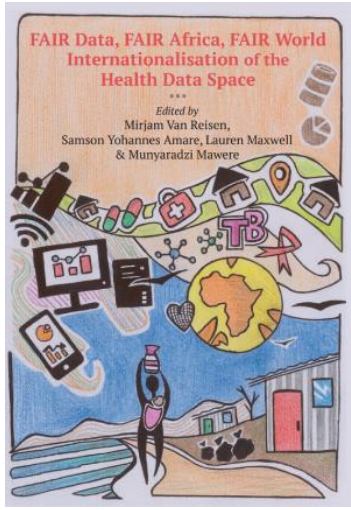
FAIR Data Implementation for Analysis of Research Data in Human Trafficking and Migration

Kai Smits, Mehul Upase, Nimish Pandey, Senjuti Bala & Mirjam van Reisen

Chapter in:
Fair Data Fair Africa Fair World:
Internationalisation of the Health Data Space



Cite as: Smits, K., Upase, M., Pandey, N., Bala, S., & Van Reisen, M. (2025). FAIR-data implementation for analysis of research data in human trafficking and migration. https://doi.org/10.5281/zenodo.15383037. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. (pp. 665–684). Bamenda: Langaa. URL: https://www.researchgate.net/publication/391750151_FAIR_Data_FAIR_Africa_FAIR_World_The_Internationalisation_of_the_Health_Data_Space

The About the Authors note can be found here: https://raee.eu/wp-content/uploads/2025/05/About-the-Authors-and-Editors.pdf
The list of figures and tables can be found here: https://raee.eu/wp-content/uploads/2025/05/List-of-Figures-and-Tables.pdf

Contents

Chapter 20

FAIR Data Implementation for Analysis of Research Data in Human Trafficking and Migration

Kai Smits, Mehul Upase, Nimish Pandey, Senjuti Bala & Mirjam van Reisen

Abstract

This study investigates the FAIRification of a dataset containing information on human trafficking. The FAIRification process involved employing standardised data formats and advanced semantic analytical tools. New insights were derived into the operational dynamics of trafficking networks. The study identified critical nodes and transit routes within these networks, providing a more detailed understanding of how trafficking operations are structured and executed. One of the key findings of this research is the feasibility of converting qualitative insights from survivor narratives into machine-actionable nodes of information. The identification of key nodes of transit routes within trafficking networks and the perpetrators involved provides researchers with valuable tools to extract meaningful insights from large and complex datasets. The FAIRification framework enhances the scalability and effectiveness of data-driven approaches in human trafficking research. While facilitating ethical data handling and ensuring the protection of sensitive information, the framework accommodates interoperability of data of varying provenance. It enables separate data storage and handling mechanisms, which is essential for maintaining the integrity and security of the data. The study concludes that the FAIRification of data will advance the human trafficking research domain.

Keywords: FAIR-data, FAIRification, data interoperability, human trafficking

Introduction

Human trafficking is a widespread criminal phenomenon that severely impacts the lives of its victims and survivors. These crimes take place in places that are hard to reach and unconnected, so-called 'black holes' in the digital landscape (Van Reisen et al., 2019). Human traffickers largely target vulnerable people, entrapping them in situations of exploitation (Shepherd et al., 2022). Refugees are one of the vulnerable populations who are targeted by traffickers. Traffickers use networks hidden within the complexities of international migration by exploiting the refugees. This means that the practices largely go undocumented. Therefore, all the research data that does exist about the problem is precious – however, due to security challenges and uncertainties concerning data sharing, the data often remains in isolation. Addressing such challenges requires innovative methodologies for human trafficking data, allowing researchers to effectively study and help combat this global issue.

Previous works, such as those by Ghardallou et al. (2022), Shelley (2010), and Laczko and Gozdziak (2005), have employed FAIRification, network analysis, and big data analytics to uncover patterns within trafficking networks. These studies highlight the potential of data-driven approaches to reveal critical nodes and operational dynamics in human trafficking networks. However, the lack of standardisation and interoperability among datasets has hindered their scalability and effectiveness.

In this chapter, we introduce a process of ontology creation and technologies to make human trafficking data more FAIR – Findable, Accessible, Interoperable, and Reusable (FAIR). We focus on datasets with testimonies of refugees who are survivors of human trafficking for ransom. Human trafficking for ransom, also called Sinai trafficking, is a form of human trafficking in which refugees are made captive and tortured on migration routes and their family members and community are extorted to pay their ransom (Van Reisen & Rijken, 2015). The FAIR principles guide our approach, making the research data more accessible, usable and secure. By applying FAIR principles we change how trafficking research data is curated, so we get a sustainable and reproducible research model that can be applied in different contexts and regions. This meets the

immediate data protection needs of the victims but also sets a framework for data sharing and collaboration and a precedent for future research.

This research addresses the need for a deeper understanding of traffickers' operational dynamics. It investigates key elements for the creation of an interoperable data system where researchers can dissect trafficker movements and identify trafficking hotspots.

Research design

The methodology involves a multi-step process designed to develop and implement a standardised, interoperable data system for human trafficking research. This process includes data collection and pre-processing, ontology creation, Resource Description Framework (RDF) data handling, tool development, strict adherence to ethical considerations, and General Data Protection Regulation (GDPR) compliance. This chapter was made possible through the work of two student Fieldlabs from 2023 and 2024.

This system, built on a qualitative dataset (with privacy and security sensitive data), shows the journey of each refugee, the travel points, the borders crossed, and the traffickers involved. The approach taken in this research includes the creation of an ontology for human trafficking and then applies various open-source tools to enable the querying and creation of knowledge graphs of the data. The approach was developed in three steps:

- Data retrieval and preparation layer, which handles data extraction, cleaning, and contextual enrichment using Python.
- Data storing layer, where the data is structured and managed using ontology and metadata frameworks in OWL and CEDAR.
- Visualisation and analysis layer, where the data is queried and presented through RDF and SPARQL endpoints in AllegroGraph.

The findings cover the data pre-processing strategy, ontology development, then the processing into RDF data and uploading in AllegroGraph. We then describe the development of analytical tools that enhance the ability of researchers to visualise the data. We then discuss the results, implications, and limitations of the study.

Dataset

The dataset reused in this chapter was first created to study human trafficking for ransom between the Horn of Africa and Europe, particularly in relation to human trafficking in Libya. The dataset is the result of the process of coding and labelling of 126 interviews held by Morgane Wirtz, Kai Smits, and Adoum Moussa. The interviews were ethnographic interviews with refugees and migrants, a majority of whom were from Eritrea. The interviews were held in Belgium, Italy, Ethiopia, Kenya, Libya, the Netherlands, Niger, Sudan, Sweden and Tunisia. The data from the interviews was analysed using a coding-labeling scheme which used purposive and open labelling. The resulting dataset was stored in a password-protected Excel sheet and was anonymised for this secondary research. The dataset has a high volume of unstructured text. The dataset was used in this chapter with the full permission of the researchers.

Data retrieval and preparation

Data retrieval and preparation included (i) data cleaning (ii) Large Language Model (LLM) structuring of data (iii) ontology creation and (iv) ensuring ontology qualities were related to the FAIR principles.

Data cleaning

The initial dataset received from the researchers contained 49 columns of unstructured data. Since the dataset had a high volume of information and a lack of structure, it required a thorough cleaning and pre-processing procedure to start the process of standardisation. An interview with one of the researchers also revealed that the dataset had standardisation issues, such as different spellings of names of locations due to Romanisation of Arabic location names and names of persons, as well as minor spelling errors. In addition, there were inconsistent fields, such as the nationality field, which could appear as 'Eritrean' or 'from Eritrea', for example. The qualitative data also included narrative sections that varied widely in detail and structure. Another problem was missing data, which was marked differently, including empty fields, but also written input such as 'not mentioned' or 'unknown'.

These problems greatly impact the dataset's usability and accuracy. Incomplete and inconsistent data hinders the ability to identify

patterns, compare results, or reach significant conclusions. Furthermore, the lack of standardisation diminishes the dataset's compatibility with external sources and tools.

To pre-process the data the following was implemented:

- Column reduction and parameter selection: We reviewed and reduced the initial 49 columns to 22 essential columns, focusing on unique identifiers, geographical locations, and trafficker information. During the second portion of the research, further columns were added on experiences and witnessed abuses, trafficker affiliations, and extortion details.
- Sequence mapping and event tracing: We implemented a system of serial numbering linked to each victim's unique identifier to maintain chronological integrity.
- Extraction of names from unstructured text: We developed a Python script to identify and extract names, creating a 'names' column.
- Geographic Coordinate Mapping: We used a Geographic Information System (GIS) framework and Google Maps Application Programming Interface (API) to standardise and verify geographical data.

Misspellings and inconsistencies in the data were corrected through dictionary lookups. The dictionaries used for standardisation were created through a mix of manual observation and automated suggestions from locally hosted language models. These LLMs scanned columns to identify unique strings and suggested potential synonyms or spelling corrections. Each suggestion was manually reviewed to ensure accuracy, especially for sensitive data like trafficker names. This process reduced inconsistencies and simplified further analysis. Furthermore, *Regex* was used to standardise numerical values such as dates and numbers expressed in currencies. Forward fill (ffill) and backward fill (bfill) operations were used in cases where data was missing.

The structured data was saved in JavaScript Object Notation for Linked Data (JSON) format, which allowed for easy conversion into RDF data. The final JSON output was formatted with indentation (use of tabs) to enhance readability, ensuring that the data could be revisited and expanded upon as needed for future queries or analyses.

Structuring of data through LLMs

Retrieval-augmented generation (RAG) through LLMs was explored to extract data directly from the interview transcripts to enhance the information available from the Excel data table. However, this approach was abandoned due to significant problems, described in the results.

Ontology creation

Ontology creation is a key step for producing linked data. The first step in the ontology creation process was looking at existing ontologies. It was determined that there were no comprehensive ontologies for human trafficking.

The ontology included important entities like Victim, Trafficker, Crime, Border Crossing, and Payment. These entities were linked together using specific relationships, showing how they interact within the dataset. The structure followed OWL (Web Ontology Language) standards, which allows for easy updates and compatibility with other ontologies if needed later on. The ontology was visually developed and validated using Protégé.

Ontology qualities based on FAIR principles

We developed an ontology framework for human trafficking using OWL within BioPortal. Bioportal is supported by the Center for Expanded Data Annotation, and Retrieval (CEDAR) for ontology management. This framework ensured that our data was organised and categorised systematically.

To implement FAIR principles, we:

- Ensured data findability through robust metadata and indexing
- Enhanced accessibility by creating user-friendly interfaces for data interaction.
- Promoted interoperability by standardising data formats and using common ontologies.
- Ensured reusability through documentation and ethical data management practices.

The FAIRification process in this research is focused on the semantic tools used for the creation of interoperability.

Data storing

The pre-processed dataset was translated into RDF data using Python's rdflib package. RDF is a widely accepted standard for data exchange. In the RDF data, each refugee's record was represented by a unique Uniform Resource Identifier (URI) that included attributes such as BordersCrossed, Gender, Nationality, and TraffickerName. The RDF data was eventually translated to Turtle format and integrated with AllegroGraph, a triple store and Knowledge Graph Platform. The data storage to AllegroGraph involved the upload of the Turtle-format in RDF data.

Data querying and visualisation

The repositing of the data in AllegroGraph allowed the performance of the following analytical querying of the data

- Utilisation of the AllegroGraph and Gephi capabilities for complex querying and data visualisation.
- Execution of Query Language and Resource Description Framework (SPARQL) queries to extract and analyse specific data segments in the AllegroGraph repository.

The data was visualised using SPARQL queries and Gruff (AllegroGraph's tool for data visualisation and queries).

To enhance user-friendly analysis of the data, we developed an interactive web application using Streamlit designed to enable dynamic exploration and visualisation of the RDF data.

In our set up the application featured interactive dropdown menus for selecting specific refugees or traffickers and visualising their profiles and movements. Additionally, we used Gephi for advanced network visualisation, creating detailed network graphs that illustrate the relationships and interactions within trafficking networks. Gephi is an open-source network analysis and visualisation software package written in Java on the NetBeans platform.

Ethical and data management considerations

Due to the sensitivity of the whole project, ethical considerations held an important position in our study. We conducted a Data Protection Impact Assessment (DPIA) to ensure compliance with GDPR and

protect the privacy and dignity of the individuals involved. Key measures included:

- Anonymising personal identifiers and using unique identification numbers.
- Implementing robust security protocols and encryption techniques.
- Ensuring transparency and ethical data handling practices at every stage of the research.

The data for the research was collected and processed under the Ethics and Data Management Committee (REDC) (identification code: REDC# 2020/01 3a) of Tilburg University. All students working with the data received anonymised data and signed an agreement that stipulated how the data should be handled.

Relevance

Human trafficking has been extensively studied from sociological, legal, and criminological perspectives. Researchers like Shelley (2010) and Laczko and Gozdziak (2005) have described the complex nature of trafficking networks, focusing on the socio-economic factors that facilitate such crimes. Recent advancements in data science, such as network analysis and big data analytics, have begun to illuminate the structural and operational dynamics of trafficking networks (UNODC, 2018; Latonero, 2011). However, the lack of standardisation and interoperability among datasets remains a significant challenge, limiting the scalability and effectiveness of these methods (Thompson & Black, 2020).

Despite the advancements, the analysis of trafficking data is challenging due to its sensitivity and the secretive nature of these hidden networks. Particularly in areas like human trafficking, ransom data is often lacking. Van Reisen and Rijken (2015) highlight the extreme violence of Sinai trafficking and the need for verifiable data to address such crimes. Van Reisen et al. (2019) highlight the effect of the lack of information on human trafficking due to it being hidden, often deliberately.

The problem that this research aims to address is therefore the lack of reliable research data on human trafficking. Where such data is present, it is isolated, and the lack of standardisation and

interoperability prevents reusing of the data. This hinders comprehensive analysis and actionable insights into trafficking networks. The sensitivity and unstructured nature of trafficking data, often gathered through interviews and testimonies, pose additional challenges. The balance between data utility and privacy, as discussed by Thompson & Black (2020), necessitates robust data protection strategies.

Our project addresses these gaps by proposing the implementation of FAIR principles in the management of human trafficking research data. By applying these principles, we aim to create a standardised, interoperable data system that enhances data sharing, collaboration, and ultimately, the effectiveness of interventions in the fight against human trafficking.

Results

Our study investigates the development of a comprehensive and interoperable data system for human trafficking research. The key findings are focused on data retrieval, data storage, and the application of analytical queries and tools for data visualisation.

Data retrieval and preparation

This section details the processes involved in preparing the data for analysis and developing a structured ontology to categorise the data. From the initial unstructured data in the Excel sheet, a Comma-Separated Values (CSV) file was created with standardised columns. This included:

- Demographic data (age, gender, nationality)
- Experiences and witnessed abuses
- Trafficker affiliations
- Border crossings and extortion details

Inconsistencies and differences in spelling in the dataset were corrected and unified. Complex fields such as payments of ransoms needed more detailed parsing. This label, for example, needed the numerical value as well as the currency label to be extracted. This was done using regex patterns. In addition, other issues such as missing data fields were resolved (see methodology for details). A sample of the final processed columns can be found in figure 1 below.

Sr. No.	Unique ID	Interviewer	Date of Interview	Gender	Nationality	Left Home Country Year	Final Locations	Borders Crossed [DONE]	City / Locations Crossed [DONE]	Locations in Libya [DONE]
16	4	Kiara Smits					Libya, Sudan, Tripoli, Lebanon, Tripoli, Libya	Libya, Sudan	Tripoli, Lebanon	Tripoli, Libya
16	4	Kiara Smits					Bani Walid, Libya, Bani Walid, Libya		Bani Walid, Libya	Bani Walid, Libya
17	5	Kiara Smits	19/12/2019	Male	Eritrean		Libya, Bani Walid, Libya, Bani Walid, Libya	Libya	Bani Walid, Libya	Bani Walid, Libya
17	5	Kiara Smits					Bani Walid, Libya			Bani Walid, Libya
17	5	Kiara Smits					Sudan,	Sudan		
18	3	Kiara Smits	17/12/2019	Male	Eritrean	2016	Eritrea, Ethiopia,	Eritrea, Ethiopia		
18	3	Kiara Smits					Ethiopia, Sudan,	Ethiopia, Sudan		
18	3	Kiara Smits					Khartoum, Sudan,		Khartoum, Sudan	
18	3	Kiara Smits					Chad, Sudan,	Chad, Sudan		
18	3	Kiara Smits					Chad, Libya, Tripoli, Lebanon, Tripoli, Libya	Chad, Libya	Tripoli, Lebanon	Tripoli, Libya
18	3	Kiara Smits					Tripoli, Lebanon, Tripoli, Libya		Tripoli, Lebanon	Tripoli, Libya
18	3	Kiara Smits					Khartoum, Sudan,		Khartoum, Sudan	
18	3	Kiara Smits					Italy, Netherlands,	Italy, Netherlands		
19	3	Kiara Smits	17/12/2019	Male	Eritrean	2009	Eritrea, Ethiopia,	Eritrea, Ethiopia		
19	3	Kiara Smits					Ethiopia, Sudan,	Ethiopia, Sudan		
19	3	Kiara Smits					Sudan,	Sudan		
19	3	Kiara Smits					Israel,	Israel		
19	3	Kiara Smits					Libya, Sudan,	Libya, Sudan		
20	13	Kiara Smits	06/01/2019	Male	Eritrean	2016	Eritrea, Ethiopia,	Eritrea, Ethiopia		
20	13	Kiara Smits					Ethiopia, Sudan,	Ethiopia, Sudan		
20	13	Kiara Smits					Khartoum, Sudan,		Khartoum, Sudan	

Figure 1. Sample of the final processed columns

The structured dataset was then saved in JSON format, creating a machine-readable as well as human-interpretable format which could be transformed into RDF data. Each interviewee's data was consolidated into a single nested dictionary containing identifiers (e.g., victim_id), age, gender, nationality, experiences of abuse, border crossings, and any financial extortion reported. The sub-dictionaries used to categorise different data aspects were nested. This was used to organise different aspects of the data, such as a list of crimes, borders crossed which captured origins and destinations, along with the corresponding ransom sums, and financial data which was stored under a distinct key, which enabled tracking of payments and currency types.

Structuring of data retrieval through LLMs

RAG through LLMs was explored to extract data from the interview transcripts directly. However, significant problems were identified with this approach. Although a privately hosted RAG was used, we felt that exposure risks for the sensitive data could insufficiently be ruled out. In addition, preliminary outputs showed inconsistencies with sensitive terms. The raw interview data also proved to be too unstructured to produce consistent results. Fine-tuning the model would have taken too much time. Therefore, the methods mentioned in the previous section for data cleaning and retrieval were used to have more control over the data extracted.

Semantic FAIR-based ontology creation

The development of a human trafficking ontology was a key step in organising the dataset for RDF transformation. The ontology was created according to the following logic:

- Victim class: this class includes properties like age, gender, nationality, and experiences of abuse. Victims were connected to traffickers through relationships like 'affiliatedWith' or 'exploitedBy'.
- Abuse hierarchy: the ontology organised abuse into categories such as physical abuse, sexual abuse, and psychological abuse. This structure, shown in the ontology graph, allowed for detailed classification while keeping queries simple.
- Trafficker class: traffickers were linked to crimes and victims, with identifiers and aliases represented as simple values and object properties.
- Border crossing events: these captured financial extortion, origin and destination locations, and currencies. The border class connected entities like victim and location, creating a clear path of movement and exploitation.
- Crime and death witnessing: events like sexual violence and death were modelled as subclasses of crime, highlighting the seriousness of incidents reported by victims.

OWL standards were used in the creation of the ontology. To ensure that synonymous terms would be included in the ontology, alternative labels were implemented through Simple Knowledge Organization System (SKOS). For example, 'PhysicalAbuse' included labels like beating, injury, and shooting as alternative labels. The ontology was mapped and validated through Protégé, which allowed for the visual checking and correcting of the ontology, as gaps were easily visible (see figure 2).

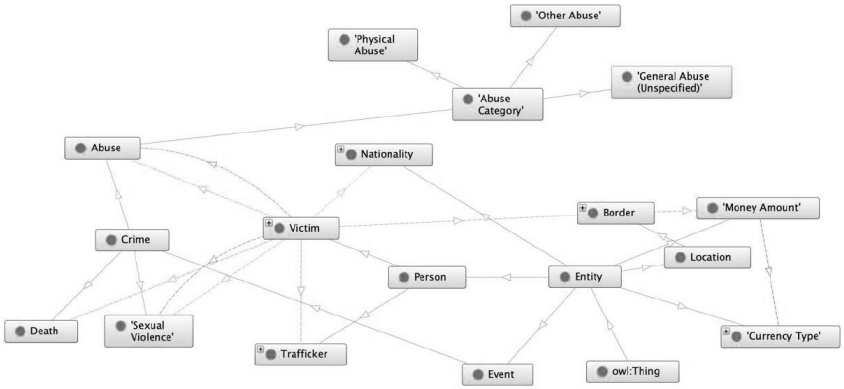


Figure 2. Mapping of the ontology

The final ontology map was then exported in JSON (see figure 3) and integrated into the RDF pipeline in the data-storing phase.

```
ex:wasTraffickedBy a owl:ObjectProperty ;
    rdfs:label "was Trafficked By" ;
    dct:description "Links a victim to the trafficker involved in the incident." ;
    rdfs:domain ex:Victim ;
    rdfs:range ex:Trafficker .

ex:experiencedSexualViolence a owl:ObjectProperty ;
    rdfs:label "experienced Sexual Violence" ;
    dct:description "Indicates the victim experienced sexual violence." ;
    rdfs:domain ex:Victim ;
    rdfs:range ex:SexualViolence .

ex:witnessedSexualViolence a owl:ObjectProperty ;
    rdfs:label "Witnessed Sexual Violence" ;
    dct:description "Indicates the victim witnessed sexual violence." ;
    rdfs:domain ex:Victim ;
    rdfs:range ex:SexualViolence .

ex:witnessedDeath a owl:ObjectProperty ;
    rdfs:label "witnessed Death" ;
    dct:description "Specifies the victim witnessed a death." ;
```

Figure 3. Ontology exported in JSON

Data storing

The cleaned and standardised dataset enabled the conversion into RDF data. Each entity in the JSON dataset was mapped to the corresponding class in the ontology so that all the terms were semantically aligned.

The RDF data was created using the RDFlib library in Python. The Python script required for this was complex, as it needed to take into account the nested data structures, missing data, and alignment with the ontology. RDFlib enabled the management of the ontology and

the data transformation. First, the ontology file was uploaded, which established the classes and relationships which would structure the data. Then, the cleaned JSON data file was uploaded. Unique identifiers were assigned to each victim, trafficker, and border-crossing event. Attributes like age and gender were processed as literals, while attributes such as traffickers and nationalities were linked through dedicated URIs. The RDF triples were stored in Turtle format.

The processed RDF data was uploaded to AllegroGraph, a triple store, to enable complex queries and data visualisation. This integration allowed for the effective management of human trafficking data within the FAIR framework.

Data querying and visualisation

The data was visualised first of all by creating knowledge graphs through Gruff. The image below (figure 4) shows the resulting graph from the selection of two victim testimonies.

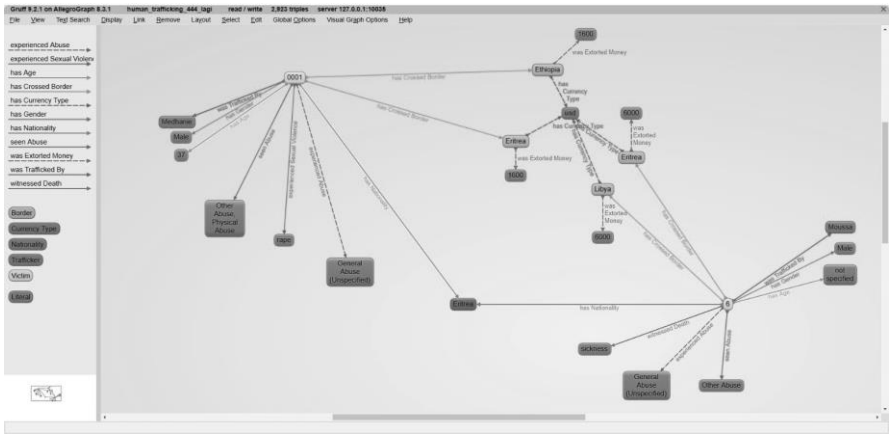


Figure 4. Visualisation of 2 testimonies through Gruff

The graph reveals detailed information and relationships between the two victims. For instance, both victims have the nationality 'Eritrea'. The graph also shows the money that was extorted from them in different locations, as well as the abuses they witnessed. It also identified who they were trafficked by.

Queries were also used to show how the data can be visualised. To gain deeper insights into the collaborative operations of traffickers,

we executed a query to identify locations where multiple traffickers worked together (figure 5).

```
PREFIX ex: <https://github.com/surya-parthipan/DSIP_FieldLab_7/tree/main/>

SELECT ?refugee ?trafficker1 ?trafficker2 ?location
WHERE {
  ?refugee ex:hasTraffickerName ?trafficker1 ;
           ex:hasTraffickerName ?trafficker2 ;
           ex:hasBordersCrossed ?location .
  FILTER (?trafficker1 != ?trafficker2)
  FILTER (STR(?trafficker1) < STR(?trafficker2))
}
GROUP BY ?refugee ?trafficker1 ?trafficker2 ?location
```

1,000 ROWS DOWNLOAD RESULTS QUERY INFORMATION

refugee	trafficker1	trafficker2	location
refugee-1043	"Muhamad"	"Tawila"	"Gambia"
refugee-1030	"Abdela"	"John"	"Agadez"
refugee-4	"Cegora"	"Kashi"	"Ethiopia"
refugee-1023	"Tuemuzgi"	"Walid"	"Egypt"
refugee-2007-1	"Tim"	"Wedi Isaac"	"Libya"
refugee-1019	"Aziz"	"El Amo"	"Tunisia"
refugee-1008	"Take"	"Walid"	"Sudan"
refugee-1027	"Aforki"	"Mahmoud Abdel Salam"	"Ethiopia"

Figure 5. Query and the result to get insights on locations where two traffickers worked together

Many other queries were tested, including for example queries on which borders were most often crossed by the interviewees, which were the highest amounts extorted in which locations, and which abuses were most observed.

To enhance the accessibility and utility of our data, we developed several analytical tools, including an interactive dashboard and visualisations using Gephi (see figure 5). These tools provide dynamic capabilities for exploring the dataset, revealing critical insights into trafficking networks.

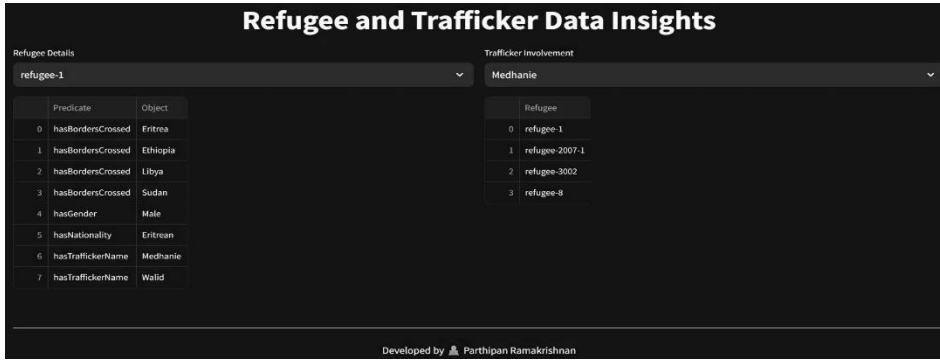


Figure 6. Interactive dashboard: Showcasing dynamic selection and visualisation of refugee data, including individual profiles, trafficker information, and border crossings

The interactive dashboard allows users to select and visualise various aspects of the dataset, such as individual profiles, trafficker

information, and border crossings. This tool facilitates a more nuanced understanding of the trafficking networks and their operations.

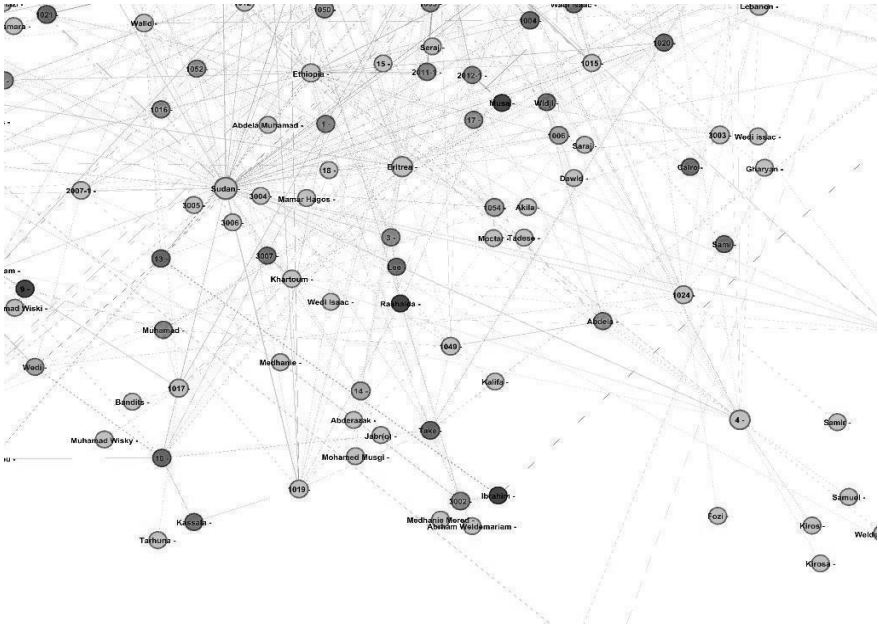


Figure 7. Interactive visualisation of human trafficking data in Gephi

Gephi was used for the network visualisation of the trafficking data. It shows an interactive visualisation, where nodes represent entities (e.g., traffickers, victims, locations), and edges represent the relationships between them. Gephi can highlight connected nodes when hovered over, allowing for a more detailed exploration of the network's structure and key relationships (see figure 6).

Discussion

Our study has demonstrated the effectiveness of applying FAIR principles to tooling used for the management of human trafficking research data. The developed data system enhances the accessibility and utility of research data by enabling a more structured and interoperable approach. The interactive tools and visualisations, such as the dashboard and Gephi network graphs, have provided new insights into the operational dynamics of human trafficking networks, revealing critical nodes and transit routes. It also made a first attempt

at developing an ontology for human trafficking, which can be built upon and validated.

The integration of semantic FAIR principles has significantly improved the accessibility and reusability of the human trafficking research data used in this research and has the potential to be expanded to other data. It resulted in the creation of a human trafficking ontology that can be validated and built upon by other domain experts.

While our project offers valuable contributions, it is not without limitations. The research data is not completely FAIR through the steps investigated in this study. The steps do not make the data files findable, and their accessibility is limited by the lack of findability. In addition, the research did not set up a framework for access control of the sensitive data – now, it relies on manual granting of access. In addition, the interoperability with other data is possible in theory but has not yet been tested. This enhances the reusability of the data. Hence the elements of accessibility, interoperability, and reusability were addressed. For the data to be completely FAIR, additional steps still have to be completed. Further steps need to be undertaken to improve the FAIRification of the data at all levels of the FAIR facets described by Wilkinson (2016), most particularly concerning the findability of the data and data catalogues through a FAIR Data Point.

The research has demonstrated the approach but only covers one dataset. In addition, the ontology that was created should be validated by other domain experts. Additionally, the complexity and sensitivity of the data posed challenges in ensuring complete accuracy during the pre-processing and analysis stages.

Our findings corroborate previous studies on the importance of network analysis in understanding human trafficking (UNODC, 2018; Latonero, 2011). Our approach to applying FAIR principles has provided a more standardised and interoperable framework, enabling more comprehensive analyses and potentially more effective interventions. For future research, it is important to validate these tools and use them on other data, which can then be linked.

The challenges that were faced in this research included balancing the need for structuring and standardisation of the data while not losing

the key meaning of the data. This required challenging data structures, particularly as the data was nested, which required complex parsing and processing in the Python script. This caused issues that needed a lot of time for debugging. Another challenge was that unstructured legacy data is quite labour-intensive to transform. However, if the process of data structuring and labelling the data semantically would be integrated into the start of data collection, it would be much more efficient to implement.

This research contributes to the field by offering a robust framework for data standardisation and interoperability, which can be adopted by future research initiatives. The framework's ability to handle heterogeneous data sources is a significant strength, as it allows researchers to integrate and analyse data from various origins. This interoperability is a cornerstone of the FAIR principles and is essential for advancing research in complex and multifaceted areas such as human trafficking.

This research underscores the transformative potential of applying FAIR principles to enhance interoperability within the human trafficking research domain. By addressing data accessibility, interoperability, and reusability, the study paves the way for more effective and ethical research practices. The adoption of FAIR guidelines not only improves the management and utility of research data but also fosters a more collaborative and insightful research environment.

Therefore, the findings of this research have broader implications for other fields that rely on complex and sensitive data. The successful application of FAIR principles in human trafficking research serves as a model for how these principles can be leveraged to enhance data management and analysis in various disciplines. As research continues to evolve, the adoption of FAIR principles will likely become increasingly important in ensuring that data is used effectively and ethically to address some of society's most pressing and sensitive challenges.

The findings from this research have practical applications for policymakers, practitioners, and researchers. The tools developed, such as the interactive dashboard and Gephi visualisations, enable

more detailed and actionable analyses of trafficking data. Policymakers can use these insights to develop targeted interventions and allocation of resources in more effective ways, while researchers can build on this framework to conduct further studies in the field. The insights gained from this research not only contribute to improving human trafficking research data management but also have the potential to make a tangible impact in the fight against human trafficking. The research can inform prevention and intervention strategies. Continued efforts in this direction will be crucial in addressing this pervasive global issue and protecting vulnerable populations.

Conclusion

This research has elucidated key elements essential for the development and application of FAIR principles to enhance the management and utility of human trafficking research data. The study conducted an in-depth investigation into the FAIRification of a dataset containing information on human trafficking. By employing standardised data formats and advanced semantic analytical tools, new insights were derived regarding the operational dynamics of trafficking networks, thereby identifying critical nodes and transit routes.

The findings of this research underscore the significant advantages associated with the adoption of FAIR guidelines in the data-handling practices within the human trafficking research domain. Specifically, the study demonstrated the feasibility of converting qualitative insights from survivor narratives into machine-actionable nodes of information, which can be seamlessly integrated with other research datasets. Key nodes of transit routes within trafficking networks and perpetrators were successfully identified, providing researchers with valuable tools to extract insights from voluminous datasets.

The research highlights the importance of FAIR principles in advancing human trafficking research. By improving data accessibility, interoperability, and reusability, the study demonstrates how these principles can be applied to enhance the management and utility of research data. The insights gained from this research have the potential to inform future studies and contribute to the

development of more effective strategies for combating human trafficking.

The proposed FAIRification framework markedly enhances the scalability and effectiveness of data-driven approaches in human trafficking research, particularly in the context of heterogeneous data sources. It facilitates ethical data handling and ensures the protection of sensitive information, accommodating data of varying provenance and enabling separate data storage and handling mechanisms. This transformation is crucial as it allows for the seamless integration of insights of one set with other datasets, thereby enriching the overall data ecosystem on human trafficking.

Acknowledgements

This chapter is partially based on the reports of student Fieldlabs for the course Data Science in Practice (DSIP) at Leiden University. The students that contributed to those reports are Mehul Upase, Nimish Pandey, Nina van Rossum, Parthipan Ramakrishnan, Senjuti Bala (Fieldlab 2023); and Sasank Amavarapu, Chiesa Serena, Ayush Sengar, Maarten Wilders, Zahra Abedi (Fieldlab 2024)

The authors would like to thank Europe External Programme in Africa (EEPA) for its support, through the projects 'Europe-Africa Response to Protection of Migrants and Refugees and Human Trafficking Data' and 'Safeguarding Vulnerable Communities: Strengthening FAIR Interoperable Data Stewardship in Humanitarian Organizations through Secure Datapods'.

This chapter is partially based on the reports of student Fieldlabs for the course Data Science in Practice (DSIP) at Leiden University. The students that contributed to those reports are Mehul Upase, Nimish Pandey, Nina van Rossum, Parthipan Ramakrishnan, Senjuti Bala (Fieldlab 2023); and Sasank Amavarapu, Chiesa Serena, Ayush Sengar, Maarten Wilders, Zahra Abedi (Fieldlab 2024)

Authors' Contributions

Kai Smits supervised the work of the students in the Fieldlabs and wrote the final version of the chapter. Smits is also one of the researchers that collected and processed the data used in this chapter.

Mehul Upase, Nimish Pandey, and Senjuti Bala all contributed to the first draft of the chapter.

The chapter draws on the work and reports of Mehul Upase, Nimish Pandey, Nina van Rossum, Parthipan Ramakrishnan, Senjuti Bala; Sasank Amavarapu, Chiesa Serena, Ayush Sengar, Maarten Wilders, and Zahra Abedi.

Ethical Considerations

Ethical considerations for this research were obtained from Tilburg University REC2017/16; REDC # 2020n13; REDC# 2020/01 3a; REDC 2020.139.

Tilburg University, Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences REDC#2020/013, June 1, 2020-May 31, 2024, on Social Dynamics of Digital Innovation in remote non-western communities. Uganda National Council for Science and Technology, Reference IS18ES, July 23, 2019-July 23, 2023.

References

Ghardallou, N., et al. (2022). FAIRification of migrants' data during the COVID-19 pandemic. *Data & Society*.

Laczko, F., & Gozdziak, E. (2005). Data and research on human trafficking: A global survey. *International Organization for Migration*.

Latonero, M. (2011). Human trafficking online: The role of social networking sites and online classifieds. *Center on Communication Leadership & Policy*.

Van Reisen, M., & Rijken, C. (2015). Sinai trafficking: Origin and definition of a new form of human trafficking. *Social Inclusion*.

Van Reisen, M., Mawere, M., Stokmans, M., Nakazibwe, P., Van Stam, G., & Ong'ayo, A. O. (2019). Black holes in the global digital landscape: The fuelling of human trafficking on the African continent. *In Mobile Africa: Human Trafficking and the Digital Divide* (pp. 3–31). Langaa RPCIG. https://doi.org/10.2307/j.ctvvh85s6.8

Shelley, L. (2010). *Human trafficking: A global perspective*. Cambridge University Press.

Shepherd, D. A., Parida, V., Williams, T., & Wincent, J. (2022). Organizing the exploitation of vulnerable people: A qualitative assessment of human trafficking. *Journal of Management, 48*(8), 2421-2457. https://doi.org/10.1177/01492063211046908

Thompson, N., & Black, J. (2020). Balancing data utility and privacy: Ethical considerations in data sharing. *Journal of Data Ethics*.

United Nations Office on Drugs and Crime. (2018). Global report on trafficking in persons 2018. UNODC.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*, 1–9. https://doi.org/10.1038/sdata.2016.18