Supporting humanitarian efforts in Afghanistan: a framework for assessing educational data quality in conflict-affected countries

Ruby Imrie

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2024

Abstract

This report presents the development and evaluation of a data quality assessment framework for educational data in conflict-affected regions, with a focus on Afghanistan. The project aims to address the challenges associated with assessing and presenting the quality of humanitarian and development data, particularly in contexts where data collection is hindered by conflict and instability. Through a comprehensive literature review, interviews with domain experts, and iterative design and evaluation processes, a web-based platform has been developed to provide users with transparent information about the quality of educational data sources. The platform features separate tabs for presenting data and data quality attributes, allowing users to filter and compare data sources based on various criteria. Evaluation sessions with both domain experts and non-experts demonstrated the usability and effectiveness of the platform, while also highlighting areas for improvement and further development. Overall, this project contributes to enhancing transparency, trust, and informed decision-making in the humanitarian and development sector by providing a user-friendly tool for evaluating the quality of educational data in conflict-affected regions.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 469605.
Date when approval was obtained: 2024-01-15
The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Ruby Imrie*)

Acknowledgements

I would like to express my gratitude to my supervisor for their guidance, support and encouragement throughout this project. I am also thankful to my co-supervisor for their insightful feedback and assistance in shaping this work. Additionally, a thank you to the interview participants for generously sharing their expertise and insights, which greatly enriched the research process.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Data plays a crucial role in understanding humanitarian crises, aiding governments, humanitarian organizations and policymakers in decision-making processes. It facilitates more efficient operations, enables accurate decision-making and allows for targeted responses. However, the quality of data can vary depending on factors such as completeness, consistency and accuracy. In conflict-affected regions, data collection faces numerous challenges, including limited access to information (Roll and Swenson [2018]) and inconsistencies in collection methods (Arooje and Burridge [2022]). These challenges are further discussed in Chapter 2. Despite these obstacles, the data remains significant in assessing and addressing humanitarian crises and guiding decision-making processes. Therefore, it is essential for users to be aware of the data's limitations rather than solely focusing on its quality.

The objective of this project was to develop a data quality assessment framework integrated into a Power BI project. This framework enhances transparency and trust between data sources and users by conveying information about data quality. By providing insights into data quality, the project facilitates informed decision-making and reduces the risk of misinterpretations. Understanding data quality empowers humanitarian workers to assess information reliability, minimizing the risk of misguided decisions and resource waste, especially in the context of limited humanitarian aid efforts. Additionally, the project broadens data exploration by visualizing data with varying quality levels, educating users and identifying areas for further research or data collection efforts. It was a priority to consider the user (humanitarian and development workers) when designing the prototype to allow ease of use and to maximise what they gain from the data quality assessment framework.

1.2 Hypothesis

Developing a data quality assessment framework will help convey to users the nature of the data they are looking at; individuals that work with humanitarian and development

data will be able to gauge the quality of the data easily ensuring transparency and trust between the data source's and the user. It will be intuitive to understand.

1.3 Objectives

· **To Research** the relevant literature on data in FCACs (Fragile and Conflict Afflicted Countries) and performed an in-depth review on the history of data in Afghanistan.

· **To Research** current methods and frameworks that are used to assess data quality in the humanitarian sector.

· **Research** key factors of data quality through literature review and interviews with humanitarian and development professionals.

· **Design** the dashboard to effectively visualize data quality for user comprehension.

· **To Implement** the methodology and User Interface (UI) of the data quality assessment.

· **To Test** the system against both experts and general users.

· **To Suggest** direction of future work.

1.4 Existing Project

This project is part of the "Education Data in Conflicted Countries" project, which aims to provide insights into the educational landscape of conflict-affected regions, with a particular focus on Afghanistan. The parent project encompasses several components, including a Power BI dashboard developed as part of an Honours project in 2023. This dashboard features various visualizations aimed at illustrating key trends and patterns in education-related data (Figure C.1). Additionally, an Excel sheet compiled by an academic partner serves as a comprehensive repository of diverse data sources pertaining to Afghanistan, categorized by data type or factors relevant to education.

This project complements the broader initiative by enhancing data quality assessment. While the parent project offers educational data sources and a Power BI dashboard, my focus is on evaluating data quality. Integrating a data quality assessment framework into the dashboard provides users with crucial insights, empowering informed decisions regarding education in conflict-affected regions like Afghanistan.

1.5 Structure

The structure of the report is as follows: Chapter 2 delves into both an in-depth literature review and a background on data visualization. Chapter 3 further explores the Humanitarian Data Quality Frameworks. Chapter 4 investigates the problem by speaking to domain experts. Chapter 5 explains the design choices made. Chapter 6 describes the implementation process of the final prototype. Chapter 7 explores the evaluations done. Chapter 8 suggests future work, summarizes the work done, and presents the conclusions drawn from the findings.

Chapter 2

Background

2.1 What makes good quality data

Data is seen to be of good quality if it is fit for purpose, meaning it meets the requirements for its intended use. While the data being accurate is of upmost importance, there are other factors that contribute to meeting user needs. The quality of data significantly depends on its journey from collection, through various stages, to being utilized, and potential revisions (Hub [2021]). It is important to recognize that data quality is not a one-size-fits-all concept. Good quality will differ depending on the data set and its use. The users standards will change depending on the criticality of the data use and access to data. Typically, data quality is measured with several factors including completeness, consistency, accuracy, timeliness and reliability. These factors are discussed in more detail later on in the paper.

Data is more likely to be better quality, more trustworthy and easier to work with if sound methods have been used throughout the whole process. There are international standards that can be followed through all stages of the process, from requirement collection and identification, design, data collection, processing and analysis to distribution and evaluation (sdm [2016]). Data should be published alongside clear and accessible documentation and metadata (UNICEF [2021]). Metadata is essentially "data about data" with the purpose of describing, defining, or annotating accompanying data. Metadata can be used to describe the structure and attributes of a database including details like file size, creation date and modification date or scientific information such the methodology used, instruments utilized, variables measured and units of measurement (Nadkarni [2011]). Data is used by the majority of organizations including public sector, humanitarian organisations and businesses. Using the right data is essential to creating effective policies, choosing the appropriate response to crisises and making the right business decisions. Data can come in various forms such as numerical statistics, textual reports, images, videos and real-time sensor readings (Organization [2021]).

When working in the humanitarian and development sectors it can often be difficult to obtain high quality data, so lower quality data has to be used which can still be useful as long as the limitations of the data are understood. Organisations that publish humanitarian data must ensure that they are a trusted source of information. An

example being UNICEF now recognises that it cannot rely on its reputation and must demonstrate that the data sets they produce are a reliable source of information. In order to achieve this they have produced a Data Quality Framework providing guidelines when publishing data with the aim to gain public confidence (UNICEF [2021]).

2.2 Data in Fragile and Conflict Afflicted Countries (FCACs)

Collecting data in a conflict afflicted country is not a simple task and can come with several challenges. This is due to the impact conflict can have on a country's information infrastructure, the decrease of human resources and the trauma and distrust individuals are left with.

Geographical areas can be rendered physically inaccessible making it impossible or too large a risk for researchers to enter and collect data. They can also be further restricted by the regulations of their research funder (such as an international donor or a university) or limited by their insurance policy, which may disallow them from travelling to certain areas. Due to the ethical obligation that researchers must avoid and mitigate risk for themselves and research partners and subjects, they are compelled to work in more accessible and safer areas and avoid data collection methods that require travelling. These accessible and safer areas tend to be more secure, built up, well equipped areas and remote rural areas end up being underrepresented. This access bias results in a difficulty in getting a robust sample size (Roll and Swenson [2018]). In addition, researchers may not be given permission to talk with primary informants and female researchers could face discrimination that could prevent them from gaining access. Even when researchers manage to reach participants,they may encounter hesitancy or reluctance stemming from trauma related to the conflict, leading individuals to withhold genuine insights during interviews. This reluctance may compromise the quality of data obtained. Nations can have political reasons for not wanting to share data, to prevent painting themselves in a bad light (Roll and Swenson [2018]).

Data is crucial in conflicted countries as it helps to paint a picture of the problem at hand and reveals the scale. The goal of every humanitarian agency is to provide relief and support to those in immediate need, often in response to natural disasters or conflicts. In the humanitarian sector data is used to make decisions (Data Driven Decision Making (DDDM)) to decide how best to respond to a situation and how to manage resources (Maxwell [2016]). The most common uses for data are operational (planning, coordinating, monitoring, evaluating) and systematic (policy making and advocating) (Buckner et al. [2022]). Often it is difficult obtaining timely data due to the difficulty in gaining the data. This often means that the data available may not be perfect. However, given the critical nature of conflict situations, any available data, regardless of its quality, becomes invaluable. In such circumstances, the urgency of the situation often compels decision-makers to utilize data of varying quality, recognizing that imperfect information is still better than no information at all. Despite this lack of high quality humanitarian data, many humanitarian crises have been responded to successfully, with many lives being improved and saved (Tapia et al. [2013]).

There is a need for better consistency and coordination across humanitarian data so that

organisations, governments, and private sector actors can work together more efficiently and rapidly. Standard data collection, formatting and assessment methods would reduce the risk of unreliable and inaccurate information and would allow direct comparisons of different data sources to be made (Shalash A [2022]). These efforts to improve the quality of the data may lead to some slower humanitarian responses at the start of the standardisation process but in the long term it would result in less time spent collating and analysing information and more time and resources spent on making the appropriate responses (Kirsch TD [2012]). The quality of the finding would be improved as they would be timely (for the Coordination of Humanitarian Affairs [OCHA]).

In recent years, new data collection methods have been innovated including mobile phone interviews, micro-blogged data collection, rapid consumption surveys and satellite images and machine learning algorithms Hoogeveen and Pape [2020]. Due to an increase in technology, data is easier to collect than ever before.

2.3 Education and data in Afghanistan

Due to varying conflicts, political regimes, foreign invasions and cultural shifts that Afghanistan has experienced for thousands of years, the state of education has been unstable throughout the country's history. These instabilities have led to setbacks that directly infringe on the fundamental human right to education (Arooje and Burridge [2022]). Before the change of governance when the Taliban took over in 2021, it had been experiencing a period of growth. Enrollment figures across various educational levels rose, and the gap between male and female students began to close. However, despite it being enforced by law that children between the ages 7-15 should be attending school, there was still a significant number of people in this category that were not attending school. This is exaggerated in remote areas and there are varying influences that impact an individual's access to information; an under-supplied demand of teachers, dangers of travelling, risk of attacks on schools, cultural belief and poverty resulting in children needing to work (Baiza [2013]). Due to the cultural and religious beliefs girls are commonly prohibited from receiving an education; girls make up the majority of the population not receiving an education (Ministry of Education and [UNICEF]). The literacy rates, are already among the lowest globally and will continue to fall without an increase in humanitarian support (UNESCO [2021]).

The general education sector consists of preschool, primary education, and secondary education, which is separated into lower and higher secondary education. Historically, schools are often segregated by gender from around the age of 16 (Baiza [2013]). Financially, the education sector's heavy reliance on international aid, which constituted nearly half of the core budget expenditures in 2020, is a pressing concern. With the recent political changes and potential financial isolation, the sustainability of the progress achieved in the education sector hangs in the balance. The international community's role is now more crucial than ever, not only in terms of financial support but also in ensuring that the right to education, especially for vulnerable groups, remains protected (UNESCO [2021]).

Afghanistan had an Education Management Information System (EMIS) prior to the

recent administrative shift, which was founded in 2005. EMIS served as a tool for collecting and managing education related data with the aim of providing accurate and timely data (EMI [2023]). Although the system had progressed over the years, there remain doubts about its data accuracy. Moreover, due to restricted internet availability and limited familiarity with the system, many individuals couldn't access it, exacerbating the issues of data promptness and thoroughness. A centralised system to monitor the educational progress of children is complicated because of the diverse education institutions, including: government-run schools, private institutions, madrassas (Islamic religious schools) and community-based schools (often backed by NGOs) (Arooje and Burridge [2022]). Consequently, data collection becomes a significant challenge. The statistics that have been gathered are difficult to authenticate due to their fragmented nature, especially in rural regions. This fragmentation often results in data that doesn't truly represent the on-ground situation, primarily because of its incompleteness. The lack of cohesion in the data emerging from the country makes it challenging to paint an accurate and comprehensive picture of the ever-changing educational scenario (Buckner et al. [2022]).

When approaching the numbers, it is crucial to do so with caution as there have been instances where there has been public discrepancies. A former education minister Minister Wardak claimed that 11.5 million children were attending school. In contrast, his successor stated that the actual number was closer to six million. This discrepancy raised concerns of inflated figures, manipulation, and mismanagement within the education ministry (Adili [2017]).

2.4 Previous data quality assessment in humanitarian and development data

As data quality is of upmost importance to anyone that works with it, various sectors (including private businesses) have developed methodologies to assess and ensure the accuracy, reliability, completeness and relevance of their data. As discussed humanitarian data has its own unique challenges, it can draw insights and inspiration from these established practices.

Data quality storytelling has emerged as a novel approach to data quality (DQ) reporting. As highlighted by Ataccama, this method uses findings from profiling and other DQ tools to assess data quality, often visualized through dashboards. These dashboards allow for a comprehensive analysis of data, revealing the root causes and impacts of poor quality data. The dynamic and engaging nature of data quality storytelling replaces the need for a traditional narrator, making the data assessment process more interactive and insightful (Ataccama [2023]). Similarly, Arcadis Gen emphasizes the importance of data visualization in assessing data quality. Their tools are designed for businesses and corporations to identify issues in their data, understand datasets, and transition into modeling in a cost-efficient manner. Data visualization not only aids in understanding data but also in identifying anomalies, missing values, and outliers, which are crucial for maintaining data integrity (Arcadisgen).

EUROSTAT's Handbook on Data Quality Assessment Methods and Tools underlines

the importance of having a clear picture of data quality. It suggests that quality reports should provide insights into user satisfaction, accuracy, timeliness, accessibility, clarity, comparability, and coherence. The handbook also emphasizes the need for both product and process quality indicators, with some being produced regularly and others only when significant changes occur. The challenge remains in developing standard quality indicators that are comparable across different periods and in creating a composite indicator that summarizes various data quality features (Bergdahl et al. [2007]).

Within the humanitarian field, several organisations have released publications on how they assess their data. The World Health Organisation (WHO) have produced the Data Quality Review (DQR) which focuses on assessing the quality of data generated by health facilities' information systems. The aim is to institutionalize a system for routine monitoring, annual reviews, and periodic in-depth assessments. Key parameters for assessment include completeness, internal consistency, external comparison, and external consistency of population data (World Health Organization [2020b]). UNICEF has developed it's Data Quality Framework to ensure the reliability, accuracy, and relevance of the data it collects and distributes. The framework emphasises the importance of having a standard for data collection methodologies, validation processes, and continuous monitoring to ensure that the data is of the highest quality. It also reveals the need for collaboration with local and international partners to enhance data quality and comparability (UNICEF [2021]).

The Humanitarian Data Exchange (HDX) platform, overseen by the UN Centre for Humanitarian Data, has embarked on a project with the Data Nutrition Project team to enhance the quality of humanitarian data sets. The Data Nutrition Project, established in 2018, focuses on enhancing transparency in data sets, taking inspiration from the nutritional information we get on food packets, and aims to create a standard label for data set quality. This initiative aimed to understand data quality perceptions in the humanitarian sector, prototype quality measures, and explore scalability, and work towards automation. It has faced several challenges; developing a scoring method that is not overly simplistic, getting a balance between scalable quantitative measures and comprehensive qualitative measures, communicating quality in a way that encourages better data practices without disincentivizing data sharing and building a quality framework that remains consistent yet adaptable to changing humanitarian. needs. It was concluded that existing quality measures can be utilised, insights from domain experts and third-party validitiors can be very valuable and the main objective of quality metrics is to improve the process of choosing and comparing data (Chmielinski et al. [2023]).

2.5 Humanitarian Data Exchange Websites

Several humanitarian data sharing platforms exist to facilitate the exchange of critical information for aid and development efforts. Among these platforms is the The Humanitarian Data Exchange (HDX), which serves as a central hub for humanitarian data, enabling organizations to share, access, and use data to inform decision-making and response efforts during emergencies and crises. Another notable platform is the International Aid Transparency Initiative (IATI), which promotes transparency and accountability in the aid sector by providing a standardized format for publishing aid

data, making it easier for stakeholders to track and monitor aid flows. Additionally, the UNICEF Data Warehouse offers a comprehensive repository of data related to children's issues, including health, education, and protection, supporting evidence-based policymaking and programmatic interventions. Furthermore, the UNESCO SDG4 Data Explorer focuses specifically on education-related data, providing insights into progress towards Sustainable Development Goal 4 (Quality Education) and supporting efforts to improve education outcomes globally. These platforms play a vital role in facilitating data sharing and collaboration among humanitarian organizations, governments, and other stakeholders, ultimately contributing to humanitarian responses and development efforts.

In addition to facilitating data sharing, these humanitarian data platforms provide valuable metadata about the datasets available on their platforms. For example, the HDX platform offers information such as the source, contributor, time period of the dataset, expected update frequency, location, field names, file types, number of rows, methodology, caveats/comments and tags. Similarly, the International Aid Transparency Initiative (IATI) provides users with metadata including the publisher, IATI identifier, last updated date, planned and actual start and end dates, activity scope, collaboration type, aid type, finance type, flow type and tied status. The UNICEF Data Warehouse provides metadata such as geographic area, indicator, unit of measure, unit multiplier, observation confidentiality, time period related to data collection activities, period for which data are provided and current age. Additionally, the SDG4 Data Explorer offers metadata on data by indicator or country, with options to view in long format, country profiles and regional averages. However, it's worth noting that metadata availability may vary, and for some indicators on the SDG4 Data Explorer, this information may be missing.

Chapter 3

Data Quality Frameworks

The initial phase of research involved reviewing and analyzing existing data quality frameworks and reports from various humanitarian and other sectors. A comprehensive list of organizations and their respective documents can be found in Table D.1. Through this process, I documented how each framework describes and measures data quality, as well as how they assess the quality of the data they handle. Definitions of each data quality principle are provided in Table 4.2. Additionally, Figure 3.1 was created to visualize the variations across organizations and identify common measurement priorities. The differences across the frameworks highlight that there is no one-size-fits-all framework, due to the inherent variations in organizational functions, data processes, and the intended purposes of the data.

The Generic Data Quality Assurance Framework produced by the UN (GSQAF) included the broadest range of data quality principles and influenced several of the other frameworks including UNICEF's data quality frameworks (UNICEF [2021]). The frameworks have the purpose of maintaining data quality within an organisation and improving transparency and trust between data providers and data users.

Figure 3.1 makes visible the variations across frameworks highlighting the variability of data quality in the humanitarian sector due to the precise uses of the data. The data can be used for different purposes such as needs assessment, resource allocation, program monitoring, and impact evaluation. This can also be due to the differences in collection methods. For example, data collected through household surveys may have different quality characteristics compared to administrative data or microblogged data collection. Another consideration is the nature of the data across various sectors within the humanitarian domain, such as health, education, food security, and shelter. For example, WHO focused on Accuracy, Reliability, Completeness, Timeliness and Consistency. Accurate health data is crucial for identifying disease trends, allocating resources effectively, and implementing targeted interventions to mitigate public health risks (World Health Organization [2022a]). Despite the variations among the frameworks, certain factors consistently emerge as critical across multiple frameworks and reports. Specifically, timeliness is identified as a crucial factor in every framework reviewed, emphasizing the importance of up-to-date data in humanitarian contexts. Additionally, accuracy appears as a fundamental consideration in all frameworks except DSEG, underlining

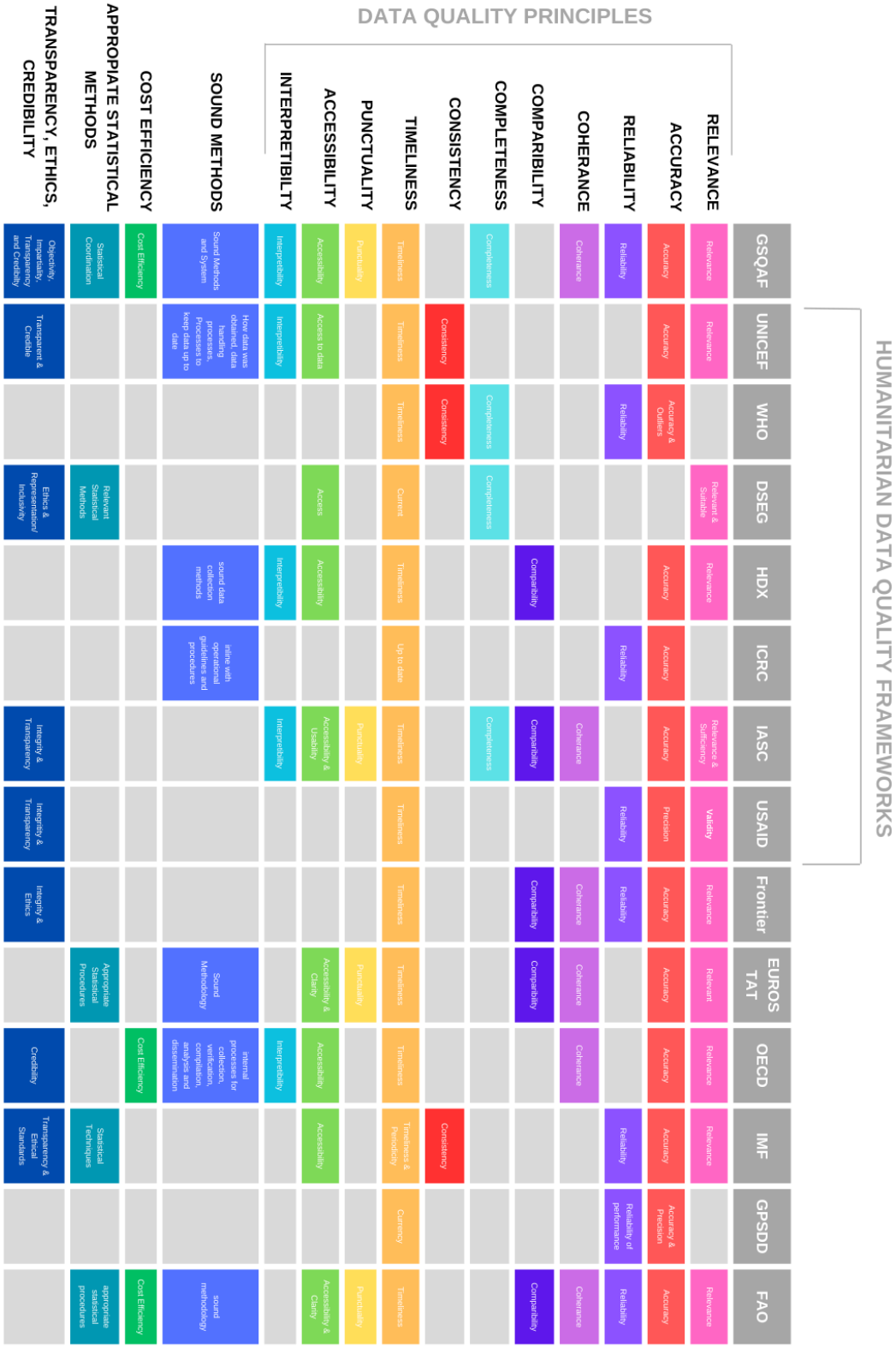


Figure 3.1: Data Quality Principles across Frameworks

its significance in ensuring the that data correctly estimates the situation it is trying to describe.

The definitions for each principle generally align across frameworks, but there are some discrepancies. For example, the GSQAF defines completeness as the ”coverage of the required population,” while UNICEF’s Data Quality Framework describes it as ”a measure of how many of the events or individuals (the ‘records’) in our target population were actually documented.” These statements emphasize slightly varying priorities. The GSQAF focuses on ensuring coverage of the intended population, whereas UNICEF’s framework emphasizes the documentation of events or individuals within that population. While both aim to ensure that data represent the intended population, their focus differs slightly. Similarly, there are variations in the definition of consistency across frameworks. For instance, UNICEF describes consistency as ”in its loosest sense implies the data are ‘at least reconcilable,’” while the WHO refers to consistency in terms of internal and external factors. Internal consistency involves accuracy assessments by comparing source data against reported values and checking consistency over time by comparing data element values at different periods. External consistency includes cross-checks between different data sources with similar information. The main difference lies in the scope of consistency assessment: while UNICEF’s definition focuses on the broad concept of reconcilability, WHO’s framework delineates consistency into internal and external dimensions, encompassing accuracy assessments, comparisons over time, and cross-checks between different data sources. Despite these variations, the overarching goal of ensuring data accuracy and reliability remains consistent across frameworks.

The frameworks together outline that the main contributors to data quality are standardized and sound collection methods, comprehensive documentation, consistent data processing procedures, robust data validation techniques and transparent reporting practices. Additionally, they emphasize the importance of regular data quality assessments, stakeholder engagement, and continuous improvement processes to ensure the reliability, accuracy and usability of the data. For instance, the World Health Organization (WHO) implements a two-stage data quality assessment process comprising a “Discrete desk review of data quality” and a “Site assessment of data quality: data verification and system assessment.” The data lifecycle comprises several stages: data collection, processing, uploading, access and utilization. According to the HDX, different data quality principles align with these stages. For instance, relevance, timeliness, and accuracy are primarily associated with data collection and processing, which are the responsibilities of the data provider. Punctuality in uploading data relies on both the data provider and the data platform. Accessibility, coherence/comparability, and interoperability are associated with data discovery, access and utilization, which falls under the purview of the data publisher (Chmielinski et al. [2023]).

From these findings, I have identified the data quality principles valued important by humanitarian organizations and observed how their differences underscore the complexity of data quality in this sector. The frameworks indicate a shared commitment to data quality among organizations, yet each organization prioritizes and implements data quality assessment differently. This suggests that the identity of the data provider can serve as a significant indicator of data quality when assessing a data source.

Chapter 4

Scoping the Problem

4.1 Interviews

The next stage of my research was to conduct interviews with individuals who work with humanitarian and development data with the aim of learning about how they currently perceive data quality, which quality principles are considered most crucial by individuals in this sector when evaluating data quality in their work and explore the preferences, ideas and expectations regarding a data quality assessment visualisation. I chose to do semi structured interviews meaning that I had prepared a list of open ended question but asked additional follow up questions based on the participants previous answers in a conversational manner (Adams [2015]). The benefits of this are that the participants normally feel more relaxed and that you can gain more detailed insights by asking follow up questions (Martin and Hanington [2012]).

The questions/prompts that I had prepared before the interview were as follows:

1. Please give me a quick background summary of your work.
2. In your work how do you interact with data?
3. When selecting data what do you consider?
4. What does data quality mean to you and what makes good quality data?
5. Rank these data quality principles in order of importance to you. See Table 4.2 for principles.
6. What would you consider timely data?
7. When you are evaluating if data is accurate, what do you look for?
8. What would you find more useful in order to understand the data quality; a scoring system or dashboard of information?
9. If you were to view a dashboard providing information about the data quality of a dataset, what key aspects would you prefer to be presented with?

The participants answers to these questions would prompt me to ask several follow up questions thus providing more insights and a deeper understanding of their expertise and opinions. For question 5, I asked participants to sort cards featuring different data principles from most to least important. This was inspired by the research method of card sorting. However I asked participants to rank the cards in order of importance rather than in to different categories, as it is a good way to discover how participants view and organise concepts (Conrad and Tucker [2019]). Asking participants to think-aloud while ranking the data principles allowed me to hear why they were making their decisions (Nielsen [2012]). For questions 8 and 9, I showed some visuals of existing data quality (Figure G.1) and dashboard systems and some idea sketches that I had prepared to help explain my ideas. These ideas included some scoring systems or data quality dashboards. These are discussed further in Section 5 and shown in Figures 5.1, 5.2 and H.1 .

4.1.1 Interview Strengths and Limitations

As discussed, semi-structured interviews offer numerous advantages in gathering rich and detailed data. However, they come with certain drawbacks that must be considered. One significant limitation is the time-consuming nature of the extensive preparation, conducting the interviews and analyzing the data. Consequently, researchers are often limited to a small number of interviews, impacting the overall sample size and potentially limiting the generalisability of findings. Additionally, semi-structured interviews can pose challenges in standardizing data collection and comparison due to their subjective nature and the varying responses obtained from participants (Adams [2015]). Alongside the time consuming nature, my network outreach was limited to existing contacts and online searches for relevant individuals.

I conducted interviews with four domain experts who specialize in humanitarian and development data, representing a diverse range of backgrounds and expertise.Table 4.1 provides their professions and job descriptions. Participant 1, with an academic and research background, served as both a co-supervisor and lead of the parent project, providing valuable insights tailored to the specific requirements of this project. However, their involvement in the project may have influenced their responses, potentially introducing some bias. Participant 2 contributes a wealth of experience in research, analytics, and program management across various sectors. Participant 3 offers specialized knowledge in gender and education in crisis situations, while Participant 4 brings practical experience from their role as a charity co-founder. Together, this group provides a well-rounded perspective of the challenges and considerations in working with humanitarian and development data, making them a valuable base for understanding the needs of the target audience. However, the interviews lack direct representation from certain stakeholder groups such as government officials, local NGOs, or community leaders who also play crucial roles in data collection and decision-making processes in humanitarian and development contexts.

Participants	Job Description
Participant 1	A PhD student with a research focus on assessing the quality of data related to educational factors in conflict-affected regions. They have a masters degree in global health challenges and have experience working on pro bono projects, including providing support to charities in Southern African countries and volunteering for local community groups. They act as my co-supervisor and the lead of the parent project that this project is part of.
Participant 2	Currently employed as a data scientist at a research organization and holds a senior research and teaching position at a prestigious university. They possess a diverse professional background encompassing research, analytics, and program management in sectors such as public health, humanitarian aid, and community development. They have made substantial contributions to various projects, including those related to COVID-19 response and policy. Additionally, they are involved in leading analytics and software development efforts for a community-focused initiative. Their previous roles include positions at organizations focusing on child protection and emergency response.
Participant 3	Currently employed as a consultant specializing in Gender and Education in Crisis, with a focus on gender-related challenges in education within complex environments. They have significant experience and skills in data analysis and education development. Their prior positions include roles at various organizations, including education-focused entities, where they concentrated on crafting data-driven strategies to address educational issues.
Participant 4	Co-founder and director of a charity. The organization collaborates closely with communities to discover and apply creative solutions aimed at alleviating poverty, enhancing access to education, and improving healthcare services.

Table 4.1: Participants and their job descriptions

4.2 Results

4.2.1 Uses of Humanitarian Data

Participants in the interviews provided insights into the diverse ways data is utilized within their respective roles . Participant 1 emphasized the role of data in storytelling, highlighting how they utilize data to communicate the narrative of the current situation in Afghanistan. Participant 2 mentioned that data is used for analysis across various domains, for example nutritional programs use data for routine monitoring and analysing to make improvements in the audit cycle. Participant 3 uses data in their work on

development projects, particularly in creating global reports focusing on girls' education in conflict and crisis situations, utilizing global datasets for comprehensive analysis. Reports like these are used to influence decision making. The organisation led by Participant 4 has a volunteer who conducts field trips to gather data through surveys. This collected data is essential for overseeing, assessing success, measuring impact, monitoring progress, and managing the overall operation of their project. They also acknowledged that visualising data can be useful from a fundraising angle. Overall, these accounts underscore the multifaceted role of data in addressing various challenges and driving decision-making processes across different domains.

4.2.2 Considerations when selecting data

The participants revealed what they first take into consideration when selecting data. Participant 1 mentioned the importance of the data being up to date. In their case this was extremely pertinent as the data recorded before the Taliban took over in 2021 tells a completely different story. They also mentioned looking at who produced the data but try to remain open minded and not prejudge. Large organizations often aggregate data from numerous countries, which can sometimes result in data that appears to be less up-to-date and relevant due to the sheer volume of information being processed. They emphasised that the date had more importance to them than who produced it.

Participant 3 also works with education data and also stated that the date was vital information to them, for a similar reason in that government policy can change dramatically. Another aspect they look for is who is missing from the data, for example, what schools have been included and excluded. Education providers span a wide spectrum, ranging from formal institutions such as government-run schools, private schools, and international schools to informal education settings like madrasas (Islamic religious schools), community-based schools and home schooling. Additionally, there are non-governmental organizations (NGOs) and community-based organizations (CBOs) that often play significant roles in providing educational services, particularly in remote or underserved areas where formal education infrastructure is lacking. This diverse landscape of education providers contributes to the challenge of standardizing and regulating the education sector effectively. In addition, Participant 3 looks for disagreeing statistics across different data sources. For instance, they shared an example of a project they are involved in, focusing on Nigerian education statistics, where they encountered significant inconsistencies in the reported numbers. Looking at three different government sources, budget enrollment rates vary from 30% to 60% to 90% in one state.

Participant 2's initial approach involves identifying the denominator, which they accomplish either by extracting it from available datasets or by generating it themself. In this context, denominators refer to the baseline numbers or population figures used as the basis for estimating the scale of a particular issue or the target population for an intervention or program.

Participant 4 only interacted with data that was collected by their own organisation so they already had confidence in the data.

4.2.3 Challenges

The participants shared what they find most challenging when sourcing and finding data. The initial challenge is often with finding the relevant data. Data comes from lots of different sources and there is no one place with all the data. Participant 1 remarked that the data can come in lots of different forms such as databases or in reports & presentations where you need to pick out the data points. Another problem they encounter is that there is often no guarantee that you will get an updated version of the database or report. Having updated versions of the data is crucial in order to compare or track changes. As noted by Participant 2, even when aware of the data's location, accessing it can be challenging due to potential restrictions, as some data may be protected. This can be due to political reasons but can also occur when the data is not controversial. Another challenge arises when conflict affects local areas, making it difficult to access these regions and obtain data. As a result, national datasets may not accurately represent the situation in these areas.

Participant 2, in their professional capacity, is affiliated with organizations that have the capacity to conduct their own research if necessary, especially when relevant existing research is not available. However, not all organizations possess the resources to conduct independent research and must rely on existing data. Navigating databases created by others can pose challenges, particularly when variable names are misleading or ambiguous, or when database relationships are complex.

Participant 3 revealed that one of the biggest challenges for them was piecing together incomplete information from different data sources in order to get a complete picture. They also reiterated the importance of knowing the date of the data , mentioning that there are now models that can predict the future, but if based on data that preceded government policy changes or pre-Taliban then this data is useless. Often, the approval processes implemented by organizations supplying humanitarian data are excessively burdensome, resulting in significant delays between data collection and release. Consequently, by the time the data becomes available, its relevance and timeliness may have significantly diminished.

Another difficulty is working out who/what has been included in the data. As discussed earlier, there is a plethora of educational providers in Afghanistan and lots of data sources don't invlude them all. For example, EMIS doesn't cover the informal education providers. Administrative enrollment data can be very politicized so the numbers can be either considerably over or under representative of the actual numbers of children you would see in the class room. Often school grant systems are based on enrollment numbers, which encourages schools to over report their numbers. Conversely, in regions like Somaliland, where there is a prevalence of private education, schools are taxed based on enrollment figures. As a result, these schools may underreport their enrollment numbers to minimize taxation burdens. House-hold surveys can be an effective method to get information about informal education providers however not all areas get included in household surveys due to security concerns. In addition sometimes the data is not fully representative as it may not include refugees, Internally Displaced Persons (IDP's) or take into consideration non traditional families. Participant 3 noted that in Afghanistan disabled children often won't be included in the surveys as families don't always see

them as children. Thus, any data sources on children with disabilities tends to be under reported. The reliability of displacement data can also be very hard to assess as there is uncertainty surrounding the circumstances of when someone identifies as an IDP. In certain scenarios, the availability of additional funding may incentivize individuals to provide entries. However, displaced individuals may also hesitate to disclose their place of origin due to concerns about being perceived as anti-government, particularly in cases such as persons fleeing Darfur. Therefore, the participant tries to look at both kinds of data sources to cover as much ground as possible. Participant 4 noted that national-level data often presents a different narrative than what is observed when visiting specific areas within a country.

4.2.4 Data Quality Principle Rankings

The next discussion in the interviews was to gauge how participants perceive data quality and the data quality principles. This is where they were asked to rank the data principles in order of importance to them. To do this I used an OpinionX survey which computes a combined ranking based on each participants results. It is important to note that participants emphasized their ranking of certain principles being lower down does not imply a lack of concern for them. Rather, it reflects their recognition of the challenges posed by these factors and their ability to navigate around these difficulties in their work.

Combined Ranking of Data Principles by Participants

1. Accuracy
2. Relevance
3. Accessibility
4. Timeliness
5. Reliability
6. Comparibility
7. Completeness
8. Coherence
9. Consistency
10. Interpretability

Whilst ranking the data quality, the participants made comments on how they regard these data quality principles and how they assess these data quality principles in their own work. These comments are summarised in Table 4.2. Some participants made comments about principles that weren't included in this list. Participant 3 said that transparency was also important, noting that it is okay if a data source isn't representative of a whole situation, for example, doesn't include every geographical province, or every type of school provider, as long as it is open about who it does and doesn't represent.

The inclusion of metadata can help ensure transparency and can set out what is included in the data.

4.2.5 Scoring System vs Dashboard

When shown some previous data quality visualisations and some sketches of my ideas, there was an agreement across participants that they would find a dashboard more useful than a scoring system. Participant 1 mentioned that although a scoring system is a good visual, they liked that a dashboard allowed the user to make their own decision. As humanitarian data differs so much across fields and countries it would be difficult to develop a scoring system that was representative of these differences. Participant 2 similarly said that the dashboard provides some judgement but not a definitive *good* or *bad*. It makes the user aware of the limitations of the data whilst giving credit to the audiences own knowledge and opinions. Participant 3 concurred that you have to take a pragmatic approach with humanitarian data and take what you can get. A bad scoring would encourage users to dismiss the data when in fact there would still likely be value in the data. Participant 1 had concerns for this reason with using colour coding to indicate good and bad within the dashboard as this has a lot of prejudgement attached. Participant 4 was also in agreement stating that having all this information about a humanitarian data source in one dashboard would be of immense value.

4.2.6 Ideas & Expectations

Participants discussed their desires when it came to a data quality visualisation. Participant 1 said the key piece of information that they would like to be able to find quickly was the dates that the data was representing. They also thought it would be useful to know the completeness and coverage in terms of which provinces, age groupings and data fields are represented in the data. It would also be useful to highlight gaps in years or data, as this can make it hard for trend analysis. They also wanted to be able to see the source of the data. Coverage, date and data source were also mentioned by Participant 2. One of the other key points for Participant 2 was knowing whether the data was sex and disability aggregated. They also want to be able to quickly find the age group and who completed the survey (e.g. head of house hold or head teacher) . They also mentioned comparability, highlighting that it would be useful to get an insight into the data fields as attendance data can't be compared to enrollment data. Again, participant 3 mentioned they would like to be able to see the source of the data. They voiced that it would be very useful to know if the data was consistent with other data sources or had conflicts.

Table 4.2: Comparison of Data Quality Principles between Framework Definitions and Domain Experts' Perceptions

Data Quality Principle	**Principle Definition taken from Frameworks**	**How Domain Experts Gauge This**
Relevance	"The relevance of a data product is the degree to which the data serve to address the purposes for which they are sought by users.	Whether they can get the answers they are looking for from the data. They all explained that when selecting data this is something they consider automatically without really thinking about" (GDQ [2015]).
Accuracy	"The accuracy of a data product is the degree to which the data correctly estimate or describe the quantities or characteristics they are designed to measure." (GDQ [2015]).	How true the data is to the event that is occurring. In order to gauge the accuracy they looked at: the data sources, the collection method. They would use triangulation to check if the data was consistent with other sources.
Reliability	"The reliability of data is the closeness of the initially released values to subsequently revised values that are released.	This was viewed the same as accuracy by domain experts" (GDQ [2015]).
Coherence	"The coherence of a data product reflects the degree to which it is logically connected and mutually consistent with other data products." (GDQ [2015]).	Coherence impacts comparability of different data sources. It was mentioned that coherence can usually be mitigated but makes the data analysis process a lot harder.
Comparability	Not defined in Frameworks.	"Being able to compare across different data sources. Factors like using a standard age range aligning to official school years contributes to this. If you can't compare data it is pretty useless as a statistic on its own doesn't tell you anything."
Completeness	"This is a measure of how many of the events or individuals (the 'records') in our target population were actually documented." (UNICEF [2021]).	Domain experts would evaluate the data based on factors such as the number of sources, the extent of geographical coverage, and the time span covered. They would also scrutinize what is included as well as what is excluded from the dataset.

Data Quality Principle	Frameworks Specification	How Domain Experts Gauge This
Consistency	"Internal consistency – accuracy is examined by recounting the source data and checking against the value to be reported; consistency over time is examined by comparing the value of a data element to the value of the same data element at earlier time periods; External consistency - including cross-checks between different data sources with the same/similar information." (World Health Organization [2020a]).	Whether a data source tells a consistent story and experts also check if the data source is consistent with other data sources (triangulation). Using consistent age groups and data fields was also mentioned.
Timeliness	"The timeliness of a data product is the length of time between its availability and the event or phenomenon it describes." (GDQ [2015]).	How up to date the data is. This was of great importance to experts as the situation is constantly changing in conflicted countries. Therefore untimely data can lead to outdated analysis. The length of time that they considered timely was very dependent on context.
Punctuality	"The punctuality of a data product implies the existence of and adherence to a data product dissemination schedule." (GDQ [2015]).	Domain experts were unsure how this differed from Timeliness. After hearing the definition they revealed they don't believe data is ever punctual.
Accessibility	"The accessibility of a data product reflects how readily the data can be discovered, located and accessed from within the Agency data holdings." (GDQ [2015]).	How easy it was to find/get the data. One domain expert mentioned that it can sometimes be hard because some data is protected this can sometimes be for political reasons, but often the data is not even controversial.
Interpretability	"The interpretability (sometimes called clarity) of a data product reflects the ease with which users can understand and properly use the data." (GDQ [2015]).	How easy the data is to understand. It can often be confusing to understand a database that someone else has created.

4.3 Findings

The interviews with participants shed light on the diverse applications of humanitarian data within their respective roles. From storytelling and program analysis to development project management, data serves as a critical tool for understanding complex situations, monitoring progress, and influencing decision-making processes.

When selecting data, participants highlighted several key considerations. These included

the importance of data timeliness, relevance, and accuracy, particularly in contexts of rapid change or policy shifts. They also emphasized the significance of understanding the context of data production and the need for comprehensive coverage and consistency across different sources. Challenges such as data accessibility, completeness, and navigating complex databases were also noted.

Participants identified several challenges in selecting and accessing humanitarian data. These included difficulties in finding relevant data sources, accessing updated versions of data, and navigating complex databases. Additionally, issues such as data protection, incomplete information, and discrepancies across sources posed significant challenges to data analysis and interpretation. The participants highlighted the importance of transparency, metadata, and comprehensive coverage in addressing these challenges and ensuring the reliability of humanitarian data.

Through the interviews, participants ranked data quality principles based on their perceived importance. Accuracy, relevance, and accessibility emerged as the top-ranked principles, reflecting their critical role in informing decision-making processes. Timeliness and reliability followed closely, highlighting the importance of up-to-date and trustworthy data in humanitarian contexts. While comparability, completeness, coherence, consistency and interpretability weren't ranked as highly, participants acknowledged their importance but believed that they could be worked around.

Participants expressed a preference for data visualization through dashboards over scoring systems. They believed that dashboards provided more flexibility and allowed users to make informed decisions based on their own knowledge and perspectives. Concerns were raised regarding the potential biases associated with scoring systems and the importance of acknowledging the limitations of available data. A consensus was reached on the value of consolidating humanitarian data into a comprehensive dashboard for enhanced accessibility and usability. Participants shared their desires for data quality visualization, highlighting the importance of key information such as data dates, completeness, coverage, and data sources. They emphasized the need for clear indicators of data consistency and conflicts, as well as features enabling quick comparisons and insights into data fields. Participants underscored the need for user-friendly interfaces and comprehensive metadata for facilitating effective data analysis and decision-making processes.

From these findings, I have identified a clear need for improved availability of data quality metrics. The system I have developed in response aims to streamline the selection of humanitarian data sources by providing a collection of data sources in one website accompanied by comprehensive insights into key factors such as accuracy, relevance, and accessibility for each data source. By providing humanitarian workers with the information that they normally spend valuable time searching for, the system empowers them to make informed decisions efficiently, ultimately facilitating them to use their time more effectively. By aggregating data quality metrics into a user-friendly dashboard, users will be empowered to make informed decisions based on their unique requirements and contexts. This approach addresses concerns raised by participants regarding the limitations of scoring systems and the importance of acknowledging data quality limitations.

Furthermore, the system will be accessible through links embedded within visualizations created in Power BI. This integration enables users of the Power BI platform to gain deeper insights into the data's origins, enhancing transparency and enabling them to consider data quality when interpreting the visualizations.

Overall, the system represents a step towards enhancing data quality governance in humanitarian contexts. Enabling stakeholders to access data quality metrics will encourage informed decision-making and foster greater trust in the data used to address critical humanitarian challenges.

From the interviews, I have gathered and refined the requirements of what my system should be expected to do.

4.3.1 User Requirements

1. When looking at the Power BI the user should be able to find a dashboard displaying information about the data quality of the source used to create the visualisation.
2. When viewing the dashboard, the user should be able to find or access:
   (a) The date the data is representing, the date the data was published and how often the data will be repeated.
   (b) Links to different editions of the data (from different years/ time periods).
   (c) The data owner, source and collection methods and processes and quality checks.
   (d) Whether the data is internally consistent/conflicting with other data sources.
   (e) The completeness and geographical coverage of the data.
   (f) The format of the data.
   (g) The confidence interval and margin of error, if available.
   (h) Details about the data, such as whether it is aggregated by sex, age ranges covered, and the scope of inclusion, including enrollment and attendance data, are provided.
   (i) Whether the data is publicly available and where it can be accessed from.
   (j) A summary of the contents of the data source.
3. From this dashboard the user should be able to make their own informed decision on the quality of the data.

Chapter 5

Designing the Solution

5.1 Wireframing & Idea Gathering

Throughout my design process, I've adopted a highly iterative approach, embracing numerous changes along the way to achieve the best possible outcomes. I found that working iteratively allowed me to refine my ideas continuously, incorporating valuable feedback and insights as I progressed.

Wireframing became an integral part of my design workflow, serving as a foundation for testing out different concepts and guiding my implementation decisions. By creating wireframes, I was able to visually explore various layout options, experiment with different user interface elements, and map out the user journey effectively. This process helped me make informed design choices that aligned closely with project goals and user needs.

One of the key advantages of this iterative approach was its ease of communication with my academic partners, supervisors and interview participants. Sharing wireframes throughout the process allowed me to gather valuable feedback from them, enabling me to make necessary adjustments and refinements before proceeding further.

When designing the solution, I prioritized user requirements and adhered to standard design principles, including Nielsen's 10 Usability Heuristics for User Interface Design (Nielsen [1994]) and the Gestalt Principles (Interaction Design Foundation - IxDF [2016, August 30]), to ensure an optimal user experience.

5.1.1 Scoring vs Dashboard

My initial design idea was to have an overall score for the data that would appear with each visualisation within the Power Bi dashboard. This score would be indicted by a coloured bar representing the data score. Then, when the user would hover over this, they could see how the data scored across categories (see Figure 5.1). This was inspired by some of the data quality assessments that business provide, as seen in Figure G.1 in the Appendix. However, from the interviews there was a much stronger preference towards a dashboard of data quality information.

(a) Scoring System Idea

(b) Scoring System Idea in Power BI

Figure 5.1: Scoring System Idea

Therefore I moved towards the idea of having a dashboard that would present all the information to the user for them to assess the data quality themselves. Wireframes of these designs can be seen in Figure 5.2.

5.2 Power BI vs External Website

I considered two options for presenting the dashboard: embedding it as a pop-up within Power BI or hosting it on an external website. While embedding the dashboard within Power BI had the advantage of consolidating all information in one platform, I ultimately chose to develop an external website. This decision was motivated by several factors.

Firstly, hosting the dashboard on an external website allowed for greater flexibility and customization in displaying information. This realization of limitations of Power BI stemmed from experimentation with Power BI and reading of the documentation (Microsoft). This flexibility of an external website enabled me to incorporate additional details and interactive elements, enhancing the overall user experience and allowing more information to be given. Secondly, by centralizing all data sources within a single website, I addressed a common challenge raised during the interviews. Users expressed the need for a unified platform where they could access, search, filter, and compare various data sources conveniently. Hosting the data sources on a designated website fulfilled this requirement, transforming the website into a comprehensive data selection tool for humanitarian professionals. However, one potential disadvantage of hosting the dashboard on an external website is the requirement for users to navigate away from their familiar Power BI environment, which may be an inconvenience for some users.

5.3 Design Choices

This section outlines the design choices I made and the reasons behind them. For the best user experience, it is recommended to view the website in full screen on your device. You can access the most up-to-date version of the site here: https://rubyimrie.github.io/education-data/, or view screenshots in Figure 5.3.

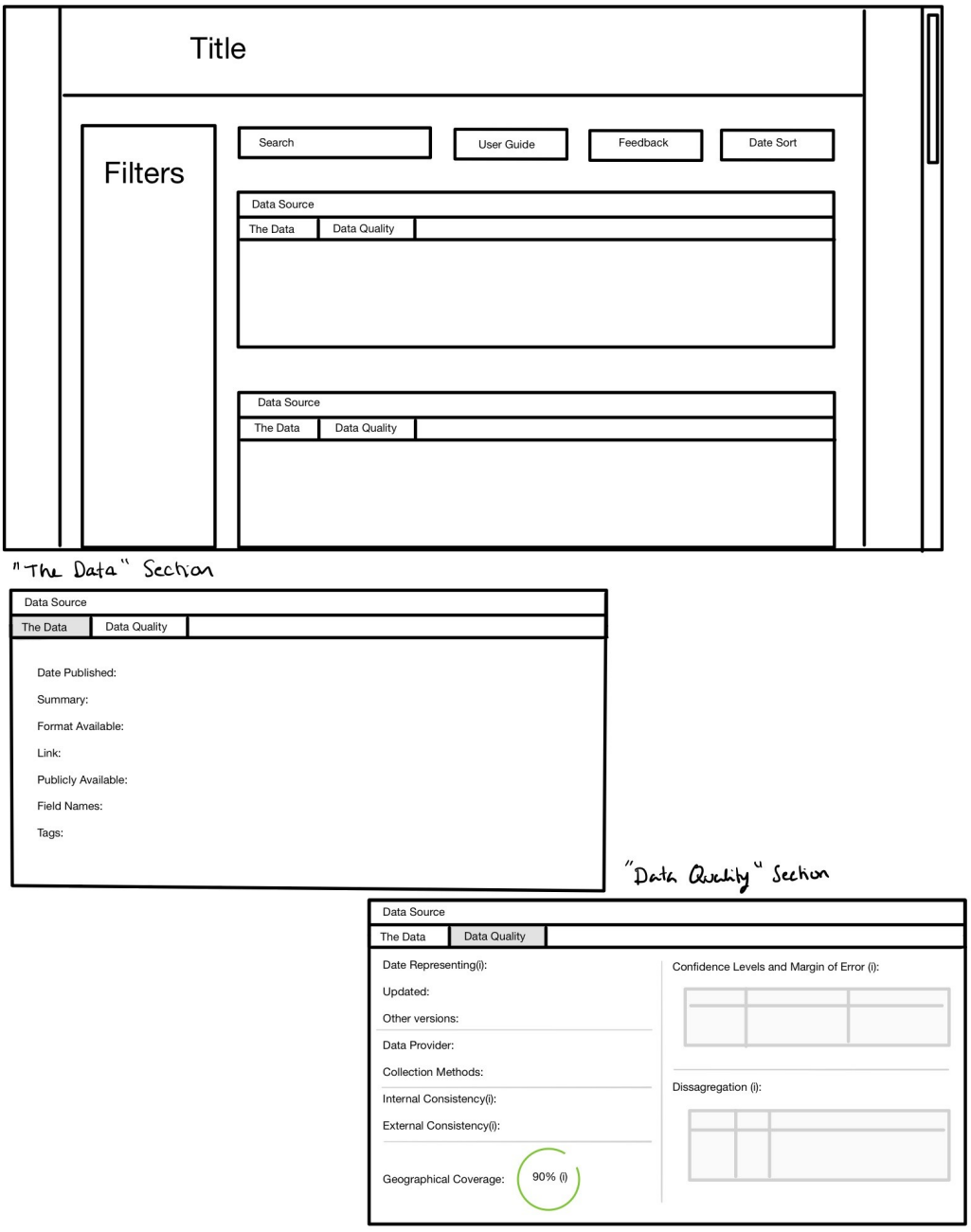
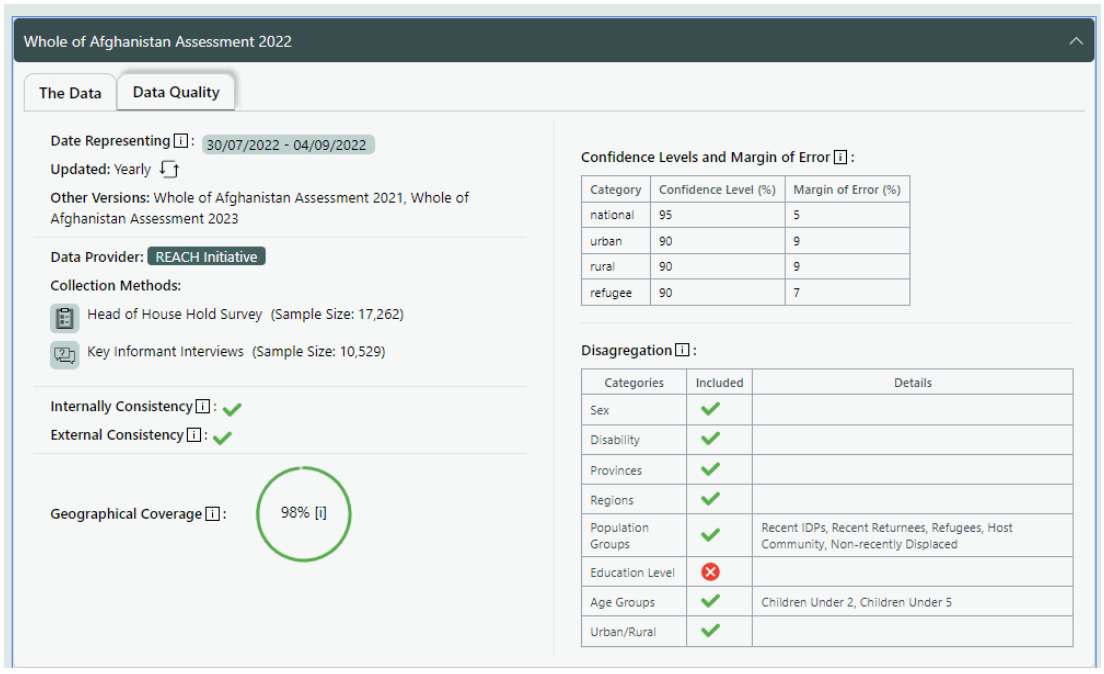


Figure 5.2: Dashboard Wireframes

(a) Website with Data Info Tab Selected



(b) Website with Data Info Tab Selected

Figure 5.3: Screenshots of Website

5.3.1 Separating “The Data” and “Data Quality”

To streamline the presentation and reduce clutter, I opted to separate the information into two tabs. The “The Data” tab includes details such as Date Published, Summary, Format Available, Link, Publicly Available, Field Names and Tags. Conversely, the “Data Quality” tab focuses on Date Representing, Updated, Other Versions, Data Provider, Collection Methods, Internal and External Consistency, Geographical Coverage, Confidence Levels, Margins of Error and Disaggregation. This division ensures that users can easily access and navigate relevant information based on their specific requirements. It also aligns with Nielsen’s heuristic of ”Match between system and the real world,” (Nielsen [1994]) as the 2 tabs mirror the design of file tabs. Moreover, this separation adheres to the Gestalt principle of Common Region, which suggests that elements that are enclosed within the same visual boundary are perceived as related or belonging together (Interaction Design Foundation - IxDF [2016, August 30]).

5.3.2 Internal and External Consistency

Participants of the interviews mentioned that they use Triangulation to gauge how accurate the data is. Triangulation is a research method commonly used in qualitative studies to enhance the credibility and validity of findings by cross-verifying information from multiple sources or methods. It involves the use of multiple data sources, data collection techniques, researchers, or theories to converge on a more comprehensive understanding of a situation. By triangulating data, researchers can strengthen the reliability and trustworthiness of their conclusions, as discrepancies or inconsistencies across various sources can be examined and resolved, leading to a more nuanced and robust interpretation of the research findings (Carter et al. [2019]).

By checking if the data is internally consistent and consistent with other sources, this can be presented to the user so they don’t have to do these checks by themselves. Of course it can’t be guaranteed that it is not conflicting with all other existing data sources but can compare to the other data sources on the website.

5.3.3 Geographical Coverage

During the interviews, it became evident that understanding the geographical coverage of the research was crucial. To address this, a pie chart has been implemented to visually represent the percentage of provinces, regions, or cities covered by each data source. Users can hover over the chart segments to glean additional insights into the areas not included and the reasons behind their exclusion. This provides a deeper understanding of the data’s scope and limitations.

5.3.4 Filters, Searching & Sorting

The website incorporates a comprehensive filter section, positioned on the left-hand side of the page. This filter section is sticky, meaning it stays in the same place while scrolling through data sources, and it features collapsible sections for improved navigation and space optimization in accordance with standard design practices Vassilatos and

Crawshaw [2023]. This intuitive placement ensures easy access and seamless navigation for users but does have the disadvantage of using a lot of screen space. The filters are in numerical and alphabetical order to ensure it is easy to find what the user is looking for. By offering a wide range of filtering options, users can precisely refine their search criteria to find the most **relevant** data sources.

The filter options are categorized into key parameters, including:

- Year: Users can filter data sources based on the year of data collection or publication, allowing for temporal specificity and the finding of **timely** data.
- Scope: This category enables users to narrow down their search based on the scope or geographical **coverage** of the data.
- Collection Method: Users can select data sources based on the method used for data collection, such as surveys, administrative records, or other mechanisms.
- Public Availability: Users have the option to filter data sources based on their availability to the public, ensuring **transparency** and **accessibility**.
- Data Provider: This filter allows users to identify data sources based on the organization responsible for providing the data.
- Update Frequency: Users can filter data sources based on how frequently they are updated, ensuring access to the most current information.
- Disaggregates By: This category enables users to filter data sources based on specific disaggregation parameters, such as demographic groups or geographic regions.
- Tags: Users can filter data sources based on **relevant** tags or keywords associated with the dataset, facilitating targeted searches.

The inclusion of a reset button streamlines the filtering process, allowing users to easily revert to the default settings without the need to manually deselect their selections. This feature enhances user experience by providing a convenient mechanism to refine search criteria or explore different filtering combinations.

Additionally, the website features a search bar positioned at the top left corner, facilitating quick and direct access for users who already have a specific data source in mind. Users can simply enter their search query, and if a matching data source is found, it is promptly displayed. In cases where no matching results are found, a message indicating this is presented, ensuring transparency and guiding users accordingly.

Furthermore, during user interviews, the **timeliness** of the data emerged as a recurring theme. To address this concern, the data sources are initially sorted with the newest data displayed first. However, recognizing the diverse preferences of users, this sorting criterion can be easily customized to display the oldest data first if preferred by the user. This flexibility is provided through the ”Sort By” selection located at the top right corner of the page, enabling users to tailor their browsing experience according to their specific needs and preferences.

5.3.5 Icons

Icons are incorporated throughout the website, for example, when indicating different collection methods, as they adhere to Nielsen's design heuristic of "recognition rather than recall." By utilizing icons, users can quickly recognize and understand the meaning associated with each method without relying solely on textual descriptions. This enhances usability and reduces cognitive load, allowing for a more intuitive and efficient user experience. Additionally, icons add visual interest and clarity to the interface, contributing to the overall aesthetics and user engagement of the website.

5.3.6 User Guide

A User Guide (see I) page has been incorporated into the website to assist end users in maximizing their experience and benefits from the platform. This practice aligns with good software development principles by ensuring users have clear instructions on navigating the site's features and functionalities to optimize their usage and achieve their objectives efficiently. Furthermore, integrating a User Guide page adheres to Nielsen's Heuristic 10 - Help and Documentation, which emphasizes the importance of providing adequate support and resources for users to resolve issues and answer questions independently, enhancing the overall usability and user satisfaction of the website.

5.3.7 Feedback Page

A feedback page is included enabling users to provide valuable input, suggestions, and report any errors they encounter in the information presented. This feature aims to foster user engagement, improve content accuracy, and enhance the overall user experience by addressing user concerns and continuously refining the platform based on user feedback.

5.3.8 Summary and Tags

The summary and tags are essential components that enable users to determine the **relevance** of the data. The summary provides a concise overview of the contents of the data source, while the tags display key terms associated with the data. These tags function as buttons; when clicked, they filter and display only the data sources that share the same tag. This feature allows users to easily find and **compare** other relevant data sources, enhancing their ability to make informed decisions. When hovering over the tags, the colour darkens to indicate that this is a clickable element. The tags in the filters section have the same design to indicate that they will perform the same function, aligning with Gestalt principles that suggest elements perceived as similar are often grouped together mentally by users (Interaction Design Foundation - IxDF [2016, August 30]). This design was inspired by HDX's data sharing platform.

5.3.9 Disaggregation

Participants highlighted the importance of understanding the disaggregation criteria, as it delineates the analytical possibilities of the data, thus enhancing humanitarian response effectiveness (Benelli et al. [2012]). Understanding disaggregation also aids users in assessing the **relevance** of the data source. To facilitate this understanding, A table presenting potential disaggregation topics in included, annotated with either a tick or cross. This visual cue enables users to quickly ascertain the disaggregation status. The topics encompass Sex, Disability, Provinces, Regions, Population Groups, Recent IDPs, Education Level, Age Groups, and Urban/Rural. Additional information is provided where necessary, such as for population groups, age groups, and education levels.

5.3.10 Confidence Levels and Margin of Error

Including information on confidence levels and margins of error in the dashboard provides users with crucial insights into the **reliability** and **accuracy** of the data presented. By understanding the level of confidence associated with the data estimates and the range within which the true value is likely to lie, users can make more informed decisions and interpretations (Hunter [2016]). This information enhances the **transparency** and **credibility** of the data, enabling users to assess its trustworthiness and suitability for their specific needs.

5.3.11 Hoverable Information Icons for Definitions

Through my research outlined in Chapters 3 and 4, discrepancies emerged between the interpretations of data quality principles outlined in humanitarian data quality frameworks and the perceptions of humanitarian workers. To mitigate potential miscommunications on the site, I opted to provide additional explanations for these terms to clarify their intended meanings. These explanations are presented as tooltips, accessible upon hovering over the respective terms, thus avoiding clutter on the page while enabling users to easily access supplementary information. This design approach is in line with Heuristic 8: Aesthetic and Minimalist Design, from Nielsen's 10 usability heuristics, which emphasizes the importance of maintaining a visually pleasing and uncluttered interface (Nielsen [1994]). Hovering, a standard interaction method, allows users to discover more details about each concept (Mantzios [2023]). An (i) is used to indicate there is additional information about a field.

5.3.12 Colour Choices

I used a Tailwind color generator to pick a colour scheme in order to maintain consistency and coherence in the color scheme of the website, ensuring a visually appealing and professional appearance. By harmonizing colors, the website not only looks aesthetically pleasing but also conveys a sense of cohesion and professionalism to the users. Specifically, I opted for a blue color theme to align with the existing Power BI dashboard system, fostering a seamless and unified user experience across different platforms and tools.

5.3.13 Integration with Power BI

For access to the website from the Power BI Dashboard, I opted to utilize the standard information button provided by Power BI. Users are familiar with this button as a means to access additional information (see Figure 5.4). Upon hovering over the button, a message prompts users to click for more details about the data source and data quality. When clicking the link they will be taken to that data source on the website. This integration ensures a user-friendly experience and facilitates easy navigation between the dashboard and the website.

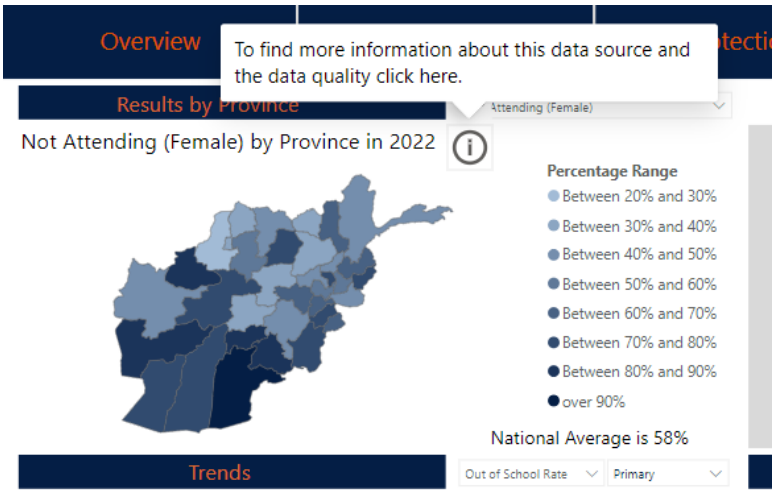


Figure 5.4: Accessing the website from the Power BI Dashboard

5.4 Limitations of Design

Not all data sources may provide the necessary information to populate the dashboard comprehensively, which could affect its utility and completeness. Additionally, there are certain design decisions that warrant further refinement, such as the use of background colors to emphasize certain information. While this visual cue is intended to draw attention to important details, there's a risk that users may mistake these colored elements for interactive buttons, potentially leading to confusion. This issue will be monitored during testing to ensure optimal user understanding and experience. Further attention is required to ensure that the website is accessible to individuals with diverse needs, including those with color blindness or poor eyesight, emphasizing the importance of incorporating accessibility features to promote inclusivity and usability for all users.

Chapter 6

Implementing the Final Prototype

6.1 Vue JavaScript Framework

I chose to implement the website using the JavaScript framework Vue. Given my prior experience with Vue in previous projects, I was already familiar with it and it allowed me to create reusable components, facilitating a more efficient development process. Vue's component-based architecture streamlined the organization of my codebase, enhancing maintainability and scalability (Saks [2019]). Vue has plentiful documentation which provided comprehensive guidance and support throughout the development process, ensuring smooth implementation and troubleshooting whenever needed (Vue.js [n.d.]). Additionally, Vue's seamless integration with Tailwind CSS offered extensive styling capabilities while maintaining code clarity.

6.2 Github

I used Github when developing the website. This was because it allowed me to version control my code, track changes and manage project tasks effectively using features like branches and pull requests. I then used Github pages to host the website as it provides free hosting for static websites directly from the Github repository, simplifying the deployment process and enabling easy sharing and access to the website without the need for additional hosting services (git).

6.3 Finding the Information

I chose the first data source to include from the Power BI system - "Whole of Afghanistan Assessment 2022" produced by Inter-Cluster Coordination Team, Reach Initiative [2022] . I gathered the necessary information from an accompanying report to the data set and information on the HDX platform. For internal conflicts I refer to conflicts discovered during the creation of the Power BI Dashboard. Before testing with users I added some other mock data sources in order to fully test the functionality of filtering and searching. These are indicted by a "(mock)" in the title of the data source.

Chapter 7

Evaluation

7.1 Experts

I conducted user testing sessions with two experts, using screen sharing to observe how they performed tasks and interacted with the system. After completing the tasks, I held semi structured discussions with them to gather their insights and opinions. The tasks included filtering to locate a specific data source, accessing information about the data quality of that particular source, comprehending the provided information, and resetting the filters. Additionally, participants were given time to explore the website autonomously. Additionally, I used the NASA Task Load Index to measure their cognitive workload during the sessions. It's important to note that this part of the evaluation was limited due to time constraints and available contacts so it involved only two expert participants.

The experts who participated in the testing sessions were both individuals who had previously been involved in the earlier interviews. In this section they will be referred to as Expert 1 (Participant 1 in 4.1) and Expert 2 (Participant 3 in Table 4.1). Expert 1 serves as an academic partner leading the parent project. I conducted Expert 1's session at an earlier stage to address any issues or follow up on suggestions, allowing for iterative improvements. This approach proved beneficial as it enabled me to gather more diverse feedback and enhance the project iteratively, avoiding redundancy in addressing recurring issues.

During our initial discussion, I asked them about what they thought the purpose of the website was. Expert 1 expressed uncertainty, indicating that, without prior knowledge, they wouldn't be immediately aware of its purpose. In response, I included a description section to address this ambiguity. When I asked Expert 2 during the later session (after these changes had been made) they immediately understood and said, "it provides data on education in Afghanistan" Next, I requested that they filter the data by the provider REACH initiative and represent the year 2022. Both participants successfully completed the filtering task. Then I instructed them to locate information about the data quality. Expert 1 hesitated initially, taking a moment before realizing that the "Data Quality" label was clickable. To address this usability issue, I improved the design to make it more intuitive. In the subsequent session following these changes, Expert 2 promptly

clicked on the Data Quality tab.

When examining the Data Quality and The Data dashboard, both experts expressed its usefulness. Expert 1 mentioned that it streamlines the process of assessing confidence in a new data source as it saves them from finding that information themselves, while Expert 2 particularly appreciated the clarity and ease of understanding of the Disaggregation table. Geographical coverage also caught their attention as notably beneficial, and Expert 1 emphasized the importance of prominently highlighting the represented date, as it is typically the first piece of information they seek. Additionally, Expert 1 found the tags feature helpful for discovering similar data sources, although Expert 2 suggested renaming it to "Indicators" for better alignment with the Sustainable Development Goals (SDG) Indicators, thus enhancing its utility as a trusted resource. Expert 1 also noted the importance of knowing who is responsible for determining confidence levels, margins of error, and external/internal consistency, whether it was me (the creator of the website) or the data provider. To address this, I incorporated this information into hoverable tooltips. It was noted that displaying the Data Provider was helpful as it allowed users to make their own judgment regarding the trustworthiness of the organization. However, Expert 2 pointed out that they were not familiar with all the organizations listed, suggesting that an external link to the organization's page would be beneficial for further clarification and context. Overall, they found the displayed information to be presented in a clear, structured manner that was easy to comprehend. They appreciated that it provided them with the necessary information enabling them to make their own judgment regarding the usability of the data source.

Both users were able to reset the filters easily. When asked if they would like any additional filters included, they appreciated the ability to filter the existing options. Expert 1 suggested that expanding the filtering options to include additional data quality characteristics, such as geographical coverage and internal and external consistency, would enhance the ability to identify other data sources of similar quality.

In their closing remarks, both experts expressed that the website was a valuable resource that they could see themselves using. They found it intuitive overall. Expert 1 praised the user interface, stating, "it looks good, it's neat, it's structured well." Expert 2 particularly liked the calming colors and remarked that the design was business-like and wasn't overly in your face. Expert 1 envisioned the website evolving into something larger, fitting well within the long-term perspective of the parent project. Expert 2 emphasized the usefulness of having multiple data sources consolidated into one platform but cautioned that its value depends on offering a collection of data sources not readily available elsewhere. They stressed the importance of such a resource, especially in regions like Afghanistan where conventional data sources are limited, particularly concerning internally displaced persons (IDP) data, as often you have to look at data sources from different sectors so being included in this website is very useful. Expert 2 expressed concerns about the website's future, highlighting the need to handle sensitive information, especially concerning Non-Governmental Organizations (NGOs). However, they noted that currently, the website doesn't contain any information that raises concerns.

After our discussion, users completed a NASA Task Load Index questionnaire regarding

the task of finding a particular data source and comprehending the data quality. The NASA Task Load Index (as seen in Figure E.1 in the Appendix)is a subjective assessment tool used to evaluate perceived workload experienced by individuals during tasks (NASA Human Systems Integration Division [2006]). Users rated each section, and their results were inputted into the Nasa Task Load Index app to calculate the overall weighted rating (the app allows you to assign different weights to different categories.) For example, in this instance, Physical Demand shouldn't be taken into consideration. Expert 1 had a final score of 7.33, whereas Expert 2's was 32.0. Expert 1's score of 7.33 suggests a relatively low workload, indicating that they found the task of navigating the website and assessing data quality to be relatively straightforward and manageable. Conversely, Expert 2's score of 32.0 indicates a higher perceived workload, implying that they may have found the task more challenging or mentally taxing. However as the rating is out of 100 this is still a relatively low rating. The differences in scores could also stem from varying perceptions of mental demand, temporal demand, performance, effort, and frustration levels experienced during the task as this is what the questionnaire takes into consideration.

7.2 General Users

Following the initial user testing sessions with domain experts, further testing was conducted with ten users lacking a humanitarian background. The objective was to assess the website's usability and gauge the comprehensibility of the information provided for non-experts. The questionnaire sent to these users can be found in the Appendix F.1, it prompted participants to filter data to locate a specific Data Source and examine the "The Data" and "Data Quality" tabs. Participants were asked to evaluate the ease of understanding of the information presented and assess the perceived quality of the data source. The questionnaire included a mix of yes/no questions, rating scales, and open-ended responses to gather comprehensive feedback from the participants. There were eleven participants ranging in age from 18 to 51, comprising 3 informatics students, 6 students from other disciplines, and 2 non-students.

Upon initial use, the first user to submit feedback noted that the text appeared very light against a white background. This was not apparent on my own device, underscoring the necessity of checking compatibility across various browsers and devices. I promptly addressed this issue before further users tested the website.

All participants successfully completed the filtering task, and feedback indicating a high level of comprehensibility for the "The Data" and "Data Quality". Specifically, 5 participants found the information "easily" understandable, while 5 found it "very easily" understandable. One participant described it as "okay." Participants appreciated the segmented breakdown of information, allowing them to explore further details if needed. The use of tables was highlighted for its clarity in presenting data, while icons facilitated quick understanding. However, one participant noted inconsistent casing across the text, particularly in the tags. All 11 participants expressed that the data source ("Whole of Afghanistan Assessment 2022") they examined was considered "good," indicating engagement with the content. However, it's challenging to definitively assess the accuracy of this perception, as discussed throughout this paper.

Participants were then encouraged to explore the site freely and provide feedback on its ease of use and intuitiveness. Out of the participants, 7 found it ”very easy,” 3 found it ”easy,” and 1 rated it as ”okay.” Feedback highlighted the site’s navigational ease, effective use of a professional color scheme, and clear, legible layout. One participant appreciated the wide range of filters available, facilitating easy comparison of data sources and their quality. Additionally, they liked the feature that allows access to the link to trace the data source. Some suggestions for improvement were also noted; for instance, one participant suggested reducing the gap between the boxes at the top and expanding the search tab to occupy more space.

The user testing sessions, encompassing both domain experts and non-experts, provided valuable insights into the usability and effectiveness of the website. Overall, participants found the interface intuitive, navigation easy, and information comprehensible. The positive feedback regarding the clarity of data presentation and ease of filtering underscores the success of the design in meeting user requirements. However, suggestions for minor improvements, such as addressing text legibility issues, optimizing space utilization, improving alignment with humanitarian standard terms (e.g. SDG indicators), were duly noted for potential future enhancements. These findings validate the hypothesis that a well-designed data quality assessment methodology can effectively convey the nature of data to users, fostering transparency and trust in the system. Moving forward, incorporating feedback from both expert and non-expert users will be integral to refining the website and ensuring its continued usability and relevance.

7.3 Limitations of the Evaluation

Limitations in the evaluation process could have constrained the depth and breadth of insights obtained from domain experts and general users, potentially impacting the robustness of the findings. The limited number and size of interviews may have restricted the diversity of perspectives within the humanitarian and development sectors and could result in some issues being overlooked that could have emerged with a larger and more varied participant pool. In addition, participants may have biases.

In addition the current testing of the website at this stage was limited by the small number of data sources that have currently been uploaded, primarily due to time constraints during the project’s development. This limitation prevented comprehensive testing of the website’s functionality and effectiveness in handling a larger range of data, therefore all potential challenges or benefits may not have been discovered.

To address these limitations and enhance the impact of the research, future endeavors could focus on expanding the scope and diversity of interviews to capture a more comprehensive range of perspectives and experiences within the humanitarian and development sectors. Additionally, allocating more resources and time to the development phase would facilitate the inclusion of a broader array of data sources, enabling more thorough testing and refinement of the website’s functionality and usability. Collaboration with stakeholders and partners could also provide valuable insights and resources to address these challenges and improve the long-term viability and effectiveness of the website as a tool for data quality assessment in humanitarian contexts.

Chapter 8

Conclusion & Future Work

8.1 Future Work

The primary objective of this project was to conduct research and devise a method for presenting various data sources while also highlighting their data quality aspects. The overarching goal was to streamline the selection process of humanitarian data sources by offering a centralized repository within a single website, accompanied by comprehensive insights into critical factors such as accuracy, relevance, and accessibility for each data source. The following section entails some continuations and future work that could be implemented in the future.

8.1.1 Expansion of Data Sources

The current prototype does not include all the potential data sources listed in the Excel sheet of data sources created for the parent project. This leaves room for the addition of more data sources to enrich the platform's utility and value. Continuously updating the platform with new data sources as they become available will enhance its relevance and make it a more comprehensive and valuable resource for humanitarian professionals and researchers alike.

8.1.2 Expansion to Different Countries and Topics

If the Power BI Dashboard were expanded to encompass additional topics related to various countries, users would gain the ability to filter and access individual country dashboards. Then the same could be applied to this project's website as it is not limited to data sources exclusively related to education in Afghanistan. There is potential to expand its scope to include various topics and countries. This expansion could be facilitated by incorporating countries into the existing filters, allowing users to explore a broader range of data sources spanning different geographical regions and thematic areas.

8.1.3 Automation

Presently, data is manually added to the website. An enhancement to the system could involve automating the process of data collection through web scraping. This automation would enable the automatic addition of new data sources to the platform, streamlining the process and ensuring that the website remains up-to-date with the latest information.

8.1.4 Personalization

Implementing user accounts could enhance the website's functionality by allowing users to log in and save their preferred data sources or search criteria. This feature could enable users to create personalized searches and receive notifications when new data sources matching their interests are added, ensuring they have access to the most relevant and up-to-date information tailored to their needs.

8.1.5 Mobile Optimization

While the current design focuses on desktop usability, this choice was made because it is a resource people would use for humanitarian work so would likely be on laptops/desktops. Optimizing the website for mobile devices could greatly enhance its accessibility and usability, especially for users who need to access data on the go. By ensuring that the website is responsive and adapts well to smaller screens, users can conveniently access and utilize the data sources and filtering features from their smartphones or tablets, enabling them to work efficiently regardless of their location.

8.1.6 Accessibility Testing

Whilst some user testing has been conducted on the website further accessibility testing should be conducted. Accessibility testing is crucial to ensure that the website meets the necessary standards and is usable by all individuals, regardless of their abilities or disabilities Accessibility community [2023]. Conducting comprehensive accessibility testing involves evaluating various aspects of the website, including its compatibility with assistive technologies such as screen readers, keyboard navigation, color contrast, and alternative text for images. Additionally, usability testing with individuals representing diverse demographics can help identify any barriers to access and usability. By prioritizing accessibility testing, the website can ensure inclusivity and provide equal access to information for all users BrowserStack [2024].

8.1.7 Suggestions from Evaluation

Suggestions from the evaluation process include aligning wording with standard terms such as SDG indicators, enlarging the search bar, and adding filters for data quality attributes like geographical coverage.

8.2 Conclusion

In this paper, it was proposed that that a framework could be established for presenting information about the quality of data relating to education in Afghanistan that will convey the nature of the data allowing those who work with humanitarian data to easily evaluate the data quality and gauge what the data represents. This should help those individuals to select data that is of a satisfactory quality and relevant. For users accessing the website from Power BI, they will gain awareness of any strengths or limitations of the data used, thereby ensuring trust and transparency between the system and the user. Throughout this report, the necessity for such a system has become evident.

Through a comprehensive literature review, focusing on Fragile and Conflict-Affected Countries (FCACs) and the historical context of data collection in Afghanistan, this study delved into the existing challenges associated with gathering education data in Afghanistan and its implications for data quality. Additionally, the concept of data quality was examined, along with the current methodologies used by both humanitarian and non-humanitarian organizations for assessing data quality. Furthermore, existing data sharing platforms were investigated to provide a broader understanding of the landscape surrounding data sharing in humanitarian contexts.

Through an investigation into data quality frameworks developed by various organizations, this study explored the principles of data quality, their definitions, and how they are perceived to contribute to overall data quality. Additionally, it examined the methodologies used by these organizations to ensure and assess data quality, providing insights into the the various stages of data production.

Through interviews with domain experts, I uncovered the varied applications of humanitarian data, the hurdles they face when sourcing and selecting data, their perspectives on and methods for assessing data quality, and their expectations from a data exchange platform presenting information on data quality. These interviews led to the identification of a distinct set of user requirements.

I proceeded with the process of designing the system, with an iterative process employing wire-frame techniques to explore various options while continually considering the insights from the interviews and user requirements. Throughout this process, I adhered to standard design practices and heuristics to ensure the development of a user-friendly interface.

Subsequently, I adhered to my designs and proceeded to implement the system, selecting data sources from the Excel sheet of data sources provided in the parent project.

Evaluation took place in several stages, commencing with user testing sessions with domain experts, where participants were tasked with completing tasks while their interactions with the system were observed. This was followed by in-depth discussions about the website and the administration of a NASA Task Load Index questionnaire. Subsequently, additional testing was conducted with a broader selection of users who were tasked with completing tasks and exploring the website followed by answering questions. The evaluation revealed that the system was deemed useful, usable, and intuitive, and users could evaluate the quality of data using the system, although there

were identified areas for improvement.

This project benefits the ”Education Data in Conflicted Countries” project by providing comprehensive information about the quality of the data sources accessible from the Power BI Dashboard. By offering insights into aspects such as reliability, accuracy, and coverage of these data sources, this project enhances the transparency and trustworthiness of the overall platform. Moreover, the user-friendly interface and interactive features make it a valuable resource for users seeking to explore and understand education data in Afghanistan more deeply. Additionally, by providing a framework for a centralized repository for all available information on education in Afghanistan, this project transforms the current Excel sheet of data sources into a more accessible and interactive resource, thereby facilitating data-driven decision-making and analysis for humanitarian professionals and stakeholders.

The hypothesis of this report was that developing a data quality assessment framework will help convey to users the nature of the data they are looking at; Individuals that work with humanitarian and development data will be able to gauge the quality of the data easily ensuring transparency and trust between the data source’s and the user. It will be intuitive to understand.

The hypothesis proposed at the outset of this dissertation appears to hold true based on the findings from the evaluation. Users reported that the system proved to be a valuable resource, offering ease of use and facilitating the evaluation of data quality. However, to validate the hypothesis further and ascertain its efficacy in real-world scenarios, it would be imperative to expand the dataset by incorporating additional data sources. This broader dataset would provide a more comprehensive understanding of how the system performs across various contexts and datasets, offering insights into its scalability and applicability in diverse settings.

Bibliography

GitHub Documentation. URL https://docs.github.com/en.

Generic Data Quality Assurance Framework for a UN Agency. Internal Document, 2015.

Statistical data and metadata exchange, 2016. URL https://sdmx.org/?sdmx_news=un-statistical-commission-sdmx-is-preferred-standard. Accessed: 26 October 2023.

Afghanistan - educational management information systems, 2023.

Accessibility community. Testing for accessibility, October 2023. URL https://www.gov.uk/service-manual/helping-people-to-use-your-service/testing-for-accessibility.

William C. Adams. Conducting semi-structured interviews. In Kathryn E. Newcomer, Harry P. Hatry, and Joseph S. Wholey, editors, *Handbook of Practical Program Evaluation, Fourth*, chapter 19. Wiley, 2015. doi: 10.1002/9781119171386.ch19. URL https://doi.org/10.1002/9781119171386.ch19.

Ali Yawar Adili. A success story marred by ghost numbers: Afghanistan's inconsistent education statistics, March 2017. URL https://www.afghanistan-analysts.org/en/reports/economy-development-environment/a-success-story-marred-by-ghost-numbers-afghanistans-inconsistent-education-

Arcadisgen. Data quality visualization. https://www.arcadisgen.com/en/marketplace/products/data-quality-visualization. Arcadis Gen.

Razia Arooje and Nina Burridge. School education in afghanistan overcoming the challenges of a fragile state, February 2022. URL https://www.researchgate.net/publication/358536594_School_Education_in_Afghanistan_Overcoming_the_Challenges_of_a_Fragile_State/citations.

Ataccama. Data quality storytelling: A better way to do dq reporting. https://www.ataccama.com/blog/data-quality-storytelling-a-better-way-to-do-dq-reporting, June 2023. Ataccama.

Yahia Baiza. *Education in Afghanistan: Developments, Influences and Legacies Since 1901*. Routledge, 2013. ISBN 1134120826, 9781134120826. URL

https://books.google.co.uk/books/about/Education_in_Afghanistan.html?id=dauAAAAAQBAJ&redir_esc=y.

Prisca Benelli, Dyan Mazurana, and Peter Walker. Using sex and age disaggregated data to improve humanitarian response in emergencies. *Gender & Development*, 20 (2):219–232, 2012. doi: 10.1080/13552074.2012.687219.

Mats Bergdahl, Manfred Ehling, Eva Elvers, Erika Földesi, Thomas Körner, Andrea Kron, Peter Lohauß, Kornelia Mag, Vera Morais, Anja Nimmergut, Hans Viggo Sæbø, Ulrike Timm, and Maria João Zilhão. *Handbook on Data Quality Assessment Methods and Tools*. European Commission, Eurostat, 2007.

BrowserStack. What is accessibility testing? examples, types, metrics, February 2024. URL https://www.browserstack.com/accessibility-testing/what-is-accessibility-testing.

Elizabeth Buckner, Daniel Shephard, and Anne Smiley. Beyond numbers: The use and usefulness of data for education in emergencies. *Journal on Education in Emergencies*, 8(1), Mar 2022. ISSN 2518-6833. doi: https://doi.org/10.33682/tgfd-m9eg. URL https://archive.nyu.edu/handle/2451/63610.

RN PhD Carter, Nancy, RN PhD Bryant-Lukosius, Denise, RN PhD DiCenso, Alba, PhD Blythe, Jennifer, and MBChB MEd MRCP FRCP(c) Neville, Alan J. The use of triangulation in qualitative research. 2019.

Kasia Chmielinski, Matt Taylor, Sarah Newman, Jessica Yurkofsky, and Chelsea Qiu. Quality measures for humanitarian data: Sprint report. Sprint report, Humanitarian Data Exchange, April 2023. URL https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/dac85080-6beb-4c3d-ad62-be0d7aecf08a/download/2023_hdx_quality-measures-sprint-report_final.pdf.

L. Y. Conrad and V. M. Tucker. Making it tangible: Hybrid card sorting within qualitative interviews. *Journal of Documentation*, 75(2):397–416, 2019. doi: 10.1108/JD-06-2018-0091. URL https://doi.org/10.1108/JD-06-2018-0091.

Data Science & Ethics Group. A framework for the ethical use of advanced data science methods in the humanitarian sector, 2020.

EUROSTAT. *Handbook on Data Quality Assessment Methods and Tools*. 2007.

Food and Agriculture Organization of the United Nations. *The FAO Statistics Quality Assurance Framework*. Rome, 2014.

United Nations Office for the Coordination of Humanitarian Affairs (OCHA). New way of working, 2017. URL https://www.unocha.org/publications/report/world/new-way-working. Accessed: 25 October 2017.

Global Partnership for Sustainable Development Data. Recipes for embedding quality assurance in data collection and analysis, 2023. URL https://www.data4sdgs.org/recipes-embedding-quality-assurance-data-collection-and-analysis.

Johannes Hoogeveen and Utz Pape, editors. *Data Collection in Fragile States: Innovations from Africa and Beyond*. Palgrave Macmillan, 2020.

Government Data Quality Hub. What is data quality?, 2021. URL https://www.gov.uk/government/news/what-is-data-quality. Accessed: 26 October 2023].

Pamela Hunter. Margin of error and confidence levels made simple, 2016. URL https://psy311.wichita.edu/wp-content/uploads/2020/07/Margin-of-error-made-simple.pdf. Retrieved February 25th, 2016.

ICRC. *Handbook on Data Protection in Humanitarian Action*. Second edition, 2018.

Inter-Agency Standing Committee. *Operational Guidance: Data Responsibility in Humanitarian Action*. 2021.

Inter-Cluster Coordination Team, Reach Initiative. Whole of afghanistan assessment 2022: Key findings presentation. Presentation, September 2022.

Interaction Design Foundation - IxDF. What are the gestalt principles?, 2016, August 30. URL https://www.interaction-design.org/literature/topics/gestalt-principles.

International Aid Transparency Initiative. International aid transparency initiative. Website. URL https://iatistandard.org/en/.

International Monetary Fund. *Data Quality Assessment Framework – Generic Framework*. 2012.

Burkle FM Canny W Purdin S Lin W Sauer L Kirsch TD, Perrin P. Requirements for independent community-based quality assessment and accountability practices in humanitarian assistance and disaster relief activities. *Prehosp Disaster Med*, 2012.

Matthaios Mantzios. Hover to reveal: Enhancing product discovery through interactive design, October 2023. URL https://uxplanet.org/hover-to-reveal-enhancing-product-discovery-through-interactive-design-a4277

Bella Martin and Bruce Hanington. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Rockport Publishers, Osceola, 2012. ISBN 9781610581998.

Rotz D. Garcia C. Maxwell, N. L. Data and decision making: Same organization, different perceptions; different organizations, different perceptions. *American Journal of Evaluation*, 2016. URL https://doi.org/10.1177/1098214015623634.

Microsoft. Power BI Documentation. URL https://learn.microsoft.com/en-us/power-bi/.

Islamic Republic of Afghanistan Ministry of Education and United Nations Children's Fund (UNICEF). All children in school and learning - afghanistan country study, 2018. URL https://www.right-to-education.org/resource/all-children-school-and-learning-afghanistan-country-study.

Prakash M. Nadkarni. What is metadata? In *Metadata-driven Software Systems in Biomedicine*, Health Informatics. Springer, London, 2011. ISBN 978-0-85729-

509-5. doi: 10.1007/978-0-85729-510-1_1. URL https://doi.org/10.1007/978-0-85729-510-1_1.

NASA Human Systems Integration Division. Task Load Index (TLX), 2006. URL https://humansystems.arc.nasa.gov/groups/TLX/. Curator: Phil So, NASA Official: Brian Gore, Last Updated: December 15, 2020.

Jakob Nielsen. 10 usability heuristics for user interface design, April 1994. URL https://www.nngroup.com/articles/ten-usability-heuristics/. Updated Jan. 30, 2024.

Jakob Nielsen. Thinking aloud: The #1 usability tool. *Nielsen Norman Group*, January 15 2012. URL https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/. Topics: User Testing.

OECD. *Quality Framework and Guidelines for OECD Statistical Activities*. 2011.

World Health Organization. Data quality review. module 3: Site assessment of data quality: data verification and system assessment, 2021. URL https://cdn.who.int/media/docs/default-source/data-quality-pages/2021_dqa_module-3_site-assessment-framework.pdf?sfvrsn=4e936e2f_9. Accessed: 27 October 2023.

Kate Roll and Geoffrey Swenson. Fieldwork after conflict: contextualising the challenges of access and data quality. *Disasters Overseas Development Institute*, 2018. URL https://pubmed.ncbi.nlm.nih.gov/30537279/.

Elar Saks. Javascript frameworks: Angular vs react vs vue. 2019.

Kelly D Elmusharaf K Shalash A, Abu-Rmeileh NME. The need for standardised methods of data collection, sharing of data and agency coordination in humanitarian settings. *BMJ Glob Health*, 2022.

Andrea Tapia, Kathleen Moore, and N.J. Johnson. Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. *Proceedings of the 10th International ISCRAM Conference*, May 2013. doi: https://doi.org/[DOINumber]. URL https://www.researchgate.net/publication/290086707_Beyond_the_trustworthy_tweet_A_deeper_understanding_of_microblogged_data_use_by_disaster_response_and_humanitarian_relief_organizations.

Frontier Technologies. Frontier data study: Releasing the power of digital data development, October 2020.

The Humanitarian Data Exchange. Welcome - humanitarian data exchange. Website. URL https://data.humdata.org/.

UNESCO. SDG4 Data Explorer. Website. URL https://tcg.uis.unesco.org/data-resources-new/.

UNESCO. The right to education: what's at stake in afghanistan? a 20-year review, 2021. URL https://unesdoc.unesco.org/ark:/48223/pf0000378911.

UNICEF. Unicef data warehouse. Website. URL https://data.unicef.org/resources/data_explorer/unicef_f/?dq=....&startPeriod=2018&endPeriod=2023.

UNICEF. Data quality framework, 2021. URL https://data.unicef.org/resources/data-quality-framework/. Accessed: 26 October 2023.

United States Agency for International Development (USAID). *How to Conduct a Data Quality Assessment (DQA)*. 2012.

Fanny Vassilatos and Ceara Crawshaw. Enterprise filtering, April 2023. URL https://www.pencilandpaper.io/articles/ux-pattern-analysis-enterprise-filtering.

Vue.js. Vue documentation, n.d. URL https://vuejs.org/guide/introduction.html.

World Health Organization. *Data quality review: a toolkit for facility data quality assessment. Module 3. Site assessment of data quality and system capabilities*. 2020a.

World Health Organization. *Data Quality Review: A Toolkit for Facility Data Quality Assessment. Module 3: Site Assessment of Data Quality and System Capabilities*. 2020b.

World Health Organization. *Data quality assurance. Module 1. Framework and metrics*. 2022a. This publication is the update of the document published in 2017 entitled "Data quality review: a toolkit for facility data quality assessment. Module 1. Framework and metrics".

World Health Organization. *Data quality assurance. Module 2 Discrete desk review of data quality*. 2022b.

Appendix A

Participants' information sheet

Participant information sheet is on following pages.

Participant Information Sheet

Project title:	Developing the ability to assess and visualise the data quality of incomplete data
Principal investigator:	Fiona McNeill
Researcher collecting data:	Ruby Imrie
Funder (if applicable):	

This study was certified according to the Informatics Research Ethics Process, reference number 469605. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

Researcher:

- Ruby Imrie – Student at the University of Edinburgh completing Undergraduate Honours Project

Supervisors:

- Fiona McNeill - Reader in Computing Education at University of Edinburgh
- Amanda Meyer - Researching data access and interpretation for the factors impacting education, for children living in conflict affected areas.

What is the purpose of the study?

The purpose of this study is to gather valuable insights from professionals in humanitarian organizations regarding their perspectives on data quality assessment. Through these interviews, we aim to:

1. Understand how individuals working in the humanitarian sector currently assess data quality.
2. Identify the factors that are considered most crucial by humanitarian professionals when evaluating data quality in their work.
3. Explore the preferences and expectations of participants regarding a potential data quality assessment system.

THE UNIVERSITY of EDINBURGH informatics

Figure A.1: Participant Information Sheet page 1

These interviews play a role in informing the development of a comprehensive methodology and user interface for assessing and visualizing data quality

Why have I been asked to take part?

The research target group is people who work for humanitarian organisations.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until February 2024 without giving a reason. After this point, personal data will be deleted and anonymised data will be combined such that it is impossible to remove individual information from the analysis. Your rights will not be affected. If you wish to withdraw, contact the PI. We will keep copies of your original consent, and of your withdrawal request.

What will happen if I decide to take part?

- Kinds of data being collected:
    - Information related to your experience in assessing data quality in the humanitarian sector.
    - Factors you consider crucial for data quality assessment.
    - Preferences and expectations regarding a data assessment system.
- Means of collection:
    - Semi-structured interview, allowing for open discussion and exploration of your perspectives.
- Duration of session:
    - The interview is expected to take approximately 45 to 60 minutes.
- Participant audio/video recording:
    - Video and Audio recording will be conducted during the interview and rewatched for accurate documentation and analysis.
- How often, where, when:
    - The interviews will take place on Microsoft Teams or in person. This will be a one off interview or you will be contacted to see if you would participate in the finished product.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

Are there any benefits associated with taking part?

No

THE UNIVERSITY of EDINBURGH informatics

Figure A.2: Participant Information Sheet page 2

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Quotes or key findings will be anonymized: We will remove any information that could, in our assessment, allow anyone to identify you. With your consent, information can also be used for future research. Your data may be archived for a maximum of 4 years. All potentially identifiable data will be deleted within this timeframe if it has not already been deleted as part of anonymization.

Data protection and confidentiality.

Your data will be processed in accordance with Data Protection Law. All information collected about you will be kept strictly confidential. Your data will be referred to by a unique participant number rather than by name. Your data will only be viewed by the researcher/research team (Researcher: Ruby Imrie and Supervisors: Fiona McNeill and Amanda Meyer).

All electronic data will be stored on a password-protected encrypted computer, on the School of Informatics' secure file servers, or on the University's secure encrypted cloud storage services (DataShare, ownCloud, or Sharepoint) and all paper records will be stored in a locked filing cabinet in the PI's office. Your consent information will be kept separately from your responses in order to minimise risk.

What are my data protection rights?

The University of Edinburgh is a Data Controller for the information you provide.You have the right to access information held about you. Your right of access can be exercised in accordance Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, Ruby Imrie (s2056921@ed.ac.uk)

If you wish to make a complaint about the study, please contact

THE UNIVERSITY of EDINBURGH informatics

Figure A.3: Participant Information Sheet page 3

inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

[**NB:** the PI should notify the Ethics panel on inf-ethics@ed.ac.uk to upload any updated PIS to the website]

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact Ruby Imrie (s2056921@ed.ac.uk) .

General information.

For general information about how we use your data, go to: edin.ac/privacy-research

THE UNIVERSITY of EDINBURGH informatics

Figure A.4: Participant Information Sheet page 4

Appendix B

Participants' consent form

Participant consent form in on following page.

Participant number:_____

Participant Consent Form

Project title:	Developing the ability to assess and visualise the data
Principal investigator (PI):	Fiona McNeill
Researcher:	Ruby Imrie
PI contact details:	f.j.mcneill@ed.ac.uk

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1.	I agree to being audio recorded.	☐ **Yes**	☐ **No**
2.	I agree to being video recorded.	☐ **Yes**	☐ **No**
3.	I allow my data to be used in future ethically approved research.	☐ **Yes**	☐ **No**
4.	I agree to take part in this study.	☐ **Yes**	☐ **No**

Name of person giving consent	Date dd/mm/yy	Signature
Name of person taking consent	Date dd/mm/yy	Signature

THE UNIVERSITY of EDINBURGH informatics

Figure B.1: Participant Consent Form

Appendix C

Power BI Dashboard

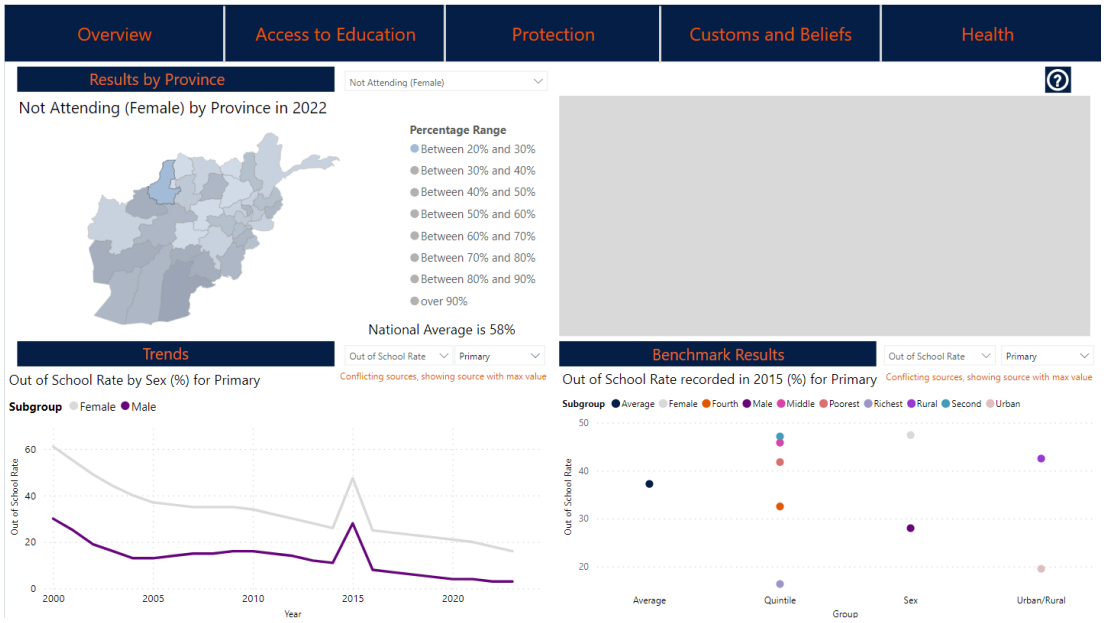


Figure C.1: Power BI Dashboard from Existing Project

Appendix D

Data Quality Framework Documents

Table D.1: Framework Documents

Organisation Name/Acronym	Framework Document Name
GSQAF	Generic Data Quality Assurance Framework (GDQ [2015])
UNICEF	Data Quality Framework (UNICEF [2021])
WHO	Data quality assurance: module 1: framework and metrics (World Health Organization [2022a]), Data quality assurance: Module 2 Discrete desk review of data quality Module 2 Discrete desk review of data quality (World Health Organization [2022b]),Data quality assurance: Module 3 Site assessment of data quality: data verification and system assessment (Organization [2021])
DSEG	A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector (Data Science & Ethics Group [2020])
HDX	Quality Measurements for Humanitarian Data - Sprint Report (Data Science & Ethics Group [2020])
ICRC	Handbook on data protection in humanitarian action (ICRC [2018])
IASC	Operational Guidance: Data Responsibility in Humanitarian Action (Inter-Agency Standing Committee [2021])
USAID	How to Conduct A Data Quality Assessment (DQA) (United States Agency for International Development (USAID) [2012])
Frontier	Frontier Data Study: releasing the power of digital data development (Technologies [2020])
EUROSTAT	Handbook on Data Quality Assessment Methods and Tools (EUROSTAT [2007])
OECD	Quality Framework and Guidelines for OECD Statistical Activities (OECD [2011])

Continued on next page

Table D.1 – *Continued from previous page*

Organisation Name/Acronym	**Framework Document Name**
IMF	Data Quality Assessment Framework — Generic Framework (International Monetary Fund [2012])
GPSDD	Recipes for embedding quality assurance in data collection and analysis (Global Partnership for Sustainable Development Data [2023])
FAO	The FAO Statistics Quality Assurance Framework (Food and Agriculture Organization of the United Nations [2014])

Appendix E

Nasa Task Load Index

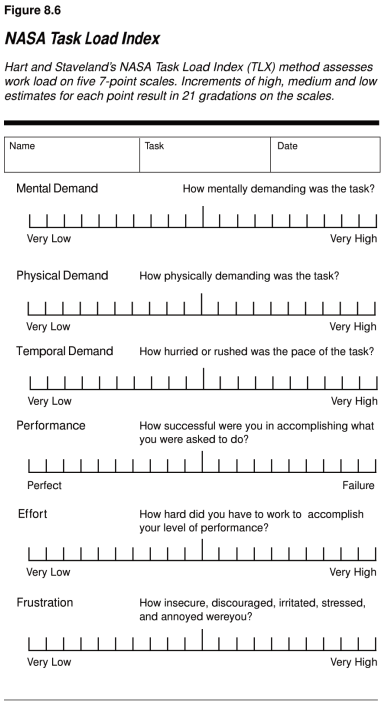


Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

Physical Demand — How physically demanding was the task?
Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

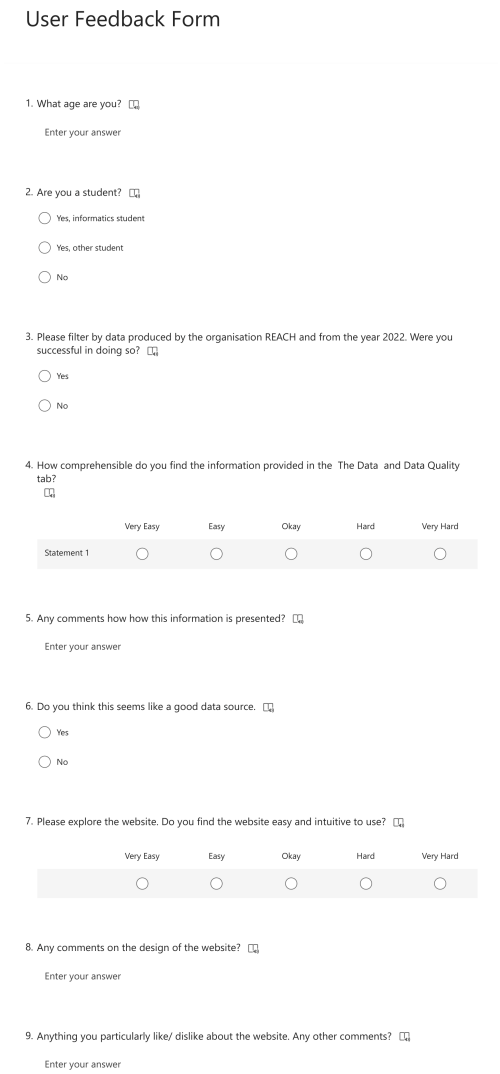
Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High

Figure E.1: Nasa Task Load Index Form

Appendix F

User Feedback Form



User Feedback Form

1. What age are you?

Enter your answer

2. Are you a student?

- Yes, informatics student
- Yes, other student
- No

3. Please filter by data produced by the organisation REACH and from the year 2022. Were you successful in doing so?

- Yes
- No

4. How comprehensible do you find the information provided in the The Data and Data Quality tab?

	Very Easy	Easy	Okay	Hard	Very Hard
Statement 1					

5. Any comments how how this information is presented?

Enter your answer

6. Do you think this seems like a good data source.

- Yes
- No

7. Please explore the website. Do you find the website easy and intuitive to use?

	Very Easy	Easy	Okay	Hard	Very Hard

8. Any comments on the design of the website?

Enter your answer

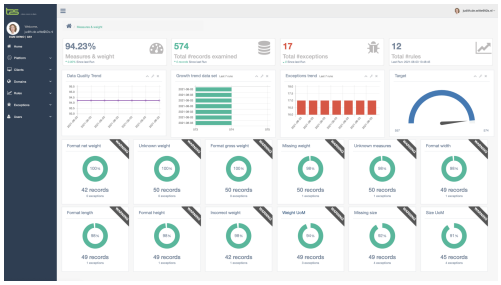
9. Anything you particularly like/ dislike about the website. Any other comments?

Enter your answer

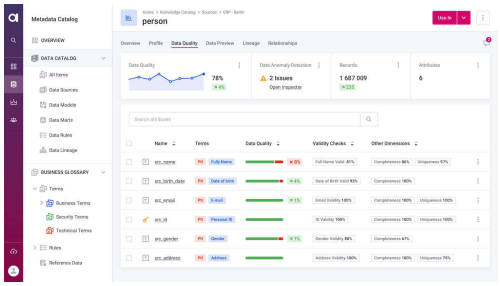
Figure F.1: User Feedback Form

Appendix G

Previous Data Quality Assessment Scoring



(a) T2S Scoring Dashboard from https://www.t2s.nl/data-quality-dashboard/



(b) Ataccama Data Quality from https://www.ataccama.com/platform/data-quality

Figure G.1: Previous Data Quality Assessment Scoring Examples

Appendix H

More Wireframe Iterations

(a) Accuracy Section of Dashboard

(b) Timeliness Section of Dashboard

Figure H.1: More Iterations of Wireframes that were shown in Interviews

Appendix I

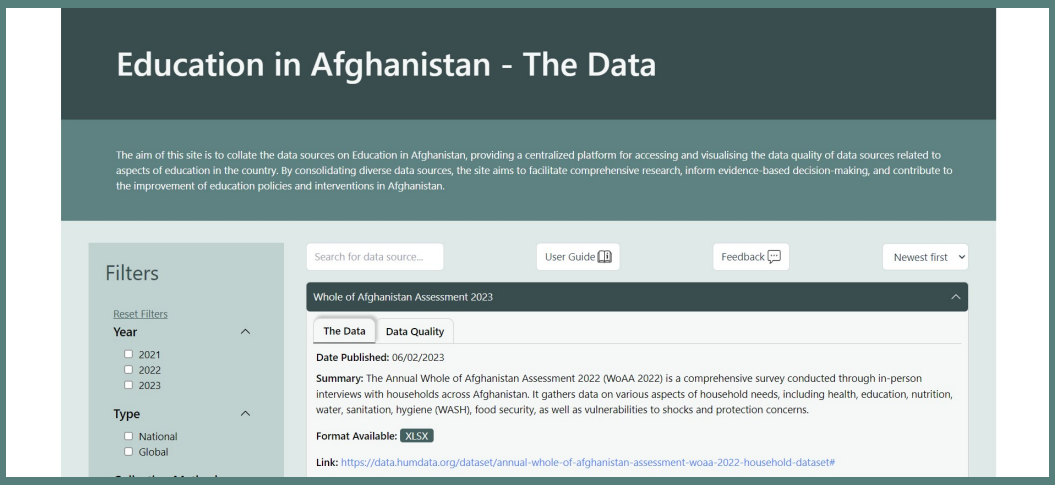
User Guide

User guide is attached across following pages.

Figure I.1: User Guide for website (page 1)

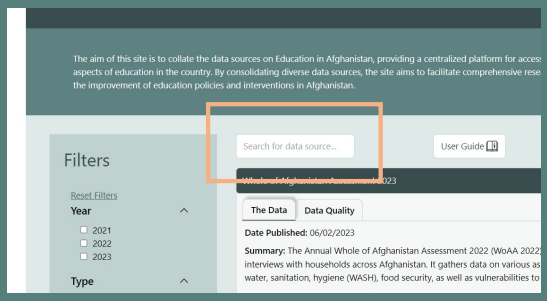
User Guide

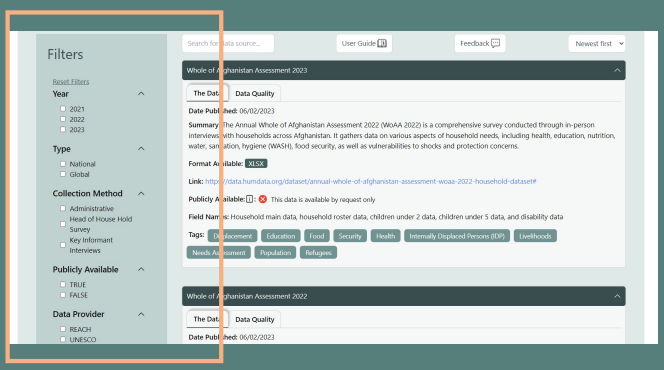
Welcome to the Education Data website! This user guide will help you navigate and make the most out of our platform designed to help find data sources that can provide insights into educational landscapes, particularly in conflict-affected regions like Afghanistan.



Searching for a data source

If you already know the name of the data source you are looking for use the search bar located is located top left above all the data sources. If a data source matches your search then it will appear.



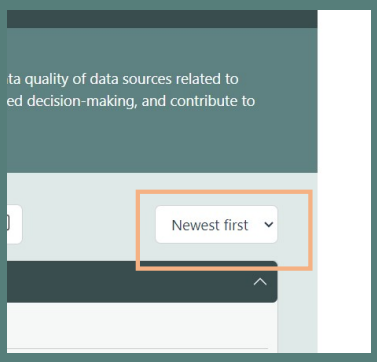


Filtering

To find a data source that is relevant t and suitable to your needs use the filter by Year, Type, Collection Method, Publicly Available, Data Provider, Updated, Disaggregated By and Tags. When you click a filter it data sources results will update automatically. Use the "Reset Filters" button to reset your search.

Each filter section is collapsable to optimize space usage on the screen.

Figure I.2: User Guide for website (page 2)

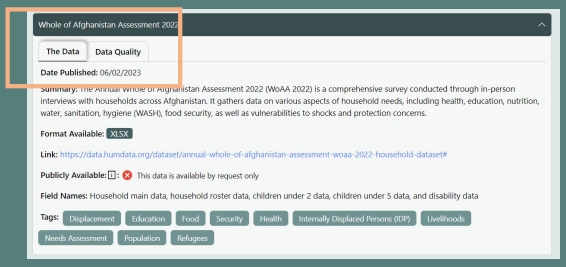


Sorting

Initially the data sources appear newest first in order to aid you to find the most timely data. However if you wish to select to view the oldest data source first you can select this using the drop down bar top left of the page.

User Guide and Feedback

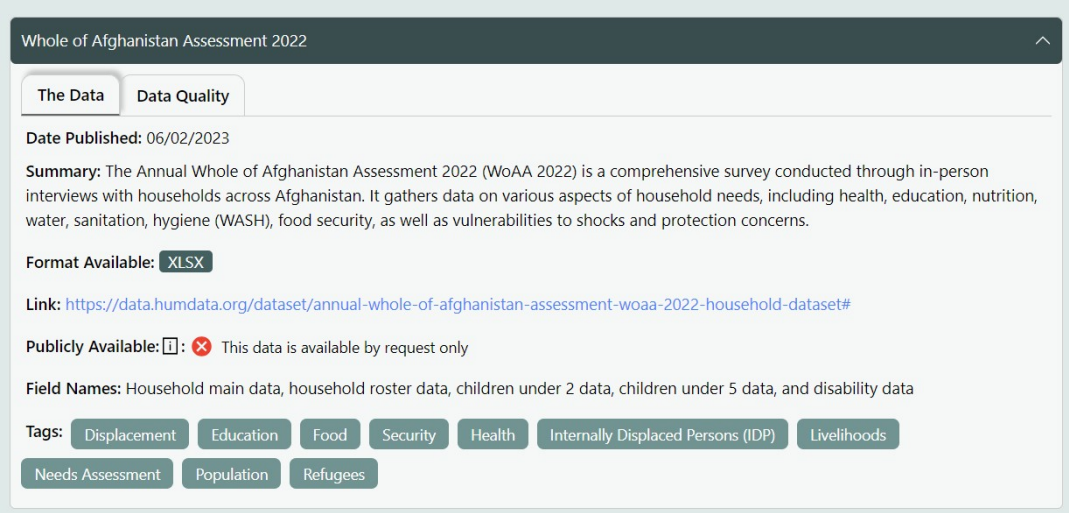
This user guide can be found at the top of the page below the title and description, providing comprehensive assistance on navigating the website's features and functionalities. Should you have any suggestions, questions, or concerns, there is a dedicated feedback form accessible through the "Feedback" option. Simply fill in your name, email, and feedback message, then click "Submit" to send your input to the website administrators. To return to the homepage from either the user guide or feedback pages, simply click on the "Home" button located at the top left



Tab Selection

The information about each data source is split across a "The Data" and a "Data Quality" tab. "The Data" tab includes detailed information regarding the dataset such as its publication date, summary, format, a link to access, whether its publicly available, field names and tags. On the other hand, the "Data Quality" tab provides insights into collection methods, consistency checks, geographical coverage, confidence levels, margin of error, disaggregation categories, and more. The tabs are located at the top of the data source page, allowing users to seamlessly swap between them to access the specific information they require.

Figure I.3: User Guide for website (page 3)



Understanding the “The Data” Tab

Date published

This section indicates the date when the dataset was originally published or made available to the public. It helps users assess the timeliness of the data.

Summary

A concise overview of the dataset, summarizing its purpose, methodology, and the type of information it contains.

Format Available

Specifies the format in which the dataset is available for download. Common formats include XLSX (Excel), CSV (Comma-Separated Values), JSON (JavaScript Object Notation), etc. This information helps users understand how they can access and utilize the data.

Link

Direct link to the dataset, where users can access or download it. This link facilitates easy access to the dataset without the need for extensive searching.

Publicly Available

Indicates whether the dataset is publicly accessible or requires special permission or requests for access. This information ensures transparency regarding data accessibility and usage restrictions. To access of request the data use the above link.

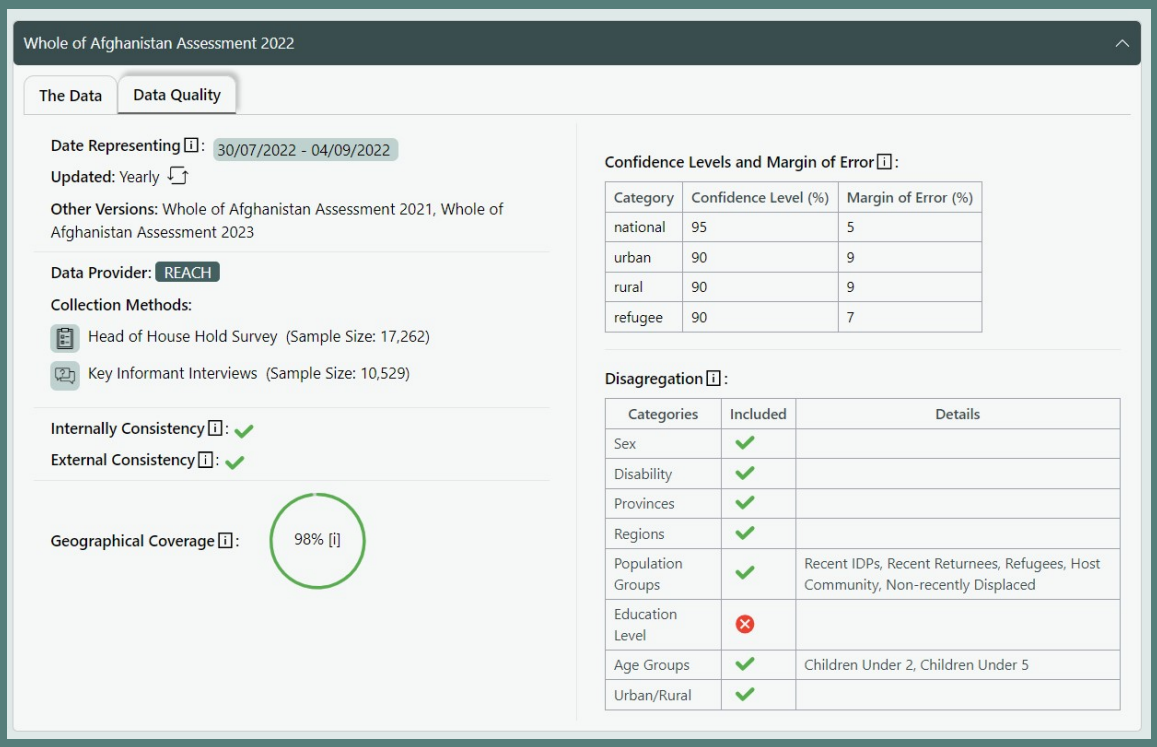
Field Names

Provides a list of the main data fields or categories included in the dataset. It helps users understand the structure of the data and identify specific variables of interest.

Tags

Keywords or labels assigned to the dataset, indicating its thematic focus or key topics. Tags aid users in quickly identifying datasets relevant to their areas of interest. The tags are clickable. When you click on a specific tag, the system filters the displayed data sources to show only those that are tagged with the selected keyword. To reset the tag filter and view all available data sources again, users can simply click the "Reset Filters" button.

Figure I.4: User Guide for website (page 4)



Understanding the “Data Quality” Tab

Date Representing:

The Data Quality tab provides information about the timeframe represented in the dataset.

Updated

This section informs users about the frequency of updates to the dataset, such as whether it is updated yearly, quarterly, or on another schedule.

Other Versions:

Users can find information about other versions or editions of the dataset, such as the Whole of Afghanistan Assessment 2021 or 2023.

Data Provider

This section specifies the organization or entity responsible for collecting and providing the dataset, such as REACH.

Collection Methods

Users can learn about the various methods used to collect data for the dataset, including surveys, head of household surveys, key informant interviews, and any sample sizes associated with these methods.

Consistency

Information about the internal and external consistency of the dataset may be provided here, ensuring users understand the reliability and validity of the data. Please note the data source has only been compared to data sources on this site that represent the same time frame.

Figure I.5: User Guide for website (page 5)

Geographical Coverage

This section indicates the extent of geographical coverage provided by the dataset, expressed as a percentage represented by a pie chart. Hover over the pie chart to find out more about the geographic coverage i.e. any areas that have been excluded.

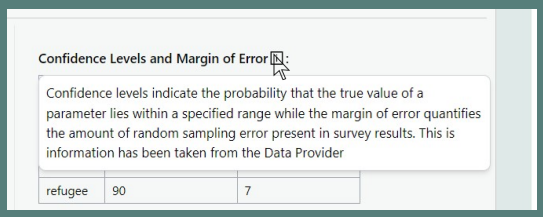
Confidence Levels and Margin of Error

Users can find details about the confidence levels and margin of error associated with different categories within the dataset, such as national, urban, rural, and refugee populations.

Confidence levels indicate the probability that the true value of a parameter lies within a specified range while the margin of error quantifies the amount of random sampling error present in survey results.

Disaggregation

Information about how the data is disaggregated or broken down into smaller categories may be provided here. This can include disaggregation by sex, disability, provinces, regions, population groups, education levels, age groups, and urban/rural areas.



Information Icons

Additionally, within the “The Data” and "Data Quality" tab, users can hover over the "i" icons to access tooltips providing further clarification on specific terms or data points. These tooltips offer valuable contextual information to enhance understanding and interpretation of the dataset's quality indica-

That’s every thing you need to know! Now you should be equipped to explore the website freely.