| **Assignment 4** | Exploring the problem through interviews |
| --- | --- |
| **Group** | FieldLab 8 - Analytics of COVID trends data |
| **Members** | Milou Schamhart, Hielke Muizelaar, Christie Bavelaar, Danielle Rachman, Kenneth Fargose |

## Introduction

This report provides insights into the interviews that we conducted to further explore the problem that we have been presented with. Our first task is to understand the bulk input tool being used to import the data into a FAIR format. We mainly need to focus on identifying and solving the challenges with the tool and thereafter, figuring out ways to resolve these challenges. Secondly, we are tasked with creating visualisations of interview and social media data related to COVID19.

First, we needed to identify the people who could assist us in further exploring this problem. We set out to understand the purpose of the bulk input tool, what problem such a tool aims to solve. Since this tool has already been deployed and is in use. To understand the tool and the challenges it faces, we interviewed one of the people responsible for creating and developing the tool. Secondly, we wanted to explore the current methods being used to digitise and FAIRify data in Africa. For this, we interviewed one of the professionals who facilitated the development and implementation of CEDAR and the bulk input tools.

The main points analysed from the interviews were the purpose, current use of the tool and its possible future. Thereafter, to understand the challenges experienced with the approach.

## Findings

### Purpose:

**CEDAR**

The main purpose of using the CEDAR tool is for template creation and to facilitate digitalisation of data. The main benefits of CEDAR is to facilitate the digitalisation of data. The interviewee mentioned that the majority of health facilities are paper based. This tool aims to promote the transition to digital record keeping. It also facilitates the FAIRification of data. Additionally, the interviewee mentioned that CEDAR allows for better data retention by the facility, security of data and improved data protection. In paper-based records, data breaches cannot be monitored and managed as they can through a digital platform. Additionally, the professionals can have instant and easy access to the data. Another reason discussed with the interviewee is that the healthcare facilities reported that the paper files take up valuable space in the facility which could better be utilised for patient beds. Lastly, the data can be analysed and thereafter, support decision making. The CEDAR tool should, therefore, be used by all health facilities in the VODAN network.

**Bulk Input tool**

In order to upload data, the user was required to upload the data individually in CEDAR. The bulk input tool aims to solve this problem by enabling the users to upload data in bulk. Thus, the tool facilitates data input and analysis of large data sets with the aim to save time and increase efficiency. Additionally, the bulk Input tool follows the FAIR format. This ensures that data is in the FAIR format. However, the interviewee also reported that this tool is aimed at facilities that already have data available digitally and thus, only few facilities are able to use this tool at this point.

### Design process:

**Bulk Input tool**

The tool was mainly designed and developed by a team of data scientists. The tool was based on the CEDAR tool and the API structure. The tool was presented to the VODAN community and feedback was provided in these meetings.

**Vocabularies in CEDAR templates**

The interviewee reported that the vocabularies were created in collaboration with the local professionals. Together, they reviewed the current registries and prioritised the templates that needed to be created for the relevant registries.

### Current use:

**Bulk Input tool**

There is no data on how many countries are actively using the tool. The use cases for this tool still remain unknown. As this is just the initial phases for the tool, the team is still figuring out the use cases and applications for the tool.

**CEDAR**

The tool is currently being used by a few of the healthcare facilities in the VODAN network. CEDAR needed to be localised in order to meet the local needs.

### 

### Challenges:

**Bulk Input tool**

The generalisability of the tool is a challenge. Different countries have varying data formats, vocabularies and semantics and thus, interoperability is limited.

An interesting observation is that the data scientist interviewed mentioned challenges for future expansion of the tool however, the interviewee did not note any current issues with the tool. Conversely, a medical doctor whose team uses the tool stated that the team has challenges using the bulk input tool, namely, at the final step, the number of files that CEDAR reports as successfully uploaded does not match the number of records uploaded. Additionally, they cannot find the instances on CEDAR. The reports indicate a misalignment between the teams responsible for developing the tool and those using the tool.

**CEDAR**

The interviewee reported that transitioning from paper-based to digital has been a challenge. Many facilities are used to their way of working and it is difficult to shift to a different way of working. Additionally, the implementation of the tool requires technical expertise which are not always available locally. The interviewee reported that the team has been addressing these issues by providing training sessions. Moreover, teams require time to enter the data digitally and often, they do not have sufficient time to enter the patient data.

## Conclusion

The first step of our exploration phase aimed to gain a deeper understanding of the bulk input tool and the processes for data collection and analysis. Based on the interviews, it can be seen that within African countries, data are difficult to access. The majority of the records are documented on paper. This brings forward many challenges. The interviewees expressed that this was the motivation for developing tools and processes to facilitate data collection in a FAIR structure. The interviews revealed that progress has been made in digitising health data however, challenges still exist. Thus, in parallel to facilitating data digitilisation, alternative sources for data analysis may assist in identifying health trends. This has emphasised the problem that we need to work on. Firstly, ensuring that data follows FAIR. The challenges with usability and applicability of the relevant tools need to be addressed to increase adoption. Secondly, visualisation of the data from interviews and media sources will provide an indication if such sources are useful and reliable. These data sources could potentially be used to provide insights into populations where data are limited.