

# Expanding Non-Patient COVID-19 Data: Towards the FAIRification of Migrants' Data in Tunisia, Libya and Niger

Mariem Ghardallou<sup>1†</sup>, Morgane Wirtz<sup>2</sup>, Sakinat Folorunso<sup>3,7</sup>, Zohra Touati<sup>1</sup>,  
Ezekiel Ogundepo<sup>4</sup>, Klara Smits<sup>2</sup>, Ali Mtiraoui<sup>1</sup>, Mirjam van Reisen<sup>2,5,6,7</sup>

<sup>1</sup>University of Sousse, Faculty of Medicine of Sousse, Department of Community Medicine, Research Laboratory LR12ES03, Tunisia, 4002, Sousse

<sup>2</sup>Tilburg University, 1002 Tunis, the Netherlands

<sup>3</sup>Department of Mathematical Sciences, Olabisi Onabanjo University, P.M.B 2002, Ago-Iwoye, Ogun State, Nigeria 120005, Nigeria

<sup>4</sup>Data Science Nigeria, Lagos 105102

<sup>5</sup>Leiden University, 1310 Leiden, the Netherlands

<sup>6</sup>Leiden University Medical Centre, Leiden University, 1310 Leiden, the Netherlands

<sup>7</sup>Virus Outbreak Data Network-Africa

**Keywords:** FAIRification process; FAIR Guidelines; FAIR data; Tunisia; Niger; Libya; COVID; Migrants; VODAN-Africa

Citation: Ghardallou, M., Wirtz, M., Folorunso, S., Touati, Z., Ogundepo, E., Smits, K., Mtiraoui, A., Van Reisen, M.: Expanding non-patient COVID-19 data: Towards the FAIRification of migrants' data in Tunisia, Libya and Niger. *Data Intelligence* 4(4), 955–970 (2022). doi: 10.1162/dint\_a\_00181

Submitted: March 10, 2021; Revised: June 10, 2022; Accepted: July 15, 2022

---

## ABSTRACT

This article describes the FAIRification process (which involves making data Findable, Accessible, Interoperable and Reusable—or FAIR—for both machines and humans) for data related to the impact of COVID-19 on migrants, refugees and asylum seekers in Tunisia, Libya and Niger, according to the scheme adopted by GO FAIR. This process was divided into three phases: pre-FAIRification, FAIRification and post-FAIRification. Each phase consisted of seven steps. In the first phase, 118 in-depth interviews and 565 press articles and research reports were collected by students and researchers at the University of Sousse in Tunisia and researchers in Niger. These interviews, articles and reports constitute the dataset for this research. In the second phase, the data were sorted and converted into a machine actionable format and published on a FAIR Data Point hosted at the University of Sousse. In the third phase, an assessment of the implementation of the

---

<sup>†</sup> Corresponding author: Mariem Ghardallou, University of Sousse (Email: ghardallou.m@gmail.com; ORCID: 0000-0001-9289-722X).

FAIR Guidelines was undertaken. Certain barriers and challenges were faced in this process and solutions were found. For FAIR data curation, certain changes need to be made to the technical process. People need to be convinced to make these changes and that the implementation of FAIR will generate a long-term return on investment. Although the implementation of FAIR Guidelines is not straightforward, making our resources FAIR is essential to achieving better science together.

## **ACRONYMS**

CEDAR	Center for Expanded Data Annotation and Retrieval
CRF	case report form
DCAT	Data Catalog Vocabulary
DSW	Data Stewardship Wizard
eCRF	electronic case report form
IOM	International Organization for Migration
FAIR	Findable, Accessible, Interoperable, Reusable
FDP	FAIR Data Point
OWL	Web Ontology Language
RDF	Resource Description Framework
VODAN	Virus Outbreak Data Access Network

---

## **1. INTRODUCTION**

Data on the prevalence of COVID-19 in Africa is limited and incomplete [1]. Vulnerable communities are often not included in this data. In particular, migrants and refugees are not represented, as they usually stay under the radar, have little money and resources, and, therefore, tend to avoid health facilities. As a result, reporting on the prevalence of COVID-19 by national health systems does not include these communities [2]. However, as refugees and migrants are highly mobile, for ethical and practical reasons, it is important to ensure their access to services and to include them in data on the prevalence of COVID-19 [3].

When clinical data is not available, as is the case for migrants and refugees, we need to depend on research data, which is often accessed on 'open data' platforms such as HumData (<https://data.humdata.org/>). However, such platforms have certain limitations, for instance, data on these platforms is accessible to everyone, but only upon publication. With FAIR, data is accessible by appropriate people (defined individually or as a group of people) and in an appropriate way (with licences) during its whole lifecycle. Accessibility depends on the purpose of data [4]. To address this, a recent international initiative interested in data publishing and management proposed that all data should be 'FAIR': Findable, Accessible, Interoperable, and Reusable. Since this initiative was started, many efforts have been made towards this, with the development of FAIR metrics, structures and tools that aim to facilitate the 'FAIRification' of data. A FAIRification workflow has emerged and matured through projects in many different fields, such as nanotechnology [5] and biopharmaceuticals [6], which presents new avenues to strengthen data ownership,

interoperability and provenance. As a result, a domain independent workflow has evolved for research data. During the COVID-19 pandemic, the FAIR Guidelines gained prominence in relation to the management of data related to the crisis [7, 8] and population health [8], including combining clinical data and research [9].

Starting in Tunisia, then extending to Libya and northern Niger, the research presented in this article aims to describe the process of the FAIRification of data (the pre-FAIRification, FAIRification and post-FAIRification phases) related to the impact of the COVID-19 crisis on migrants, refugees and asylum seekers from sub-Saharan African countries, according to the scheme adopted by GO FAIR. The main goal of this research is to make researchers and policymakers understand the influence of the measures taken to control the virus, to give better support to this vulnerable population and, ultimately, to prepare for future pandemics.

## **2. RESEARCHING VULNERABLE GROUPS DURING THE PANDEMIC**

Worldwide, there are an estimated 272 million migrants, and some are more vulnerable than others, due to personal, social, situational and structural factors [10]. Their vulnerability may be exacerbated in crisis situations such as the COVID-19 pandemic. Persons displaced internally and across borders are particularly at risk. The majority of the 25.9 million refugees and 41.3 internally displaced persons in the world are in developing countries [10]. Tunisia is among the countries affected by this phenomenon, as it is a point of transit for migrants from Africa to Europe. In fact, Tunisia and Niger (particularly the Agadez region in northern Niger) are entry and exit gateways to Libya for sub-Saharan migrants. Migrants and refugees often travel clandestinely, from stopover to stopover, with the end goal of reaching Europe. Thus, Tunisia hosts an increasing number of migrants from sub-Saharan Africa. This population is highly mobile, with many having fled the war in Libya, usually remaining temporarily in Tunisia.

In Libya, migrants and refugees are held in inhumane conditions, including living in overcrowded and unsanitary housing, which makes them more susceptible to COVID-19 infection. Despite the pandemic, journeys to Europe continue to be organised for migrants and refugees through illegal networks, and data on these vulnerable groups is scarce. In fact, in Tunisia, Libya and the Agadez region of Niger, the irregular situation of these people places them far from official data.

Nevertheless, the 2030 Agenda for Sustainable Development recognises the contribution of migration to sustainable development. In fact, 10 out of 17 Sustainable Development Goals (SDGs) contain targets and indicators that are relevant to migration or mobility. The Agenda's core principle to 'leave no one behind'—including migrants—highlights the need to improve migration data locally, nationally, regionally and internationally [11]. Moreover, improving migration data, especially in the current COVID-19 pandemic, is a crucial step towards understanding the movement of the virus and, consequently, controlling the pandemic.

Data on migration and health may be derived from various sources, including traditional or routine data sources at the national level (e.g., civil registration, vital statistics, population censuses, and household surveys) and non-traditional or agency-based sources (e.g., BioMosaic, IOM sources, etc.) [12]. However, these resources need to be enhanced, and the collection of data on migration at the national level remains a challenge, especially for developing countries.

### 3. VODAN PROJECT: THE INTEGRATION OF RESEARCH DATA WITH CLINICAL DATA

The FAIR Guidelines were first formulated in 2014 at the Lorentz Center in Leiden, the Netherlands by a multidisciplinary team of scholars, librarians, archivists, publishers, and research funders [13]. This was followed by a series of 'Bring Your Own Data' (BYOD) workshops, which were conducted to support making data FAIR [14]. Then, in July 2016, the European Union (EU) published the Guidelines on FAIR data management in Horizon 2020 [13]. Among the ongoing initiatives working on the implementation of FAIR Guidelines, GO FAIR is the most prominent. GO FAIR is a bottom-up, stakeholder-driven, and self-governed initiative that promotes the establishment of networks between individuals, institutions and organisations by offering an open and inclusive ecosystem [15]. The Virus Outbreak Data Access Network (VODAN) Implementation Network is one of the activities carried out by GO FAIR, in collaboration with other institutions, such as the Leiden University Medical Center. It was set up to help fight the COVID-19 pandemic [15].

The architecture of the VODAN project aims to create community convergence and facilitate the interoperability of data pertaining to the (expected) incidence of COVID-19 infection linked to location, time and age group. Thus, it integrates clinical patient data and research data on COVID (Figure 1). In fact, convergence and interoperability between semantic metadata in the same domain can produce a situation in which the metadata from different data sources can be visited for inspection and the analysis of such data carried out.

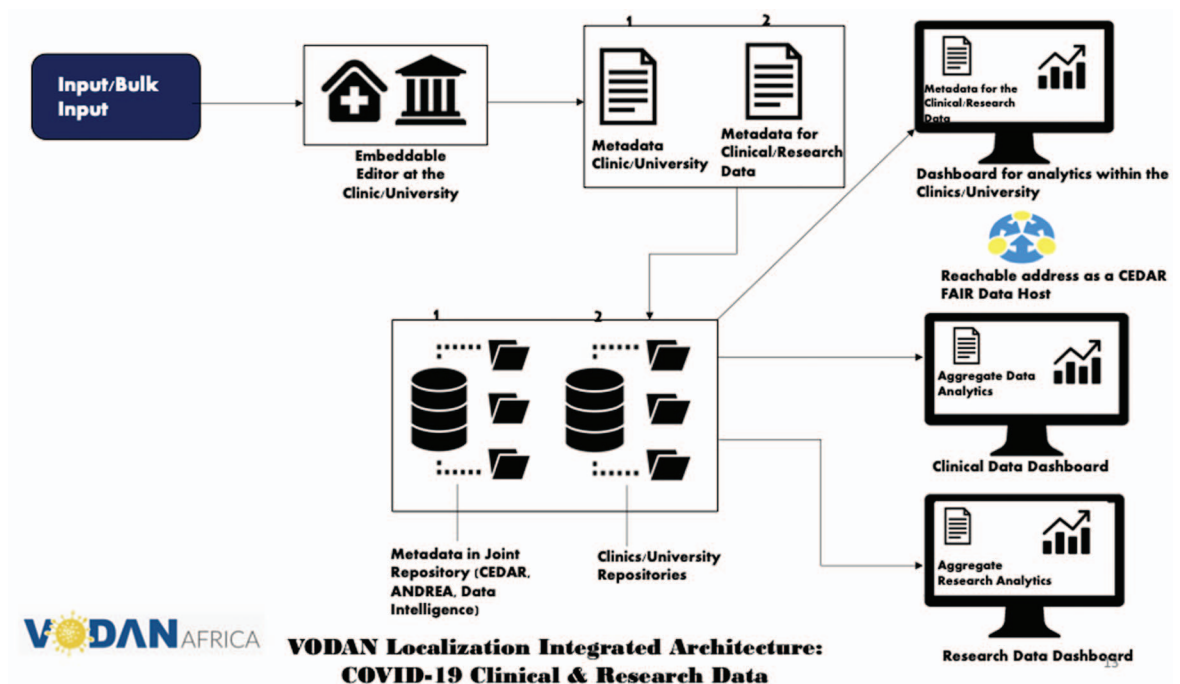


Figure 1. The VODAN-Africa-CEDAR integrated localisation architecture for clinical patient data and research data on COVID prevalence [16].

The data visiting is controlled in this architecture by pre-agreement about access to data for research and clinical data dashboards, and these streams can—if agreement is obtained—be combined in one data-analytical pipeline.

#### 4. METHOD

The study is a descriptive work of the FAIRification process of research data collected as part of the VODAN project. This research data concerns the impact of the COVID-19 crisis on migrants, refugees and asylum seekers from sub-Saharan African countries and was obtained from two sources: press articles published during the pandemic from May to mid-September 2020 and interviews conducted between May and June 2020.

The data FAIRification process was elaborated according to the workflow performed by Jacobsen et al. [17]. It is a generic and domain independent workflow that can be used as a guide or template, and then adapted to the specificities of each research domain.

This workflow is divided into three phases—pre-FAIRification, FAIRification and post-FAIRification—and these phases are divided into seven steps. The details of this generic workflow are presented in Figure 2.

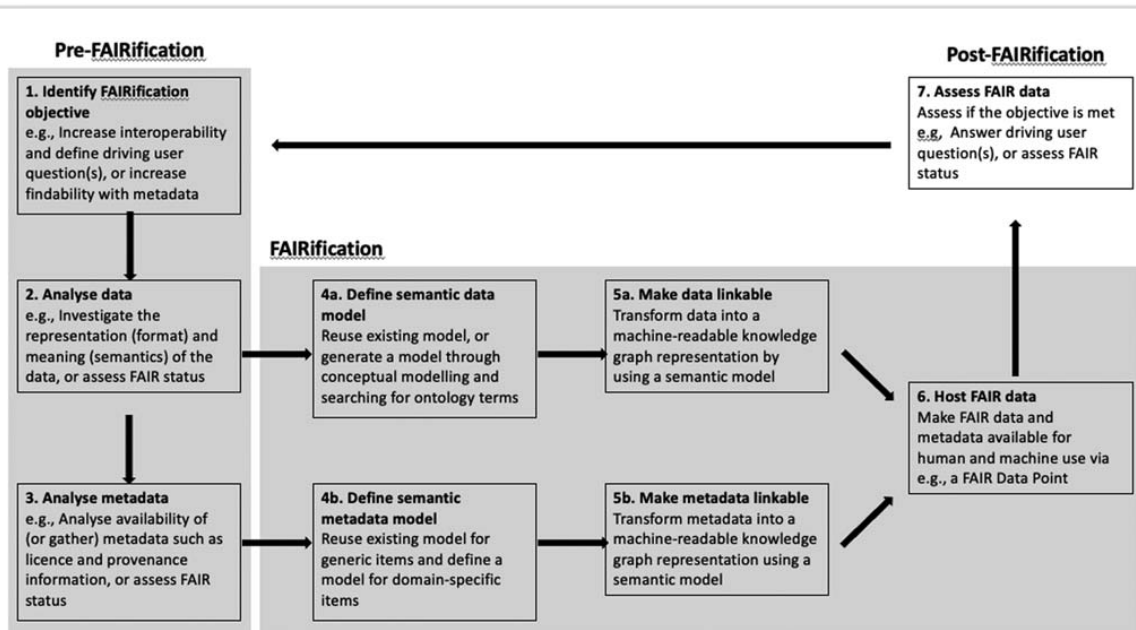


Figure 2. A generic workflow for the data FAIRification process [17].

To deploy FAIR Guidelines, in the first step of the project, the Implementation Network GO FAIR-VODAN Africa- Leiden University Medical Center used a set of tools called VODAN in a Box [18]. This was developed by the Dutch Techcentre for Life Sciences and based on the FAIRification process proposed by GO FAIR. VODAN in a Box is used to create, publish, find and annotate datasets [13]. This set of tools provides two significant services:

- A Data Stewardship Wizard (DSW) adjusted to serve as a Wizard for filling and maintaining electronic case report forms (eCRFs) and other templates
- A FAIR Data Point (FDP) to maintain metadata about eCRFs or templates created in DSW

And, to support these, other services are included, such as AllegroGraph triple store for eCRF data and queries. Then, to allow more healthcare templates and research templates to be added, the architectural design was adjusted, with the integration of the platform of the Center for Expanded Data Annotation and Retrieval (CEDAR). Thus, the VODAN-FDP became the CEDAR FAIR Data Station (Figure 3).

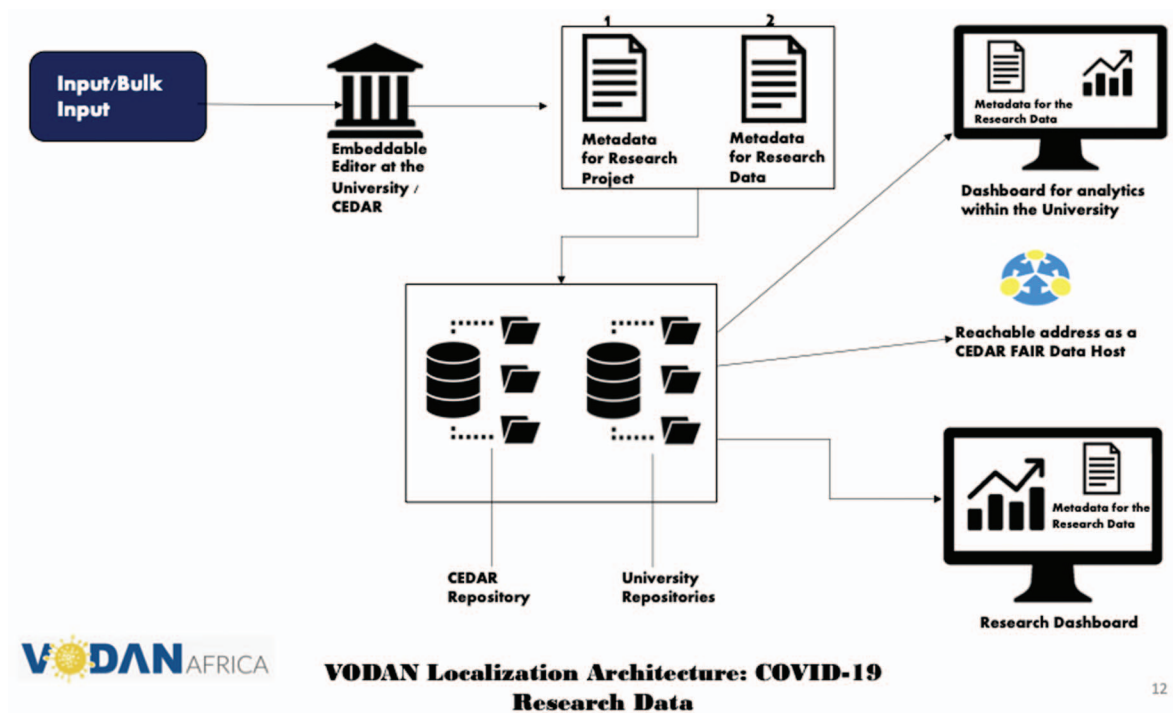


Figure 3. The VODAN-Africa-CEDAR localisation architecture for research data [18].

## 5. RESULTS

### 5.1 Pre-FAIRification Phase

This phase consists of steps 1 to 3.



### *5.1.1 Step 1. Identify FAIRification Objective*

This step involved the retrieval of non-FAIR data. At the beginning, a data management plan was created to specify what kinds of data would be used in the project, and how we would process, store and archive the data. The objective of the research was to shed light on the impact of coronavirus on sub-Saharan Africans hosted in Tunisia, Libya and Niger. Hence, to draw on the reality of the situation, two sources of information were used: interviews and press articles published during the crisis. A total of 124 in-depth qualitative interviews were conducted by two journalists and researchers based in Tunisia and a researcher in Niger. The fieldwork was carried out between May and June 2020. In Tunisia, the interviews were conducted in Zarzis, Medinine, Djerba, Sfax, Kerkennah, Sousse, Kelibia and Tunis. In Niger, the fieldwork was carried out exclusively in Agadez, which is the last stopover before crossing the Sahara to Libya and Algeria.

The people interviewed were from a wide range of backgrounds: migrants hosted by international organisations such as the United Nations High Commission for Refugees (UNHCR) and the International Organization for Migration (IOM), independent migrants with various levels of vulnerability, authorities, representatives from non-governmental organisations and international bodies, and members of civil society, as well as smugglers and fishermen.

In addition, from May to mid-September 2020, the publication of press articles on the situation of migrants in the Central Mediterranean region was closely monitored. Daily, students of African studies at the University of Sousse and the Director of the *Air Info* newspaper in Agadez listed the articles, videos and podcasts published on the research subject. Their focus was mainly on local media (in Arabic, Hausa and French), but they also extended their research to the international press (in French, English, Italian and Dutch). A total of 565 press articles were collected. In this first step, links to all press articles were summarised in a word document and for interviews in audio records.

### *5.1.2 Step 2. Analyse Data*

All digital resources can be considered to be data. In this case, we focused on data in the restricted sense of the term. To ensure the interoperability of the collected data, the content of the data was inspected to identify the different concepts and the structure of the data elements. Then, the data extracted from the press articles and interviews were repurposed and validated to be more useful for research. During this step, data fields, types (quantitative, qualitative) and values were characterised and media or interviews concepts (such as data fields for media frames or sources, etc.) were extracted. Then the relationships and the expected values of this data were validated according to the appropriate semantic model (see Step 4).

Regarding data interviews and to facilitate their sharing in the future with respect for privacy issues, a de-identification process was developed based on the Health Insurance Portability and Accountability Act (HIPAA) de-identification standard for protected health information [16]. Accordingly, individual identifiers such as names or geographic subdivisions were removed from the dataset to minimise the risk of identification of data subjects.

All data were entered into Excel sheets and, to ensure the FAIR features, a unique identifier was assigned to each data distribution, including for media reports (e.g., <https://fdp.uc.rnu.tn/distribution/4a4484ea-82a9-4f48-805b-773d53da8080>) and interviews (FAIR Principle F1 [19]). These identifiers created for the VODAN-Africa project will survive the termination of the project (under Tunisian law, these identifiers can be maintained based on an agreement to be renewed every five years). The Centre de Calcul El Khawarizmi (CCK) [20] is the legal institution responsible for the longevity of these identifiers (FAIR Principle F1 [21]). This centre is the Internet service provider for all establishments, agencies and administrations under the Ministry of Higher Education and Scientific Research in Tunisia.

For fully mechanised access, this identifier (F1) follows a globally-accepted schema that is tied to a standardised, high-level communication protocol, the Hypertext Transfer Protocol (HTTP) (FAIR Principles A1.1 and A1.2 [21]). This protocol is open, free, universally implementable and was mainly used to ensure the retrievability of press articles. However, a non-mechanised access protocol was applied for interview metadata, because it is more sensitive. Hence, the user seeking access needs to ask permission from the principal investigator before it is delivered, according to the requester's profile. Accordingly, a contact email address has been made available to researchers ([vodan.tunisia@gmail.com](mailto:vodan.tunisia@gmail.com)).

### 5.1.3 Step 3. Analyse Metadata

Metadata is any description of a resource that can serve the purpose of enabling findability, reusability, interpretation or assessment of that resource [22]. The metadata to be gathered for this research were identified and FAIR features verified (FAIR Principles F2 and R1, R1.1, R1.2, R1.3 [21]). Thus, for press articles and interviews, many labels and descriptors were attached to the data to make it rich and, therefore, optimise findability. These descriptors were also defined to serve the intended research purpose (such as event, date, type, geolocation, country, etc.) and to meet the requirements of the other FAIR Principles (licence, copyright, provenance, etc.). Hence, the licence cc-by-nc-nd3.0 was applied to describe the conditions under which the metadata can be used (cc-by-nc-nd-sa). Provenance descriptions were implemented following the community specific templates approach. A new metadata template was authored using the CEDAR system.

In addition, the Data Documentation Initiative (DDI) standard was used for metadata. Then, a machine actionable metadata model was used to explicitly link the data resource to its metadata. This link was provided by the FDP, which was structured according to the Data Catalog Vocabulary (DCAT, version 2) [23]. Hence, the basic metadata structure was organised into four layers: data repository, catalog(s), dataset and distribution:

- Repository metadata (9 components): title, description, publisher, version, language, licence, start date, last update, institution country
- Catalog metadata (5 components): title, description, publisher, version, language
- Dataset metadata (12 components): title, description, publisher, version, language, licence, issued, modified, keywords, theme, contact point, landing page
- Distribution metadata (10 components): title, description, licence, issued, modified, download URL, access URL, media type, format, byte size



The use of DCAT not only enables the description of the metadata schemata (FAIR Principle F2 [21]), but also provides unique identifiers for potentially multiple layers of metadata (FAIR Principle F3 [21]). Moreover, the metadata will still be accessible even when the data is no longer available (FAIR Principle A2 [21]), as described in the data management plan. In fact, all metadata will be published along with the report of the project on the Tilburg University Globalization, Accessibility, Innovation and Care (GAIC) website (<https://tilburguniversity.edu/about/schools/humanities/departments/dcu/gaic-network>).

## 5.2 FAIRification Phase

This phase consists of steps 4 to 6.

### 5.2.1 Step 4. Define Semantic Data and Metadata Model

As there was no available semantic model (knowledge representation) for our data and metadata, a new one was generated. A conceptual model was generated and specific ontologies and terms will be defined and created (FAIR Principles I1 and I2 [21]; see figures 4 and 5). The Wikidata ontology project [24] can be used as a source of some terms when creating vocabularies. Then, the new terms can be transferred and submitted to the Bio portal using an Excel spreadsheet template and SKOS Play Convert. Thus, the created semantic model can be used as a template in later steps to transform data and metadata into a machine readable format.

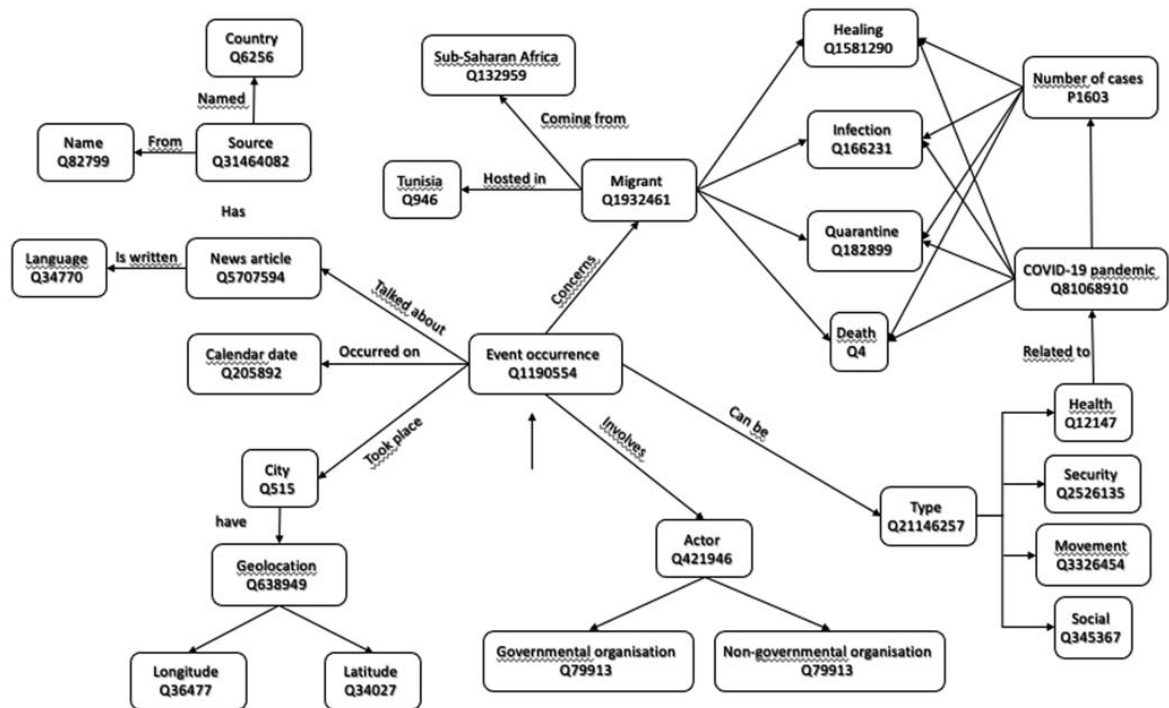


Figure 4. Press articles semantic data model (first draft).

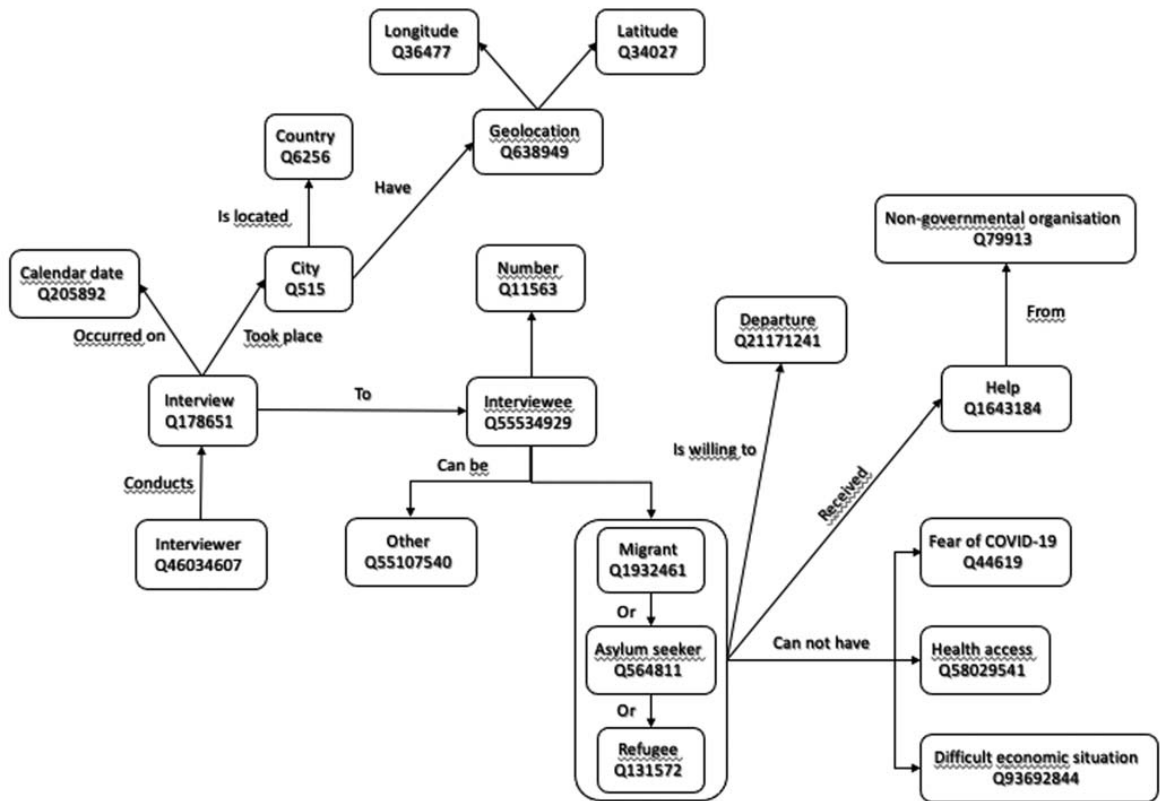


Figure 5. Interviews semantic data model (first draft).

### 5.2.2 Step 5. Make Data and Metadata Linkable

To make data and metadata linkable, the previous semantic data model was embedded in CEDAR to provide a linkable machine readable global framework, which is the Resource Description Framework (RDF) (FAIR Principle I1 [21]).

### 5.2.3 Step 6. Host FAIR Data

To make FAIR data available for consumption, a CEDAR FDP or station was deployed. This FDP is hosted by the University of Sousse site (Tunisia). An FDP is software that allows data owners to expose datasets in a FAIR manner and allows data users to discover the metadata about the offered datasets and, if the licence conditions allow, to access that data. It is based on DCAT, as described previously. In this way, the data in the project stay where they are collected, but are still accessible. An agreement with Google project will be established to make this data point reachable.

### 5.3 Post-FAIRification Phase

This phase consists of step 7.

#### 5.3.1 Step 7. Assess of FAIR Data

Given that the FAIRification process is currently being implemented, an assessment can be performed only after the realisation of this implementation. For this evaluation, a core set of semi-quantitative metrics [25] that have universal applicability can be applied. This assessment will be reported in a future publication.

## 6. DISCUSSION

The FAIR Guidelines define a set of characteristics that data resources, tools, vocabularies and infrastructures should have in order to assist discovery and reuse by third parties. These characteristics are related, but independent and separable. Each principle is broken down further into a core set of measurable criteria. The FAIR Guidelines have the advantage of having a low barrier-to-entry. Thus, they are minimally defined, which is why data producers, publishers and stewards can easily adhere to them, and they can be adhered to incrementally. In addition, the FAIR Guidelines are also modular, so they can be applied in any combination and, as a consequence, they can suit a wide range of applications and special circumstances [21].

The FAIR Guidelines have to have a common high-level interpretation to achieve global implementation. There are many successful experiences of the implementation of FAIR Guidelines in repositories, such as Dataverse [26], an open-source data repository software installed in dozens of institutions globally to support public community repositories and institutional research data repositories. In addition, there are also several projects for which FAIR is a key objective. These projects may provide valuable advice and guidance for those wishing to become more 'FAIR'. For instance, we can mention the biomedical and healthCARE Data Discovery Index Ecosystem, bioCADDIE [27], a consortium that works to develop a Data Discovery Index (DDI) prototype, which is set to be as transformative and impactful for data as PubMed has been for biomedical literature, or the Center for Expanded Data Annotation and Retrieval (CEDAR) [28], which was used for the present research to improve the 'FAIRness' of migrant data.

In support of communities and resources that are already pursuing FAIR objectives, several documents were published in this vein to facilitate the harmonisation of FAIR implementation choices between and within work groups, such as the 'FAIR Principles: Interpretations and Implementation Considerations' [22].

In the present research, and to implement the FAIR Guidelines, we reused some already existing solutions, but faced challenges in other situations. Certainly, the implementation of the FAIR Guidelines is not straightforward, but we have to overcome these challenges to make our resources FAIR to achieve better science together. Accordingly, the process of making data resources FAIR ('FAIRification') was broken down into steps, allowing the different dimensions of FAIRness to be distinguished according to the considered resource and the cost-benefit to the implementer and their community stakeholders [21].

In this article, we described the FAIRification process for migrants' data. To our knowledge, it is the first experience of FAIRification in a Tunisian academic institution. During this process, we tried to respect the FAIR implementation considerations [22]. The generic workflow proposed by Jacobsen et al. was applied for this purpose [17]. This workflow has the advantage that it is applicable in any domain. Hence, it makes FAIRification easier [17]. However, some challenges and specific requirements have to be considered when dealing with certain kinds of data. In fact, some data are more sensitive than others and require special ethical considerations. This is true of the data gathered from the interviews. Given the vulnerability of the migrant population (their illegal status), data de-identification was undertaken. This challenge, which relates to privacy, is also present when processing health data. This is why, Sinaci et al. [13], from the Software Research Development and Consultancy Corporation in Turkey, adapted the FAIRification process proposed by GO FAIR in response to health data requirements. For this, they applied some restrictions on existing steps and introduced new steps for ethical and legal requirements. As a result, the new workflow takes into account specific functionalities for data curation, data validation, data de-identification/pseudonymisation and data versioning [13].

As mentioned previously, during the implementation process we tried to reuse the existing FAIR solutions, if they were applicable. However, for some steps and principles, the solutions developed in other settings were not appropriate, such as the component Data Stewardship Wizard, which is a semantic data model based on the World Health Organization's (WHO)'s COVID-19 CRF [18]. The narrow focus of the tool limited its use. To overcome this limit of the template, DWS recently developed the Template Development Kit (DSW TDK), which is a command-line tool to make the work on templates efficient [29]. In the same vein, and to respond to the growing trend among investigators in various fields to define templates that structure metadata, an open source workbench based on semantic web technologies was launched within CEDAR [13, 30]. After consideration of the tools, the research team decided to use the CEDAR templates for the present research instead of the DSW for FAIRification.

As we were obliged to create new templates adapted to research data in a general way (non-clinical data), we faced a problem related to language and vocabularies, which have to follow the FAIR Guidelines. In fact, it is recommended to use language and thesauri that are universally accessible, such as the RDF, to represent knowledge on the Web in a machine accessible format and the Web Ontology Language (OWL) for ontologies. In our case, we already chose to use a publicly accessible registry, namely, Wikidata, which contains interesting schema information that can be expressed naturally using RDF and OWL [31].

In addition to these barriers faced in our research, many other obstacles are reported in the literature. In fact, FAIR implementation requires financial investment, training and the construction of technical infrastructure [6]. Thus, many changes are required at different levels (in relation to people, processes, technology and data), which can cost money. To convince people to make these changes, stakeholders need to be convinced that FAIR implementation will generate a long-term return on investment.

Moreover, a cultural change is needed. This can be achieved by incentives such as peer recognition and financial rewards [6]. Stakeholders also need to consider the required expertise—they have to boost capacity building on FAIR data stewardship over the long term [17]. Finally, all of these requirements for FAIR implementation cannot be achieved without the commitment of and investment by senior management.

## **7. CONCLUSIONS**

The FAIR Guidelines are meant to be a guide for all researchers to enable digital resources to become more findable, accessible, interoperable and reusable for both machines and humans. While there have been a number of recent, often domain-focused, publications advocating for specific improvements in practices relating to data management and data archiving, FAIR differs in that it describes concise, domain-independent, high-level principles that can be applied to a wide range of disciplines, such as the present research on migrants during the COVID-19 crisis.

Many technical barriers needed to be overcome. In this research, VODAN in Box was used at the beginning of the project. This toolset allows for the creation, publishing, finding and annotating of FAIR datasets. It can be deployed wherever the user wants (in a cloud provider, in a server or on a local machine). In this research it facilitated the imputing of data related to virus outbreaks and the publication of metadata describing these datasets, but was not flexible enough to use research data on the same topic but with a different structure.

Despite of these barriers, the implementation of the FAIR Guidelines has many benefits. As the FAIR Guidelines enhance the findability, accessibility, interoperability and reusability of data, the analysis of data from multiple sources becomes more efficient and error free. With FAIR, the whole process of running data becomes streamlined, data wrangling activities are minimised and scientific queries answered more quickly. Although the implementation of FAIR Guidelines is a challenging process, if each community shares their solutions for dealing with the barriers, this process will become easier.

## **ACKNOWLEDGMENTS**

We acknowledge and greatly appreciate the support of the Virus Outbreak and Disease Network (VODAN)-Africa for the training and workshops on creating the science of FAIR data in Africa and especially in Tunisia. We also sincerely thank the Tunisian team of VODAN-Africa and the University of Sousse (Tunisia) for providing data stewardship. Thanks are also due to Misha Stocker for managing and coordinating this Special Issue (Volume 4) and Susan Sellars for copyediting and proofreading. This project is supported by funding from NWO, domain Social Sciences and Humanities under the 'Corona Fast-track Data' call for proposals, file no. 440.20.012. Finally we acknowledge VODAN-Africa, the Philips Foundation, the Dutch Development Bank FMO, CORDAID, and the GO FAIR Foundation for supporting this research.

## **AUTHORS' CONTRIBUTIONS**

Mariem Ghardallou (Mariem.Ghardallou@famso.u-sousse.tn, 0000-0001-9289-722X): Principal investigator, conceptualization and writing. Morgane Wirtz (morgane.wz@gmail.com, 0000-0002-2514-5797): Investigation and data collection. Klara Smits (klarasmits@gmail.com, 0000-0003-1713-7057): Data Collection. Sakinat Folurunso (sakinat.folorunso@ouagoye.edu.ng, 0000-0002-7058-8618): Validation (literature review). Ezekiel Ogundepo (ogundepoezekiel@gmail.com, 0000-0003-3974-2733): Validation (literature review). Zohra Touati (zohra.touati@live.fr, 0000-0002-8329-706X): Methodology. Ali Mtiraoui (ali.mtiraoui@gmail.com, 0000-0002-6990-955X): Supervision. Mirjam van Reisen (mirjamvanreisen@gmail.com, 0000-0003-0627-8014 ): Conceptualisation, Supervision, Acquisition and Review.

## **CONFLICT OF INTEREST**

All of the authors declare that they have no competing interests.

## **ETHICAL STATEMENT**

Tilburg University (REDD-Cie)—preliminary approval obtained

## **REFERENCES**

- [1] Houreld, K., Lewis, D.: In Africa, lack of coronavirus data raises fears of 'silent epidemic' [Online]. Reuters (2020). Available at: <https://www.reuters.com/article/us-health-coronavirus-africa-data-insigh-idUSKBN24910L>. Accessed 15 June 2021
- [2] OECD: What is the impact of the COVID-19 pandemic on immigrants and their children? [Online]. Organisation for Economic Co-operation and Development (OECD) (19 October 2020). Available at: <https://www.oecd.org/coronavirus/policy-responses/what-is-the-impact-of-the-covid-19-pandemic-on-immigrants-and-their-children-e7cbb7de/>. Accessed 15 June 2021
- [3] UNHCR: Inclusion of refugees into national health systems. Public Health Services Survey. United Nations High Commissioner for Refugees (January 2020). Available at: <https://app.powerbi.com/view?r=eyJrljoiMWQ0OGM4YWVtNzYxZS00MTVlTk4ZTI0tMjk4YzU5NTkwYjhhIiwidCI6ImU1YzM3OTgxLTY2NjQtNDZlZjY1NDNkMmFmODBiZSIsImMiOj9&pageName=ReportSection>. Accessed 15 June 2021
- [4] Purnama Jati, P.H., Lin, Y., Nodehi, S., Cahyono, D.B., Van Reisen, M.: FAIR versus open data: A comparison of objectives and principles. *Data Intelligence* 4(4), 865–881 (2022)
- [5] Kochev, N., Jeliaskova, N., Paskaleva, V., Tancheva, G., Iliev, L., Ritchie, P., Jeliaskova, V: Your spreadsheets can be FAIR: A tool and FAIRification workflow for the eNanoMapper Database. *Nanomaterials* 10(10), 1908 (2020). <https://doi.org/10.3390/nano10101908>
- [6] Wise, J., de Barron, A.G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., et al.: Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discovery Today* 24(4), 933–938 (2019). <https://doi.org/10.1016/j.drudis.2019.01.008>
- [7] OECD: Enhancing access to research data during crisis [Online]. Virtual Workshop Hosted by Research Data Alliance (RDA) (23 April 2021). Available at: <https://www.oecd.org/sti/inno/enhance-access-research-data-during-crises.htm>. Accessed 15 June 2021



- [8] ZonMw: Open Science in COVID-19 Research [Online]. ZonMw (2021). Available at: <https://www.zonmw.nl/en/research-and-results/fair-data-and-data-management/open-science-in-covid-19-research/>. Accessed 15 June 2021
- [9] Nabudere, H.: The Digital REACH Initiative Roadmap (Slide 27) [Online]. SlidePlayer (28 March 2018). Available at: <https://slideplayer.com/slide/14530772/>. Accessed 15 June 2021
- [10] IOM: Migration factsheet No. 6—The impact of COVID-19 on migrants. International Organization for Migration (IOM) (n.d.). Available at: [https://www.iom.int/sites/default/files/our\\_work/ICP/MPR/migration\\_factsheet\\_6\\_covid-19\\_and\\_migrants.pdf](https://www.iom.int/sites/default/files/our_work/ICP/MPR/migration_factsheet_6_covid-19_and_migrants.pdf). Accessed 9 March 2021
- [11] Migration Data Portal: Migration data for the Sustainable Development Goals (SDGs) [Online]. Migration Data Portal (n.d.). Available at: <https://migrationdataportal.org/sdgs?node=0>. Accessed 16 March 2020
- [12] Migration Data Portal: Migration and health [Online]. (n.d.). Available at: <https://migrationdataportal.org/themes/migration-and-health>. Accessed 9 March 2021
- [13] Sinaci, A.A., Núñez-Benjumea, F.J., Gencturk, M., Jauer, M. -L., Deserno, T., Chronaki, C., et al.: From raw data to FAIR Data: The FAIRification workflow for health research. *Methods of Information in Medicine* 59(6), e21–e32 (2020). <https://doi.org/10.1055/s-0040-1713684>
- [14] Hoof, R., Goble, C., Evelo, C., Roos, M., Sansone, S., Ehrhart, F., et al.: ELIXIR-EXCELERATE deliverable D5.3: Bring Your Own Data (BYOD) [Online]. Zenodo (2019). Available at: <https://zenodo.org/record/3207809#.YEBu7NzBXIU>. Accessed 9 March 2021
- [15] GO FAIR: GO FAIR initiative [Online]. (n.d.). Available at: <https://www.go-fair.org/go-fair-initiative/>. Accessed 27 December 2020
- [16] Van Reisen, M., Oladipo, F., Stokmans, M., Mpezamihgo, M., Folorunso, S., Schultes, E., et al.: Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research. *Advanced Genetics* 2(2), e10050 (2021). <https://doi.org/10.1002/ggn2.10050>
- [17] Jacobsen, A., Kaliyaperumal, R., da Silva Santos, L.O.B., Mons, B., Schultes, E., Roos, M., et al.: A generic workflow for the data FAIRification process. *Data Intelligence* 2(1–2), 56–65 (2020). [https://doi.org/10.1162/dint\\_a\\_00028](https://doi.org/10.1162/dint_a_00028)
- [18] VODAN: About the VODAN in a Box [Online]. (n.d.). Available at: <https://docs.vodan.fairdatapoint.org/en/latest/about/about.html>. Accessed 9 January 2021
- [19] US Department of Health and Human Services: Guidance regarding methods for de-identification of protected health information in accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Available at: [https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveridentities/De-identification/hhs\\_deid\\_guidance.pdf](https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/understanding/coveridentities/De-identification/hhs_deid_guidance.pdf). Accessed 16 March 2021
- [20] Ministry of Higher Education and Scientific Research: Centre de Calcul El-Khawarizmi [Online]. Tunisia (n.d.). Available at: <http://www.cck.rnu.tn/>. Accessed 14 January 2021
- [21] Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., et al.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1), 1–9 (2016). <https://doi.org/10.1038/sdata.2016.18>
- [22] Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., et al.: FAIR Principles: Interpretations and implementation considerations. *Data Intelligence* 2(1–2), 10–29 (2020)
- [23] W3C: Data Catalog Vocabulary (DCAT) [Online]. (n.d.). Available at: <https://www.w3.org/TR/2014/REC-vocab-dcat-20140116/#class-catalog>. Accessed 9 January 2021
- [24] Wikidata: Welcome to Wikidata [Online]. (n.d.). Available at: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page). Accessed 9 January 2021
- [25] Wilkinson, M.D., Sansone, S.-A., Schultes, E., Doorn, P., da Silva Santos, L.O.B., Dumontier, M.A.: Design framework and exemplar metrics for FAIRness. *Scientific Data* 5 (2018)

- [26] Dataverse Project: Open source research data repository software [Online]. (n.d.). Available at: <https://dataverse.org/home>. Accessed 1 March 2021
- [27] DataMed: Home—bioCADDIE Data Discovery Index [Online]. (n.d.). Available at: <https://datamed.org/>. Accessed 1 March 2021
- [28] CEDAR: Better metadata means better science—About [Online]. (2021). Available at: <https://metadatascenter.org/>. Accessed 1 March 2021
- [29] Suchánek, M.: DSW Template development kit: The tutorial [Online]. Data Stewardship Wizard (2020). Available at: <https://www.youtube.com/watch?v=FFElv-e24NE&feature=youtu.be>. Accessed 9 March 2021
- [30] Musen, M.A., Bean, C.A., Cheung, K.-H., Dumontier, M., Durante, K.A., Gevaert, O., et al.: The Center for Expanded Data Annotation and Retrieval. *Journal of the American Medical Informatics Association* 22(6), 1148–1152 (2015). <https://doi.org/10.1093/jamia/ocv048>
- [31] Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., Vrandečić, D.: Introducing Wikidata to the Linked Data Web. In: P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, et al. (eds), *Semantic Web—ISWC 2014, Lecture Notes in Computer Science*, 8796, 50–65, Springer International Publishing, Cham, Switzerland (2014) [https://doi.org/10.1007/978-3-319-11964-9\\_4](https://doi.org/10.1007/978-3-319-11964-9_4)