Ref. Ares(2021)5931564 - 29/09/2021

Interconnecting Histories and Archives for Migrant Agency

www.ithacahorizon.eu

DELIVERABLE 1.6

Data Management Plan

Work Package concerned: 1
Due date: 30/09/2021
Actual submission date: 29/09/2021
Concerned work package leader: UNIMORE
Deliverable responsible: UNIMORE
Dissemination level: Confidential
Authors: Michaël Gasperoni, Marco Iacovella, Matteo Al Kalak

This project received funding from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement n° 101004539.

Legal Notice
The sole responsibility for the content of this publication lies with the author/s. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein.

Project full title: ITHACA, Interconnecting Histories and Archives for Migrant Agency: Entangled Narratives Across Europe and the Mediterranean Region
Start date: January 1, 2021
Duration: 51 months
Project funding scheme: H2020-SC6-MIGRATION-09-2020
Coordinator: University of Modena and Reggio Emilia
Primary Coordinator Contact: Matteo Al Kalak
E-mail: matteo.alkalak@unimore.it

List of Abbreviations and Acronyms

WP	**Work-Package**
T	**Task**
SAB	**Scientific Advisory Board**

Document review history

Date	Author/s	Feedback/Review	Version
05/07/2021	Michaël Gasperoni, Marco Iacovella, Matteo Al Kalak		D7.4_V1
20/07/2021	Marco Iacovella, Matteo Al Kalak		D7.4_V2
18/09/2021	Marco Iacovella, Matteo Al Kalak		D7.4_V3

TABLE OF CONTENTS

1. Type of study

The ITHACA project focuses on narratives on migrations in the past and present, analysing them in a rigorous historical framework, through an interdisciplinary methodology, and adopting a comparative and transnational approach. The project is based on the idea of creating a digital platform (the ITHACA Platform) that brings together – through a clear system, easy searchability and a user-friendly interface – migration narratives from the 15th century to the present. The Platform, at the heart of the project, intends to present an innovative analysis of narratives on migrations, whilst considering different causes, actors involved, geopolitical scenarios and migration routes. In particular, the project takes into account migration narratives related to religious causes, humanitarian crises, political reasons, decolonization processes, environmental and climate causes, with sources and case-study concerning both the past and the present.
The type of study crosses archival, historical, ethno-anthropological and sociological methodologies, retracing the main historical roots of the construction of the discourse on migration and of migrants from the early modern times to the contemporary age.

2. Data description

2.1. Personal data

The ITHACA project deals with personal data of dead and living people. The ITHACA superarchive gathers names, surnames, correspondences, reports, certificates and various records on/of people involved in migration across Europe and the Mediterranean region.

The ITHACA superarchive will collect data divided into two main categories:

- Data issued by archival search, mainly related to dead people (15th-19th century), whose relatives can be still alive
- Data deriving from ethno-anthro-sociological activities and oral history methodology, collected from individuals during in-person and online workshops, interviews, focus groups, trainings and engagement meetings.

A third form of personal data collection is related to the communication and dissemination online and to in-person activities all along the project (workshops, seminars, conferences, ITHACA diary contest etc). These data won't be included in the superarchive.

2.1.1. Special categories of personal data/sensitive data

According to the Article 4(13), (14) and (15) and Article 9 and Recitals (51) to (56) of the GDPR, personal data to be considered "sensitive" is:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

Special categories of personal data are those collected by the ITHACA project during research activities with/on vulnerable people (such as refugees, undocumented migrants and LGBTIQ+ people). Data related to racial or ethnic origin, political opinions, religious or philosophical beliefs, and data concerning a person's sex life or sexual orientation will be collected too.

No trade-union membership; genetic data, biometric data processed solely to identify a human being; and health-related data will be collected.
The anonymisation and pseudonymisation techniques will concern all the personal data collected during the various activities foreseen in the archival research project (archival research, interviews, focus groups, ethnography, audio and visual recordings, workshops, etc.). The same techniques will also applied to data collected during the phase of contacting research participants (i.e. names, phone numbers and address) and the data collected and audio-recorded or video recorded during the research activities – i.e. interviews and focus group – and during the dissemination phase. In those cases, however, where participants expressed either in the informed consent or orally their will to be named and recognised with their individual and personal identity, researchers will acknowledge this choice and make sure that it is appropriately reflected into the project, its materials and dissemination practices.

2.1.2. Pseudonymised data

In order to respond to the principle of minimization of risks for the participants to the project, preventing all forms of potential harm they can suffer from having information they provide be linked to them, rigorous pseudonymisation techniques will be implemented.

As described in the ITHACA Protection of personal data deliverable (D7.4 POPD), phone numbers of interviewees will be saved on the research's mobile/devices under pseudonyms and will be erased after the research activity.

In publications and dissemination activities, the names of research participants' names will be replaced with pseudonyms that will be agreed with research participants themselves, except in those cases where the choice of being named has been expressly formulated by participants. In this way, data can no longer be attributed to a specific data subject without the use of additional information.

2.1.3. Anonymous data/anonymised data

Research participants will have the possibility to decide which kind of data to eliminate or to anonymize after his/her participation to the research activity.
Name and surnames of vulnerable categories participating to the project will be anonymised through anonymisation techniques in order to protect the safety and privacy of research participants as well as the confidentiality of the information provided.
In this way, the data subject to anonymisation is no longer identifiable.

2.1.4. Aggregate data

No aggregate data will be collected or produced.

2.1.5. Confidential data

No confidential data (i.e. investigations, data protected by intellectual property rights, passwords, financial information, national safety information) will be collected or produced.

2.2. Non-personal data

According to the GDPR Regulation, non-personal data are data which originally did not relate to an identified or identifiable natural person, or data on maintenance needs for industrial machines; or data which was initially personal data, but later made anonymous.
The ITHACA project will collect and produce data related to this second category.

3. Metadata

3.1. Metadata standards and data documentation

Starting from the Dublin Core standard, the main categories of metadata to be applied and implemented during the ITHACA project research and dissemination activities are the following:

- information about records: concerning the description of the sources used for the research
- information about the contents: concerning the migration narratives explained or reported by the sources.

A first definition of the different categories to be use for the creation of metadata is resumed in the following tabs.

	Information about records															
Categories	Original ID / DOI	Title	Abstract	Language	Resource type	Place	Format	Publisher	Issued	Modified	Time Period	Annex name	Creator	Contributor	Resource is part of (Title of Dataset)	Notes

Information about content								
(story) Places	(story) Time	Nationality / ethnicity	Gender	Age	Sexual orientation	Migratory / legal status	Education	Profession / activity

A more precise standardisation of the metadata collected as well as the application of each category to the different narratives to be analyzed during the project will be identified during the WP2, 3 and 4, with a co-creation methodology.

In order to guarantee accuracy and appropriateness of the definitions used during the creation of metadata, the title/label used for each category will be deeply discussed within the Consortium too.

4. Data collection and generation

4.1. Data provenance

Data provenance within the ITHACA project is archival research, ethno-anthro-sociological activities and oral history methodology, collected from individuals during in-person and online workshops, interviews, focus groups, trainings and engagement meetings, dissemination and communication activities.
Data will be collected by the ITHACA staff members.

4.2. Format

The ITHACA project will use these data formats:

- Microsoft Office 2010 (or later) or LibreOffice for text-based documents (or any other compatible version) .ods, .doc, .docx, .xls, .xlsx, .ppt, .pptx. Also, especially where larger datasets need to be dealt with, .csv and .txt file formats will be used. All finished and approved documents will also be made available as .pdf documents.
- Microsoft Office 2010 (or later) or Libre office for Table or databases (or any other compatible version) .csv, .accdb, .xls, .xlsx, .ods, .dbase.
- Illustrations and graphic design will make use of GIMP or Photoshop (format: different types possible, mostly .png), and will be made available as .jpg, .psd, .tiff and .ai files.
- GIMP or Photoshop for digitization, scanning or photography pictures and will be made available as .jpg.
- MP3 or WAV for audio files
- MP4, MOV or WMV for video files
- xml, csv and xslt format.

4.3. New dataset value

The ITHACA project will create the ITHACA superarchive, i.e. a new dataset on narratives on and of migration.

5. Data management and storage

Data management within the ITHACA project is part of the WP1. UNIMORE as project coordinator, is responsible for FAIR (Findable, Accessible, Interoperable and Reusable data) data management, following the Guidelines on FAIR Data Management in Horizon 2020. As lead beneficiaries of the WP2 and WP3, Sorbonne University and AMM collaborate to the FAIR data management.

For both open and non-open data, the aim is to preserve the data and make it readily available to the interested parties for the whole duration of the project and beyond.
A public Application Programing Interface (API) and an Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) will be provided to registered users allowing them the access to the platform. In addition a web site is already developed to allow the member's ITHACA project to visualize data sets.
The following measured will be implemented to guarantee a proper management of data:

- Access control list for user and data authentication. Depending on the dissemination level of the information an Access Control List will be implemented reflecting there for each user the data sets that can be accessed.
- Implementation of an alert system that informs in real time of the violation of procedures or about hacking attempts.
- Identification of a person who is responsible for keeping safe the information stored.

The non-open research data will be archived and stored short-term in the NFS storage of Sorbonne University Huma-Num Virtual Machine administered by Sorbonne University. The Sorbonne university virtual machine is currently being employed to coordinate the project's superarchive and to store all the digital material connected to ITHACA.

No data embargo period will be applied to the open deliverables from the ITHACA project. The ITHACA superarchive is currently under construction and closed. The Share-docs storage repository is limited to the consortium PIs; a user account login is required to access the data.

In order to make public available data interoperable, all publicly available data within the ITHACA project are made available in text formats, namely PDF/A, audio or video formats, namely MP3 and MP4, and as xml, csv and xslt in the ITHACA superarchive.

Non-public data from the project will remain available to the consortium partners after the end of the project in the Share-docs repository.

There is no time-limit on the availability of the open data from the ITHACA project, made available on the project website for an unlimited time-period.

6. Data security

6.1. Main risks to data security and formal information/data security standards

The coordinating organisation of the ITHACA project, UNIMORE, is subject to the laws and guidelines that are relevant for this project in Italy, and the General Data Protection Regulation (GDPR), as well as all the consortium partners. In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected.

In terms of data security, according to the ITHACA POPD – Requirement n°10 Data processing activities (DELIVERABLE 7.7), the ITHACA research teams have to follow a specific protocol for the confidential and anonymous collection and treatment of all collected data.

1. Collected datasets are separated from personal identity information during the phase of collection or transcription.
2. Anonymization is ensured in different ways in relation to the different research tools.

2.1 Questionnaires are collected ensuring complete anonymization; they can be collected in folders where the participants deposit the questionnaires autonomously, through an anonymized online site or trough return-envelopes.

2.2 In audio-recordings and video-recordings, all names of the interviewees are deleted in all transcriptions and replaced by a number combined to three letters, indicating gender, role and country.

3. All personal information is kept in a separate file and is no longer linkable to the results (it is treated in an aggregated way).
4. Raw data are only accessible to authorized researchers affiliated to ITHACA project.

5. Processed data do not include reference to personal data, in order to prevent identification.
6. Without consent of the interviewees, the original audio-recordings and video-recordings of interviews and focus groups are not used.
7. Some parts of video-recorded or audio-recorded activities can be included in the restricted data of the website at the following conditions of consent.

7.1 If a restricted use is only permitted by the subjects, the data are encrypted using available technology.

7.2 If consent is denied, the recordings will not be included and used.

8. Individuals cannot be singled out in datasets; two records cannot be linked within datasets or between two separate datasets; no information can be inferred in these ways.
9. No sensitive personal data is shared with third parties.
10. Incidental findings concerning sensitive data will be deleted.

In a first step, the only data collected will be the personal data (name, surname, email address) of the project participants in order to access online tools. This collection will be done by creating a human-id account, in compliance with terms and rules of Huma-Num. A certified repository will be used: a private account, protected by username and password, on sharedocs.huma-num.fr, French research infrastructure providing data repositories to scholars. By these account only selected staff members can access to the common cloud Sharedocs storage space. Sharedocs is accessible to Web et clients WebDAV. It is used for stocking high amount of data. The repository has no costs. No fixed ending date of this repository is foreseen.

In a second step, currently underway, the goal is to find a secure storage space, to receive sensitive data that will satisfy the following requirements:

- Regular vulnerability scans to prevent the most common security flaws that lead to data breaches.
- Penetration tests to search for vulnerabilities in the network or application, to avoid possible hacking.
- Perimetral protection of the network forbids those services considered dangerous or vehicle for infections.
- Protocols for navigation: https - internet browsing; ftp - file transfer services; ssh -secure access to terminal.
- Data transfers take place via secure on-line channels where the data are encrypted, rather than copied for transportation.
- Access to all places and servers, used to host hardware and software on which personal data is stored, are restricted to staff members of the consortium partners.
- The database servers are located behind a firewall with default rules to deny all traffic.
- The firewall is open only for specific applications or web servers, and firewall rules do not allow direct client access.
- Database accounts used by the staff are individual accounts, rather than shared group accounts.
- The database is remotely accessible via a secure encrypted link (e.g. IPSEC or SSL VPN tunnel) with access control in place.

In parallel, about security, ITHACA project intends to follow general principles:

- Sensitive personal data that participants provide to authorised researchers are only accessible to research teams. The staff members are coached and trained before being allowed to access confidential or personal data.
- Restricted data are never sent via email.

- Restricted data are encrypted during transmission over the network, using encryption measures strong enough to minimize the risk of the data's exposure.
- Redundancy of restricted data are eliminated.
- Server-side scripts such as PHP, JSP, or ASP.NET are excluded.
- All personal and sensitive data held electronically are stored centrally and only by authorised researchers.
- Sensitive personal and confidential data are never stored on portable devices.
- All portable devices, used to collect data during the research activities, are password-protected to prevent unauthorised use and unauthorised access to the database.
- The data models adopted to encode the archival metadata in XML format are: EAD, to encode the description of the archival resources; EAC-CPF, to encode the description of the archival authority records.
- Accounts to the database are only provided to authorised staff members.
- Application code is reviewed for SQL injection vulnerabilities.
- No Spyware is allowed on the application, web or database servers.
- Secure authentication to the database is used.
- External service providers employed by Departments/Centres are subject to strict procedures for accessing sensitive personal data established through formal contracts.

For the activation and access of Superarchive foreseen in WP4, users will be required to provide data only concerning name/names and email address (login procedure).

7. Data preservation strategy

7.1. Data quality and standards

Data will be preserved on the ITHACA staff member laptops and on USB and HDD memories. Photoreproductions of archival documents, audio and video registrations will be preserved in the ITHACA staff member laptops and in external memories, protected by passwords.

A certified repository will be used: a private account, protected by username and password, on sharedocs.huma-num.fr, French research infrastructure providing data repositories to scholars.

Sharedocs is accessible to Web et clients WebDAV. It is used for stocking high amount of data.

The repository will have no costs. No fixed ending date of this repository is foreseen.

8. Data sharing and access

Concerning data sharing and re-use, all deliverables are available for download and re-use on the ITHACA project website as soon as possible after being submitted to the European Commission. The ITHACA website is covered by a Creative Commons CC-By 4.0 license, requiring the users to credit the ITHACA project and the European Commission as funding agency if any data from this is referred to or reused externally.
ITHACA publications will be available in Open Access (Gold or Green). Details on publications, journals, conferences are available in the ITHACA Publication Plan.

9. Responsibilities

The coordinating organisation of the ITHACA project, UNIMORE, remains the first responsible of the data management. Daily operations are run by ITHACA data manager, to be appointed by SU.

The Data Controller is ITHACA Coordinator, University of Modena and Reggio Emilia.

In accordance with Article 37 et seq. of the GDPR, the University has appointed

avv. Vittorio Colomba as Data Protection Officer, e-mail: dpo@unimore.it
PEC: dpo@pec.unimore.it

For the purpose of the Research, the Principal Investigator of the Research is Prof. Matteo Al Kalak, University of Modena and Reggio Emilia UNIMORE, Italy, e-mail: matteo.alkalak@unimore.it