: (Application Load Testing Tools for API Endpoints with loader.io 2017)]

5.3 Comparison with open data for Africa's web application

Finally, a comparison was made with food security data visualisations provided by open data for Africa. Open data for Africa on its microsite for Ethiopia has an extensive list of datasets on different sectors including food security.[11]

Under the topic of Food security, it provides two data sets namely "FAO Food Security Data by Food Groups/Items, June 2012" and "Regional Population Distribution and Daily per

[10] http://www.loader.io
[11] http://ethiopia.opendataforafrica.org

Capita Calorie Intake, 2000-2011, Ethiopia". After selecting a data set, it takes the user to a data visualisation tool. a user can interact with the data visualisation tool by selecting different variables.

The variables included in the dataset for food security are food consumption quantities, food production and trade, and diet composition. It is listed that the source of this data is the food and agricultural agency of the united nations. The dates range from the year 1992 to 2007 and it this data visualisation was published in 2012.

The data set for population distribution and calorie intake has regional population distribution by sex and residence, and regional daily per capita calorie intake across time. The webpage states that the source of the data is the central statistical agency of Ethiopia. The dates range from the year 2000 to 2011, and the data visualisation was published in 2013.

It is visible that there has been a tremendous effort in making this visualization and open data available for the public, but the primary setback is the frequency of updating the data sets. The web page mentions that the next release date is the beginning of the year 2018. That means there is a 4 to 6 years duration between updates on the data visualisations.

The practical part of this thesis tried to remove this significant delay between updates by utilising a data API provided by IFPRI. Unlike the case with the open data for Africa, when the data provider updates its data source the data visualisations on the web application provided will automatically update.

6 Conclusion

The primary objective of this thesis was to determine how open data can facilitate the development efforts of Ethiopia. This thesis has analysed the benefits of open government data, its implications on developing countries and has come to this conclusion.

The level of poverty in most of the African countries does not come as news to many. However, the development initiatives do not get as much coverage as the broad underdevelopment. The story of Ethiopia is not so different from other fellow developing countries in the continent. From vastly growing small and medium enterprises to massive projects like the grand renaissance dam boost the country's development. The strong urge to become a lower middle-income economy drives the government and people of Ethiopia in making the country one of the fastest developing countries in the world.

One can admire this continuous effort of development but eradicating poverty is not an easy and straightforward endeavour. First and foremost, it needs a continuous engagement of citizens and government. Governments should change their policy on engaging more citizens in government processes and hear the voices of the people. Inequalities, as it is present now, will worsen with the lack of engagement and government transparency. Good governance, although many would argue its existence, is diminishing and inconsistent. This problem makes the civil societies and media including data backed journalism to abandon their duties of passing unbiased information to the people.

In the case of Ethiopia, the government has a tight grip on information technology resources, mainstream as well as social media platforms as it feels threatened by the growing use of internet and technology.

A considerable number of researchers in developmental studies would agree that extreme poverty is a result of bad governance, corruption, inequality, and lack of accountability. These can be avoided, to some extent, by the adoption of open government and fostering the use of open data. This thesis discussed and analysed that the main bottleneck for adopting an open policy in Ethiopia will be the current infrastructural capacity constraint in Ethiopia.

Adopting open government cannot solve the poverty issue by itself, but can provide much potential in the struggle for development. Checks and balances are core principles of fighting corruption. Budget transparency and open contracts can, for example, let citizens oversee the usage of their countries resources. Which in turn makes it harder for government officials and involved parties to commit corruption.

The first partial objective of this thesis is to review public access to government data. The thesis has identified the main source of Ethiopian government data to be the central statistical agency of Ethiopia. The agency is responsible for collecting data form all other government agencies and distribute data to interested parties

The second partial objective of this thesis was to amylase the available datasets in Ethiopia and evaluate their level of openness. This thesis has reviewed the level of openness of available data sources in Ethiopia and has found it to be poor compared with the standards of open data. Based on the five-star deployment scheme, the sources in Ethiopia provide a two-star web data. Ethiopia is also classified as a capacity constrained country by the open data barometer. The efforts of the government for opening data to the public has shown that there is a potential for adoption and the need for a dedicated party to drive this endeavour to become successful has been identified.

The last partial objective was to analyse and design a web application for the agricultural sector in Ethiopia. Open data-driven web application, openethiopia.info, was designed as a practical implementation using open datasets on food security. The thesis managed to show an easy approach on how to add a new value to open government data. The implementation is limited to the use of two data sources. Further development can be done by integrating more sources and providing a statistical analysis of the output of the data visualisations. The target user group of the application provided is limited to the portion of the population which has access to data or broadband internet.

In conclusion, the thesis highlighted how the adoption of open data, open principles and later an open government would contribute to the country's development. Open data counters the failure in dealing with the roots of poverty and push the country to join the lower-middle income economies by the target year of 2025.

7 References

1 A Primer on Machine Readability for Online Documents and Data - Data.gov, 201. *Data.gov* [online],

2 ATTARD, JUDIE, ORLANDI, FABRIZIO, SCERRI, SIMON and AUER, SÖREN, 2015, A systematic review of open government data initiatives. *Government Information Quarterly*. 2015. Vol. 32, no. 4, p. 399-418. DOI 10.1016/j.giq.2015.07.006. Elsevier BV

3 TAUBERER, JOSHUA, 201, Open Government Data: The Book. *Opengovdata.io* [online]. 2016. [Accessed 5 April 2016]. Available from: https://opengovdata.io/

4 GROUP, OPEN, 2016, Open Definition 2.1 - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge. *Opendefinition.org* [online]. 2016. [Accessed 5 February 2016]. Available from: http://opendefinition.org/od/2.1/en/

5 Goldstein, Bre, Lauren Dyson, and Abhi Nemani. Beyond Transparency. San Francisco, Calif.: Code for America Press, 2013. Print.

6 HERZOG, TIMOTHY, 2017, Technology Options | Data. *Opendatatoolkit.worldbank.org* [online]. 2017. [Accessed 14 February 2017]. Available from: http://opendatatoolkit.worldbank.org/en/technology.html

7 BERNERS-LEE, TIM, 2017, Linked Data - Design Issues. *W3.org* [online]. 2017. [Accessed 11 April 2017]. Available from: https://www.w3.org/DesignIssues/LinkedData.html

8 KORN, NAOMI and OPPENHEIM, CHARLES, 2016, Licensing Open Data: A Practical Guide. *Discovery.ac.uk* [online]. 2016. [Accessed 14 March 2016]. Available from: http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf

9 *Open Data Goldbook For Data Manager And Data Holders*. 1st ed. European Data Portal, 2016. [Online]. [Accessed 16 Mar. 2016].

10 THE STAR, CODE FOR KENYA, 2016, StarHealth | Kenya's leading health portal by the Star - Home. *Health.the-star.co.ke* [online]. 2016. [Accessed 27 March 2016]. Available from: https://health.the-star.co.ke/

11 Yannoukakou, Aikaterini, and Iliana Araka. "Access To Government Information: Right To Information And Open Government Data Synergy". *Procedia - Social and Behavioral Sciences* 147 (2014): 332-340. [Online].

12 COETZEE, JACQUES, 20167, Nigeria launches Open Data initiative to stir innovation, investment and growth – Ventureburn. *Ventureburn* [online]. 2016. [Accessed 23 November 2016]. Available from: http://ventureburn.com/2014/01/nigeria-launches-open-data-initiative-to-stir-innovation-investment-and-growth/

13 E., EYO, 2016, NODA. *Opendata.com.ng* [online]. 2016. [Accessed 10 December 2016]. Available from: https://opendata.com.ng/

14 Ghana Open Data Initiative | Ghana Open Data Initiative, 2017. *Data.gov.gh* [online],

15 DAVIES, TIM, 2017, Rankings. *Opendatabarometer.org* [online]. 2017. [Accessed 13 February 2017]. Available from: http://opendatabarometer.org/2ndEdition/analysis/rankings.html

16 REFUGEES, UNITED, 2017, Refworld | Ethiopia: Proclamation No. 590/2008 of 2008, Freedom of the Mass Media and Access to Information Proclamation. *Refworld* [online]. 2017. [Accessed 12 February 2017]. Available from: http://www.refworld.org/docid/4ba7a6bf2.html

17 Ethiopian government's witchhunt against privately-owned media | Reporters without borders, 2017. *RSF* [online], [Accessed 27 April 2017]
18 World Development Indicators | The World Bank, 2017. *Wdi.worldbank.org* [online],
19 ITU | 2017 Global ICT Development Index, 2017. *Itu.int* [online],
20 FBC - Ethiopia launches open data on agriculture, socioeconomic wellbeing, 2017. *Fanabc.com* [online], [Accessed 27 April 2017]
21 Where My Money Dey?, 2016. *Wmmd.codeforafrica.org* [online], [Accessed 27 April 2016]
22 PesaCheck, 2017. *PesaCheck* [online], [Accessed 27 April 2016]
23 HURUmap Kenya: Making Census Data Easy to Use, 2017. *Kenya.hurumap.org* [online], [Accessed 27 April 2016]
24 Kenya Open Data, 2016. *Opendata.go.ke* [online], [Accessed 27 April 2016]
25 Home - African Information Highway Portal, 2016. *Knoema* [online], [Accessed 27 April 2016]
26 openAFRICA, 2016. *Africaopendata.org* [online], [Accessed 27 March 2016]
27 Welcome - Ethiopian Government Open Data Portal, 2017. *Data.gov.et* [online], [Accessed 15 June 2016]
28 The Open Data Handbook, 2016. *Opendatahandbook.org* [online], [Accessed 22 September 2016]
29 FAOSTAT, 20176. *Fao.org* [online], [Accessed 12 September 2016]
30 Calorie Supply Per Capita, Crops Equivalent | Food Security Portal, 2017. *Foodsecurityportal.org* [online], [Accessed 12 September 2016]
31 Receipts of Food Aid (cereals) | Food Security Portal, 2017. *Foodsecurityportal.org* [online], [Accessed 12 September 2016]
32 Countries fao-calorie-supply- | Food Security Portal, 2017. *Foodsecurityportal.org* [online], [Accessed 12 September 2016]
33 Application Load Testing Tools for API Endpoints with loader.io, 2017. *Loader.io* [online], [Accessed 10 November 2016]
34 Demo of CORS Anywhere, 2016. *Robwu.nl* [online], [Accessed 27 March 2016]
35 Cross-Origin Resource Sharing (CORS), 2016. *Mozilla Developer Network*[online], [Accessed 10 November 2016]

8 Appendix

1. The complete source code for the practical part of this thesis is found on the following GitHub repository: https://github.com/leulab/Open_Data_Ethiopia
2. The implementation of the web application is provided here: http://openethiopia.info/