Robin Larsson

Uganda Open Development Partnership Platform

How can the platform be improved?

Informatics
Bachelor's Thesis

Semester: June 2013
Supervisor: Remigijus Gustas
Abstract

Poverty and corruption are problems that developing countries fight everyday. Politicians and leaders abuse their influence to gain property and wealth by unacceptable means. Open data is a way to challenge these problems and make a positive change in the society.

Open data can be freely accessed, used, reused and re-distributed by anyone. Open development is about providing the means for organizations to share open data. The process for open development in Uganda has just begun and there are many reasons to have an open development platform that gathers open data, which the society demands when the government are not willing to share their data. The Uganda government denied invitation to join Open Government Partnership, that would mean a commitment to open development. This leaves non government organizations to take the first step to introduce open data to the society and other organizations. The Uganda Open Development Partnership Platform is an initiative by non government organizations that combines their knowledge to promote open development. The development of this platform has just begun and it needs further assistance to be ready for the public. The available data sets and documents on the platform are released in proprietary formats, without the alternative of open formats. A portal for open development that releases documents and data sets in proprietary formats alone are something that collides in the platform's purpose of being a portal for promoting openness.

The purpose of this report is to promote open data with an overview about the subject and explain improvement proposals on flaws in the Uganda Open Development Partnership Platform. The initiatives and organizations that practice open data can get initial guidelines from this report on how to apply open data. The research should display the current state of the platform to identify the flaws and get an understanding how the platform works. The research was performed through interviews that were conducted in Kampala, Uganda, for three weeks in the beginning of 2013. This gave the chance to meet numerous citizens of this developing country which offered information of interest for the research. Observations were made by visiting the partners of the platform, in order to see how they work with current means of visualization and to get an understanding of what can be improved. The Uganda Open Development Partnership Platform can be improved with the presented proposals to introduce the open license, multiple formats for material and structured data sets.
# Table of Contents

1 Introduction .......................................................................................................................... 5
   1.1 Background ....................................................................................................................... 5
   1.2 Purpose ............................................................................................................................... 6
   1.3 Research Objectives ......................................................................................................... 6
   1.4 Target Audience ............................................................................................................... 6
   1.5 Statement of the problem .................................................................................................. 6
   1.6 Scope of the study ............................................................................................................. 7
   1.7 Limitation ......................................................................................................................... 7

2 Methodology ........................................................................................................................... 8
   2.1 Choice of Subject ............................................................................................................... 8
   2.2 Research Types ................................................................................................................. 8
   2.3 Data Analysis Techniques .............................................................................................. 8
   2.4 Data Collection ................................................................................................................ 9
   2.5 Choice of Respondents .................................................................................................... 10
   2.6 Procedure ......................................................................................................................... 10
   2.7 Methodology Problems ................................................................................................... 10
   2.8 Reliability and Validity .................................................................................................... 11

3 Theory ..................................................................................................................................... 13
   3.1 Technologies and Software ............................................................................................ 13
      3.1.1 Remote Application Programming Interface (Web Service API) ......................... 13
      3.1.2 Data Formats ............................................................................................................ 13
      3.1.3 Software .................................................................................................................. 15
      3.1.4 Type of Visualization ............................................................................................. 16
   3.2 Open Data ....................................................................................................................... 18
      3.2.1 Open Definition ........................................................................................................ 19
      3.2.2 Open Concepts ......................................................................................................... 19
      3.2.3 Open Data Benefits ............................................................................................... 20
      3.2.4 Open Data Initiatives ............................................................................................. 20
      3.2.5 Type of Open Data ................................................................................................. 21
      3.2.6 Open Formats vs. Proprietary Formats .................................................................. 21
      3.2.7 Open License ......................................................................................................... 22
      3.2.8 Problems with Open Data ..................................................................................... 23
      3.2.9 Open Data in Uganda ............................................................................................ 23
   3.3 Uganda Open Development Partnership Platform ......................................................... 25
      3.3.1 Open Data Case ....................................................................................................... 25
      3.3.2 Partners ................................................................................................................... 25
List of Tables

Table 1: List of Acronyms..................................................................................................................3
Table 2: Example of data set structure in a spreadsheet.................................................................14
Table 3: Forest Cover 1990 and 2005 data from data.ug.................................................................17
Table 4: Overview of respondents..................................................................................................28
Table 5: Identification of data.ug data sets visualization and type.....................................................31
Table 6: Part of Polished Data Set of Agriculture Exports 2006 to 2010 (Uganda Opendev 2012a)......32
Table 7: Part of Transposed Data Set of Agriculture Exports 2006 to 2010........................................33
Table 8: Part of original data set Quantity of fish produced in Ugandan lakes 1961-2010 (Uganda Opendev 2012a)......................................................................................................................34
Table 9: Part of transposed data set Quantity of fish produced in Ugandan lakes 1961-2010..............35

List of Figures

Figure 1: A column graph created with data set forest cover from data.ug.................................17
Figure 2: A pie chart created with data set forest cover from data.ug.............................................18
Figure 3: A map over number of reported issues per category for Uganda Watch 2011.
Figure 4: Graph in columns from data set of Agriculture Exports from 2006 to 2010.
Figure 5: Graph in lines and dots from data set of Agriculture Exports from 2006 to 2010. ..........34
Figure 6: Chart made based on data set from open development platform....................................35
Figure 7: Analysis model diagram..................................................................................................37
Figure 8: Creative Commons Attribution 3.0 Unported License.
   Source: https://opendata.go.ke. [2013-03-28]............................................................................40
Figure 9: Creative Commons Attribution-ShareAlike 3.0 Unported License
   Source: http://www.opendevelopmentcambodia.net. [2013-03-28]............................................40
### Table 1: List of Acronyms

<table>
<thead>
<tr>
<th>Shortening</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC</td>
<td>Association for Progressive Communications</td>
</tr>
<tr>
<td>ATIA</td>
<td>Access to Information Act</td>
</tr>
<tr>
<td>CC</td>
<td>Creative Commons</td>
</tr>
<tr>
<td>CIPESA</td>
<td>The Collaboration on International ICT Policy in East and Southern Africa</td>
</tr>
<tr>
<td>CKAN</td>
<td>Comprehensive Knowledge Archive Network</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-Separated Values</td>
</tr>
<tr>
<td>DRT</td>
<td>Development Research and Training</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Standard Web Resource Protocol</td>
</tr>
<tr>
<td>JSON</td>
<td>JavaScript Object Notation</td>
</tr>
<tr>
<td>NGO</td>
<td>Non Government Organization</td>
</tr>
<tr>
<td>ODC-by</td>
<td>Open Data Commons Attribution License</td>
</tr>
<tr>
<td>ODS</td>
<td>OpenDocument Spreadsheet</td>
</tr>
<tr>
<td>ODT</td>
<td>OpenDocument Text document</td>
</tr>
<tr>
<td>OGP</td>
<td>Open Government Partnership</td>
</tr>
<tr>
<td>OKFN</td>
<td>Open Knowledge Foundation</td>
</tr>
<tr>
<td>PDF</td>
<td>Portable Document Format</td>
</tr>
<tr>
<td>REST</td>
<td>Representational State Transfer</td>
</tr>
<tr>
<td>SOAP</td>
<td>Simple Object Access Protocol</td>
</tr>
<tr>
<td>Sida</td>
<td>Swedish International Development Cooperation Agency</td>
</tr>
<tr>
<td>SPIDER</td>
<td>Swedish Program for ICT in Developing Regions</td>
</tr>
<tr>
<td>UBOS</td>
<td>Uganda Bureau of Statistics</td>
</tr>
<tr>
<td>UNICEF</td>
<td>The United Nations Children's Fund</td>
</tr>
<tr>
<td>UMDPC</td>
<td>Uganda Medical and Dentist Practitioners Council</td>
</tr>
<tr>
<td>Remote API</td>
<td>Remote Application Programming Interface</td>
</tr>
<tr>
<td>XLS</td>
<td>Microsoft Excel Spreadsheet</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
1 Introduction

The introduction chapter describes the background problems that open data can solve. The problems with open data in Uganda give context to the purpose. The research objective presents the research question. The question function as a ground for the statement of the problem that describes the flaws found during examination of Uganda Open Development Partnership Platform. The last part of the introduction is the scope and limitation that defines the boundaries of the study.

1.1 Background

Poverty and corruption are problems that developing countries fight everyday (Sida 2012). Politicians and leaders abuse their influence to gain property and wealth by unacceptable means. Open data is a way to challenge these problems and make a positive change in the society. The problems with corruption would get a closer step to a solution, for example, by having an open budget and property borders defined on maps. The effect of less corruption would be decreased poverty, since government and aid organizations' fundings would end up where it belongs, and a lot of it are meant for the poor. The Swedish Program for ICT in Developing Regions [SPIDER] (2012) contributes to the well developed platform of open development Cambodia and describes the problem, with open development as a solution.

“Lack of transparency in land conversion and industrial development endangers Cambodia’s poor, hampering advocacy and resource stewardship. Government and industry data tend to be inconsistent and incomplete and many organizations self-censor to avoid conflict with the government. The project will develop an Open Development Information Hub, an open data resource to assess and systematically track land conversion to enable evidence based analysis and timely and fact based advocacy.”

Open data can be freely accessed, used, reused and re-distributed by anyone. Open development is about providing the means for organizations to share open data. The process for open development in Uganda has just begun to take form. There are many reasons to have an open development platform that gathers data which the society demands, when the government is rarely willing to share. The available governmental data are spread in different ministries which makes it hard to gather. The data are hard to reuse and analyze because the shared material contains information in various shapes and forms. Uganda Open Development Partnership Platform [Uganda Opendev] (2013) describes the problems with denied access to information and bulky non user-friendly material.

“The Ugandan government refuses to acknowledge open development for unknown reasons. The Uganda government declined in September 2011 to join Open Government Partnership (OGP). The government was invited to join OGP at the same time as five other African countries joined the partnership. Kenya, Liberia, Ghana, South Africa and Tanzania, joined the OGP (Association for Progressive Communications & The Collaboration on International ICT Policy in East and Southern Africa [APC & CIPESA] 2012:9; Development Research and Training [DRT] 2012:1). Development Research and Training (DRT) is a non profit organization that perform policy research and analysis to benefit the poor.
DRT is investigating further the question why the invitation was declined by the government. Since the government does not promote openness leaves non-government organizations (NGOs) to take the first step to introduce open development to the public and other organizations. The Uganda Open Development Partnership Platform has just begun their development, as it was launched September 2012 (Uganda Opendev 2012a). This young initiative is in need of further assistance to be ready for the public.

1.2 Purpose

The purpose of this report is to promote open data with an overview about the subject and explain improvement proposals major flaws in the Uganda Open Development Partnership Platform. The initiatives and organizations that use open data can from this report get initial guidelines on how to practice the basics of open data. The research should display the current state of the platform to identify the flaws and get an understanding on how the platform works.

1.3 Research Objectives

The objectives with this report is to discover and display the current state of the Uganda Open Development Platform and propose the improvements that can be made. This includes finding out how the presentation of data sets are visualized, in what formats data sets are released, how data sets are licensed, what open data has done for Uganda and the problems the partnership has with open development. These areas can be summarized in a major question, that will be answered in this report.

– What is the current state of the Uganda Open Development Partnership Platform and how can the platform be improved?

1.4 Target Audience

The audience for this report is the governments, organizations and other stakeholders that use, or promote openness and want to get started with open development. The Uganda Open Development Partnership Platform should see this report as a proposal on how further development of the platform should be driven. The report can also be useful without an open data interest, to learn about visualization of data and how to use different data formats.

1.5 Statement of the problem

The Uganda Open Development Partnership Platform has many challenges with being open, referred to data being freely accessed, used, reused and re-distributed. Wrong structured data sets makes it hard to create good visualizations with the tools available in the platform (Kavi 2013).

The lack of variety in available formats for resources makes it harder to use the available material in other context. The use of proprietary formats creates the problem when opening the material with incompatible editors, since the proprietary formats are only fully compatible with the editor it was intended for. This forces users to obtain the software with the proprietary formats (Stallman 2007). The Uganda Open Development Partnership Platform is supposed to promote openness, but uses the proprietary format XLS (Excel Spreadsheet) for data sets, the static PDF format for most documents and the proprietary DOC format (Word text editor) for few editable documents.

Lack of license associated with the material creates uncertainty among the visitors of the platform, as
they do not know how the data can be used, reused or redistributed (Open Knowledge Foundation [OKFN] 2009).

1.6 Scope of the study

Open data is a wide subject, where many organizations produce policies and collect data. The open development platform focuses on gathering and presenting these documents and data sets, with the goal of spreading them. The focus of this report is similar to the open development platform, because the objective of this report is to improve the platform. The report's largest scope is structuring and visualizing data sets through graphs, in collaboration with the partnership. The improvements achieved for open data will be displayed in this report that presents what changes have been performed on the Open Development Partnership Platform. The improvements will be focused on open license, structure of data sets, visualization of data sets and usage of open formats. The open license's influence on the use of material has only been speculated in the theory. Unfortunately, no research has been performed to check these theories.

1.7 Limitation

A question that this report will not discuss is why the government refuses to acknowledge open development. This question is mentioned to explain why the Uganda Open Development Partnership Platform is needed and what its function is in the society.

One of the questions around open data is if open data can be useful in fight with the corruption and influence the poverty. This is a complex question and it will not be answered in this report. A case study on the possibility to use open data would be interesting to display since that would show what happens with data after it has been shared. This report's scope is limited to focus on the platform and its capability to share material.
2 Methodology

The methodology chapter describes how the study was performed to fulfill the purpose of the report. The chapter describes how collection of the material was conducted and techniques used to analyze the gathered data. This chapter is useful for those who want to redo the study for another initiative or the same initiative, to see if the results are different or has changed. The methodology problems can be useful for researchers in a similar study, to avoid falling in the same pitfalls and other problematic occurrences that could have affected the method.

2.1 Choice of Subject

The author was drawn to the subject when analyzing the project Open Development Cambodia. The project helps farmers who are fighting for the borders of their land. Families were previously driven from their homes without warning by influential persons that wanted the lands for their own purpose. The initiative helped by visualizing the land borders on maps to display further problematic occurrences and place a definition of the farmers land (Sida 2012). The author's previous encounter with Uganda made the choice to focus on similar subject as Open Development Cambodia and investigate the needs of open data in Uganda. The author's interest is to help developing countries to fight corruption and poverty by the expertise of using information technologies.

2.2 Research Types

The material has been gathered through interviews with people related to open data and people involved in the open development platform. Interview is one of few options to gather data on a new specific subject such as open development in Uganda. Questions in an interview should be of low structure, but still prepared so the interviewer has enough based questions to answer the purpose of the research.

Observations were performed in Uganda to discover how organizations exerts open development. An interview is not always enough to get the whole picture. Watching how an operator is practicing its work can communicate another picture from another angle and give more insight in the subject (Patel & Davidsson 2011:91).

Investigation was made to find suited projects for visualization with graph. Design evaluation method in form of architecture analysis was used to study how well different projects software and services function with the platform's available data sets. The design evaluation method used for the platform's existing tool was a static analysis to examine if structure of the tools static qualities could be used together with any data set. This leads to providing an analytical optimization to demonstrate how to operate properties of existing tool and define boundaries in the tools behavior (Hevner et al. 2004:86).

Short-life documents were used in form of Internet resources that were found when browsing the web (Patel & Davidsson 2011:67). Information in websites from open data initiatives, projects, services and articles were found useful to the study.

2.3 Data Analysis Techniques

The research types used for this study is qualitative research methods that results in text material, which means qualitative techniques should be used for data analysis. According to Patel & Davidsson
(2011:133) parts of the analysis of the material should be done during the data collection, to have interpretations written down, instead of trying to interpret later from ambiguous notes and fraction of memories from interviews.

The improvement proposals for the platform should display how they were presented to the partnership and be separated from the results that describes how the platform responded to the improvements. The interviews are described in the results as qualitative focused interviews, where the author's interpretation is first presented and then illustrated with quotation from the person interviewed (Patel & Davidson 2011:133).

2.4 Data Collection

One way to find a reliable source is to follow blogs and recommendations from leading organizations that promote open development. Open Government Partnership (OGP) (APC & CIPESA 2012:9) and Open Knowledge Foundation (OKFN) (Molloy 2011) are two reliable sources of material for data collection since they are the leading organizations in development of policies for open data practice.

Wiki How (WikiHow 2013a) and similar web pages that can be altered by anyone are not reliable sources. However, these web pages works well as source when the material is a guide that contains mostly images. This report has been using Wiki How as references to explain how a system works.

Appendix is placed in the end of the report. A more interested reader gets the opportunity to read more by referring to an appendix in the text. The blog entries were added to the appendix because they can change swiftly, before being indexed and stored in the state they were interpret.

Key words used for data collection when searching for information are listed below with explanation of the words meaning.

- Open data – data that are free to use, reuse and accessible for anyone.
- Open government data – open data are mostly used by governments.
- Open development – holds the meaning of open data, open source, open knowledge and openness.
- Data visualization – display data in a table, map or graph to make data easier for other people to interpret.
- Open source visualization – free tools for visualization.
- Open data initiatives – organizations that are promoting open data.
- Transparency – political decision and budgets should be public available.
- Accountability – people should feel responsible for their actions. The word is often seen connected to positive results with open data.
- Openness – abstract term of open.
- Open license – publish data with open license to make the data open.
- Open standards – open data should be practiced and in what formats to publish data.
- Open formats – formats data should the data sets be released in.
- Data sets – collected data stored in a file.
- East Africa – East Africa is a key word to find information regarding the area of Uganda and neighbor countries, which is useful when comparing Uganda to other, similar countries.
- Uganda – add Uganda to any other key word to find Uganda specific information.
2.5 Choice of Respondents

Respondents were found through looking for the latest open data news and leading open data initiatives in Uganda. Persons of significance that could contribute to the material for this study were found through the web. One interview was booked by contacting an interviewed person in an article, which lead to organizations and subjects related to the open development platform that were very good assets for this study.

In 2010 was the M4D 2010 conference, in Kampala, Uganda, that was attended by the author. Connections were made at the conference that proved useful when searching for respondents. Contacts at Karlstad University that had done work in similar areas were another good asset, since they contributed with experience and information that lead to more respondents. One meeting gave recommendations for further contacts that became respondents for the research.

2.6 Procedure

Meetings were planned by initial contact through email, to introduce the interviewed person about the author and that the purpose of the meeting was to write a thesis based on the gathered information. The interviews were performed by traveling to Kampala, Uganda, for three weeks in the beginning of 2013. This gave the chance to meet numerous citizens of this developing country which offered information of interest for the research. Observations were made by visiting the partners of the platform, in order to see how they work with current means of visualization, to get an understanding of what can be improved.

When interviews and observations had been made and information about current state had been gathered, then more data was gathered on how to visualize and use data. Each year is an international event, the Open Data Days (opendataday.org). This year the event was arranged for the first time in Kampala on the 23th of February (HacksHackers Kampala 2013). After preparations in the subject of visualization, the author attended the event to exchange knowledge and learn more about open data initiatives in Uganda. The event was useful to learn how to work with data, how to use data and different ways of visualization of data. Interesting experience like Open Data Days and stories told live about problems makes open data more understandable.

Notes were written in a diary at the moment the material was gathered during the visit in Uganda. Interpretations of interviews and observations were documented when they were fresh and later summarized in this report. The work was continued after leaving Uganda, with creation of guides and blog contributions about improvements to be published on the opendev.ug website.

2.7 Methodology Problems

There is information about the open development platform on their website, opendev.ug, that could have been used as a ground for the interviews. Unfortunately, the website was discovered during the stay in Uganda and was not properly analyzed before the interviews took place. The questions used in the interviews would have contained more questions related to the respondent that could be connected to the platform.

Attempts were made to book interviews in advance before traveling to Uganda, to use the time as effective as possible while staying in Uganda. Unfortunately, it was very hard to communicate through email from another continent. The feeling was that people were not interested, while the hope was that the contacted people did not want to plan so far ahead. When arriving in Kampala, the interviews started
to blossom, although as predicted, only a couple of days in advance could be planned. The author's energy and time could have been saved if the culture of planning was known in advance and there would have been no need to hunt the respondents so hard.

Less effort was required to get the interviews planned while staying in Uganda. The problem was getting to the interviews in time and finding the location. Homes have often been placed randomly. Kampala has close to no standard at all for the structure of buildings outside the city center. City plans has not been made for the whole city, and this makes it hard to find the location for an interview. The traffic would often get stuck in a condition called “traffic jam” and that could ruin the plans for a whole day, which was a normal and acceptable excuse for being late. This occurred once on the way to a meeting with a European, who had other meetings planned afterwards and counted on the author's arrival without the excuse of a traffic jam. This is why the precaution was taken to only book one meeting a day.

2.8 Reliability and Validity

Reliability means that areas of the research should be measured with the correct tool that is intended for the purpose. To have a high reliability in the tool used for the research means the same result should be provided when repeating the measurement, independently of the person using the tool (Professional Testing Inc. 2006). The measurement should not be affected by other elements than the one being researched (Patel & Davidsson 2011:103).

This study is about open data in an open development platform. Since the area is new, the information available on the web are thin and only few people can provide the information required for this study. The possible respondents for interviews are people with knowledge of open data and are applying open data in their organization. The questions should be formulated so the answers from the respondent can lead to new questions in the interview, which a survey can not provide because it is answered distance apart from anyone who can affect it. This makes the approach on the questions to be low grade of standardization and structure (Patel & Davidsson 2011:75). Open data is a new subject and the knowledge is thin, so the questions in an interview could lead to new knowledge for the respondent and open it's eyes to areas that could change how they adapt open data and other related subjects for their organization. The respondents might receive new information that affects their organization. Suppose that another researcher attempts to recreate this study by using the same respondents and using similar questions would result in different answers. The reliability for this study will be complicated to measure since the criteria changes.

Validity refers to measure what is meant to be researched, and not anything else (Professional Testing Inc. 2006). The researcher should be clear with what the study is about. Patel & Davidsson (2011:102) mentions two ways to measure validity, content validity and concurrent validity. Content validity is a logic analysis of the content for the tool, to see if the content is wide enough to capture the answers required for the research. For example the questions in a survey must be limited and short enough for the participants to be patient enough to answer and the survey questions must be formulated well enough that the answers covers the purpose of the research. Additionally, the interview questions should be prepared so the interviewer have enough base questions to collect answers for covering the purpose of the research. Concurrent validity is testing with a criteria to see the result of a variable in a tool to be compared with the result of same variable in another situation or context.

The questions concurrent validity are raised for the research since the results are the same from observations, interviews and short-life documents. The author's contributions for improvements of the
platform are proved when the partnership includes the contributions on their website. This would mean that interpretations of flaws in the platform discovered from interviews and observations were correct, which proves a high content validity of the research.
3 Theory

Theory describes how different parts of the report's subject are connected to create a better understanding of the research problem. The theory is divided into three parts to display the difference and relation to each part. The first part describes the general technologies and software used in context with open data. The second part describes what open data is and how open data can be used. The third part describes the Uganda Open Development Partnership Platform's current state of using open data and also the author's improvement proposals.

3.1 Technologies and Software

General technologies and software that are used in context to perform open data.

3.1.1 Remote Application Programming Interface (Web Service API)

Remote API are libraries with information about data that are available to fetch with web service technology such as Simple Object Access Protocol (SOAP) or Representational State Transfer (REST) (Suda 2003:1). A system can be set up to fetch data from one or multiple systems to another system, based on request from the user or automatically be updated on frequently bases. Remote API gives the possibility to create a visualization that is updated frequently with new information, which means a data set does not need the manual process of download and import every time there is need for an update of the visualized data.

3.1.2 Data Formats

Data formats are used to define a standard on how data can be used between different systems (Dietrich et al. 2012). This makes it possible to create a system based on theory, without having all the data sources available. The knowledge of what formats to expect would limit the development when the system only need to handle a few formats. Data formats used in open data context for different purpose are explained and compared in this section with a recommended structure.

Portable Document Format (PDF)

A common way to store and share material is through the PDF format, since many editors can save documents in this universal format (Open Formats 2004). PDF readers are free to use, which means everyone can open documents with this type of format. The difference to other formats is that the PDF format is static and cannot be edited, therefore editing programs with their own settings cannot ruin the document structure.

Comma-Separated Values (CSV)

The most plain format is the CSV format, that can be opened in most text editors and is used in many systems for import and export (Dietrich et al. 2012:17). The recommended structure for CSV files is to have each record on an own row, where values are separated by chosen character, mostly by either comma, semicolon or tab. The first row should be the headers for each column and each of the following rows representing a record (Kavi 2013). Another chosen character represents line breaks for separating the rows. Since only few characters are structuring the data it means the content does not use much
space, which makes CSV a lightweight format. One disadvantage with this format is that the tab character can be used for separating values. The tab character is interpreted differently by text editors and can be read as a space instead of a tab. This would make it hard to divide values with spaces from values intended to be separated with a tab character. A disadvantage with using the comma as value separator is that the value itself can contain a comma character for numbers with decimal but also in text (Dietrich et al. 2012:17). This would lead to value separating problem for all editors and all systems that supports CSV for data imports. The following example is of data from data set on data.ug. The data has been placed in the CSV format where values are comma-separated and lines are separated by line breaks. The comma can be used because the data only contains whole numbers and singular words.

Year,Victoria,Albert,Kyoga,ED/GE,Wamala,AL.NILE,OTHERS
1961,2550,1180,680,1250,100,0,180
1962,2340,1220,1320,1210,200,0,360

In this example, the first row is the column headers and the first header is the year, followed by represented lakes. The following rows are containing the values of captured fish in each lake.

**Microsoft Excel Spreadsheet (XLS) and Open Document Spreadsheet (ODS)**

XLS and ODS are very common formats to use for storing data in columns. The software Excel is widely used because many people have access to the spreadsheet software through public computers (Microsoft 2013). Excel Spreadsheet (XLS) uses the XLS format for the 1997/2003 version of the software and the Excel Spreadsheet eXtensible Markup Language (XLSX) format for the 2007/2010 version. The Excel Spreadsheet software is part of the Microsoft Office suit. An alternative spreadsheet software is the free version, Open Office Calculation (Calc) that uses the ODS format (Open Formats 2004) and is part of the free open source software suit, Open Office (WikiHow 2013b).

The recommended structure for a spreadsheet is to have the first row to be the headers for columns and have values on the following rows (Table 2). This structure make the format readable by both humans and systems that are supporting spreadsheet formats import.

<table>
<thead>
<tr>
<th>Row 1</th>
<th>Header 1</th>
<th>Header 2</th>
<th>Header 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 2</td>
<td>value1</td>
<td>value2</td>
<td>value3</td>
</tr>
<tr>
<td>Row 3</td>
<td>value4</td>
<td>value5</td>
<td>value6</td>
</tr>
</tbody>
</table>

The strengths with spreadsheets are that the column structure is human readable and the spreadsheet software has export possibilities to export content into other formats. One standard format to export content with is CSV, because it is a format that many systems can handle (Dietrich et al. 2012:17). A disadvantage with spreadsheets as data container is that more storage is required to store data compared to store data in the CSV format due to document background information such as formulas and text formatting.

**eXtensible Markup Language (XML)**

XML is a data container format that is mostly used by sending data to other systems through web services. This format can also be used to store data in a file. The strength of XML is the possibility to
specify own tags and choose name tags that suits the project (Suda 2003:2). Own specified tags becomes both system and human readable, which means that it is easy to write a XML document by hand. New standards will be developed from this language since it is structured by the developer. The web standards eXtensible HyperText Markup Language (XHTML), Web Service Definition Language (WSDL) and Resource Description Framework (RDF) have been created based on XML. The negative about XML is that tags specified by the creator might lead to a large and complex data container, compared to other data container formats. However, a standard should be used that defines how data containers are structured, to limit the length of tags capability and make it possible to create systems that handles data sets from different sources (Suda 2003).

**JavaScript Object Notation (JSON)**

JSON is lightweight when it comes to store and send data (Crockford 2006). This format can be edited directly in a text editor without formatting possibilities. The problem is that JSON is rarely human readable since it uses multiple different characters to separate columns, values, objects and groups of data. There is mostly no space between the characters when JSON is used to transfer data, which makes it hard to see what group belongs to which column.

### 3.1.3 Software

**Software for reading documents and data in different formats and projects that make it easier to use data.**

**Microsoft Office**

Microsoft Office is a software suite that contains the Word text editor. The Word software uses the document (DOC) format for 1997/2003 version of the software and document eXtensible Markup Language (DOCX) format for 2007/2010 version. The software suit also includes the Excel Spreadsheet software. The software suit is widely used because many people have access to the software through public computers (Microsoft 2013).

**Open Office**

The free and open source software suit, Open Office, is containing the Writer software for documents, which uses the Open Document Text (ODT) format, and the Calculation software for spreadsheets. The formats used in this software suit are based on the Open Document Format (Open Formats 2004). The profit with an open format is that any other software can implement the method to read and write in the same format. Libre Office (WikiHow 2013b) is a popular copy of the original open source software Open Office. One reason for Libre Office's popularity might be because it is in later versions installed with the popular open source operating system, Ubuntu.

**Miso Dataset**

The Miso Dataset project is an open source data handler with the goal to make it easier to import data from different sources (Graul 2012). The software can read data from the formats CSV, JSON, remote API and Google Spreadsheet, to combine into a single object. The object can be used in a script and can in combination with other software create, for example, a visualization with table, graph, chart or map.
Dataset is part of the Miso Project that is created by The Guardian and Bocoup. The project was started because existing tools for data handling can be hard to use and take a long time to implement (Graul 2012).

**Highcharts**

Interactive visualizations can be created with the framework Highcharts based on a few parameters for x-axis, y-axis and the headers to display. The software is developed by Highsoft Solutions AS and is released with a Creative Commons Attribution-NonCommercial 3.0 License. This license defines the software as free for personal or non-commercial use, without the author's permission (Creative Commons 2007). Users can use the interactive visualization offline since Highcharts is created in JavaScript and does not require a server to run (Highcharts 2013). There are many different types of visualizations, line, pie and column are examples of types that can be a representation for the data. The visualization is interactive in the way that users can click on any of the headers to remove the representation from the visualization. The same action would redisplay the representation. The next step after the user has chosen the interesting headers is to download the visualization as a picture or send directly to the printer. The cursor can be placed over a representation to display the values in numbers and the name of the header. A visualization can be created by using an example from the software website as a ground and exchange few parameters to make the visualization software use the correct x-axis title, x-axis value, y-axis title, y-axis value, overview title, subtitle and headers to be represented in the visualization (Highcharts 2013).

**Open Aid**

The Open Aid project is a distribution of Drupal 7, an open source content management system (CMS) (Reynen 2012). Many useful modules are included as default with the Open Aid project. This provides initiatives with a quick start, when the design they need for their information pages are ready to be filled with information about the initiative. Open Aid has already about page, blog, resources for documents and data sets, partners page and contact page, ready after installation and quick configuration (Reynen 2012).

**Comprehensive Knowledge Archive Network (CKAN)**

CKAN is an open source software created as a tool for handling and tracking data sets (CKAN 2013). The software is developed by OKFN and released under licensed terms of Affero GNU GPL v3.0. The software features are to upload, share, track and analyze data sets. The analysis are made with the included data explorer Recline, a JavaScript library used to display data sets in tables, create visualization in interactive graphs and maps based on uploaded data sets, directly in the web browser (Wainwright 2012). A remote API can be set up in CKAN to make it possible for other systems to fetch data from data sets available in the project.

### 3.1.4 Type of Visualization

There are many ways to visualize in order to make sense of data. The use of data is the reader's interpretation of the visualization. The representation of data can be useful in different purposes depending on the content in the data set. This section describes different types of visualization and when they are useful.
A text is used to describe the data with words of publisher's interpretations to help the reader to make sense of data. Also, data can be visualized in text as unstructured data, but that would be very hard to read and analyze. The data in the first row of the data set forest cover from data.ug (Table 3) has been described by the author in an informative text. To describe the whole data set would lead to analyzed data with completed interpretations but make it complicated to compare the values.

Broad leaved land cover had 18.682 hectares in 1990 and only 14.841 hectares in 2005. The change between 1990 and 2005 is -3.841, -21 percentage. The percentage change for each year is -1.37.

A table is data structured in columns that is useful when there is a small amount of data that can be compared in an overview and analyzed directly when the table is interpret by the reader. To compare and analyze data in columns that has a structure is a lot easier than comparing and analyzing data in an unstructured text (Table 3).

**Table 3: Forest Cover 1990 and 2005 data from data.ug.**

<table>
<thead>
<tr>
<th>Land cover</th>
<th>2005 (hectares)</th>
<th>1990 (hectares)</th>
<th>Change (1990-2005)</th>
<th>Change in Area over 15 years</th>
<th>Annual percentage change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broad leaved</td>
<td>14.841</td>
<td>18.682</td>
<td>-3.841</td>
<td>-21</td>
<td>-1.37</td>
</tr>
<tr>
<td>Conifer</td>
<td>18.767</td>
<td>16.384</td>
<td>2.383</td>
<td>15</td>
<td>0.97</td>
</tr>
<tr>
<td>THF well</td>
<td>542.787</td>
<td>651.11</td>
<td>-108.323</td>
<td>-17</td>
<td>-1.11</td>
</tr>
<tr>
<td>THF low</td>
<td>201.644</td>
<td>273.062</td>
<td>-71.417</td>
<td>-26</td>
<td>-1.74</td>
</tr>
</tbody>
</table>

A graph is useful to visualize data when there are large amounts of ambiguous data that needs to be compared. Example of basic types of graph are line and column. Line graph is useful when there is multiple columns to be compared against one column, such as a period of time. Column graph is useful when few columns needs to be compared against each other without the aspect of another value (Figure 1).

![Forest Cover](image)

*Figure 1: A column graph created with data set forest cover from data.ug.*

17
A chart is useful to compare data with only one point of time involved. Pie chart is example of a common way to visualize this kind of data and useful when few columns needs to be compared. The pie consists of a round circle divided in colored “cake” pieces to display the value in a visible form (Figure 2).

![Forest Cover 2005](image)

**Figure 2**: A pie chart created with data set forest cover from data.ug.

A map is a good way to visualize because people get more connected to data that can been seen around their own community (HacksHackers Kampala 2013). To visualize on maps requires data collected with coordinates of longitude and latitude or name of the area. The coordinates provides display of the exact location for each record. Circles are often used when data is connected to an exact location to show the spreading pattern on a map. The data can instead of coordinates contain the location by name of the area, to visualize data connected to the area. Colored areas connected to a number interval or graph can be used to display data on a named area, as seen in the project Uganda Watch (2011) (Figure 3).

![Map over number of reported issues per category for Uganda Watch 2011. Source: http://www.ugandawatch.org](image)

**Figure 3**: A map over number of reported issues per category for Uganda Watch 2011. Source: [http://www.ugandawatch.org](http://www.ugandawatch.org) [2013-05-27]

### 3.2 Open Data

The basics for open data are that data should be free, reusable and easy to download from a web page.
or web service (Dietrich et al. 2012:6; Uganda Opendev 2012b:5). Open data is often referred to open government data, that data collected by public funded organizations should be open and accessible by anyone. The type of data that is published regarding the government is in open data context often divided into the categories of agriculture, education, health, finance, energy and justice. The open data section describes what open data is and what it means to a society.

3.2.1 Open Definition

Open Definition (2006) is an initiative by the Open Knowledge Foundation (OKFN). Their definition is used and mentioned at many other initiatives websites. The definition from Open Definition described in one sentence “A piece of data or content is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike”. The Open Definition (2006) has a full text and the most important parts of open data are listed:

- Can be freely accessed, used, reused and re-distributed by anyone.
- Is non-personal data that should not contain any personal information.
- Does not compromise national security restrictions.
- Should allow combination of data sets from different sources and systems.

3.2.2 Open Concepts

This section covers description of different concepts used in the report that might need to be explained in their context with open data. The concepts explained are the key words when working with open data and are often used when talking about the beneficial results and why open data is needed.

Open development

Open development holds the meaning of grouping open data, open source, open knowledge, openness, transparency and accountability together in the same context. Open development is about providing and sharing information using information technologies among information sharing channels (CIPESA 2012). A government should aim for an open government with open development since all resources spent in open development is an investment (Uganda Opendev 2012a).

Transparency and openness

Transparency and openness means that a government should publish their decisions and present how money is divided in different projects. It should be possible to follow up and track fundings and find out how they are used (CIPESA 2012). There should be no secrets in a government about political decisions or data that can be used as a ground for other decisions. Organizations that base their decisions on public data should not have any problems to retrieve the data they require to make their decisions. If data is easy to retrieve then decisions can be made faster which results in saved time and money in development (APC & CIPESA 2012:9; DRT 2012:2).

Accountability

Accountability refer to that people should care about what they do in their work and feel that their actions matters and are noticed (DRT 2012:3). Example of low accountability is politicians who have a lot of power in the societies and use their influence to benefit for themselves and do not care about that
their actions can have a bad effect on the rest of the society. Another example is people that are working
with collecting data should feel responsibility and motivation why they should collect data. They might
only collect data because their closest boss tells them it is important. If they know how data are being
used, then they might start to care and do a better job (APC & CIPESA 2012:35).

Data Source and Data Sets

The data source is the source of data collected and published as data sets (Dietrich et al. 2012:6). The
sources can be organizations, governments or other parties who released their data in any format. Data
set is a group of data for a specific purpose, that has been collected and stored into a container. Values in
a data set can for example be grouped by a header and each row represents a new record. Example of
data set is the collection of data about all the schools in a country, with data about name, city, type of
school, number of students, number of toilets and students' results. The data should be raw data that has
not been structured in any other way than table or comma separated. Data sets can be stored in many
different formats, for example PDF, XLS, CSV, XML and JSON. Any other data sources' data sets can
be included to complement if a data set is insufficient when the data is being used and visualized (APC
& CIPESA 2012:10).

3.2.3 Open Data Benefits

To let data free and let people be informed will help to improve the transparency and accountability of
the society. A lot of decisions are based on data and if data is easy accessible then decisions can be made
faster, money and time in development could be saved as a result (APC & CIPESA 2012:9; DRT
2012:2). OKFestival (2012) states "Open knowledge – from open data, to open educational resources, to
open software – can play a powerful role in supporting sustainable global development".

An organization, government or other stakeholder may not afford to build an application for their data,
but stakeholders should at least build a way for the public to access their data (Dietrich et al. 2012:5). A
developer who is learning something new and at the same time wants to do something useful, could
practice by creating an application based on real data. The developer learn valuable experience and the
stakeholder obtains a new application.

OGP (2012b) has created a video to inspire what a government can accomplish by open data. In Brazil,
an open budget system was launched where officials posted expenses, leading to the result that
corruption fell. In the United Kingdom, heart surgery success rates were published, patients started to
compare the results and hospitals started to compete. The result was that survival rates were improved
by 50%. Tanzania open development, created a project where the public can report broken water
supplies and have the government respond to take action (OGP 2012b).

3.2.4 Open Data Initiatives

There are many organizations working with promoting open data, as a mean of raising the quality and
knowledge of open data. The initiatives seen most in open data context are described in this section.

Open Government Partnership (OGP)

OGP was founded in 2011 and their goal is to make government initiatives committed to open data,
share knowledge among the governments and inspire to open data (APC & CIPESA 2012:9). In order to
join OGP there are four requirements that must be met with at least 75% of maximum points in the
measurements (OGP 2012a). The first step is to achieve fiscal transparency with open budget that is measured by using the 2010 Open Budget Index. The second step is to have an access to information law, measured by survey from Right 2 Info. The third step is to have transparency for income and assets for elected and senior public officials, measured by World Bank survey on disclosure, titled "Disclosure by Politicians". The forth step is openness to citizen participation, measured by using the 2010 Economist Intelligence Unit (EIU) Democracy Index’s Civil Liberties sub-indicator.

Open Knowledge Foundation (OKFN)

OKFN builds communities around open knowledge and has been spreading the knowledge of open data since 2004 (Molloy 2011). OKFN makes it easy for organizations to get started with open data since many projects have been started by this initiative. One project is built as open source software, called CKAN, that is a tool for handling data sets with features to share and analyze the data sets. Another well-known project is Open Definition that sets the definition for what open means and when data is open. The foundation has created two licenses in the project Open Data Commons to make data open (Open Data Commons 2009a). Public Domain Dedication and License is one of the licenses which states that the material is free to use by anyone without restriction. The other license is Open Database License that is a share-alike plus attribution license (Open Data Commons 2009b).

Open Data Days

On the 23rd February 2013 there was an international event, Open Data Days (HacksHackers Kampala 2013). The event brought developers and journalists closer together because the journalists understand what kind of data are needed and demanded from the society. The developers understands the means to make data usage possible.

3.2.5 Type of Open Data

The Open Definition (2006) states that data should be free to use, redesign and redistribute, except when it is data of personal integrity or a threat to the national security. Dietrich et al. (2012:6) states “The key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals”. Many countries have laws that allow citizens to request information from public organizations, that they have to share. Exceptions are for personal data or when spreading data is a threat to the national security (Uganda Government 2005:6; Open Definition 2006). With data accessible through web services or when meaningful data sets are published, then the data requested can be extracted by the requester and there is no need to disrupt the stakeholder with the data request.

3.2.6 Open Formats vs. Proprietary Formats

The accessibility part in the definition of open data can be reached with open formats. APC & CIPESA (2012:28) define an open data format as a format that are machine readable and allowing data to be extracted. An open format is a format with public specification on how the format can be used for compatibility between different software, or if the format has transparency to be opened in plain text editors (Open Formats 2004; Dietrich et al. 2012:18). APC & CIPESA (2012:28) also states that Excel is an open format since it can be reused. Contradictorily, Open Formats (2004) recommends to avoid usage of the Excel format since it is a proprietary format.
The Word and Excel software formats are not open formats and that results in errors when other software try to open documents with these formats. For example, when opening a document in proprietary format with another editor results in the images and text often rearranges with disappearing fonts, which makes the document unreadable. The proprietary formats can only be fully compatible with one editor and this forces users to buy the software with the proprietary formats (Stallman 2007). Further more there is a problem between the DOC and DOCX format, Stallman (2007) describes both of these problems.

“Most computer users use Microsoft Word. That is unfortunate for them, since Word is proprietary software, denying its users the freedom to study, change, copy, and redistribute it. And because Microsoft changes the Word file format with each release, its users are locked into a system that compels them to buy each upgrade whether they want a change or not. They may even find, several years from now, that the Word documents they are writing this year can no longer be read with the version of Word they use then.”

The product that uses proprietary formats usually has problems opening the open formats. It is not recommended to open CSV files in Excel to edit because it can change the internal data structure and leading to a failing import. The open format editors can however save documents in the proprietary formats (WikiHow 2013b).

3.2.7 Open License

*The data on a website does not become open because it is free to download. For data to be counted as open it requires data to be reusable as well as reachable. An open, simple and short license should be published on the website, so visitors easily can read the license and use published data without hesitation (OKFN 2009).*

**Creative Commons (CC)**

CC is an initiative to help creating an open license to release associated with the material. The creator of the material can choose:

- if the material should be able to be used under the terms of a non-commercial license
- if other users should be able to modify the material before use
- if the jurisdiction should be international or national.

There is a permanent term of use for the material, which is to keep the attribute intact, meaning that the text chosen by the creator should remain next to the material without modification. In 2009 the amount of CC licenses were estimated to 350 millions. One organization that is using the license since 2009 as main license is Wikipedia, the free encyclopedia (Creative Commons 2007).

**Open Data Commons Attribution License (ODC-by)**

ODC-by is a share-alike plus attribution, which means that the user is free to use the material if they agree to mention the source, contribute the changes back to the owner and share material the same or similar way as the source (Open Data Commons 2009b). The difference in licenses with share-alike and these licenses without share-alike may be affecting the use of the material (Ito 2004). The share-alike part of the license put the restriction to not be able to mix the work with other licenses that is not similar, for example proprietary licenses (Ito 2004; Pollock 2010).
3.2.8 Problems with Open Data

There is a large step to commitment in contributing with open data and some individuals in organizations does not want to take that step. Bbosa 1 says that organizations do not want to share data because there is fear of data being released to the public. Increased knowledge would lead to people asking more questions about the decisions, economy and what is happening around current affairs in the society. The effect that is feared by some organizations are increased demands that more should be done with their budget. The organizations might hide facts that they do not want to be asked about and be revealed. Bbosa also says that the problem is that few people know what to ask for and therefore there is low demand on data. The government would have to take actions if people acquires knowledge that they can demand (APC & CIPESA 2012). Bbosa thinks that fear of new technology is a challenge because introducing new technology might lead to some jobs that will no longer be necessary and therefore people are against new technology.

Problems with open data after commitment to openness are that the ministries might only give a paper version published in a book or a PDF document (Uganda Opendev 2012b:9). Others can only access the information by retrieving a copy of the book. Data is not recommended to be shared with the PDF format since it is a static format (Open Formats 2004). Still, there are occasions when data are shared in the PDF format. Creating a system that automatically can read data from PDF documents are not an optimal solution and complicated to implement (Dietrich et al. 2012:9). The most simple way to extract data from PDF documents is to copy and paste the content into another format (Molloy 2011).

Data collection might result in errors and faulty data because of human errors. The collected data can also be made up and invented because the person that is supposed to retrieve the data might not be able or willing to collect the correct data and instead provide random numbers or statements (Molloy 2011). Another problem is the lack of knowledge on how to use data. Citizens will not have accountability without the knowledge to use the data (Roberts 2012). Few people are able to access the data in its raw form. These few will have the chance to spread the data to the rest of the society. A government cannot afford to create applications that cover all the society's requirements. Instead they should reach out to the few who want to contribute to accountability and transparency. An open government’s responsibility is to reach other contributors and enable the means for them to contribute (Dietrich et al. 2012:5).

3.2.9 Open Data in Uganda

This section is about the usage of open data in Uganda and how open data has helped to make a positive change by increasing the accountability, transparency and public contribution from the citizens of the society.

Access To Information Act (ATIA)

Many countries have laws that let the public access information that were collected by public fundings, which is a first step to open data. ATIA is the Ugandan act that allow access to government information for the society. The act was proposed in 1995 and became an act in 2005. The Access to Information Act (2005:6) Act No. 6. explains the rules for access to information.

"Every citizen has a right of access to information and records in the possession of the State or any public body, except where the release of the information is likely to prejudice the security or sovereignty of the State or

1 Francis Bbosa Statistician DRT, interviewed 18th February 2013.
Uganda Medical and Dentist Practitioners Council (UMDPC)

The UMDPC is in charge of registrations for medical and dentist workers. In 2011 they released registers containing data about the medical workers (UMDPC 2013). A project was launched based on the registers, where it is possible to check if a doctor has a license or not. The check is performed by sending an SMS with the doctors full name to a short-number. Doctor's license registrations raised after the registers were released, Dr. Ssentongo states “When I came to UMDPC 1.5 years ago, it was 25% of all doctors who had a license, now it is up to 59%, because the first thing I did was to release the registers”. Dr. Ssentongo says that releasing the registers had effect on the accountability since people are tending to avoid doctors who do not have a license. Improved reporting on misbehavior and absent license are leading to doctors without license fear to be discovered.

Dr. Ssentongo says it is free to use the registers in any way and applications are free to be build based on these registers. This means the registers can be counted as free data, but not open data since it does not have a license associated with the registers. Dr. Ssentongo says the registers were released because journalists wanted to know information about the doctors and frequently asked the council for information. Much of the requested information can be found in the registers and majority of journalist calls have decreased.

The problem except absent license for the registers to be counted as open data is that open data should not be connected to any individuals (Dietrich et al. 2012:6). In this case it is all about the individual doctor, when data with name, address, phone number, email and more are listed in the registers. No harm comes to the doctors who have a license, except for the contact information that marketing companies can use to reach the doctors, since the registers are a list of phone numbers, emails and addresses that are accessible.

Medical Tracking (mTrac)

The mTrac project is a Ministry of Health initiative with The United Nations Children's Fund (UNICEF) as partner. The project enables the citizens to report all kinds of problems in the health sector. Some examples of problems are absent doctors, empty drug stocks and misbehavior in health facilities (UNICEF 2013). The society can report health problems through free toll SMS service, resulting in becoming a registered user and receive further summarized reports containing number of reported diseases and updates on drug stocks.

This project does not contribute to open data but it is a project that could contribute. The data is collected with government funding and by the help of the societies contribution to be gathered and analyzed by the Ministry of Health (mTrac 2013).

Uganda Watch

The Uganda Watch 2011 is a project where citizens can report problems with the election process and receive information about the election. Example of categories for citizen to report are imbalanced

---

2 Dr. Katumba Ssentongo Council Registrar UMDPC, Interviewed 5th March 2013
journalism, questions and complains about the election process, related incidents, vote bribes, violence and violation of laws (Kalemera et al. 2012:26). The project can also contribute to positive feedback on the election. Uganda Watch (2011) describes the project as “Citizens can SMS 6090 to report election abuses, ask questions, or praise individuals or groups that are contributing to a good election”.

The reports from the citizens are published on the Uganda Watch website and mapped to the location of the problem. A graph visualization is displayed with number of reports for each category on the area from where the report was sent. The reports are investigated by the Democracy Monitoring Group (DEMGroup) field staff to get more information from the reporter and to verify the situation. After the story has been verified it is marked as such on the website (Uganda Watch 2011).

### 3.3 Uganda Open Development Partnership Platform

*This chapter describes the current situation of open data for the Uganda Open Development Partnership Platform and how it can be improved from the current state.*

#### 3.3.1 Open Data Case

The open development platform is a portal for collecting existing data sets and documents to share with the society. The partnership was formed after an open development workshop in 11-12\(^{th}\) September 2012 with a total of eleven partners (Uganda Opendev 2012a). The workshop had a hundred representatives from different stakeholders of government, civilians, media, organizations, development partners and universities. They gathered to discuss long term focus strategies and explore opportunities to promote open development in Uganda (Uganda Opendev 2012b:1).

#### 3.3.2 Partners

*This section describes partners collaborating with the platform and their role in the collaboration.*

**Uganda Bureau of Statistics (UBOS)**

UBOS is the principal of data collecting, processing, analyzing and disseminating in Uganda (Uganda Opendev 2012b:15). Uganda Opendev (2012b:17) describes this partners role and function as “At the technical level, UBOS could contribute toward and support the development process and coordination in the use of geospatial mapping on the open data platform”. All ministries should send their data to UBOS, who gather and summarize the data. UBOS as a partner could publish their data on the platform, but they are awaiting the maintenance and stability of the platform before further actions (Uganda Opendev 2012b:15).

**Collaboration on International ICT Policy in East and Southern Africa (CIPESA)**

CIPESA is a non profit organization who works with freedom of expression and accessing information and policies (Kalemera et al. 2012:28). CIPESA has written many reports related to open data that displays their knowledge in the subject, for example Uganda Open Government data readiness study (APC & CIPESA 2012) and How ICT Tools Are Promoting Citizen Participation in Uganda (Kalemera et al. 2012).

**Development Research and Training’s (DRT)**
DRT’s goal is to help the poor by analyzing the needs and providing them with training and development. This partner is the main contributor and pulls the strings in the development of the platform (DRT 2012:4). DRT did already have a plan of their ideas on the objectives for the platform in beginning of 2012, before the initial workshop took place (DRT 2012).

Fruits of Thought

Fruits of Thought introduces new software to the partnership, for example, mapping, in the initiative Mapping Day (Uganda Opended 2012a). The company behind the Fruits of Thought is Mountbatten Ltd. Reiner Battenberg, the founder of Mountbatten Ltd, attended the open development workshop to contribute with experience and knowledge. Uganda Opended (2012b:17) describes Reiner’s part in the platform as “Reiner committed to promoting awareness about the open data platform amongst the tech provider community and the student fraternity to promote apps development and innovation leveraging on the open data”.

3.3.3 Platform Solution

The platform consists of two websites to cover the needs of open development. These websites are set up by using open source projects Open Aid as information, data sharing portal for the opendev.ug website and CKAN for handling the data sets on the data.ug website.

The opendev.ug website has the following site map:

- About – containing the sub pages Problem Statement, Opportunities, Goals, Purpose & Objectives, Theory of Change and Key programme components.
- Blog – blog entries from the partnership, activist members and guest contributions.
- Resources – a list of data sets and documents published on opendev.ug.
- Datasets – a list of data sets published on data.ug.
- Partners – list of partners in the partnership.
- Contact – contact info, address, map and a contact form to submit messages.

The data.ug website has recently been set up and has not been fully configured, so it still has the default look of CKAN. Three groups of users have been created, “opendev”, “External Sources” and “UBOS”. A group is in control of their own material to freely add, update and delete the material associated to that group. Anyone can register, login and upload material that will be marked as an unknown source, without being controlled by an administrator.

The platform’s solution has been setup by a local supplier company who installed Open Aid and CKAN for the partnership. The supplier has developed many other successful sites contributing to openness.

3.3.4 Data Sets

Currently there are a total of 49 data sets available from different data sources (DataUg 2013). There are only a few different data formats available for download of the data sets. None of the data sets exists in multiple formats for the same data, except for the data set managed by UBOS. Throughout 49 data sets available 36 are in the format XLS, five are HTML, three are CSV and one each is XLSX, Shapefile (SHP) and PDF. Out of these formats are the XLS and CSV formats the only formats that the graph builder supports. The data sets available in these formats are not structured to be used in the platform’s graph builder tools in an appropriate way, since they need to be transposed before being properly used.
The license is defined on detail level for two data sets, uploaded by external sources.

### 3.3.5 Data Sources

The sources of the available data sets is mostly the partners themselves that currently provided 37 data sets under the group “opendev”. The source is not stated for which partner was involved with providing which data set or which license the data set has. UBOS is a partner of the platform and should be included in the group, instead they conflictingly place data sets in their separate group. Currently UBOS has one data set in the group “UBOS”. This group want to be in control of their own data sets because they are unsure about the maintenance and sustainability of the platform (Uganda OpenDev 2012b:15). Five data sets are provided by unknown sources. External sources provided six data sets in a group of references to sources outside of Uganda.
4 Empirical Results

This chapter describes interview respondents, analysis of data from interviews, observations and contributions from this study. The results are divided into three parts to display the contributions, proposals and current state of the platform. The first part are interviews described as qualitative focused interviews. The second part contains the analysis of improvements for suggested proposals. The third part describes the procedure for visualization of data.

4.1 Presentation of Respondents

Francis Bbosa is working as a statistician for the organization Development Research and Training (DRT), situated in Kampala, Uganda, and was interviewed the 18th of February 2013. Bbosa holds a bachelor's degree in Statistics at Makerere University, Kampala.

Andrew Kawooya is working as a senior programme officer of governance and transparency for the organization DRT, situated in Kampala, Uganda, and was interviewed the 18th of February 2013.

Lillian Nalwoga is working as a policy officer for the organization CIPESA, situated in Kampala, Uganda, and was interviewed at 19th of February 2013. Lillian holds a bachelor's degree in Development Studies at Makerere University, Kampala and a Postgraduate Diploma in Project Management.

Person 1 from the supplier company was interviewed the 25th of February 2013. The supplier is involved in many projects related to open data in Uganda, and now they are delivering opendev.ug and data.ug. The respondent is anonymous because it is unclear if the name can be used.

Table 4 is a summarized overview of the respondents.

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Organization</th>
<th>Role</th>
<th>Interviewed</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Francis Bbosa</td>
<td>DRT</td>
<td>Statistician</td>
<td>18th of February 2013</td>
<td>Kampala, Uganda</td>
</tr>
<tr>
<td>Andrew Kawooya</td>
<td>DRT</td>
<td>Programme officer</td>
<td>18th of February 2013</td>
<td>Kampala, Uganda</td>
</tr>
<tr>
<td>Lillian Nalwoga</td>
<td>CIPESA</td>
<td>Policy officer</td>
<td>19th of February 2013</td>
<td>Kampala, Uganda</td>
</tr>
<tr>
<td>Person 1</td>
<td>Supplier</td>
<td>Supplier</td>
<td>25th of February 2013</td>
<td>Kampala, Uganda</td>
</tr>
</tbody>
</table>

4.2 General Issues Around Open Data

This section describes the analysis of the interviews performed for this report in purpose of describing the current situation of the open development platform and possible improvements. The interviews are described as qualitative focused interviews where the interpretation is first presented and then illustrated with citation from the respondent.

The organization DRT is working with analyzing the needs of data to demand, and data that is demanded from the society. There are multiple sources to analyze when choosing which data to gather. Kawooya states “We try to listen to what people want, through radio, television and in the fields when meeting people”. The sources Kawooya mentioned are either one-way mass communication or person to person communication. The next step would be expanding the source of assays to online social networking
services like Twitter and Facebook to hear the voices of people and simultaneously provide them more information about the platform and open data. This expansion to mass-to-mass communications would lead to more of the public participation in contributing with steering further focus for data collection.

The goal for the platform is not to collect data, instead the platform focuses on gathering data from multiple sources. Kawooya states “UBOS has much data, but we also gather data from different ministries, like the ministry of agriculture, and of course the World Bank”.

The interesting data might not always be easy accessible, since the data are often placed in tables shared in the PDF format along with text and images. The partnership must extract the interesting data and place it in logical data sets. Nalwoga explains their problem with data as “Data are not accessible. A lot of data are sent with PDF, but only a small amount of the data in the PDF are interesting and there is no good way to retrieve the relevant data from those documents”.

The focus for analysis of data sets to be published are steered by two factors, DRT statisticians and the steering committee. Kawooya states “We have statisticians working with analysis and selecting data from raw data sets, but also data that has already been semi-analyzed. There is also a steering committee that discusses where our focus should be”. The semi-analyzed data that Kawooya mentions refers to data stored in PDF documents that have to be analyzed before extracting the data. Kawooya says that the steering committee consist of members from the partnership.

It is noticed that programming knowledge are missing in DRT as employees have searched for an open source project for visualization, without finding any suitable software. This is probably because it is hard to find suitable software without the right knowledge, as Bbosa describes the situation.

“We have tried looking for a suitable open source project, but there is nothing. We have checked a few projects, but we do not have the required programming skills needed to get started with an open source projects that require some starting phase.”

Kawooya explains the missing knowledge as he states “We need a more interactive website and help with visualizing data, and also analyze the needs of which data to collect and publish”.

The supplier and the partnership are not always coming along. This is clear when Person 1 describes their independent actions that affects the partnership.

“We installed CKAN because of curiosity and did not discuss with anyone from the partnership, so now they are a little bit confused. The partnership wants one website to maintain, but CKAN has so much that is already developed.”

One benefit with the CKAN project is that the graph builder tool Recline is included in the project, so graphs can be built directly in the web browser based on uploaded data sets. When the partnership explore CKAN and notice the benefits of having a powerful open source project as a ground for the platform, it will be approved. This is noticed when the author introduces the graph builder tool to DRT as Bbosa states “The graph builder is fantastic”.

Person 1 explains that the goal for the supplier is to get the platform started to provide the means for others to work with open data. Person 1 states “We just want to get the data out and get things working. We want to enable the means so others can publish their data”.

29
4.3 Uganda Open Development Partnership Platform

A few guides were created based on the improvement proposals and given to the partnership to see if they would be published on opendev.ug as blog contributions by the author. This chapter presents why the focus was on the chosen improvement proposals.

4.3.1 Visualization and Data Sets

DRT has a way to visualize their data by using Excel. Bbosa says “Excel can visualize data in a nice and interactive way”. This way to visualize only require a data set in Excel without being raw data since the user can select the rows and columns to be used for the visualization. The possibility to select data from an interface results in users not understanding why other formats or raw data are required when the data sets are used in automatic processes. Bbosa says that data sets are placed in Excel and from there the data can be converted into other formats.

“We place our data in Excel because everyone has Excel today and everyone can open that document. From Excel it is possible to convert the data to the formats that you want to have, for example CSV, if someone would have use of it.”

Using Excel for creating interactive visualizations are a manual process since the creator has to export the visualization and then publish online. An interactive way to visualize is missing on the Uganda Open Development Partnership Platform. One important feature is for users to choose, for example, the headers to display from a data set. Bbosa states “We would like to have something similar on our new website that is under development”, meaning similar to Excels interactive visualization.

The author's observations on the Open Data Days created a better understanding of how data are used by the journalists and how data can be visualized. There seems to be a wide use of existing tools among the participants to visualize the data in static graphs. Interactive visualizations were not seen at all. The author introduced Google's tool for visualization and displayed the upload of data sets and usage of selected columns in interactive graphs.

A guide (Appendix 3) was created by this report's author to spread the knowledge about transposing data, so data sets can be used when building graphs with the graph builder included in data.ug. The section 4.4 Procedure of Visualization, will display why it is important to transpose data sets correctly.

4.3.2 Multiple Formats

Person 1 delivers the platform to the partnership and understands the needs of an open development platform. The person informs that the recently installed data set handler, CKAN, supports multiple formats of data sets to be uploaded and shared.

The author's improvement proposals for multiple formats have been sent to the partnership for publishing on the opendev.ug blog (Appendix 4). Bbosa agrees with the text as he states “Very interesting and informative write up especially for the non-technical personnel”. Hopefully the blog entry will change the fact that Bbosa thinks that everyone has Excel and believes that all users can open data sets in the Excel formats.

4.3.3 Open License

The data sets on the opendev.ug website are not connected to any license and it is unclear how the material can be used. However, the data are free to use as Bbosa states “No, our data does not have a
license, but we are aware of Creative Commons. The data sets published on our website are free to use, but we have no information about license on the website”. That DRT knows about CC means that they are familiar with licenses, but clearly not familiar enough to publish any information about how their data can be used. Bbosa see no big problems with the absent license, since Access to Information Act (ATIA) says that data must be free in organizations. Bbosa says the only problem is that ATIA is not always followed by all organizations.

The author's improvement proposals for open licenses have been sent to the partnership for publishing on the opendev.ug blog (Appendix 5). The proposal should provide an insight in why it is important to have an open license associated with the material that should be considered open.

### 4.4 Procedure of Visualization

This section presents an evaluation in two parts, the exploration of the platform's existing tool to create a graph and the procedure of using the platform's data sets for visualization with external software.

#### 4.4.1 Analysis of Data Sets

The partnership needs help to identify how the data sets can be visualized. The author's proposal of visualization for few data sets have been mapped and structured in Table 5. There are four columns structuring the mapping, the first column holds the name of the identified data set, followed by the recommended visualization in graph, table, map or text. The type of representation depends on the visualization type and can be line, column, pie, area or spot. The recommendation of visualization is motivated in the forth column.

<table>
<thead>
<tr>
<th>Name</th>
<th>Recommended Visualization</th>
<th>Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agricultural exports 2006-2010 with graph</td>
<td>Graph</td>
<td>Line</td>
<td>Compare multiple column values for each year.</td>
</tr>
<tr>
<td>Fish Production in Columns</td>
<td>Graph</td>
<td>Column, Line</td>
<td>Compare multiple column values for each year. Column type is possible when only a few years are available.</td>
</tr>
<tr>
<td>Forest cover in Uganda 1990-2005</td>
<td>Graph, Chart</td>
<td>Column, Pie</td>
<td>Multiple values compared to each other. Two graphs can be created because mix of values and percentage.</td>
</tr>
<tr>
<td>UCE in 2012, results</td>
<td>Graph</td>
<td>Column</td>
<td>Multiple values compared to each other.</td>
</tr>
<tr>
<td>Primary education accessibility indicators by district, (2008 – 2010)</td>
<td>Table</td>
<td>Column</td>
<td>Different column values compared with many schools. Probably all schools are interesting to compare, too many for a graph.</td>
</tr>
<tr>
<td>Latrine coverage by district</td>
<td>Graph</td>
<td>Column, Line</td>
<td>Compare multiple column values for each year. Column type is possible when only a few years are available.</td>
</tr>
</tbody>
</table>
4.4.2 Existing Visualization Tool

Table 6 is part of a polished data set released by the partnership (Uganda Opendev 2012a) and can only be used as ground for a graph that is hard to analyze. Since the x-axis can in most graph builders only exists of one column, leaves the “Commodity” column the only option as the x-axis. The values for that column will serve as labels on the x-axis that marks the column representations in the graph. The years of the data set must in this case work as headers that group the values displayed as columns.

Table 6: Part of Polished Data Set of Agriculture Exports 2006 to 2010 (Uganda Opendev 2012a)

<table>
<thead>
<tr>
<th>Commodity</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>19.7</td>
<td>19.9</td>
<td>23.4</td>
<td>17.9</td>
<td>17.5</td>
</tr>
<tr>
<td>Cotton</td>
<td>2.1</td>
<td>1.5</td>
<td>0.8</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Tea</td>
<td>5.3</td>
<td>3.6</td>
<td>2.7</td>
<td>3.8</td>
<td>4.2</td>
</tr>
<tr>
<td>Tobacco</td>
<td>2.8</td>
<td>5</td>
<td>3.9</td>
<td>3.6</td>
<td>4.2</td>
</tr>
<tr>
<td>Fish</td>
<td>15.2</td>
<td>9.3</td>
<td>7.2</td>
<td>6.6</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Example of a graph from the data set in Table 6 created with the platform's tool can be seen in Figure 4. The graph is displaying a limit of three different years, since more years would make it hard to compare the values for each year. Even with the limitation of three year is the red color of year 2010 not visible at all. Usually column graphs have one label displayed on the x-axis for each column. Commodities are used as labels for the x-axis and there are many more columns than labels. It is not possible to match a label to a column. Overall, it is impossible to analyze all of the presented data in this graph.

*Agriculture Exports 2006, 2008 and 2010*

![Graph in columns from data set of Agriculture Exports from 2006 to 2010.](http://data.ug/)

A transposed data set means that rows and columns have switched place and been transformed, as seen in Table 7. This makes it possible for the graph builder to use the year column as the x-axis, instead of the commodities and have the years displayed as labels for the representation. The benefit is that the graph can include all the years from 2006 to 2010 and showing the commodity values for each year. The difference is to limit the amount of commodities displayed, not the years. Still, it is more interesting to
display the values for all the years available and select few interesting commodities to compare rather than the opposite.

**Table 7: Part of Transposed Data Set of Agriculture Exports 2006 to 2010**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coffee</th>
<th>Cotton</th>
<th>Tea</th>
<th>Tobacco</th>
<th>Fish</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>19.7</td>
<td>2.1</td>
<td>5.3</td>
<td>2.8</td>
<td>15.2</td>
</tr>
<tr>
<td>2007</td>
<td>19.9</td>
<td>1.5</td>
<td>3.6</td>
<td>5</td>
<td>9.3</td>
</tr>
<tr>
<td>2008</td>
<td>23.4</td>
<td>0.8</td>
<td>2.7</td>
<td>3.9</td>
<td>7.2</td>
</tr>
<tr>
<td>2009</td>
<td>17.9</td>
<td>1.5</td>
<td>3.8</td>
<td>3.6</td>
<td>6.6</td>
</tr>
<tr>
<td>2010</td>
<td>17.5</td>
<td>1.2</td>
<td>4.2</td>
<td>4.2</td>
<td>7.9</td>
</tr>
</tbody>
</table>

The goal is to compare the cost of different commodities during a period of time. The amount of data available makes it excellent to use a graph of the type line for visualization.

1. The graph is created by uploading a polished data set with column structure in the CSV or XLS format. It is required to login in order to use the graph builder. The login can be done either by creating an account or use an OpenID.

2. At the graph builder (Appendix 3), the “Year” column is selected to be represented as the x-axis, resulting in five labels displayed, one for each year available in the data set. Multiple series are added to the y-axis, as it is possible to select multiple columns to represent the y-axis. The columns for the y-axis chosen in this example are the most interesting commodities to be compared over a period of time.

3. The graph is by default set to a value of 15.000, but the biggest value in the data is of a higher value. Filter can be useful when there is a need to create a range for the values in the y-axis. The result of the default value of 15.000 is that whole lines are not visible because the biggest value for the data is higher than the default value. In this example the filter is set to the highest value for this data, so all the lines are visible.

4. The last step when satisfied with the graph is to use the embed button to get the code that can be used on another website to display the graph.

A graph created in the platform's builder can be seen in Figure 5, displaying a limit of six different products to make it possible to analyze and compare the values for each year. For example, Coffee is the most exported commodity with a increased export in 2007 and decreased export in 2008. The fish and gold exports are interesting to compare since they decreased while other commodities increased the same year.
The negative part in this graph builder is that titles can not be added to explain the x and y-axis labels. Another problem is the data for the years does not have a decimal, still a decimal is added when the values are displayed in the graph.

4.4.3 External Visualization Software

The limitations and negative parts of the platform's graph builder can be solved through using other software. The next example graph has been built by combining the data reader, the Miso Project Dataset and graph software, Highcharts. The graph was created based on one data set available on opendev.ug resource page.

The original data set has more information than necessary, mentioning the lakes as headers. The values are placed horizontal and the lakes name are placed inappropriate for machines to read (Table 8). A script must be created to fit the exact structure of this data set in order to read data and that is not sustainable to do for a single data set.

Table 8: Part of original data set Quantity of fish produced in Ugandan lakes 1961-2010 (Uganda Opendedev 2012a)

<table>
<thead>
<tr>
<th>Year</th>
<th>Lakes</th>
<th>1961</th>
<th>1962</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Victoria</td>
<td>2550</td>
<td>2340</td>
</tr>
<tr>
<td>1</td>
<td>Albert</td>
<td>1180</td>
<td>1220</td>
</tr>
<tr>
<td>2</td>
<td>Kyoga</td>
<td>680</td>
<td>1320</td>
</tr>
<tr>
<td>3</td>
<td>ED/GE</td>
<td>1250</td>
<td>1210</td>
</tr>
<tr>
<td>4</td>
<td>Wamala</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>
In order for the Miso Project to read the data it must be polished and transposed to have the structure with one horizontal row of column names followed by values placed vertically where each row representing a record of that year (Table 9).

### Table 9: Part of transposed data set

<table>
<thead>
<tr>
<th>Year</th>
<th>Victoria</th>
<th>Albert</th>
<th>Kyoga</th>
<th>ED/GE</th>
<th>Wamala</th>
<th>AL.NILE</th>
<th>OTHERS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1961</td>
<td>2550</td>
<td>1180</td>
<td>680</td>
<td>1250</td>
<td>100</td>
<td>..</td>
<td>180</td>
</tr>
<tr>
<td>1962</td>
<td>2340</td>
<td>1220</td>
<td>1320</td>
<td>1210</td>
<td>200</td>
<td>..</td>
<td>360</td>
</tr>
</tbody>
</table>

The data set has been stored in the CSV format to be compatible with the Miso Dataset software. After the Miso Dataset has read the data, it can be used with the graph software, Highcharts. The graph was created by using an example on the creators website as a ground. Only a few parameters were exchanged to make the graph software use the correct format for x-axis, y-axis values and headers to be displayed as lines in the graph. The rest of the parameters used in the example are x title, y-title, graph-title and subtitle.

Figure 6 is a line and spot graph and is interactive in the way that the user can click on any of the lakes to remove the representation line from the graph. The next step after the user has chosen the interesting headers, is to download the graph as a picture or make a print.

![Figure 6: Chart made based on data set from open development platform.](chart.png)

The following code will include the Miso Dataset in a web document.

```html
<script src='http://misoproject.com/js/miso.ds.deps.0.4.0.js'></script>
```

The Miso Dataset code required to read data is very simple, the constructor only require two parameters. The first parameter is the source of the data in form of CSV or JSON and the second is the delimiter of the data from the source.

```javascript
var ds = new Miso.Dataset({
    url: dataset.csv,
    delimiter: ",",
});
```
The following code will start to fetch the data. The code within the success function will be run if the fetch is successful. A column from the source can be fetched by calling the “column” function with parameter of the column name followed by the variable to fetch. In this case we fetch the column “Year” with related data.

```
    ds.fetch({
        success: function() {
            this.column("Year").data
        }
    });
```

The following code will include the Highcharts software in a web document.

```
    <script src="http://code.highcharts.com/stock/highstock.js"></script>
    <script src="http://code.highcharts.com/stock/modules/exporting.js"></script>
    <div id="container" style="height: 500px"></div>
```

The Highchart's constructor has many optional parameters and a few required. One required parameters is “chart”, to set chart type and define where the chart should be placed within the HTML document. The second required parameter is “series”, that stores the data for each column that should be visualized. The optional parameters used in the example are “title”, “subtitle”, “xAxis” title, column format, “yAxis” title, “yAxis” color, “tooltip” for displayed text on mouse over above line, and legend to define the display of column labels.

```
    chart = new Highcharts.Chart({
        chart: {
            renderTo: 'container',
            type: 'line',
            marginRight: 130,
            marginBottom: 25
        },
        series: [{
            name: 'Victoria',
            data: this.column('Victoria').data
        }]
    });
```

Each of the columns to display needs to be defined in the series parameter. The author uses the power of the Miso Dataset when using a loop to fetch all the columns dynamically and store them in a container for later use when using the series parameter. The code excludes the “Year” column since it cannot be compared with the other values in the data set.

```
    this.eachColumn(function(colName, colObject, index) {
        if(colName != "Year") {
            seriesData[seriesData.length] = {
                name: colName,
                data: this.column(colName).data
            }
        }
    });
```
5 Analysis

The analysis chapter connects the theory with the results together with the author's interpretations. The analysis model in Figure 7 is ground structure for the chapter.

5.1 Analysis Model

The analysis model should be a reflection of the reality in an descriptive, yet not to detailed, overview. The model should be a limitation on the gathered data and ground for the empiric procedure leading to analysis of the results. The image below display the focus of the analysis in open license, multiple formats, visualizations and data sets.

![Figure 7: Analysis model diagram.](image)

The condition for the research to proceed is to have knowledge of open data, including redistribute data with open license and data made accessible with multiple formats. The evaluation conditions required are material of data sets from the opendev.ug website as well as tools for visualization from existing graph builder on data.ug website and external software. The evaluation of existing graph building tool and external software is performed as architecture and static analysis by using available data sets. The effect of evaluation are findings of suitable, optimized tools and data sets for further means of visualization. The evaluation has lead to finished visualizations that can be published on the platform.

5.2 Multiple Formats

It is beneficial for open development to release documents and data sets in editable formats, so others can download, change, reuse and redistribute material on different places in other context. Multiple formats available for download opens up for the user to choose the option that suits them the most for their editor (Open Formats 2004). The Uganda Open Development Partnership Platform is supposed to promote openness, but uses the proprietary format XLS (Excel) for data sets and DOC (Word) for documents (Appendix 3).
The Uganda Open Development Partnership Platform rarely releases raw data and only non open formats exists for most of the available data sets. A document has been published in the DOC format and the available data sets are released in the XLS format. Both of these formats are proprietary formats (Open Formats 2004). Most of the documents at the platform's website are policies and guides released in the read-only PDF format (APC & CIPESA 2012:28). Text can be copied from PDF documents, but since headers, images and line breaks can not be copied makes the content non-reusable (Open Formats 2004). Discussions with DRT employees about the lack of multiple formats and data structure, gave the indication that they do not have the knowledge of using other formats or structure data for systems. The conclusion drawn is that they have no understanding why other formats or raw data are required. The reason that data sets are shared in a proprietary format is because DRT believe that data can be converted by anyone into other formats by using Excel. However, the reality is that everyone does not have the Excel software and may have problems with opening the data sets.

The Excel software are widely spread because many people have access to them on public computers. Many schools, communities and governments have joined the Microsoft Education Alliance Agreements, which means the software are installed at computers for the students and workers to learn and use (Microsoft 2013). This makes people familiar with the software and creates the illusion that the software exists everywhere for everyone. The alliance only includes schools, communities and government, the rest have to buy the software to read the documents in these formats (Stallman 2007). Some users choose not to have the software and for those there are free open options available. Example of open software with text editor and spreadsheet software are Open Office and Libre Office, which uses the Open Document Format (WikiHow 2013b).

The read-only format PDF is useful for those who do not have knowledge of or do not have the possibility to use any text editor that supports the editable formats. A read-only document reader does not rearrange text or photos and a document can be shared and viewed with original structure since the document will be displayed the same for all viewers (WikiHow 2013a).

The partnership has previously been informed about openness in an open discussion at the open development workshop. Uganda Opendev (2012b:5) summarizes the discussion as “Sharing information is not enough; data should be accessible and reusable. Datasets are as important as the reports produced. Openness is about allowing people to use and re-use data”. Still, the discussion seem to have left out the part on how to practically reach accessibility and reusability (Open Formats 2004). Uganda Opendev (2012b:9) describes that one challenge with data accessibility and reusability is that no policy for standard formats have been agreed and the government's guideline can not fully be used for material sharing purposes. This statement display that standard formats were discussed but the workshop could not conclude to use any standard formats.

APC & CIPESA (2012:28) are aware of open formats as they write about it in their report. This does not mean that the partnership holds any practical knowledge.

“Most of the websites have some data that can be accessed although majority of it is not in a re-usable state, with exception of the Ministry of Finance Planning and Economic Development (www.finance.go.ug) that has data in open format – although a great deal of its data is also not in open format.”

Their definition of open format includes Excel. They start by using the term “open data formats” and later only the term “open format”. Open data formats can include Excel since it can be reused, but the term “open format” should not be mixed with a proprietary format such as Excel (Open Formats 2004;
Dietrich et al. 2012:18). To call Excel an open data format is on the edge, since data might not be available in a few years, it might not be editable at all due to changes in software, after all, it is a proprietary format with only one product owner.

“Open Data formats are formats that are machine readable and allow computer applications to extract data from them. Excel format is an example of an open format because one can manipulate it.”

The supplier who delivers the platform displays their knowledge when they proposed to include data.ug to the platform solution. Their motivation is that the open source project, CKAN has a lot to offer that will provide the means to evolve the open development (CKAN 2013; Wainwright, M. 2012).

5.3 Open License

The material does not become open and free to use without restrictions because it is possible for anyone to download, copy and use (Dietrich et al. 2012:6; Uganda Opendev 2012b:5). For material to be counted as open requires it to be reusable as well as reachable (Open Definition 2006). An open, simple and short license should be published on the website to define the border of use for the material, so visitors easily can read the license and use published data without hesitation (OKFN 2009). OKFN (2009) describes the problem with an absent license.

“Licensing is important because it reduces uncertainty. Without a license you don’t know where you, as a user, stand: when are you allowed to use this data? Are you allowed to give to others? To distribute your own changes, etc?”

The data handler site, data.ug, uses the Open Database License, since it is placed in their footer (DataUg 2013). By having this license means the material is free to use for non-commercial use with the restriction that the attribute is intact (Open Data Commons 2009b). This license is placed by default in the footer when installing the open source system, CKAN, so it is unclear if the license was an active choice. However, the opendev.ug website is missing a license associated to their published data sets and it is unclear how the material can be used. Two of the data sets uploaded by external sources have a license connected on a detail level (DataUg 2013). The direct connection between opendev.ug and data.ug that has been established to display data sets directly on opendev.ug does not display the license chosen on detail level for these data sets.

The report's empirical chapter presents that even if the platform's material does not have a license, the material are still free to use. There are knowledge about Creative Commons (CC) and familiarities with licenses, but the partnership is not familiar enough to publish information about how their data can be used. Bbosa see no big problems with an absent license, since the Access to Information Act (ATIA) says that data must be free in organizations and government.

The partnership has previously been introduced to open licenses. Uganda Opendev (2012b:7) describes that the experience from Kenya Open Data initiative lead to use an open license that is not share-alike, so material can be used commercially. This proves an active choice of license from the Kenya initiative.

“The license under which data is released allows anybody to use some or all or new combinations of data as they like (e.g. perform analyses or build applications) and then distribute these new works either commercially or for free.”

There is no indications that the Kenya experience has been analyzed by the partnership. More focus
should have been placed on the use of a license and whether this information was properly received and analyzed.

Figure 8 is the license that Open Data Kenya has placed in their footer together with the text “Creative Commons Attribution 3.0 Unported License”. The figure 8 is a link to the CC website which describes the license in detail. The part of the figure with a symbol of a human and text “BY”, means by the following conditions, that the use of the material must have an attribution naming the source of the material (Creative Commons 2007; Uganda Opendev 2012b:7).

![Figure 8: Creative Commons Attribution 3.0 Unported License.

The Open Development Cambodia website also has the license in the footer. The CC license in Figure 9 has the text “Creative Commons Attribution-ShareAlike 3.0 Unported License”. The figure 9 is a link to the CC website that explains the content of this license in more details. The license differ from Figure 8 with a repeat symbol and the text “SA”, standing for share-alike license. This means that the material should have the same or similar license when reusing and redistributing (CC 2007).

![Figure 9: Creative Commons Attribution-ShareAlike 3.0 Unported License

The difference in these licenses may be affecting the use of the material (Ito 2004). For example, someone who creates a product to earn money may not want to use the material if the share-alike condition is included. Having the “SA” part of the license means competitors can use the material to get started quickly with their business. It would be enough for the user to mention the source, so the competitors can receive the base material that has been gathered by public funding and not make the commercial companies to share their work in the same way. The “SA” part of the license put the restriction to not mix the work with other licenses that is not similar, for example, proprietary licenses (Ito 2004; Pollock 2010).

5.4 Data Sets

Releasing the data sets in multiple formats as raw data makes it usable in more places (Open Formats 2004) and data can without difficulty be visualized in a graph. To make this possible, many of the current data sets on opendev.ug must be transposed and adapted to the column structure with values placed vertically. The goal with structuring data sets is for the data to be compatible with other systems (Dietrich et al. 2012).

The original data set used for the existing graph builder tools is a polished data set released by the partnership (Uganda Opendev 2012a). This data set can only be used as a ground for graphs that are hard to analyze. Since the x-axis can in most graph builders only exist of one column (Wainwright 2012), leaves a column with long vocable the only option for the x-axis. The values for that column will serve as labels on the x-axis that marks the representations in the graph. The problem is that the x-axis is
limited to size and can not display all interesting values to be compared. The data set has collected years that would serve as a suitable x-axis because it always have the same length for values. In this case the years are instead used as headers, structured horizontally that group the values.

A transposed data set means that rows and columns have switched place and been transformed. This makes it possible for graph builders to use the year column as the x-axis representation, instead of the column with long vocable. Since it is more interesting to display the values for all the years available and select a few interesting vocable to compare over a period of time, rather than the opposite. The website data.ug that includes the graph builder Recline can use the transposed data sets in a graph, directly in the web browser (Wainwright 2012).

The original data set used for external graph builder software has more information than necessary. The values are placed horizontal and the headers are placed inappropriate for systems to read. A script must be created to fit the exact structure of this data set in order to read data and that is not sustainable to do for a single data set (Dietrich et al. 2012).

In order for the Miso Dataset to be compatible with the data set requires that the data are polished and transposed. The target structure is one horizontal row of column names followed by values placed vertically, where each row representing a record of that year (Kavi 2013). The data is then exported in CSV format to be used in Miso Dataset.

### 5.5 Visualizations

This section presents the evaluation in two parts, the architectural analysis and exploration of the platform's existing tool with a static analysis to create a graph. The second part is the procedure of using the platform's data sets for visualization with external software in an architectural analysis.

One organization of the partnership mostly visualize their data by using Excel, because it is interactive, fast and easy to use. This way to visualize only require a data set in Excel, without being raw data, since the user can from the interface, select the rows and columns to be used for the visualization. This results in users do not understanding why other formats or raw data are required when using data sets in automatic visualization tools. However, a blog entry from Fruits of Thought has been published on the opendev.ug website about polishing data to raw data. Polishing is made to achieve data that are easier to analyze and visualize.

A similar, interactive way to visualize like Excel is missing on the platform. Using Excel for creating visualizations are a manual process since the creator has to export the visualization and then publish online. There should be a more effective procedure that can be performed online. Users should be able to choose, for example, the headers to display from a data set.

#### 5.5.1 Existing Tool

The opendev.ug website is connected to site data.ug, which is based on the CKAN project. The included tool for graph building, Recline, enables the transposed data sets to be put in a graph, directly in the web browser (Wainwright 2012).

The data often requires explanation of what the data represents. Labels are listed with a color that is matching a color in the representation of the visualization. The axes usually have labels to explain their values. The limitations in the Recline graph builder is that labels can not be added to explain the x and
y-axis values.
Another problem is that the default range of the y-axis is not set to the maximum value of the used data, leading to the representations with higher values going out of range of the graph. This problem can be solved by changing the y-axis range manually. The following problem after changing the range is that data without a decimal still are displayed with an added decimal. This is very illogical when using numbers that are obviously not containing a decimal, for example years. The data set can consist of many years that all needs to be included in the graph to make sense of the data. In the case when all x-axis values can not be placed as the x-axis in their original form leads to limitation of the used data and with a decimal added creates even more limitations.

5.5.2 External Software
The limitations of the platform's graph builder can be overcome through using other software to get more flexibility. A suitable combination is the data reader, the Miso Project Dataset and the graph software, Highcharts. The evaluation was performed through architectural analysis to find suitable software and static analysis was based on one of data sets available on opendev.ug resource page.
The Miso Project Dataset is an open source data handler with the goal to make it easier to import data from different sources. The software is compatible with data from CSV, JSON, remote API and Google Spreadsheet, to combine into a single object that can be used with, for example, Highcharts (Graul 2012).
Highcharts is a framework to create interactive visualizations. The interaction is perceived through that users clicking on any header in order to remove the representation from the visualization, the same action would redisplay the representation. This is one feature that was requested from the partnership. The user chooses the interesting headers and then download the visualization as a picture or send it directly to the printer. The cursor can be put over a part of the visualization's representations to display the values in numbers and the name of the header. There are many different types of visualizations, charts, line, pie and column, are only a few of all types that can be a representation for the data (Highcharts 2013).
Novice development skills are required to create own visualizations. The examples from the Miso Project's website makes it very easy to create same visualization with another data set (Miso Project 2012). The result displays a combination of these projects, based on an example from Miso Datasets website. The example was rewritten to dynamically use the column names as x-axis. This dynamical example can easily be switched to use data from another data set.

5.6 Effects
This chapter is presenting the effects of the research that have had influence on the target audience, and the contribution from the author. Interpretations of the material and results from evaluation of tools and software have lead to contribution of visualizations and optimization.

5.6.1 Optimization
The results of evaluation are findings of suitable, optimized tools and data sets for further means of visualizations.

• The optimization of data sets are how to transpose and structure the data. This was demonstrated when displaying the example procedure of visualization in the empirical results.
• The procedure also displayed how to use existing tool to display a higher number than the default number in the Recline graph builder, so the whole graph is displayed. This also display the limits of the existing tool as it is not possible to display a label for any of the axes or display values without adding comma.

• Show how external tools can be used to extend the visualization to be even more optimal for a specific data set.

• The dynamical example can easily switch to use data from another data set and be connected to a interactive interface. It is one of the partnership's goals to implement an interactive graph on the platform.

5.6.2 Visualizations

The visualization created through external software has been done with data from fish production and been mentioned on the opendev.ug website as a blog entry (Appendix 1). The existing tool resulted in finished visualizations from data about fish production, agriculture products and forest coverage. These visualizations have been shared with the partnership.

• The fish production and agriculture products have large amounts of ambiguous data that needs to be compared. Line graph has been used because line is useful when there is multiple columns to be compared against one parameter, as a period of time.

• The forest coverage have been visualized with a graph in column, since it contains few columns that needs to be compared against each other. A graph is used for this data set to give a stronger vision, even if there is not too much data to compare in a table.
6 Conclusion

This chapter starts with answering the purpose of the report with the conclusion of the described results. This is followed by the recommendation in first hand to the Uganda Open Development Partnership Platform but also to other open development initiatives that is in the beginning of their development to promote openness in the society. The recommendations and conclusions can be used for initiatives that want to improve their platform by comparing with the report's recommendations. A discussion is presented that describes the circumstances which could have affected the outcome and result of this report. Next section. Further studies in the area open data are presented in the end of the chapter.

6.1 Closure

To use the open source projects Open Aid and CKAN were a good combination to get a quick start with the platform development. The use of Open Aid project gave the platform the modules it needs for the data sharing part of the solution, as it already had the pages about, blog, resources, partners and contact ready from the initial phase. The CKAN project was a natural step to include in the platform to maintain the data sets and create interactive visualization. The project can upload new data sets, keep track of data sets and create graphs based on the uploaded data sets.

After a visit at DRT office, it has been discovered that programming knowledge and other higher technical competence is missing within DRT. A lack of technical expertise among the partnership makes the platform's supplier the only source for new ideas in further technical development. The gap in technology skills between partnership and supplier is great and creates a misunderstanding between the two parties. Introducing a third party of technical expertise could be a bridge between supplier and partnership that could strengthen the knowledge and create trust between the parties.

The opendev.ug website uses the proprietary format XLS (Excel Spreadsheet) for data sets and DOC (Word) for documents. A portal for open development that releases documents and data sets in proprietary formats are something that collides in the platform's purpose of being a portal for promoting openness.

Visualization of the platform's data sets requires data sets to be transposed before they can be used with existing tool. The existing tool have limitations that can be overcome by using the combination of external software Miso Dataset and Highcharts, that has been demonstrated in the results.

The Uganda Open Development Partnership Platform can be improved with the proposals from this report on how to use open license, multiple formats for material and transposed data sets. The following proposals have been made:

- Documents published on the platform should have multiple formats available for download. The users should be able to choose the format that suits them the most to be compatible with their own text editor. At least one format should be an editable, open format, so others can reuse and redistribute the document with their own changes. At least one format should be read-only format, so the document can be opened to be viewed without any changes from the difference in editors readings of the format.

- Publications of data sets should be available for download in multiple formats. The users should be able to choose the format that suits them the most to be compatible with other systems. At
least one of the formats should be in the CSV format, since most text editors is compatible with this format. The data sets in the CSV format should have data structured with values placed vertically, under a row with headers for each column.

- Both data sets and documents should have an open license for the material. Material is not free to use without restrictions just because it is possible for anyone to download or copy. An open license, for example Creative Commons placed visibly associated with the document or data set would define the border of use for the material.

- Data sets uploaded on data.ug are proposed to be structured and adapted to the column structure with rising values placed vertically, which is more likely to be compatible with other systems. The data.ug website is based on the CKAN project that includes the graph builder tool, Recline. The tool creates the possibility of proposed structured data sets to be used in a graph, directly in the web browser.

Articles and guides have been written by the author based on the knowledge gained in this study. The created material have been sent to the partnership for publishing on the opendev.ug blog.

- One guide was created to spread the knowledge about transposing data to reach a suitable structure of data sets. The second part of the guide is to display how data sets can be used for building graphs with the graph builder Recline, included in CKAK project (Appendix 3).

- The author's improvement proposal for multiple formats (Appendix 4) explains why it is important to release material in at least three different formats, one open format, one well spread format and one static format.

- The author's improvement proposal for open licenses (Appendix 5) should provide an insight in why it is important to have an open license associated with the material that should be considered open. The license usage of other open development initiatives are explained as comparison to get an understanding what the differences are in their licenses.

Except for the improvement proposals are the following recommendations for the partnership:

- The third party that would serve as a bridge of technical expertise and trust between the platform and the supplier of the platform could be established by students of information technology faculties from nearby Universities and their collaborators. Students will remain neutral, because they write thesis's and articles that will be judged by other students and teachers that could discover whether a student took non-objective side on the recommendations they gave to the partnership. It would also be beneficial for the partnership of the platform to hire personnel with technical competence. The technical personnel could focus on communications with the other technical parties to create an understanding in further development of the platform.

- Set up the API that comes along the CKAN project. The API would make it possible for other platforms to fetch data and metadata of available data sets from the data.ug website. The platform should inform visitors about the API, because metadata would spread the word on other websites about the data that can be fetched from the opendev.ug website. This leads to more visitors and increased use of the platform's data.

- Select a few interesting and valid graphs from the data handler website data.ug to be published next to the data sets on the platform site opendev.ug. This will give more distance between the two websites, as users have less need to visit data.ug and only visit this website when looking for
more information. The data.ug website should provide more information about the data sets published on opendev.ug, also provide management of data sets and possibility to build new graphs.

- Use the improvement proposals to create more awareness around open data for both non-government organizations and different government agencies.

### 6.2 Discussion

There is a lot of information about the open development platform on their website opendev.ug that could have been used as a ground for the interview. However, the website was discovered during the time in Uganda and was not properly analyzed before the interviews took place. The interviews could have been more prepared if this website was discovered before having the interviews and that could have lead to different results of the study.

The report could have been affected by the interviews made. A recording device was lacking when the interviews were performed and what was said at the interview were written down on notes by the interviewer to later be interpreted by memory of the interviewer. This means that interviewers own words may have been mixed together with the interviewed into the quotations used in this report.

The flaws found when visiting the partners of the platform and the platform itself has affected this report in a major scale. This is because correcting these flaws is the next step for the platform and the focus of this report is the purpose to improve the platform.

### 6.3 Further Studies

After asking if graphs will be fetched from data.ug and be published on opendev.ug, Bbosa says “Putting graphs on opendev.ug will just be duplication of the same graphs on data.ug which does not make sense”. This is a good point for the moment when data sets are provided by known providers and confirmed valid. However, anyone can upload data sets and graphs on data.ug, without confirmation, that can lead to many graphs uploaded by unknown users. The interactive graph builder tool is useful for visitors to build new graphs based on their own requirements. What about visitors who do not understand the graph building tool and do not want to explore the piles of existing graphs created by other users. It is possible as a normal user to upload any data set or graph, based on real or made up data and give it an interesting name. Visitors might find the material without realizing the validity and that is the reason to publish visualizations on data.ug.

Further studies could be made in the area of why there are hindrances to open data. Especially why the government did not join the Open Government Partnership when all five other invited African countries joined after they were invited at the same time as Uganda (APC & CIPESA 2012:1).

An area that can be researched is how open data changes society and prevent problems with corruption and poverty. This could be studied in a historical case like SPIDER (2012) and Sida (2012), or the effects of the contribution of this study. Changes for these problems would probably take time to get effect and would need a longer period of time than a bachelor's thesis to prove that the contribution of the study had affect.

Choosing what data to collect and then collecting the data is the initial steps to open data. Providing means for organizations to collect data is essential since much can be improved in this area. Examples of improvements are to create applications with better user interfaces and find valid and reliable ways to
gather data. Problems that arise after collecting data is where this data should be stored and how it should be refined before reaching the users.

Android smart phones is spreading widely in developing countries as they become cheaper (Brunette et al. 2013:63). Smartphones spreads and that will open up the market for all kind of mobile applications, not just for collecting data. One further step for the Uganda Open Development Partnership Platform would be to develop an Android application that displays the available data sets and visualizations.
7 Acknowledgment

This report is a result of bachelor's thesis in the faculty of Arts and Social Science at Karlstad University.

Thanks to DRT, CIPESA, Makerere University and all the other organizations and people for kind and warm welcome received in Uganda. Thanks to all those people who made it possible so the information of this report could be gathered. Thanks to Caroline Wamala and John Sören Pettersson who gave wise advices about researching for thesis and experiences in Uganda that lead the right way in many decisions that was taken during the whole work with this study. The author also want to thank supervisor Remigijus Gustas for advices, the author's family for the support and all other people related to the research that helped along the way.

Thank you.

Robin Larsson
8 References


Microsoft (2013). Education Alliance Agreements. [Online] Available: 
[2013-04-30]


[2013-04-11]

Open Data Commons (2009a). Open Data Commons Public Domain Dedication and License (PDDL). 

http://opendatacommons.org/licenses/by/summary/ [2013-03-18]

[2013-06-05]


http://www.opengovpartnership.org/eligibility [2013-03-18]

http://www.youtube.com/watch?v=Bq_ZW1l1ZXA0 [2013-01-27]


Forskningsmetodikens grunder: Att planera, genomföra och rapportera en undersökning. Lund: 
Studentlitteratur AB, pp. 67-133.

Professional Testing Inc. (2006). How do you Determine if a Test has Validity, Reliability, Fairness, and 
Legal Defensibility? [Online] Available: 

Pollock, R. (2010). Why Share-alike Licenses are Open but Non-commercial Ones Aren’t. [Online], 1-2, 
Available: 
http://blog.okfn.org/2010/06/24/why-share-alike-licenses-are-open-but-non-commercial-ones-arent
/ [2013-04-24]


9 Appendix 1 Blog Entry for Visualization

Source: [http://www.opendev.ug/blog/how-share-your-uganda-opendata-projects](http://www.opendev.ug/blog/how-share-your-uganda-opendata-projects) [2013-03-28]

**How to share your Uganda OpenData projects**

**BLOG POST February 23, 2013**

We are very exited to share so much OpenData about Uganda and are looking forward to see what people will create with the data. We have created a few examples, like these graphs about Fish Production figures in Uganda.

If you have created anything online with datasets that are published on data.ug we would like to know from you. If you published a blog post, a visualisation, a map or anything online, just send us an email to maps@data.ug or graphs@data.ug. Please mention the dataset you used to create your work, so we can add it to the Apps & Ideas section of that dataset (like here for Fish Production). If you register on data.ug you can do this yourself as well.

If you want us to publish your Graph or map for you, no problem. Just zip all your files and email them to the maps@data.ug or graphs@data.ug and we will publish them. Please mention the dataset you used again.

If you are a bit more technical, we are maintaining a list of maps & graphs in our Github repository. If you fork the project and to a pull request, we will also publish your data.

10 Appendix 2 Blog Entry for Polishing Data


**Polishing spreadsheet data before upload to data.ug**

**BLOG POST April 2, 2013**

The document attached explains how to polish a spreadsheet before uploading to data.ug. With polished spreadsheet data, the dataset becomes easier to comprehend at a glance and is therefore easier to analize and visualize.

The steps explained in this document relate directly to the raw spreadsheets at data.ug but may go beyond the highlighted steps. The level of polishing to be done depends on the layout of a given spreadsheet and are not limited to the ones mentioned here.

The keys steps for now, include:

• Aligning spreadsheet data.
• Creating a single spreadsheet from multiple sheets.
• Deleting empty rows/columns and duplicate rows.

When a user intends to upload a spreadsheet to data.ug, it would be wise to follow the above steps. A detailed explanation is attached.

Files:
11 Appendix 3 Blog Entry for Visualization in data.ug


Data visualisation using graphs

BLOG POST April 17, 2013

To visualize is an important step in the open data process because this is the step when the use of data is displayed and interpret by the viewer. It is hard to make sense and overview of data structured in a table. It becomes much easier to understand if the data is visualized in a graph. It is useful to display data in a graph visualization when tables have large amounts of ambiguous data that needs to be compared.

The site data.ug, http://data.ug is built on the open source system CKAN for handling data sets. CKAN uses the JavaScript library ReclineJS, http://reclinejs.com/ that is free to use under the MIT license. This is a tool for creating graphs based on column structured data sets.

Files:

stepstograph.pdf

12 Appendix 4 Blog Entry for Multiple Formats


Release documents and data sets in multiple formats

BLOG POST May 22, 2013

It is beneficial for open development to release documents and data sets in editable formats, so others can download, change, reuse and redistribute them on different places in other context. If multiple formats are available for download it opens up for the user to choose the option that suits them the most for their editor (Open Formats 2004).

The Word text editor which uses the doc (1997/2003) and docx (2007/2010) formats and Excel Spreadsheet which uses the xls (1997/2003) and xlsx (2007/2010) formats are widely used because many people have access to them on public computers. Many schools, communities and governments have joined the Microsoft Education Alliance Agreements (Microsoft n.d) which means the software are installed at computers for the students and workers to learn and use. This makes people familiar with the software and creates the illusion that the software exists everywhere for everyone. The alliance only includes schools, communities and government, the rest have to buy the software to read the documents in those formats.

Some users chose not to have the software and for those there are free open options available. Example of open software with text editor and spreadsheet software are OpenOffice and LibreOffice, which uses the open document format.

The Word and Excel software formats are not open formats and that results in errors when other
software try to open documents with those formats. For example when opening a closed format
document with another editor results with the images and text often are rearranged with disappearing
fonts, and this makes the document unreadable. The closed formats can only be fully compatible with
one editor, this forces users to buy the software with the closed formats (Stallman 2007). Further more
there is a problem between the doc and docx format, Stallman (2007) describes both of these problems:

“Most computer users use Microsoft Word. That is unfortunate for them, since Word is proprietary
software, denying its users the freedom to study, change, copy, and redistribute it. And because
Microsoft changes the Word file format with each release, its users are locked into a system that compels
them to buy each upgrade whether they want a change or not. They may even find, several years from
now, that the Word documents they are writing this year can no longer be read with the version of Word
they use then.”

The product that uses closed formats usually have problems opening the open formats. Lime Survey
Manual (2009) recommends not to open CSV files in Excel: “In general we do not recommend to use
Excel to edit the exported .csv files because Excel does some changes to the internal data structure
which results in a failing import.”. The open format editors can however save documents in the closed
formats (WikiHow n.db).

The read only format PDF is useful for those who do not have knowledge of or do not have the
possibility to use any text editor that support the editable formats. A read only document reader does not
rearrange text or photos and documents can be shared as original because the document will be
displayed the same for all viewers (WikiHow n.da).

Stallman, R. M. (2007). We Can Put an End to Word

Open Formats (2004). Which open formats can I

WikiHow (n.da). How to Convert a Microsoft Word Document to PDF
3-04-24]

WikiHow (n.db). How to Create Documents in OpenOffice.org That Will Work With Other
h-Other-Software [2013-04-24]

13 Appendix 5 Blog Entry for Open License


Why Licensing is important in Open Data sharing

BLOG POST May 10, 2013

Material is not free to use without restrictions because it is possible for anyone to download or copy
from the owner. By using an open license, for example Creative Commons (CC) and placing the license
visible near the downloadable document or data set defines the border of use for the material. OKFN
Blog (2009) describes the problem with an absent license:

“Licensing is important because it reduces uncertainty. Without a license you don’t know where you, as
a user, stand: when are you allowed to use this data? Are you allowed to give to others? To distribute your own changes, etc?"

One way to solve this problem is to place the license to always be visible in the footer. Open Data Kenya website, https://opendata.go.ke/, has placed a CC license in their footer together with the text “Creative Commons Attribution 3.0 Unported License”, that is a link to the CC page to describe the license in detail. The part of the image with a symbol of a human and text “BY”, means by the following conditions, that the use of the material must have an attribution naming the source of the material (Creative Commons 2007).

The Open Development Cambodia website, http://www.opendevelopmentcambodia.net/, has also placed the license in the footer. Their CC license has the text “Creative Commons Attribution-ShareAlike 3.0 Unported License”, that is a link to the CC page explaining exactly the content of this license. The license differ from Open Data Kenya license with a repeat symbol and the text “SA”, standing for share-alike license, meaning the material should have the same or similar license when reusing and redistributing.

The difference in these licenses may be affecting the use of the material (Ito 2004). For example, someone who creates a product to earn money may not want to use the material if the share-alike condition is included. Having the SA part of the license means competitors can use the material and get started quickly with their business. It would be enough for the user to mention the source, so the competitors can receive the base material that has been gathered by public funding and not make the commercial companies to share their work in the same way. The SA part of the license put the restriction to mix the work with other licenses that is not similar (Ito 2004).

Sources


