

Application of ML models on FAIR HPV-HIV-FGS Infections data in Uganda

Basajja, M¹, Akandinda, M^{2,3}, Bwaga, I^{2,3}, Van Reisen, M³

1. Leiden Institute of Advanced Computer Science, Leiden University
2. Kampala International University, Western Campus, Uganda
3. Leiden University Medical Centre, Netherlands.

Problem Statement

There is lack of knowledge on the co-infection of HIV, HPV and FGS.

No data on Female genital Schistosomiasis, Urinary schistosomiasis, intestinal schistosomiasis, HIV and HPV data is collected for this study. The data is however not connected. There is a lack of well-defined ontologies, data ownership and data analytics in most health facilities in Uganda. This has led to duplication, lack of accessibility in the long term, lack of interoperability thus poor data quality. Two data sets identified for the preliminary study:

Point of Care dataset1 containing data on HIV, schistosomiasis and HPV for a period 2013-2022 (9 years) from the MOH Uganda and three districts: Lira, Entebbe and Hoima. It was logical to select these because of they are close to waterbodies usually having snails which transmit FGS, thus easily getting infected with HIV and HPV. The dataset contains the fields: district, period, number tested for HIV, number tested positive for HIV, urinary schistosomiasis, intestinal schistosomiasis, HPV1-Dose 1, HPV2-Dose 2 and HPV3-Dose 3 from the three different districts and MOH Uganda which are relevant for the study. The data sources HMIS002 (diagnosis field), HMIS OPD 105 Health Unit Outpatient Monthly report, HMIS 033B Health Unit Weekly epidemiology surveillance report, and HMIS 108 Health Unit Inpatient monthly registers in the health facility.

HMIS Form 105: Health Unit Outpatient Monthly Report
1.3.20 Neglected Tropical Diseases (NTDs)
NT04. Urinary Schistosomiasis

NT05. Intestinal Schistosomiasis

4.0 HIV/AIDS TESTING SERVICES (HTS)
HT01. Number of individuals
(by Entry Point for Health facility testing
HT02 Number of individuals (by Entry Point for Community testing)

2.7.4 VACCINE AVAILABILITY AND WASTAGE
WT09. HPV

HMIS 108: Health Unit Inpatient Monthly Report
6.1.5 Neglected Tropical Diseases (NTD's)
NT04. Urinary Schistosomiasis
NT05. Intestinal Schistosomiasis

HMIS 033B: HEALTH UNIT WEEKLY EPIDEMIOLOGICAL SURVEILLANCE REPORT
Disease of public health importance
Schistosomiasis (SC)

Introduction

Knowledge about HPV, HIV and FGS co-infection is still revolving.

3.4% of articles reviewed describe any sort of interaction of FGS, HPV and HIV

66% recommend integrated approaches for treatment within risk

communities of co-infection. No available strategy to reduce the co-infection.

Volume data collection is a critical enabler for AI approaches such as Machine

Learning. This study makes use of existing HIV, HPV and schistosomiasis data

from three districts: Lira, Hoima and Entebbe and Ministry of Health Uganda. The

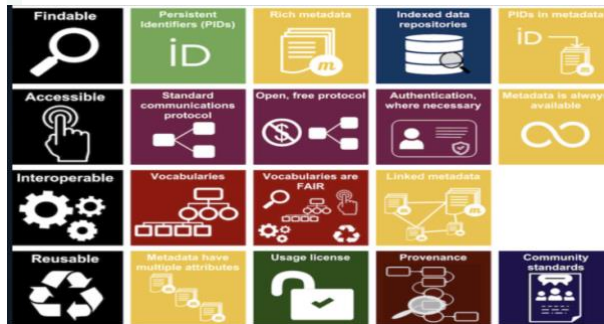
data is FAIRified using FAIRification tools to convert the data into a FAIR format

through template and ontology construction which creates the potential to make

knowledge graphs and ML model clusters and predictions from the data. Thus,

making the data FAIR does not only help for data reuse, but also for additional

analysis. Below is the figure showing the FAIR principles:



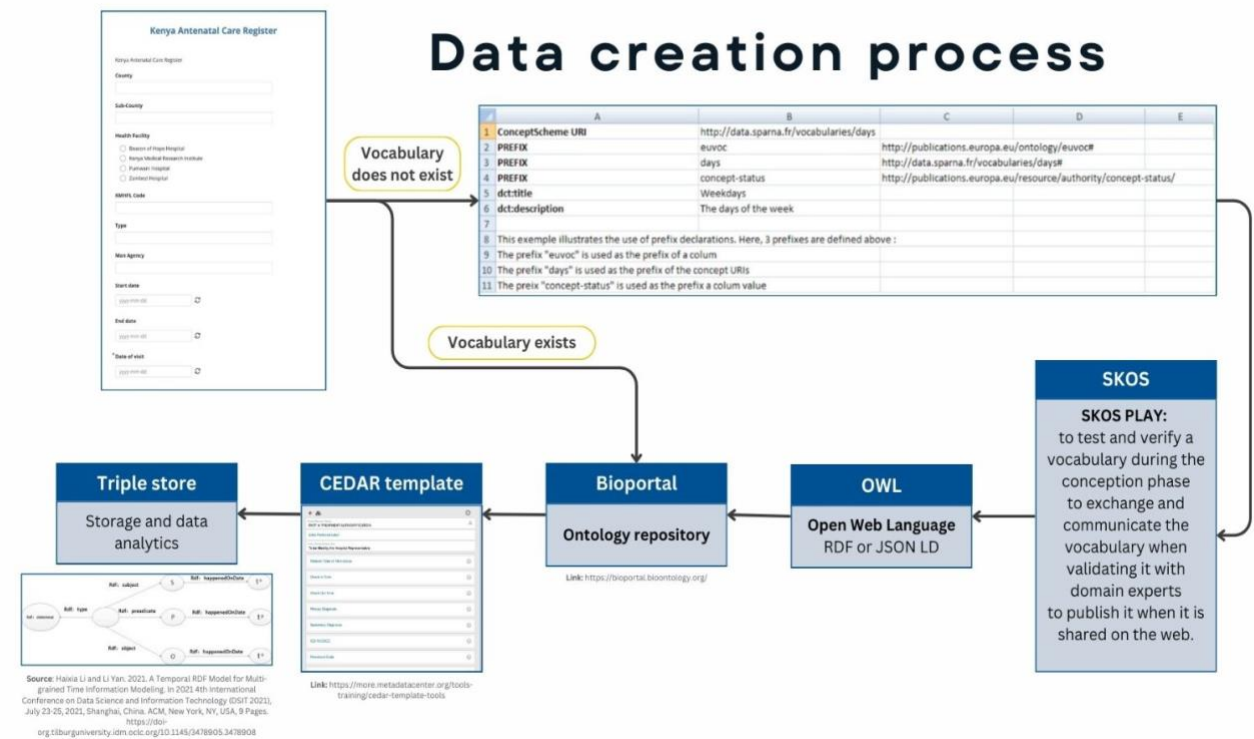
RESEARCH OBJECTIVE

To explore the use of Findable, accessible, Interoperable, Reuseable (FAIR) and semantically enriched data for identification of differentiated communities' exposure profiles to HPV, HIV and FGS. This study aims to:

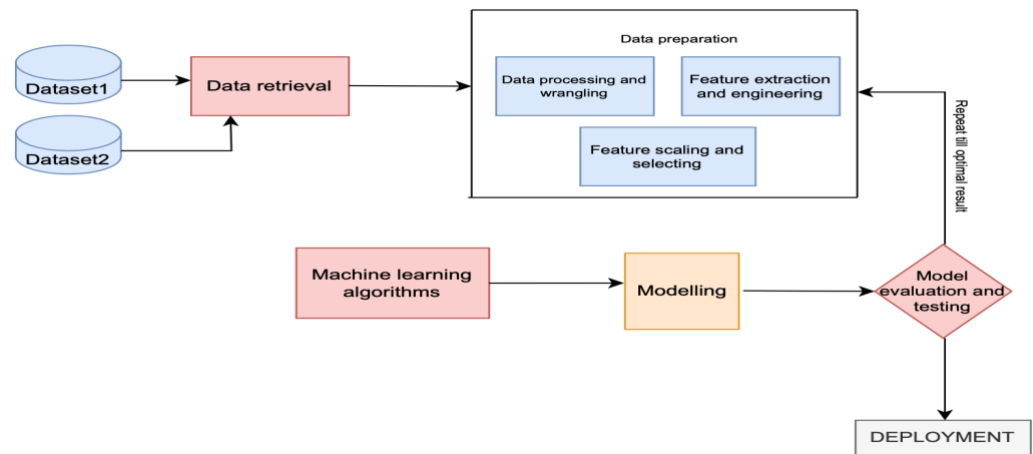
1. Make the data Findable Accessible Interoperable and Reusable under well-defined terms and represent the data using a knowledge graph.
2. Data analysis using python and machine learning models

Methodology

PHASE 1: The diagram showing the Data FAIRification process through Cedar template creation and ontology construction (Van Reisen,M)



PHASE 2: The diagram showing Data analysis using python and hierarchical clustering ML model (Basajja,M)



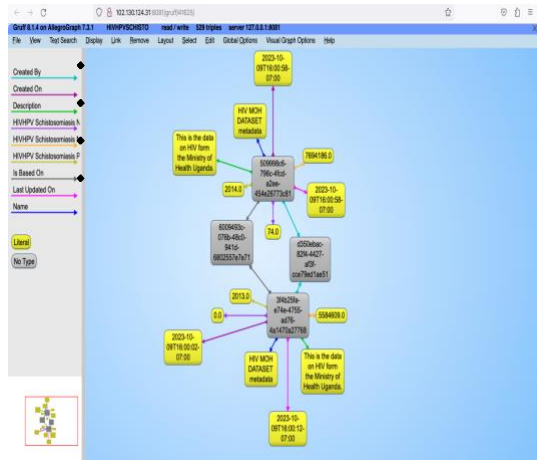
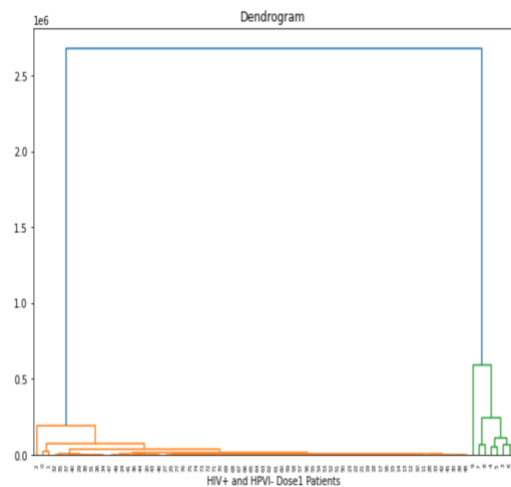
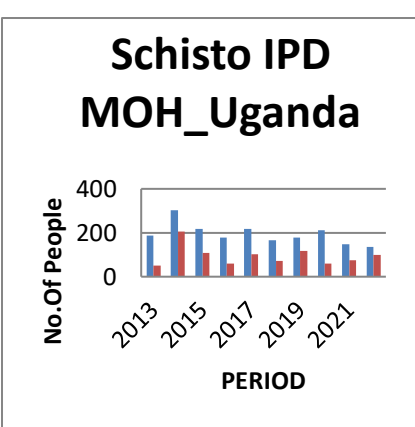
Preliminary results, Conclusion and future work

Results (Phase 1)

Digitized data collection tools created
Interoperability with well-defined Ontology for the data.
Knowledge graph created using the FAIRified data

Results (Phase 2)

As the number of vaccinated patients increases, the number of HIV and Schistosomiasis in patients decreases.
The number of intestinal schistosomiasis cases is consistently higher than urinary schistosomiasis number of patients in Uganda



Conclusion

1. The data is easily reusable after FAIRification
2. Data on schistosomiasis is poorly documented.
3. There's also inadequate data and data collection for measuring the co-infection of FGS, HPV and HIV data.
4. The data on Female genital schistosomiasis is not available in all health facilities, capitalization on FGS data
5. The current DHIS2 aggregate data is inadequate for analytics

Future work

1. Further research is required to investigate the different infection risk levels within more regions in Uganda.
2. A semantically enriched data template with different fields for FGS, HIV and HPV will be created and used for FAIRification approach at the source of creating the data instance.
3. A comparison of the FAIRified aggregate data within clinic data through the patient records will be carried out using machine learning algorithms.
4. Hierarchical clustering ML model will be used to categorize the different regions into closely related subgroups.