Visualization of the need and distribution of aid in Africa

Dennis Englund
Erik Krantz
Abstract

Visualization of the need and distribution of aid in Africa

Dennis Englund
Erik Krantz

When humanitarian aid is distributed to developing countries, crisis areas and humanitarian aid organizations it is often difficult to get a good overview of how much aid is needed compared to what is given. Organizations have a tendency to focus on either a targeted area or specific topic. This issue is further hindered by the large amount of active aid organizations which adds to the complexity of getting good overview and the fact that crisis-areas are in a constant state of change.

We intended to ease the decision making and the overall overview by supplying a website with an interactive map displaying the difference between needed- and donated financial aid of countries within the African continent. The scope can be expanded by including other parts of the world. In order to give the user an overview we collected humanitarian aid data gained from different aid organizations, where the user then could easily choose what is to be displayed, e.g food, care and educational aid. Each country then gets color coded depending on the current gap in funding. This data visualization method is called heatmap. This can then be used as a base for decision making or to show aid contributors the amount and what type of aid is needed in a designated area.

Handledare: Mats Daniels och Björn Victor
Examinator: Björn Victor
Sammanfattning

När ekonomiskt bistånd distribueras till behövande länder och hjälpsorganisationer är det ofta svårt att få en bra överblick över hur mycket som behövs gentemot vad som ges. Hjälpsorganisationer har även en tendens att enbart fokusera på ett särskilt område eller specifikt ämne. Att få en bra överblick försvåras ytterligare i och med att det existerar ett stort antal olika hjälpsorganisationer som leder till ytterligare komplexitet och att humanitära kris-zoner är i ständig förändring.

För att ge en lösning på detta är vår intention att underlätta beslutsfattande och ge en bättre överblick genom att konstruera en websida med en interaktiv karta som visar skillnaden mellan behov och givet ekonomiskt stöd i olika länder, med fokus på Afrika men med plats för ytterligare expandering. För detta har vi samlat finansiell data från olika hjälpsorganisationer, där användaren själv kan välja vad som ska visas, till exempel mat, vård och medicin. Varje land får sedan en färgnyans i proportion till den största skillnaden av bistånd. Denna metod att visualisera kallas heatmap. Heatmapen kan sedan användas som grund för beslutsfattande eller för att visa andra donatorer hur mycket och vilken typ av behov som behövs i ett specifikt område.
# Contents

1 Introduction .................................................. 1
   1.1 End-users ................................................. 1
   1.2 Original idea from SIDA ................................. 2
   1.3 Acknowledgement ......................................... 2

2 Background .................................................... 2
   2.1 Humanitarian crisis ....................................... 2
   2.2 Agenda 2030 ............................................... 3
   2.3 Data sources ............................................... 4
   2.4 Data visualization ........................................ 5
   2.5 Swedish International Development Cooperation Agency, SIDA 6

3 Purpose, aims, and motivation .............................. 7
   3.1 Issues with the handling of data ......................... 8
   3.2 Ethical challenges ......................................... 8

4 Related work .................................................. 9
   4.1 Aid databases .............................................. 9
       4.1.1 Humanitarian Data Exchange, HDX ................ 9
       4.1.2 Financial Tracking Service, FTS .................. 9
       4.1.3 The Organization for Economic Co-operation and Development, OECD 10
       4.1.4 Open Aid ............................................. 10
   4.2 Shortcomings and problems with current websites .. 10

5 Tools and techniques .......................................... 11
9.3.6 Information button ................................. 27
9.4 Technical results and discussion .................... 28
9.5 Parameters ............................................ 28

10 Conclusions .............................................. 28

11 Future work .............................................. 29
   11.1 Expanded visualization of data ..................... 29
   11.2 Data handling ...................................... 29
1 Introduction

As of 2018 roughly 10%, 736 million, of the world’s population lived in extreme poverty with an income of less than $1.90 a day estimated by a report [Was18] made by The World Bank. This rate, although better compared to the rate of 36% at 1990, was still not in line with the goal of ending global poverty by 2030 as agreed by the United Nations [Uni19a](UN) due to that the decline in poverty has slowed down. In some regions the rate even increased between 2013-2015 due to conflicts in Yemen and Syria according to the report.

The average poverty rate of African countries was estimated to be at least 10% by 2030. Since there was a clear correlation between the amount of financial aid received and a developing countries GDP growth [AL15] a good way to help resolve poverty was to give additional financial aid to low-income countries. How to effectively distribute financial aid can however be difficult as some donors of foreign aid do not coordinate with each other [FNÖ15] and most of the actual aid data only shows funding needed or funding received and does not, for instance, take into account a local refugee camp within a country.

1.1 End-users

The original idea for the project was presented to us by the Swedish International Development Cooperation Agency [SID19](SIDA) which is the Swedish government agency in charge of distributing the Swedish development assistance to developing countries which we will describe in further detail in section 2.5. Therefore, the intended user was an employee working with humanitarian aid. More so, one that had a need to continuously check and be updated on the overall distribution or current situation in a specific area. Even if a statistician could be considered as a primary example of a user we wanted to emphasize the fact that not everyone was a data analyst or statistician which could lead to different interpretations of statistics. We aimed to circumvent this by emphasizing on making a website that was easy to navigate for everyone and thereby we covered a wider array of users within organizations. Although our intended user was an employee working with humanitarian aid, the website can still be used by anyone in search for enlightenment on the situation of current funding in Africa.
1.2 Original idea from SIDA

As stated earlier the original idea for the project was based on a request made by SIDA. We proceeded by using their idea as an inspiration while we ourselves set the demands for the project. The idea they presented was a heatmap visualizing the funding gap between the countries in Africa. Although that was much in line with our website we wanted to further analyze what other websites were lacking and what needed to be improved, not only visually but also in terms of efficiency within the organizations.

1.3 Acknowledgement

Patrik Johansson also partook in this project, but not as an author of this report. He worked with the original design of the website.

2 Background

In today’s society there exists a large amount of organizations working with charity. Critically analyzing how different topics, such as hunger and active conflicts, were related could be overwhelming with the large amount of factors in play. For example determining how a a poor education in a country could dictate the mortality-rate due to a lack of knowledge within first aid. Additionally, determining to what degree they affect one another is essential. A good example of this being displayed through different graphs is Gapminder [Gap19]. In Gapminder regular users can easily see how these topics affect each other. Gapminder was an initiative made by the legendary statistician Hans Rosling with the goal to "Fight devastating ignorance with a fact-based worldview that everyone can understand".

2.1 Humanitarian crisis

On the 10’th of December 1947 the United Nations General Assembly proclaimed The Universal Declaration of Human Rights [Uni19b] (UDHR) in Paris, France. It described the fundamental human rights for all people. As of 2019 we still could not uphold the declaration as extreme poverty in the world still existed even though we could most likely end it [Sac05].

While there are several factors beyond human prevention contributing to the high amount
of poverty, e.g. viral outbreaks and drought, some may be aided by humanitarian contributions [Str07]. Therefore it was important to help solidify any decisions that may decrease the impact on infrastructure and amount of casualties of these crisis events. We felt that our project may help by providing an overview of the current situation, not only to the parties involved with making these decisions but also to those that were interested in the situation.

2.2 Agenda 2030

The Agenda for Sustainable Development was established by the UN in September 2015 with the goal to end poverty by 2030 [Uni19a]. The purpose with agenda was to create a worldwide partnership where each country has to participate and take responsibility. The agenda recognizes that to end poverty in general you need to consider several factors. With a large focus on human rights but also climate change the goals covered a wide range, which can be observed in Figure 1.
2 Background

Figure 1 The Sustainable Development goals presented in Agenda 2030.

2.3 Data sources

Determining what defined a reliable source was an important aspect. Furthermore, collecting data within a poverty stricken country is both expensive and error prone [SG15], which leads to further uncertainty in terms of reliability. In the interview with SIDA they also addressed the uncertainty with data sources, even the data provided by UN. They mentioned that "When UN releases their yearly report on global humanitarian overview they sometimes have to make estimated guesses when covering information in hard-to-reach areas due to the information being difficult to access". One other important factor that we took into consideration was how updated the information was. When was the most recent information gathered and whether or not the information was updated on a regular basis. By conducting interviews with several organizations we could further clarify what defined a reliable source.
2.4 Data visualization

In order to adhere to UDHR a lot of difficult and constructive decisions must be made. But in order to make those, the people in charge of making any kind of impacting decision, must have a clear picture of the current situation. It would not be feasible for these people to all be able to analyze the massive amount of data available.

However, if we visualized the data needed by replacing endless text and numbers with colour-charts and various graphs the decision-making and overall understanding would be improved [WWA15]. For example, we offered the user the option to select a specific category to be displayed on a map of Africa with a range of colours resembling the value ranges as show in Figure 2.

![Figure 2 The original design for displaying information with colour-ranges](image)

Our project aimed to support humanitarian aid by easing the process for the organizations in charge of the distribution of aid. Although we wanted to simplify the information displayed to a greater extent, we also wanted to leave room for the user to access more specific data if needed. Since a deeper understanding in a specific topic could be essential for making good decisions. Furthermore, by highlighting a country one would be able to access further information about said country. One design choice was a graph
containing further information over a designated time-period. This was done in a similar fashion to how UCDP have chosen to display number of conflicts on their website as shown on Figure 3.

![UCDP's, Uppsala Conflicts Data center, website.](image)

**Figure 3** UCDP’s, Uppsala Conflicts Data center, website.

### 2.5 Swedish International Development Cooperation Agency, SIDA

In 2019 the total financial aid to be distributed by SIDA was estimated to 44.2 billion SEK, which is in line with Sweden’s 1% gross nation income foreign aid policy. As of 2019 they cooperate with 35 countries, with the majority being in the African continent. We interviewed them to get a better grasp of the current situation both within the organization and on the field. We also asked for minor inputs and requirements on our website.
3 Purpose, aims, and motivation

The project’s main focus revolved around two different topics. Firstly, we wanted to find out what defined a reliable source within our chosen area. More so, analyzing how frequently the data was updated. By doing so we wanted to encourage organizations to be more source-critical since there may exist sources that are better suited to their needs.

Secondly, we wanted to provide easily accessible data in one place with a strong focus on developing a user-friendly graphical interface. The data was to be represented in a simplistic way with symbols, graphs, and maps replacing the large amount of raw text and numbers gathered from databases. This was done in order to lessen the amount of time spent on research for organizations to further reduce their workload.

Organizations that work for a similar cause may use different data and sources and therefore get differing results. Red Cross, which is an international humanitarian movement [Red19], uses their own database at International Federation of Red Cross and Red Crescent Societies [Int19](IFRC) and could therefore present data which conflicts with the data provided by SIDA. It is important to update the standard in this area since people in need of aid are dependant on a good practise within the field. If organizations do not see over potential alternatives for sources and use different data this will further increase the confusion of the overview that is accessible online.

With our project we combined several aid organizations data to cover their respective grey areas. By improving the current methods used to calculate aid distribution this will lead to a better work-environment for organizations working with aid. Moreover, by increasing the efficiency in the workplace this could lead to the workforce being better utilized [WWA15].

The project aimed towards supporting sustainable development by providing tools to improve the efficiency of organizations that are working with humanitarian aid. Furthermore, the project worked towards the sustainable development goals of Agenda 2030. Said goals that we supported were no poverty, zero hunger, good health and well-being, equality, quality education and peace justice and strong institutions [Uni19a]. Since information regarding hunger, current conflicts and financial situation to mention a few was available on the website the project covered a wide range of the goals presented in Agenda 2030.

Interviewing multiple organizations was vital for the success since the data was dependent on a variety in opinions and factual data in which the organizations trust and utilize. The main advantages of our solution was that we combined and compared several different sources and thereby provided our end-users with a more correct overview of the aid distribution.
3.1 Issues with the handling of data

While conducting interviews with SIDA and other aid-organizations, we were instantly introduced to one of the main issues with the handling of data. When organizations were updating their internal statistical data it was often done manually with excel documents covering data from multiple sources. Since it was done manually it was more error-prone to human made mistakes.

In the case of SIDA they combined data from UN, Office for the Coordination of Humanitarian Affairs [Uni19c](OCHA) with the annual report Global Humanitarian Overview(HNO) made by OCHA. Once they had collected and updated the data in an excel document they then forwarded their data to the Financial Tracking Service [OCH19a](FTS) which in turn may needed to forward the data an additional time. These chains of events leads to further issues regarding updating. If one of the organizations was slow or made a mistake during the update process it would lead to uncertainty regarding if the data was updated and if it was updated with the correct information for the organizations next in line. With more organizations within the update process the frequency of errors is increased.

Creating an automated system to replace the excel documents would be a project on its own. Instead our focus in this area was directed towards enlightenment in the subject and how the current methods could be improved.

3.2 Ethical challenges

There were some ethical challenges since it was mainly up to us to decide what factors to take into consideration when defining a good source. There was a great number of parameters to consider and to rank their relevance compared to one another was difficult. Therefore, there was a risk that we put more focus on certain parameters that would be differently prioritized by others. With less dependable sources for our data there was a risk that we would provide less reliable information. We also made no difference between needs within a parameter, for instance medical requirements might also include different degrees of severity.

We mitigated some of these issues by only collecting data from certain databases that we learned about through our interviews with SIDA and other organizations. However their insight in how data is handled is also quite limited and shows the severity of the issue but as the infrastructure of a country continues to develop this will might be less of an issue in the future [WWA15].
4 Related work

In this section we discuss others solution to either aid database summation or visualization of data and how they compare to our project.

4.1 Aid databases

There are several databases aimed at collecting and summarizing humanitarian aid contributions and distributions. Most prominent of these are the The Humanitarian Data Exchange [OCH19b](HDX), Financial tracking service [OCH19a](FTS), and The Organisation for Economic Co-operation and Development [OEC19](OECD).

4.1.1 Humanitarian Data Exchange, HDX

HDX is a platform for sharing humanitarian aid data. It uses an open data policy that makes sure everyone can contribute and extract the data that they need. It uses the Comprehensive Knowledge Archive Network [Ass19](CKAN) as its data management system, which is open-source and used by organizations and government all over the world. As of 2019 their collected database consisted of 8927 datasets spanning 248 locations world wide from 1151 sources. They also have some tools available to visualize the data available, much like our project. HDX is aimed at advanced users wanting to analyze the data themselves.

However as the amount of available data and different ways to visualize it increases so does the complexity. This results in a lot of time spent in order to get a complete overview of the data the user might be interested in. We streamlined the visualization and hid unnecessary parameters from the end-user thereby presenting a less cluttered interface. We used HDX as a database to get the datasets relevant to what we want to visualize in our project.

4.1.2 Financial Tracking Service, FTS

FTS is specialized in tracking the financial aid from donor to recipient. To visualize this they are using a world map only showing the required and funded amount. There are hardly any parameters and as such it looks clean and feels simple to use when compared to HDX.
In that aspect our project filled the complexity gap between HDX and FTS data extraction and visualization. Where FTS only show the total financial aid given or received, our project also contain information about each country, refugee-camp status, and other parameters such as health- and educational aid requirements and needs.

### 4.1.3 The Organization for Economic Co-operation and Development, OECD

OECD is not focused on humanitarian aid data but aim to help improve the lives of all people by collecting data that the end-user can analyze themselves. For instance you can check the amount of leisure time citizens get within a country or how much tax they pay. OECD describes their goal as "The common thread of our work is a shared commitment to market economies backed by democratic institutions and focused on the well being of all citizens." OECD does not have a way to visualize data themselves so its up to the user to utilize their data in terms of their needs.

In that sense they are indirectly involved with humanitarian work as they aid governments to make socioeconomic decisions to the benefit of the people. We used OECD database to provide information about each country.

### 4.1.4 Open Aid

Open Aid is a web service that provides information on Swedish aid. The information is gathered from open government data. The information service is a collaboration between SIDA and the Ministry for Foreign Affairs (Utrikesdepartementet), which is the ministry responsible for foreign policy in Sweden.

Our project was similar to the work of Open aid. We wanted to share how aid is distributed based on several open databases to provide a better overview, much similar to their standpoint. "The goal is to fight poverty as effectively as possible. To achieve this goal, aid has to be open to public control and ideas from more sources."

While working with our project we utilized their website, not only for data, but also to observe how they chose to display their data to see how it could be improved.

### 4.2 Shortcomings and problems with current websites

Even though there existed other websites that helped highlighting the problem they were often focusing on a specific area or topic thereby lacking information that could
potentially be highly relevant to their choice of topic. The websites that covered a larger amount of information were often complex to navigate through and could feel overwhelming for an average user.

Uppsala Conflict Data Program [Dep19](UCDP) is a good example of a website narrowed down to one specific topic, namely armed conflicts. Although the information is presented in a visually appealing way with numerous graphs and an interactive global map it is lacking data that can be a factor to the information they present. For example, once a country is selected the number of deceased people will be displayed based on type of violence with a graph. But other information that can be a cause to the current situation like food-shortage or funding is missing. There is however a summary about the selected country, but it is displayed in raw text. It also contains a lot of information that may be redundant for the user and could be better represented with a graph or image showing data relevant to the cause.

We aimed to solve this problem by supplying an interactive dynamic map which displays the difference between required and received funding.

5 Tools and techniques

In this section we discuss the tools used to make the project. We developed the project in Python3, Hyper Text Markup Language, Javascript and Cascading Style Sheets. To manage our database we used the open-source version of MySQL. In order to construct a visually appealing interface and intuitive data representation we interviewed potential end-users and made changes based on their critique.

5.1 Languages used

Python3 is a high-level object-oriented programming language that is easy to use and offers a good amount of scale ability. It often used in introductory language courses in several computer science or information technology majors [Per15].

Examples of other popular object-oriented languages are Java and C#. However as most of the databases handling datasets relevant to our project had developed an Python based API we opted to use Python as our back-end language and thereby ease the collection of datasets.

Hyper Text Markup Language(HTML) is used to create web pages and applications. It can be viewed as the foundation used to build a website as it can not be programmed
5 Tools and techniques

in the classical sense, for instance there is no if, else syntax often seen in other programming languages. Instead it calls other script languages if programming logic is needed. Examples of other markup languages are TeX and eXtensible Markup Language (XML).

**Javascript (JS)** is an interpreted programming language with a high level of system abstraction. JS should not be confused with Java which, although similar name and syntax, are completely different. It is often used in combination with HTML to make interactive websites by using `<script> JS script </script>`. JS is also cross-platform and as such there is no need to write it with a specific platform in mind. Examples of similar languages are CoffeeScript and TypeScript. However, these still need to be interpreted by most browsers as JS to function and as such we can eliminate this extra step by using JS. However, if we were doing a large project with a lot of collaboration involved TypeScript will probably be better as it is a fully object-oriented. This results in cleaner code than JS and as such less error prone.

**Cascading Style Sheets (CSS)** is a stylesheet language used to alter the style of other elements, e.g. text, box layout or other effects in the website. It is, together with HTML, regarded as a must-have in web design and is used by 95.8% of the websites connected to world wide web [Q-S19]. Other alternatives to CSS are Leaner Style Sheets (LESS and Extensible Stylesheet Language (XSL)). These are CSS preprocessors and is basically CSS but with added features. As the project is one webpage this would result in more unutilized code.

### 5.1.1 Libraries and frameworks used

1. **hdx-python-api** Python API used for pulling and pushing data to HDX. In this project we only pulled data from HDX. HDX is based on CKAN data management system. The API is developed in such a way to mimic the HDX website with tags etc. and hides CKAN idiosyncrasies. Every dataset in HDX can be seen as an object and can be reached the same way as on HDX’s website.

2. **Node.js** A JavaScript run-time environment used to run JavaScript before the webpage is called. We used Node.js to calculate the dynamic elements in our website.

3. **Leaflet** An open-source JavaScript library made specifically to make interactive maps. Used in the project to make the map over African countries interactive.

4. **JQuery.js** A JavaScript library with many different features, such as event handling, Ajax and more. We primarily used the Ajax feature to retrieve data from the webserver.
5. **pandas** An open-source python library, developed to help analyze data information and structure. We used this to read, manipulate and save .CSV and .JSON files collected from databases.

6. **Chart.js** A JavaScript library that can create various graphs and charts. We used this tool for the majority of our graphs and charts.

### 5.2 Interviews

As the project progressed we continuously asked intended end-users, as discussed in section 1.1, for feedback with focus on the data visualization. First we presented them with the current iteration of the project and asked for feedback regarding the look and general user experience. We then put that feedback into our backlog to be integrated with our project. This was then repeated until we were satisfied with the end-result.

### 5.3 Ease future work

In order to make it easy to implement and expand the project we followed the PEP 8 [Gui19] coding convention. PEP 8 is the standard convention used for Python and developed by Python Software for coding. This combined with a good use of commenting made the code easy to read and understand. As for the database we followed the same principle of a coding convention, in this case The SQL Style Guide [Hol19].

As we collected datasets from different databases we then needed a standardized naming convention which would not differ depending on the database we got it from.

### 6 System structure

The system is largely based around databases. However, we also wanted to make the visual product as appealing as possible and therefore the graphics and interface was a vital segment. As all relevant data are gathered from different databases, e.g HDX or OECD, we first had to collect it in order to visualize it. Most of the databases we were interested in had developed an API to ease the data collection. As most of these were made for Python or JSON we opted to write our back-end in Python and Javascript.

When we designed the project we had to take into account that it would have had to be dynamic as the data represented continuously changed. Two choices were now avail-
able, first we build the project website to request data from the databases we want to represent and thereby eliminate the need for our own database. However, as some databases do not have a way to efficiently query the data we want, this was not an optimal option. Therefore, we used our own database that queries and stores the data of interest which the website easily can acquire data from as shown in Figure 4. This resulted in a faster and more consistent website as we would always get the data needed quickly and from one source.

![Simplified dataflow of relevant datasets](image)

**Figure 4** Simplified dataflow of relevant datasets

### 6.1 Back-end

The database is continuously updated at a interval of once a month. We opted to use this update rate as we saw that most of the datasets were updated at roughly the same interval. As most of the datasets were stored as either JSON or CSV, we first downloaded all the relevant datasets. Then we categorized the information into dataframes using Pandas library so that they matched the columns in the table. Each of the parameters available on the project website has its own table in our mySQL database where the columns are countryCode, year, requirements, and funding. CountryCode is the countries ISO 3166-1 alpha-3 code e.g. AGO for Angola. Available parameters are population, food, educa-
7 Requirements and evaluation methods

The requirements on our project was that the website was visually-appealing and contained reliable data. The main focus was that SIDA employees were satisfied with our front-end solution as back-end were unspecified in that we could use any language and platform for our project. This also means that we have to update the data dynamically at a realistically fast pace, i.e. updating our database at the same pace as most databases we gather datasets from update theirs. Another goal is to ease the inclusion of new databases and features for future work.

To produce a good result and evaluate our work we needed to work with different organizations dealing with data visualization, such as UCDP, and aid organizations, such as SIDA, as well as our family members and friends. As our family and friends come from a wide range of background and age, we felt that they would provide valuable feedback. Not only for the analyzing of data but also to try out our prototypes to see what changes are needed to make the product more user-friendly. This is achieved by having interviews with said users and we also read about data visualization to better our understanding on what makes a user-friendly webdesign.

7.1 User-friendly evaluation

Other than having a discussion with SIDA we also evaluated the project by using test groups. To evaluate the website user-friendliness we used different test groups of our
friends and family where they tried to understand and navigate the website. After they tried to use the website they filled in an evaluation form. In the form they were asked to answer the following questions:

- How easy was it to navigate through the website? Did it take a lot of clicks/time to find what you were searching for?
- Was the data visualization intuitive? Were the graphs too complex or too simple?
- Did the website feel trustworthy? If not, what could be improved?
- Were there any problems with the website? Were there any misleading icons or colors?
- Would you like other parameters other than the ones supplied?
- What could be improved?
- Other opinions about the project.

Based on the feedback given we then made changes to the website and/or the database. This was done a total of 3 times as we made changes and repeated the procedure for each iteration of the project.

### 7.2 Technical evaluation and requirements

When a web page takes more than 3 seconds to load 53% of the users feel like its hindering the user experience and move on [Dan19]. This meant that the project website had to be fast in order to keep the user experience optimal and reach a broader audience.

Therefore we evaluated the project website first view load time with a load time requirement of 3 seconds or less. This was done by using the GTmetrix [GT.19] tool on our website url.

As data is an important part of the project we also have to check the consistency of our database, with a goal of no anomalies. We did this by increasing the update rate to 1 per hour for 24 hours in order to simulate 2 years of run-time. We then checked the database for any errors or data being in the wrong place through using the website and inspecting each table in the database for inconsistencies. To check for other anomalies we used the built in diagnostic check for Azure databases.
8 Heatmap implementation, dataset collection and database design

This section is dedicated to describe how we collected datasets and the database design for our mySQL server. In the following sections we have dataset collection from the HDX database as an example. The data collection was done in similar fashion for all the databases we used and hence the HDX example can be seen as the general working process in this area.

8.1 Information extraction

As HDX has a vast amount of data available they need to have a way to categorize and find datasets the user might be interested in, i.e, information about the datasets. This is called the datasets metadata and often uses tags to describe what data are held within the dataset, when the dataset was last updated, which organization the dataset came from etc. This means that the user does not have to know the specific name of a dataset and can simply search by either the organization associated, e.g "FTS", with the dataset or through tags describing the dataset e.g "Funding".

<table>
<thead>
<tr>
<th>Source</th>
<th>OCHA Financial Tracking Service</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contributor</td>
<td>OCHA FTS</td>
</tr>
<tr>
<td>Date of Dataset</td>
<td>May 20, 2019</td>
</tr>
<tr>
<td>Expected Update Frequency</td>
<td>Every day</td>
</tr>
<tr>
<td>Location</td>
<td>Kenya</td>
</tr>
<tr>
<td>Visibility</td>
<td>Public</td>
</tr>
<tr>
<td>License</td>
<td>Creative Commons Attribution for Intergovernmental Organisations</td>
</tr>
</tbody>
</table>

FTS is open to all government and private donors, funds, recipient agencies and implementing organizations wishing to report financial pledges and contributions for humanitarian action. FTS team curates the reported data by manually reviewing, validating and reconciling incoming information and updating FTS data records on an ongoing basis.

<table>
<thead>
<tr>
<th>Caveats / Comments</th>
<th>Please see About FTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tags</td>
<td>CASH ASSISTANCE</td>
</tr>
<tr>
<td></td>
<td>FINANCIAL TRACKING ...</td>
</tr>
<tr>
<td></td>
<td>FUNDING</td>
</tr>
<tr>
<td></td>
<td>HXL</td>
</tr>
</tbody>
</table>

Figure 5 Metadata in HDX.
8 Heatmap implementation, dataset collection and database design

To collect the data from HDX a Python3 script was made that used hdx-python-api’s search and download methods in order to search for datasets we are interested in and then download them. We do this by first having a text document with all the country names we want to collect datasets about and then iterate through the list.

When we had the datasets downloaded we used the Pandas library to manipulate the .CSV and .JSON files in order to categorize the information. We have to do manipulate the downloaded datasets because the metadata tags are also in the files and would cause problems if we tried to insert them into our mySQL database. We also wanted to match the header row with the tables in our mySQL database in order to ease the mySQL insertion. To achieve this we used the read_csv and json methods from Pandas depending on the file format. By doing so we could now take away the headers and tags from the dataset and categorize the data so that they had the same name as the table columns in our mySQL database tables as a dataframe. The dataframe is then inserted into our mySQL database to their corresponding tables using to_sql method.

We run this script with a time interval of once a month in order to have consistently updated data.

8.2 Database design

When we designed the database we had to take into account that it could be updated at the same time as a user was using the website. In order to ensure that the information displayed for the user is correct we had to make sure the database transactions were ACID compliant. ACID stands for Atomicity, Consistency, Isolation and Durability and is a benchmark all relational databases should fulfill [Mul19]. It makes sure minimal information corruption or loss in case of a failure during a data transaction.

In order to get the database up and running we used Microsoft’s Azure [Mic19] database for mySQL service. Microsoft Azure is a public cloud computing platform that offers a wide spectrum of service ranging from databases to machine learning computations.

We wanted it to be easy to expand the database in order to ease future work by reducing data redundancy within the database. To achieve this we used a technique called data normalization, which is a way to structure the database to ensure minimal data redundancy and logical data dependency within a database. Data normalization makes several logical rules that each table within the database adheres to and depending on how many of these rules a table fulfill they are categorized as being within a certain normal form. Most databases deem the 3rd normal form to be a sufficient requirement as this keeps anomalies that can occur when doing data insertion, updating etc. are kept to a minimum [Wik19].
As a first overview of the database we used ERDplus [ERD19] to make the entity-relationship diagrams. We also used this to check that the normal form was satisfactory, i.e. above the 3rd normal form. We then created the tables for the database according to our ER diagram which resulted in the database schema shown in Figure 6

![Database schema diagram](image)

**Figure 6** Our database schema.

We tried to keep it as simple as possible, but with room for further expansion.

### 8.3 Heatmap

A main goal of the project was to make the website interactive. This section describes how we implemented the interactive elements of the website, such as the heatmap.

To make the heatmap we used the JS library Leaflet [Aga19]. Leaflet allows us to simply put our map element within a HTML element. First we had to include Leaflet CSS in our CSS file, then call the Leaflet Javascript within the HTML by using a `<script>` tag. Leaflet let us add wanted features to the map by using layers e.g. polygons for each
country and the ability to add onclick listeners to each of these. Layers include all the things that you want to either display or make interactive e.g. country borders or refugee camps.

To make the countries interactive we used a geoJSON dataset describing country borders within Africa. We had to load all the different color settings, as discussed in Section 9.1, and information gained from the database about each country as soon as the map initialized. This led to a bit more load time when the website is first visited but as a upside the load time between parameter switches was minimal.

![Figure 7 Heatmap displaying country borders](image)

9 Results and discussion

The project depended significantly on two issues. Firstly, we need to clarify what defined a reliable source. Secondly, we want to provide information in a way that was intuitive.
9.1 Project website

In Figure 8 we can see the website as a whole. On the left we can see the different parameters and their color range as displayed on the heatmap. The heatmap can be seen on the right where the countries have different colors depending on the gap of the parameter chosen. The gap for each parameter is calculated by simply taking the required funding and subtract the donated funding for the current year for that category i.e. food, health etc. Below the heatmap are two charts, a pie chart to the left and a bar chart to the right. The pie chart show funding needed and required for the country depending on what parameter is chosen. The bar chart are showing the different needs and requirements of all the parameters as well as the population of the country selected.

Below the parameters we can see the line chart parameters and the line chart itself. All the data is shown every 5 years and includes all the countries in Africa over time. Next to the line chart is a info button where we can see when the data used in the line chart was last updated.
9.2 Reliability of a source

By shedding light on the subject we could clarify what aspects to take into consideration when choosing a specific data source.

Some data might be unnecessary to update at the same rate as others. For example, a yearly update on the size of the population can be more than enough as most countries do not even do yearly population census [Sta19], meanwhile information regarding conflicts and funding required a higher update rate. This was due to the fact that sudden conflicts can alter the situation in a country in a rapid fashion. For instance if a government within a country is deemed to corrupt funding to that might be suspended. In 2018
the United Kingdom suspended all funding to Zambia due to corruption and uncertainty regarding the use of the received funding \cite{BBC19}.

9.3 Visual tools

Our choices of visual tools and graphs were dictated by a few factors. The main focus was to combine the parameters with the most appropriate visualization.

9.3.1 Heatmap

The heatmap displayed gaps by coloring countries in a lighter-shade for lower values and increasingly darker tone for higher values. By selecting a topic of one’s choice on the left, the heatmap is updated to visualize the data with corresponding color, which can be seen in Figure 9. To avoid confusion when it came to countries in which data of a desired topic did not exist, we colored said area grey. This made it easier for the user to separate the countries with a value in the lowest range area to a country without accessible data.

![Heatmap](image)

**Figure 9** The heatmap displaying the gaps within educational funding in Africa with grey areas resembling non-existent data.
9.3.2 Line graph

Since the user may want to check the history within a particular subject we wanted to implement the visual tool that best described a change over time. Furthermore, with sudden dramatic changes, such as a natural disaster, a sole focus on the current situation can give a misleading perspective on the subject. A standard graph with the x-axis covering the time and the wanted information on the y-axis is a visual representation that is common in the everyday life and therefore familiar for the user. We used a line graph to display the trend and disregarded the choice of using an alternative method, such as bars. This was done to better visualize inclines and declines in the trend, whereas a bar graph would contain gaps which could give the impression of missing information. We can in Figure 10 observe that the line graph clearly visualizes the inclines and declines thereby aid the users ability to better grasp the greater picture.

Figure 10 Line graph displaying the trend of number of conflicts in Africa over an interval of five years
9.3.3 Pie chart

Pie charts are useful when you want to visually represent and compare percentages of a whole, in particular with a small set of parameters. In our case it was best used when comparing the ratios within a certain subject. We limited our pie chart to two parameters. With only two parameters taken into account we were able to clearly visualize whether the current funding had reached the needed funding. When observing a pie chart an important aspect is to easily be able to distinguish colors from one another. With an increased amount of parameters we would have been forced to use different shades of certain colors to be able to cover all parameters. By using darker and lighter tones of a color this could lead to confusion for the user regarding which parameter was related to which tone of the color. Due to this fact they have to look several times before fully grasping the chart. An increased amount of parameters also leads to the ratios becoming more difficult to analyze, and in such cases a graph with bars would have been more suitable. We further optimized the chart to display the percentage of the parameter that the user hovered above, as can be seen in figure 11. This was done to provide additional clarification, in particular when the current and needed funding were evenly matched.

Figure 11 The pie chart displaying the ratios of educational funding in Somalia along with percentage and factual amount of the needed funding.
9.3.4 Bar chart

The reasoning behind the use of a bar chart was similar to the choice of the pie chart. The difference being that we wanted to display information about multiple parameters. With a pie chart the user would have a troublesome time to distinguish the ratios when having more than two parameters. By dividing each parameter to a set of two bars the user could easily distinguish the parameters from another as well as observe the ratios between the two bars in the associated parameter. To provide further information we also implemented the specific value of the bar to be displayed once it was hovered over, as can be seen in Figure 12.

![Figure 12](image)

**Figure 12** The bar chart displaying several parameters of funding gaps in Somalia along with the specific value of needed funding for health.

9.3.5 Color choices

The color choices was set to red, resembling the negative values i.e the needed funding, and blue for the positive values i.e the current funding. We refrained from our initial choice of green as the color resembling positive values to adapt to people with color-blindness. By doing so we increased the number of targeted users so that more people could benefit from our website.
9.3.6 Information button

To provide verification of the data we had accessed we also implemented a button that if pressed displayed the latest date the data was updated on our website, which can be seen in Figure 13. This was done in order to increase the credibility of both the data and the website. With credible and regularly updated data we increased the chance of users returning to our website.

Figure 13 Once the button is pressed the date of the latest data update is displayed
9.4 Technical results and discussion

When we tested the database by using the method discussed in Section 7.2 we noticed no anomalies present with the current system. The query time from the database were also satisfactory. This was expected as the current database are quite small but this could change as the database grows.

9.5 Parameters

We can see the parameters we implemented in Figure 8. However, due to the quality and categorization of data in some of the datasets used, some of the data that should be taken into account given a certain parameter were not. Another problem with the some of the datasets was that they were categorized in a different language, i.e. mat(food in Swedish) instead of food, and hence we were unable to categorize it in our database to its relevant area. As most datasets we collected were of good quality this problem was a minor one, and the total gap shown is still correct as these was collected under the "other" parameter.

We also had to take into account that the datasets were stored in different file formats. Thankfully most of them were stored in .CSV or .JSON and were easily imported to our database.

10 Conclusions

The goal of the project was to make a website that was easy to understand with a map that helps highlight the gap between humanitarian aid needed versus funded. This was done by a color range from smallest gap to the largest. That means a country with a smaller gap would have a lighter shade than a country with a larger gap. The user could also set different parameters to display the gap for different needs e.g food, health and education. Different parameters are displayed in different colors.

As our intended end-user where people that might not have the technical knowledge to use a more complex website, say HDX, we hid the technical parts and thereby granted an overview of the current situation to a broader audience.

We felt that we achieved this in a good way visually. However due to the large amount of datasets available on several different websites and the difficulty in making sure that the data stored was correct, we had to trust that these websites did a sufficient quality
control of each dataset available. As such we failed to achieve the goal of us critically assess each dataset, however we only took datasets from organizations that we trusted to be true. Thereby we tried to minimize the unethical and outright wrong data taken into account in the project.

11 Future work

This section includes possible future expansion of the project. Subjects discussed are different ways to visualize data, database structure and more detailed data handling.

11.1 Expanded visualization of data

A big part of the project were to visualize data as a tool to give the user an overarching view of the current aid situation. While this can be done in a multitude of ways we chose to visualize the information through a heatmap, mostly because the original idea were to use a heatmap but also because in our interviews with SIDA it came apparent that they have an extensive use of heatmaps in their analyses and presentations.

Depending on the situation this might not always be the optimal way to visualize it. As such some future work would be in different ways to visualize the information. A good way to further highlight the current aid gap that we talked about doing is to make a cartogram. In a cartogram the size of each country would be distorted depending on the aid gap, i.e. large gap would result in a large country and vice versa. This would further emphasize countries and areas that has a large funding gap.

Another thing which we wanted to implement was to add an icon if a large refugee camp was present within the country. There is already a table column within the mySQL database for each country that could be used to flag if this occur.

11.2 Data handling

We also talked about the possibility to work with sub-national data. The necessary datasets are available on HDX but there would be a lot of datasets from many different organizations to go through and collect. During our early design phase we included sub-national in our design for the website but we quickly noticed that the huge workload would not be feasible within our time frame.
Some additional datasets from other organizations could also be included and the option select which organization the datasets are collected from. When adding this feature you have to be careful not to clutter the website in a way to keep the user-friendly navigation intact.

Another future work is to swap the mySQL database into a noSQL database as it grows bigger. This would result in a better performance once you reach a certain threshold and would also have better scaling capabilities [EDU19]. The queries and speed of data extraction would suffer as a result however so these factors have to be weighed against each other.

References


References


