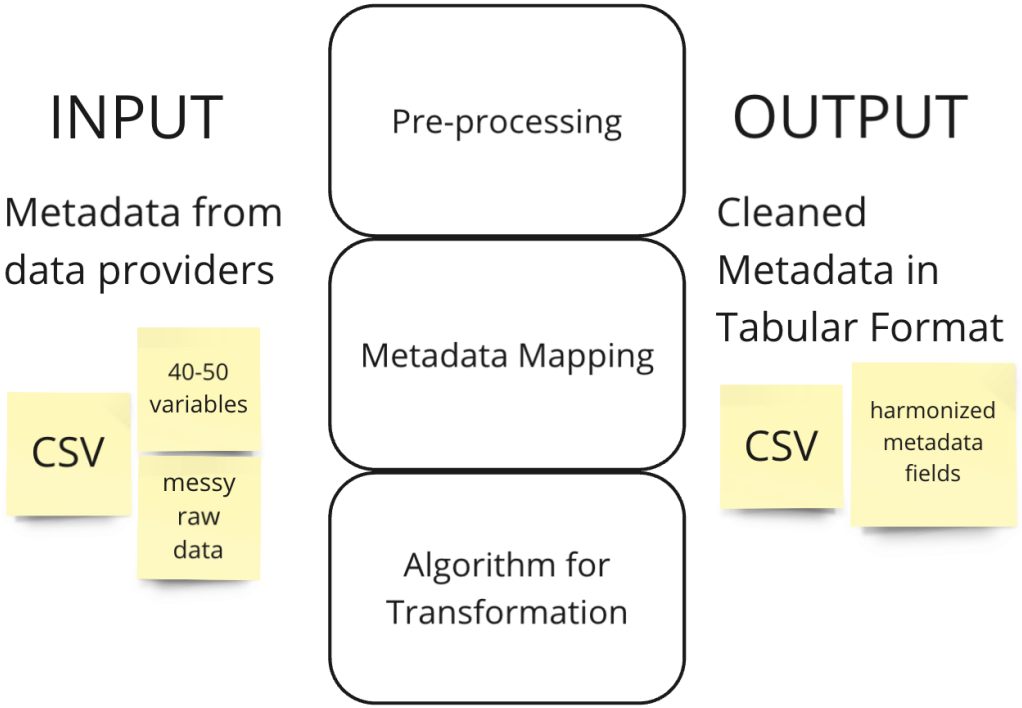
eLwazi Metadata Harmonization Tool

Wei Kheng Teh

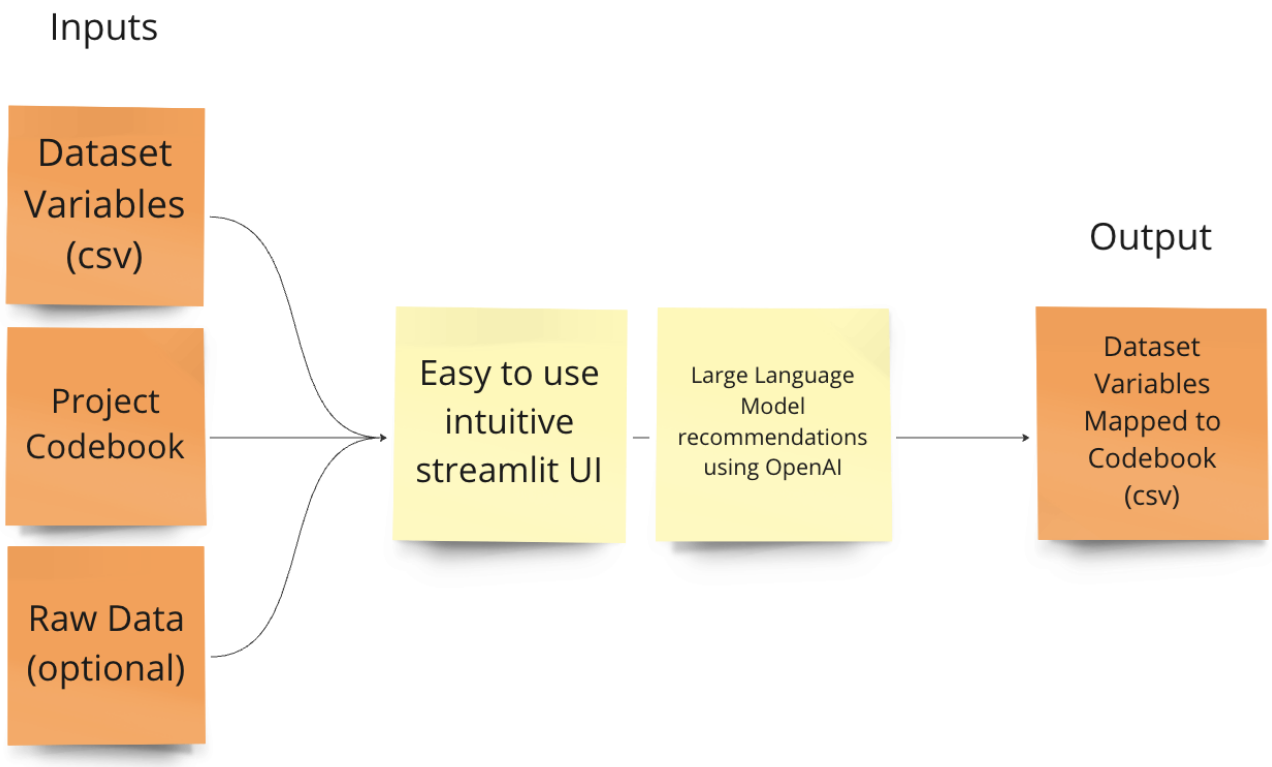
eLwazi

OPEN DATA SCIENCE PLATFORM

Metadata & Data Harmonization

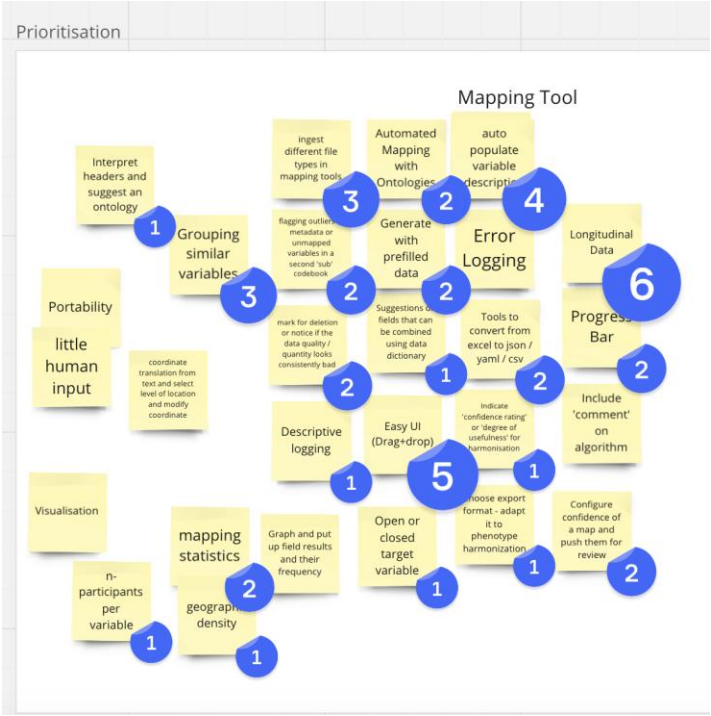


eLwazi Metadata Harmonization Tool



eLwazi Data Jamboree 2023

Collaboration with HE2AT Centre and other DMACs present



eLwazi Metadata Harmonization Tool

Development over time

OpenAI (costs) → Gemini (no cost)

Transformation mode

Relational mode

Open source code: https://github.com/elwazi/Metadata-Harmonisation-Tool

DS-I Africa
Data Science for Health Discovery and Innovation in Africa

Mapping App

Page

Map Studies

Study

Study1

View variables:

To do

About Unsort

Relational Mode Transform Mode

Please report any issues to the GitHub repository or contact peter.marsh@uct.ac.za for more information.

Select Variable To Map

weight_kg

Variable name and description:

variable_name	description
weight_kg	Weight in Kg

Example data:

```
82 ; 57 ; 99 ; 69 ; 53 ;
 ; 84 ; 53 ; 97 ; 58 ; 58 ; 95 ; 72 ; 81 ; 55 ; 51
```

Please complete the form below:

Does this map to any of these variables?

Weight in Kg - 100%

Patient ID:

None - 0%

Date:

None - 0%

Can this variable be mapped to our codebook?

- To do
- Successfully mapped
- Marked to reconsider
- Marked unmappable

Notes about this variable:

Auto Generate Transformation Instructions

Transformation instructions for this variable:

x

Type of transformation applied to this variable:

Direct

Preview of transformation:

```
82.0 ; 57.0 ; 99.0 ; 69.0 ; 53.0 ;
 ; 84.0 ; 53.0 ; 97.0 ; 58.0 ; 58.0 ; 95.0 ; 72.0 ; 81.0 ;
```

Installation Documentation

eLwazi

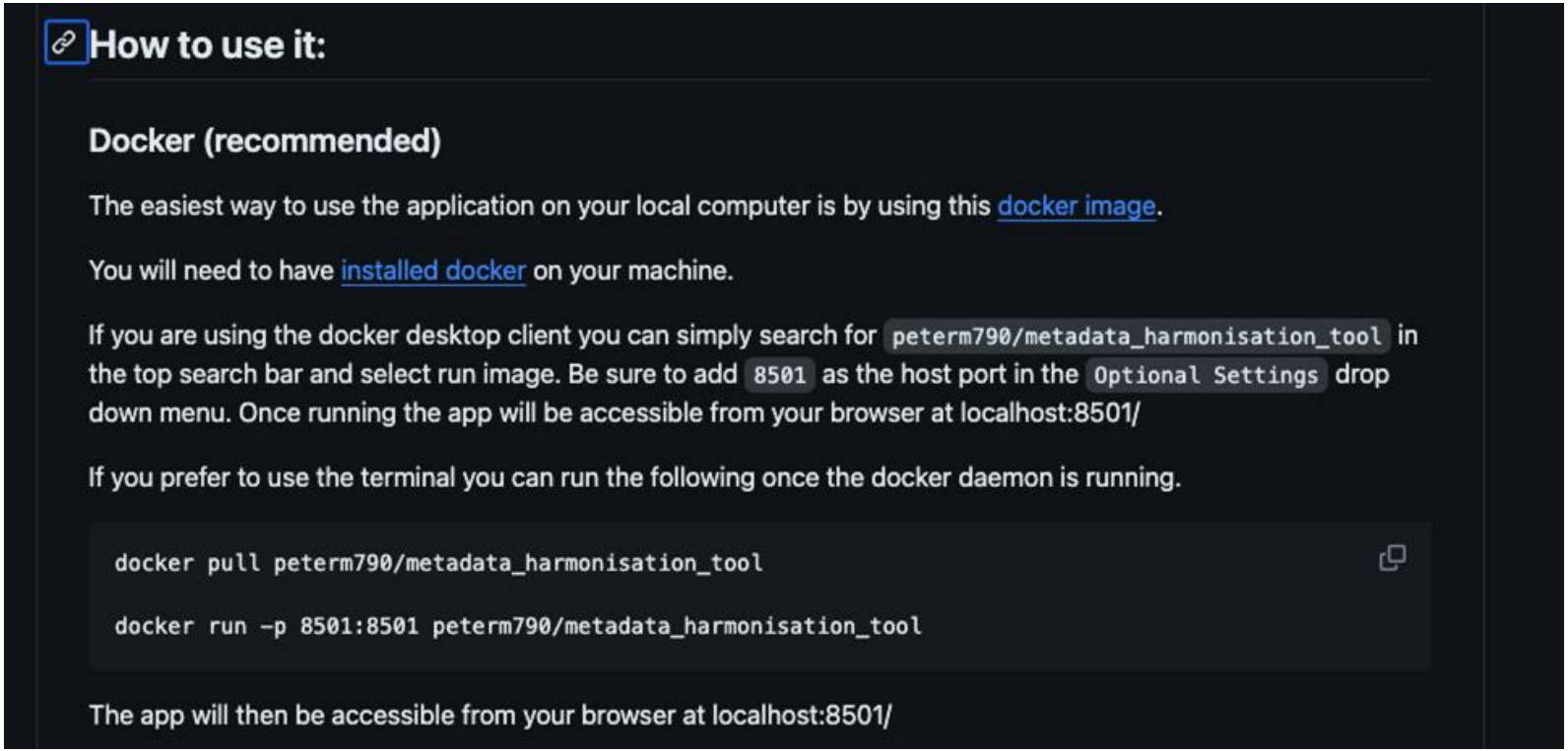
OPEN DATA SCIENCE PLATFORM

Links

Installation	https://github.com/elwazi/Metadata-Harmonisation-Tool?tab=readme-ov-file#how-to-use-it
Codebase	https://github.com/elwazi/Metadata-Harmonisation-Tool
Issues / Future Developments	https://github.com/elwazi/Metadata-Harmonisation-Tool/issues

Option 1: Docker (will not have all features)

https://hub.docker.com/r/peterm790/metadata_harmonisation_tool



How to use it:

Docker (recommended)

The easiest way to use the application on your local computer is by using this docker image.

You will need to have installed docker on your machine.

If you are using the docker desktop client you can simply search for `peterm790/metadata_harmonisation_tool` in the top search bar and select run image. Be sure to add `8501` as the host port in the `Optional Settings` drop down menu. Once running the app will be accessible from your browser at localhost:8501/

If you prefer to use the terminal you can run the following once the docker daemon is running.

```
docker pull peterm790/metadata_harmonisation_tool

docker run -p 8501:8501 peterm790/metadata_harmonisation_tool
```

The app will then be accessible from your browser at localhost:8501/

Option 2: Python Environment

Configure python environment

Alternatively if you are familiar with configuring python environments a suitable environment can be configured using conda.

If you do not have a preferred python package manager already installed I recommend installing Micromamba

```
git clone git@github.com:csag-uct/Health_Data_Harmonisation_Platform.git

cd Health_Data_Harmonisation_Platform

conda env create -f environment.yml -c conda-forge
conda activate harmonisation_env

pip install -r requirements.txt # some packages not available on conda channels

cd app/

streamlit run mapping_interface.py
```

The app will then be accessible from your browser at localhost:8501/

Please note the application requires a Unix like filesystem and so windows users will need to use WSL.

Future Steps

Development: https://github.com/elwazi/Metadata-Harmonisation-Tool/issues

- Backlog of issues
- Updating Docker Image
- Updating Documentation
- Training