

D1.1

DATA MANAGEMENT PLAN

Colleen E. Boland
Santiago A. Villar Arias

Universitat Autònoma de Barcelona

February/2021



"This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 882986".

Deliverable Factsheet	
Title and number	<i>Data Management Plan (D1.1)</i>
Work Package	WP1
Submission date	01/February/2021
Authors	Colleen E. Boland (UAB) & Santiago A. Villar Arias (UAB)
Contributor	Cristina Blasi Casagran (UAB)
Reviewers	Emma Teodoro (UAB) & Thilo Gottschalk (FIZ)
Dissemination level	PU (Public)
Deliverable type	DEM (Demonstrator, pilot, prototype)

Version Log			
Issue Date	Version nº	Author	Change
22/12/2020	V1.0	Colleen Boland Santi Villar Arias	Initial version sent to reviewers.
04/01/2021	V1.1	Emma Teodoro Thilo Gottschalk	Initial version including changes and suggestions made by reviewers
13/01/2021	V1.2	Colleen Boland Santi Villar Arias	Updated version sent to the ITFLOWS DPA for review.
26/01/2021	V1.3	Jonathan Andrew (ITFLOWS DPA)	Updated version including changes and suggestions made by the ITFLOWS DPA
01/02/2021	V2.0	Colleen Boland Santi Villar Arias	Submission to EC
23/06/2022	V3.0	Colleen Boland Santi Villar Arias	Final version after EC review

Disclaimer

This article reflects only the authors' views, and the Agency is not responsible for any use that may be made of the information it contains (Art. 29.5 Grant Agreement).

Table of Contents

Abbreviations.....3

1. Data Summary.....4

1.1 Purpose of the data collection and generation, and its relation to the objectives of the project

1.2 Types and formats of data that the project will generate and collect

1.3 Re-use of existing data

1.4 Origin of the data

1.5 Expected size of the data

1.6 Data utility

2. FAIR Data.....14

2.1 Making data findable, including provisions for metadata

2.2 Making data openly accessible

2.3 Making data interoperable

2.4 Increase data re-use

3. Allocation of resources.....22

3.1 Costs for making data FAIR in your project and how to cover them

3.2 Responsibility for data management in the project

3.3 Resources for long term preservation

4 Data security.....26

4.1 Provisions in place for data security and long term preservation and curation measures

5. Ethical aspects.....29

6. Other issues.....32

7. Further support in developing the DMP34

Abbreviations

AIDA: Asylum Information Database
API: Application Programming Interface
BUL: Brunel University London
CEPS: Centre for European Policy Studies
CERTH: Ethniko Kentro Erevnas Kai Technologikis Anaptyxis
CESSDA: Catalogue of the Consortium of European Social Science Data Archives
CIT: Cork Institute of Technology
COO: Countries of Origin
CRI: Associazione della Croce Rossa Italiana
CSD: Center for the Study of Democracy
DDD: Dipòsit Digital de Documents
DMP: Data Management Plan
DOI: Digital Object Identifier
DPA: Data Protection Advisor
EAB: Expert Advisory Board
EASO: European Asylum Support Office
EC: European Commission
EMT: EUMigraTool
EUI: European University Institute
FIZ: Karlsruhe – Leibniz-Institut für Informationsinfrastruktur
FRONTEX: European Border and Coast Guard Agency
GDPR: General Data Protection Regulation (Reg. EU 2016/679)
HDX: The Humanitarian Data Exchange
IAI: Istituto Affari Internazionali
IEB: Independent Ethics Board
IfW: Institut für Weltwirtschaft
IGC: Independent Gender Committee
ILO: International Labour Organization
IMF: International Monetary Fund
IOM: International Organization for Migration
IPC: Integrated Food Security Phase Classification
IPR: Intellectual Property Rights
OCC: Open Cultural Center
OECD: Organisation for Economic Co-operation and Development
OIT: Oxfam Italia Onlus
PWG: Policy Working Group
TB: Terabyte
TRC: Terracom AE
UAB: Universitat Autònoma de Barcelona
UB: Users Board
UNHCR: United Nations High Commissioner for Refugees
WDI: World Development Indicators
WP: Work Package
XML: Extensible Markup Language

DATA MANAGEMENT PLAN

1. Data Summary

1.1) Purpose of the data collection and generation, and its relation to the objectives of the project

By collecting and generating data, ITFLOWS partners aim to fulfil the four main purposes of the project, which are: (a) providing accurate predictions on migration; (b) providing policy solutions for the management of irregular arrivals and asylum systems in the EU; (c) proposing solutions for reducing potential conflict/tensions between migrants and EU citizens, by taking into account a wide range of human factors and using multiple sources of information; and (d) providing policy solutions for optimising refugee integration in the EU.

These objectives have been specifically designed to directly address the main concerns in managing migration towards and within the EU, and to produce ambitious but realistic solutions. During the project, ITFLOWS partners will gather, process and analyse data across a broad range of disciplines, establishing new analytical and computational predictions. In addition, ITFLOWS will use data to train and test the models for the EUMigraTool (EMT).

In particular, new knowledge and insights will be gained on drivers of migration (WP3), patterns of migration within the EU (WP4) and public attitudes towards migration in EU Member States (WP5). This will be achieved mainly via evidence-based and original research activities, such as empirical case studies in selected EU Member States (WP4), Big Data analytics of selected social networks (WP3, WP5) and interviews with migrants and asylum seekers to understand the push/pull factors as to why they choose a particular EU host country (WP3). As mentioned above, this knowledge will be necessary both to build the models for the EMT and also to design policy solutions.

Therefore, ITFLOWS partners will generate new data, and compile and analyse existing data, to accomplish objectives for each Work Package, and for the project as a whole. Extensive quantitative and qualitative data is expected to be delivered in

the format of project deliverables; as a result, the experienced ITFLOWS Consortium will provide non-digital and digital solutions for different public audiences.

1.2) Types and formats of data that the project will generate and collect

Types

Depending on each particular task and/or deliverable, types of data may vary. ITFLOWS will collect publicly available data from different sources for the purposes of the ITFLOWS project. The main types of data are:

- Indexes on economic growth, socio-economic structures and other numeric variables from selected countries of origin and transit;
- Socio-economic indicators in host EU Member States;
- Data from the World Development Indicators (WDI) of the World Bank and from the OECD;
- National-level data sources on population statistics collected and used for the acquisition of relevant demographic indicators;
- Data on displacements provided by the DTM of the International Organization for Migration;
- Validation data from the UNHCR;
- IOM Flow Monitoring Data;
- Data on public attitudes and/or sentiment toward migrants;
- Asylum application data in the Member States from data sources such as IGC, EUROSTAT, UNHCR open data and the Asylum Information Database (AIDA).

Formats

Specific formats of data will depend on the type, the amount, the intended use and the source of each particular data/dataset. An accurate definition of data formats will hence be derived from inputs from WP3, WP4 and WP5 and integrated in the DMP. At this early stage, it is difficult to accurately indicate all data formats to be used. Data formats can change over time due to computational requirements or changes in the source-data (e.g. change from one database to another, changes in the respective APIs)

As a preliminary approach, we can indicate that the data gathered throughout the project would have some of the following formats (file extensions):

Text: PDF, DOC, DOCX, TXT, among others.

Image: JPG, PNG, TIFF, BMP, among others.

Audio: AIFF, WAV, MP3, among others.

Video: GIF, MOV, MP4, among others.

Structured data: CSV, XML, JSON, among others.

Modelling data: MAT, TENSORFLOW, among others.

1.3) Re-use of existing data

Qualitative and quantitative data will be collected from existing sources, including: macro data of countries of origin and transit; macro data of EU Member States; surveys on the sentiments of EU citizens; and finally, examination of data from platforms like Google Trends and Twitter, in order to analyse migration drivers, as well as public attitudes of EU citizens towards migrants. Where possible, ITFLOWS will reuse (pre-processed) publicly available existing datasets and sources.

Some of the data sources re-used in ITFLOWS have been originally generated by the UN Refugee Agency (UNHCR), FRONTEX, the International Organisation for Migration (IOM), the European Union through Eurostat and Eurobarometer, and the Organisation for Economic Co-operation and Development (OECD), amongst others. Reusable data is most often anonymised. Where this is not the case, ITFLOWS will ensure strict compliance with GDPR requirements and provide sufficient safeguards to protect the individual rights and freedoms of potentially affected data subjects through technical (e.g. anonymisation) and organisational (e.g. concept to handle Art. 15 requests) measures.

1.4) Origin of the data

In table n^o1 we describe all types and origins of data used in ITFLOWS, and the related Work Package number.

Table n^o1

Data Type	Origin	WP
Macro qualitative (T3.1) and quantitative (T3.2) data on drivers of EU-bound irregular mixed migration in countries of origin and transit: i) indexes on economic growth, socio-economic structures and other numeric variables from selected countries of origin, data from the World Development Indicators (WDI) of the World Bank and from the OECD; ii) national-level data sources on population statistics will also be collected and used for the acquisition of relevant demographic indicators; iii) data on displacements provided by the DTM of the International Organization for Migration	Publicly Available data	WP3
Data from social media environments (T.3.3): i) tweets; ii) Google Trends	Publicly available data	WP3
Interviews (T3.4) with adult asylum seekers/ migrants	Primary data	WP3 WP4
Quantitative data of Member States: i) German Panel Survey Data on asylum seekers and refugees (T4.1); data on secondary movements of refugees within the EU: economic and social indicators (T4.2)	Publicly available data	WP4
Asylum and refugee applicable law and procedures (T4.3)	Publicly available data	WP4
Asylum application data (T.4.3) in the Member States obtained from data sources such as IGC (covers applications in 16 Member States); EUROSTAT (covers 32 countries); UNHCR open data; and the Asylum Information Database – AIDA (contains information on asylum procedures across 23 countries)	Publicly available data	WP4
Pan-European surveys: individual survey data (European Social Survey, Eurobarometer, World Value Survey), combined with aggregate-level data on countries (Eurostat, Word Bank, etc.) to identify micro- and macro-level factors affecting attitudes to immigrants/migration policies (T5.2)	Publicly available data	WP5
Big Data (T5.3): i) geo-referenced tweets for hate speech content; ii) geographically disaggregated data at the destination country level	Publicly available data	WP5

Open-intelligence data sources (T6.1): video content; web news from mainstream media sources; social media (mainly Twitter)	Publicly available data	WP6
Data gathered from the organisations that will test the tool (T7.3)	Primary data	WP7
Participatory workshops data	Primary data	WP3 WP6 WP7 WP8

Sources of publicly available data

Table n°2

Source	Type of data	Link
UNHCR	Migration - IDPs, Refugees, Asylum seekers, Situation and camp reports	https://www.unhcr.org
FRONTEX	Irregular crossing	https://frontex.europa.eu
IOM Displacement	IDPs and Returnee IDPs	https://displacement.iom.int/
IDMC	IDPs by reason (conflict vs disaster)	https://www.internal-displacement.org
HDX	Humanitarian Data Exchange database	https://data.humdata.org
ACLED	Conflict events	https://acleddata.com
EMDAT	Climate and other disasters	https://www.emdat.be/
ECMWF	Climate data	https://www.ecmwf.int
WDI	World Bank World Development indicators	https://databank.worldbank.org
WB Governance Indicators	Background and Institutional	https://info.worldbank.org
European Social Survey	Attitudinal survey	https://www.europeansocialsurvey.org/
European Value Study	Attitudinal survey	https://europeanvaluesstudy.eu/
Eurobarometer	Attitudinal survey	https://ec.europa.eu/commfrontoffice/publicopinion/index.cfm
Eurostat	GDP per capita, real GDP growth, unemployment rate	https://ec.europa.eu/eurostat/data/database
OECD	Migration statistics - Database of immigrants in OECD countries	https://stats.oecd.org/

AIDA	Migration - IDPs, Refugees, Asylum seekers	https://www.asylumineurope.org/
Eurostat	Unemployment rates, Destination, Education, Income, Gender ratio, Age of inhabitants, Marriage rate, Minimum wage, Gini-Coefficient, Lt-unemployment	https://appsso.eurostat.ec.europa.eu https://ec.europa.eu/eurostat/
IPC Food Security	Food security data	http://www.ipcinfo.org
City Population	Population data	https://www.citypopulation.de
GDELT	News event data	https://www.gdeltproject.org/
Rulers, Elections, and Irregular Governance (REIGN) dataset	Governance data	https://github.com/OEFDatascience/OEFDatascience.github.io
IMF	All economic indicators	https://data.imf.org/
ILO	All economic, demographic indicators	http://www.fao.org/economic/est/statistical-data/en/
Google Trends	Selected keywords	https://trends.google.com
IOM Displacement Tracking Matrix Surveys on Migration Flows	Reports on migrant presence & profiles, info on flows, routes, and journey characteristics	https://dtm.iom.int
EASO Country of Origin Portal	Reports on the conditions in asylum seekers' countries of origin	https://coi.easo.europa.eu
European Country of Origin Information Network	Periodical reports, analyses, media reporting, etc. on the conditions in C.O.O. of asylum seekers	https://www.ecoi.net
Konfliktbarometer	Conflict data	https://hiik.de/konfliktbarometer/
Uppsala Conflict Data Program	Conflict data	https://ucdp.uu.se/
Open Street Maps	Geospatial data	https://www.openstreetmap.org/
United Nations Development Programme (UNDP) - Population Division-	Data on migrant stock and flows	https://www.un.org/development/desa/pd/content/international-migrant-stock

Socio-economic Panel (SOEP)	Demography, Employment, Income, Taxes and Social Security, Family and Social Networks; Health and Care; Home, Amenities; Education; Attitudes, Values and Personality; Time Use and Environmental Behaviour; Integration, Migration	https://www.diw.de/
IAB-BAMF-SOEP Survey of Refugees	Survey of people seeking protection in Germany from political persecution and violent conflicts	https://www.diw.de/sixcms/detail.php?id=diw_01.c.572992.en
Demographic & Health Survey (DHS)	Data on population, health, HIV, and nutrition	https://dhsprogram.com/
Wittgenstein Centre for Demography and Global Human Capital	Demographic data, population trends and projections	https://www.wittgensteincentre.org/en/data.htm
Vienna Institute of Demography of the Austrian Academy of Sciences	Demographic data	https://www.oeaw.ac.at/vid/data/demographic-data-sheets
Migrant Integration Policy Index (MIPEX)	Data on family reunification	www.mipex.eu/family-reunion
Statista	General data	https://www.statista.com/
MigrationsKB (MGKB)	Anonymized migration-related annotated tweets	https://w3id.org/mgkb
Hatexplain	Dataset on hate speech	https://paperswithcode.com/dataset/hatexplain
HurtLex	Lexicon of offensive, aggressive, and hateful words in over 50 languages	https://github.com/valeriobasile/hurtlex
FinnSentiment	Dataset (in Finnish) annotated with sentiment polarity	https://paperswithcode.com/dataset/finnsentiment
OGTD (Offensive Greek Tweet Dataset)	Manually annotated dataset from Twitter posts (in Greek)	https://paperswithcode.com/dataset/ogtd

DALC: the Dutch Abusive Language Corpus	Dataset with tweets manually annotated for abusive language (in Dutch)	https://paperswithcode.com/paper/dalc-the-dutch-abusive-language-corpus
Multilingual and Multi-Aspect Hate Speech dataset	Dataset on hate speech	https://paperswithcode.com/dataset/mlma-hate-speech
GermEval (2018)	Manually labelled data on offending statements/ hate speech towards foreigners	https://github.com/uds-lsv/GermEval-2018-Data
European Labour Force Survey	Cross-sectional data from the European Labour Force Survey	https://ec.europa.eu/eurostat/web/lfs
European Institute for Gender Equality (EIGE)	Gender statistics on various aspects of (in)equality between women and men.	https://eige.europa.eu/gender-statistics/dgs
Statistikdatabasen	Statistical Database (Sweden)	https://www.statistikdatabasen.scb.se/pxweb/en/ssd/
Eurofound	Data on Labour market participation	https://www.eurofound.europa.eu/topic/labour-market-participation#s-05
ITFLOWS Interviews	Interviews with migrants	-----

1.5) Expected size of the data

The exact size of the data as of now (M18) remains uncertain, and will be updated in D1.4 (M36). However, we expect not to exceed a few TB (Terabytes) in total, with the EMT as the hub for all data. Data will be limited to what is necessary for the purposes laid down in the Grant Agreement as summarised in Section 1.1 of this DMP.

1.6) Data utility

Data will be crucial in producing ITFLOWS publications (reports, analysis, policy briefs) and in developing the EMT. Moreover, data generated and collected within the ITFLOWS project will be useful for ITFLOWS Partners and also for other various stakeholders, such as:

a) EU Institutions, EU Agencies and policy-makers: EU institutions, bodies and agencies will benefit from ITFLOWS data, as well as from outputs derived from data generated and collected. In addition, the ITFLOWS project has one Work Package fully devoted to Policy Reach and Recommendations (WP8), which will translate the project's accumulated results into evidence-based policy analysis and

recommendations, to be disseminated to wide and relevant audiences and professional groups in charge of policy decisions at different levels. Additionally, the Policy Working Group (PWG), an internal ITFLOWS Board comprised of policy-makers at the national, EU and international level, will be fully engaged throughout the lifetime of the project, and will provide inputs on these topics. Policy-makers that compose the PWG will receive access to new and enriched sources of information, and will also help ITFLOWS partners to increase the impact of the project through their networks and resources. By engaging these stakeholders in the project, ITFLOWS will receive important input on how to address policy/legal difficulties in the field of migration, as well as how to influence future migration regulation.

b) Local authorities, practitioners and civil society organisations: Local authorities and civil society organisations play a significant role in this project. In particular, the ITFLOWS project created an internal board, the Users Board (UB), which is in charge of, among other tasks, validating the EMT. All members of the UB are practitioners, NGOs and municipalities working either as first responders at the borders, or in the management of asylum seekers, and in the integration of refugees. Through the Users Board, municipalities and NGOs working with refugee integration will receive access to the early versions of the EMT. The information provided at early stages may already offer valuable knowledge, e.g. to identify upcoming/existing risks of tensions. The cooperation with such parties will further ensure that the EMT will be developed in a way that is tailored to their specific needs. Eventually, the EMT aims to enable municipalities and NGOs working with refugee integration to receive additional data points via real-time input and knowledge with regards to the areas at risk of developing tensions between migrants and local communities, with the purpose of finding solutions to mitigate such risks. First line practitioners and NGOs will receive access to reliable prediction data on migration in order to allocate resources and assistance accordingly.

c) Scientific Community: The scientific community, in particular researchers in the field of migration, will benefit from data and outputs generated by the ITFLOWS project, as they will have access to migration data and analysis, made publicly

available on the website and also via specific dissemination and communication events organised by ITFLOWS.

d) General public: The general public will be able to follow the project through the ITFLOWS website (www.itflows.eu), where all analyses and results are published. The website is not only fully accessible and easy to use for practitioners and researchers, but is also user-friendly for the broader public.

e) Consideration of other stakeholders: As the EMT is developed, alongside its exploitation plan, partners are in serious discussions in conjunction with feedback and monitoring from WP2 ethics partners, the Independent Ethics Board (IEB) and the Data Protection Advisor (DPA) regarding EMT use, as well as access to the related CKAN Project repository. This will be addressed by M25. In particular, they are evaluating use by other stakeholders like national authorities, including the European Migration Network. The European Migration Network (EMN), (established in by Council Decision 2008/381/EC EU and amended by Regulation EU No 516/2014 of the European Parliament and of the Council on April 014), is a network of migration and asylum experts convened to provide and compare objective knowledge production and policy-relevant information on asylum and migration-related issues in Europe. Its governance and operation involves national authorities.

In the meantime (before the Consortium consensus on EMT users in M25), the partners are also preparing a series of EMT Reports to be available to all stakeholders, with analysis of the most recent EMT output by technical and research partners, released in M25, M30 and M36.

2. FAIR Data

To the extent possible as provided for in the Grant Agreement and as meets ethical and legal requirements, the project will make sure its research analysis and results are FAIR compliant.

2.1) Making data findable, including provisions for metadata

Discoverability of data

Metadata offers detail as to how data was generated, gathered, curated or updated, in order to identify, further investigate or link data. In ensuring that data is findable, the project's final results will provide metadata that at least includes: the terms "European Union (EU)" and "Horizon 2020"; the name of the action, acronym and grant number; the publication date; length of embargo period if applicable; and DOIs. The metadata will comply with anonymisation processes described further in this document.

The project's final results (research data and analysis) will be deposited in the UAB depository (DDD) elaborated upon in 3.1. The DDD applies the relevant Spanish and EU data protection legislation and is OpenAire¹ compliant. The UAB will assign the data a digital object identifier (DOI) and a permanent URL upon deposit in the DDD. UAB will then later attempt to provide the data, including relevant DOIs, to other international research indices, including the DataCite² registry and the Catalogue of the Consortium of European Social Science Data Archives (CESSDA).

Identifiability of data

Project findings and results will include DOIs, which offer reliable and version-specific citation in publications.

Naming conventions used, search keywords approach, clear versioning approach, and standards or procedures for metadata creation applied

Should any issues regarding metadata protocol arise over the course of the ITFLOWS project, the project will resolve them by referring to the Data Documentation

¹ <https://explore.openaire.eu/>

² <http://search.datacite.org/ui>

Initiative's standards, as found in its DDI LifeCycle 3.2 document³. All metadata will at least include the components described in the first paragraph of 2.1 above. The project's final naming conventions, renaming procedures and approach to keywords will be provided in a later version of the DMP, and partners responsible for deliverables will be informed of these.

As part of making data findable, it follows that data be openly accessible, interoperable and reusable, and further detail as to how the project will proceed in ensuring this for the relevant project analysis (and as ethical and legal guidelines allow) follows below.

2.2) Making data openly accessible

Data openly available

The project's Consortium has committed to an open by default approach to the primary data it generates, as briefly summarised in Part 1 of this DMP. This includes the qualitative and quantitative primary data generated through interviews with migrants and asylum seekers, as well as through pilot tests and participatory workshops relating to the project's EMT, in territories where Users Board organisations operate. Secondary data (examined during the project's research activities) will also be shared when possible, but will remain subject to the terms and conditions of its owners.

Regarding ownership of results generated by partners outside of the scope of the project, but is used in ITFLOWS, the procedure will be as follows: the project will follow the general guidelines that the generator of results possesses full ownership of them, with joint ownership clarified to the maximum extent possible via a clear breakdown of the work, reached through consensus. In the case that individual ownership is not clear, all joint owners will be entitled to royalty-free exploitation of all results. Each partner may transfer ownership of its results, under the same obligations listed here and in the project Grant Agreement, but must notify the other partners with 45 days advance notice; the partners may object within 30 days of

³ <https://ddialliance.org/Specification/DDI-Lifecycle/3.2/>

notification if they can demonstrate it negatively affects access rights, and the transfer will remain pending agreement between the partners.

Data kept closed

In order to allow for project data to remain open access, project partners will remove personal data and confidential information from raw data before storage, ensuring complete anonymisation before open publication. Data that cannot be anonymised due to practical or technical reasons is excluded from publication to ensure sufficient protection of the fundamental rights and freedoms of the (potentially) affected data subjects.

The project's EMT, composed of both general (deep machine learning) and localised (agent-based) prediction models, will contain algorithms qualifying as intellectual property. They will be examined on a case-by-case basis, and the final decision as to making all algorithms open source will be provided in an updated version of the DMP.

All research results of ITFLOWS, in the form of deliverables, are to be made public. Further detail as to ensuring that this data meets confidentiality and privacy requirements can be found in (5) Ethical Aspects, describing processes of redaction and anonymisation.

How data will be made available

All deliverables produced using the results of ITFLOWS, from research conducted in WP2, WP3, WP4 and WP5, will be made available as open access, free of charge, via both the gold and green open access models. The green access route is the default approach. When there is a strategic rationale for doing so (e.g. significantly greater impact will be achieved through open access publication by the publisher) the gold route will be followed. Criteria of cost and value is decided on a case-by-case basis by the Steering Committee.

The UAB's institutional repository, the DDD, supports the green open access route. The UAB DDD is accessible in that it follows the Open Archives Initiative model.

Again, as outlined in 2.2, metadata of the datasets will be provided to the DDD, and later to other catalogues as is feasible.

Methods or software tools for data access

Given that published data resulting from the project will be exported to the DDI recommended XML format, the project makes data accessible to a broad range of methods and software. XML is extensible, does not require licensing fees, can be migrated across different computers and adapted to technological changes, has widespread database support, as well as is supported by various software companies and many Web browsers. For the project's subject matter, it is optimal in facilitating and adapting to human and machine documentation and reading.

Software documentation and other information needed

Various measures ensure that potential users of the EUMigraTool (EMT) developed in this project are supported in effectively engaging with the tool. The EMT receives input from the Users Board of partners and external stakeholders, as well as trains these users in how to operate the EMT. Partners from WP7 are responsible for the design of the training strategy, and for developing the training material to educate, inform and provide EMT training. This training is conducted in a tailored way for each stakeholder in the group, with formats including downloadable material, workshops, online forums, video tutorials, and so on. Trainings take place on months 5 and 12 of the project, and an online community will be developed to facilitate engagement, knowledge transfer and support. Online tests take place with these users in order to validate the EMT over the course of months 18 to 24 of the project. A User Guide is made available as part of deliverable 7.3 in month 24 of the project.

Given that the final project results and deliverables will be provided in data formats accessible via a range of methods and software as described, it is not foreseen that any specific additional software documentation will be needed to access this data, or additional information needed in order to understand this data. However, this will be re-evaluated on a continual basis over the life cycle of the project.

Repository for deposit of data, metadata, documentation and code

The UAB's DDD, where the project's results will ultimately be deposited, complies with the following frameworks and policies: the Regulation of the Digital Deposit of Documents of the UAB; the UAB Open Access Institutional Policy; the Creative Commons licenses recommended by the UAB (CC0); the Preservation policy of the Digital Depository of Documents of the UAB (DDD); Spanish National policy including "Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación" and "Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020"; Horizon 2020: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (Model Grant Agreement, article 29); Horizon 2020: Guidelines on FAIR Data Management (Model Grant Agreement, article 29); Horizon Europe: the next research and innovation framework programme; and Plan S: Making full and immediate Open Access a reality.

Again, as indicated, metadata of the datasets will be provided to the DDD, where datasets will have a DOI and permanent URL.

Ultimately, the ITFLOWS Project will strictly comply with Regulation (EU) 2016/679 (EU GDPR), observing the principle of data protection by design and default (art. 25) in particular.

To this end, WP2 implements multiple measures to monitor the data processing in the project and provide guidance for legally compliant approaches. Measures to monitor the processing currently include, inter alia: access to the CKAN⁴ instance of the project; participation in technical meetings; research on related legal and technical literature; and bilateral discussion with the partners.

Access restrictions

At the start of the project, the Consortium signed a Consortium Agreement relating to relevant management, risk and liability and intellectual property rights, as well as ownership of results. Only in the case when results must remain confidential, if advised by the project's Ethics lead partner, will exceptions be made to the open

⁴ <https://ckan.org/>

access of project data; given the anonymisation process detailed here, it is not foreseen that this would occur. In particular, part 5 of this DMP outlines when overriding ethical or legal concerns may prevent making data available. At the same time, the project's qualitative and quantitative individual level data will be made available as a "public use file," without requiring handling restrictions. Moreover, publication embargos will expire at the project's end, with the exception of any PhD theses resulting from the project, which may maintain an embargo of up to three years, starting from the award of the degree.

The project's management team will address any issues concerning access rights, including cases where partners join or leave the project during its execution. Project partner Terracom (TRC), responsible for the ITFLOWS EMT Exploitation Plan, will be responsible for issues relating to the exploitation of IPR.

In general, access to project results (in the form of deliverables) will fall under a CC0 license ("no rights reserved"), the default licensing of the UAB DDD.

2.3) Making data interoperable

Data interoperability assessment

As previously indicated, the XML format consistent with the DDI standards allows for interoperability in that it is accessible to an extensive range of methods and software, and adaptable to various technological changes. Moreover, the data and metadata vocabularies in conformance with DDI standards observed throughout the project will facilitate interoperability, as does the project's open by default approach to data. As outlined in Part 1 of the DMP, common file formats will be provided for, including text, image, audio and video.

The DDD follows the Open Archives Initiative model, which allows interoperability with the OAI-PMH metadata transmission protocol (Open Archive Initiative - Protocol for Metadata Harvesting). Documents will be visible from different platforms and collectors including Google Scholar, Base, Core, Europeana, Hispana, WorldCat, OpenAire, OpenDOAR, Recolecta and ROAR.

Standardisation of data to facilitate interoperability

Metadata provides the necessary connections between various interoperable databases. Please refer back to 2.1, which describes the minimum metadata that will be included, as well as how DDI standards will serve as a reference and guide in metadata protocol. As the project progresses and the partners reach consensus, further specifics regarding standard metadata vocabulary (establishing the standardised naming conventions, for example) for final results will be provided in the D1.4 (M36).

Data licensing for wide reuse

Data generated in the project will be made publicly available if possible and as subject to WP2 legal and ethical monitoring, following the FAIR data directive. Data licensing will take place in conformance with the Grant Agreement. Each Consortium partner may grant licenses to its results or provide the right to exploit them. They will provide royalty free access to fellow partners and affiliated entities. Sub-licensing will not be permitted, and exclusive licenses can only be granted in the event that all other partners waive their access rights.

Consortium partners will grant royalty-free access to members of the ITFLOWS Boards, NGOs and municipalities, as well as EU institutions, bodies, agencies, offices and Member States, as subject to WP2 legal and ethical monitoring and if in conformance with the Grant Agreement.

2.4) Increase data re-use

Timing of data availability for re-use

As soon as possible, and at the latest upon the end of the project, machine-readable project deliverables (results and analysis) will be placed in the DDD.

Data usability by Third Parties

As a result of making the DDD-deposited data open access, third parties will be able to access, mine, exploit, reproduce and disseminate from the open access depository, including after the end of the project. Licensing of results or right to

exploit them may not impede open access rights, as will be reflected in the Exploitation Plan for ITFLOWS (D9.4 and D9.5).

Restrictions to data reuse

The results and outcomes of ITFLOWS will comply with the restrictions on data reuse as outlined in the Terms and Conditions of the project's Grant Agreement. Firstly, it observes the provisions made for third party reuse. For up to four years after the end of the project (August 2023), the Research Executive Agency could object to a transfer of ownership or exclusive licensing of results, if: (a) it is to a third party established in a non-EU country not associated with Horizon 2020 and (b) the Agency considers that the transfer or license is not in line with EU interests regarding competitiveness or is inconsistent with ethical principles or security considerations.

Moreover, if a beneficiary intends to transfer ownership or grant an exclusive licence, it must formally notify the Agency before the intended transfer or licensing takes place, and: identify the specific results concerned; describe in detail the new owner or licensee and the planned or potential exploitation of the results; and include a reasoned assessment of the likely impact of the transfer or licence on EU competitiveness and its consistency with ethical principles and security considerations.

The Agency may request additional information, and if the Agency decides to object to a transfer or exclusive licence, it will formally notify the beneficiary concerned within 60 days of receiving notification (or any additional information it has requested). No transfer or licensing may take place in the case of: a pending Agency decision within the 60 days upon receipt of notification; if the Agency objects; and until the conditions are complied with, if the Agency objection comes with conditions.

As a general rule, data reuse is restricted to non-commercial and non-competitive use, and is only to be used for intended purposes and following the appropriate confidentiality obligations. Exceptions to commercial use or reuse will be clarified for the partners in a later version of the DMP, along with: the procedures and agreement for the management and exploitation of the knowledge resulting from the project and the project results; any potential commercial applications the

partners must be aware of; and guidance in registering IPR. Any commercial exploitation of results would take place in conformity with the interests of the European Union and in accordance with the aims of the H2020 programme.

Quality assurance process

The partner institutions are committed to data quality assurance in keeping with the scientific rigor espoused by their organisations, seeking to ensure data validity, reliability, precision, integrity and timeliness. Curation of data will be the responsibility of the Consortium partner that generated it, again following DDI guidance. Specifically, the qualitative and quantitative primary data generated through interviews with migrants and asylum seekers, as well as through participatory workshops and pilot tests of the EMT in territories where Users Board members operate, will be subject to internal review processes incorporating reviewers from outside of the individual tasks or Work Packages to systematically crosscheck in order to reduce estimation, measurement and data processing error. In bias mitigation, the Ethics lead partner, the project's IEB, Expert Advisory Board (EAB), Independent Gender Committee (IGC) and Policy Working Group (PWG) will guide in monitoring bias issues and gender discrimination. In addition, the project's DPA will also ensure quality by serving as an independent monitoring body specifically appointed for data protection issues within the project.

Length of time of data re-usability

There is no limit on length of time of data re-usability. The project's published data deposited in the DDD, as per the UAB's Institutional Policy for Open Access⁵, will be maintained perpetually.

⁵ <https://ddd.uab.cat/record/222172>

3. Allocation of resources

3.1) Costs for making data FAIR in your project and how to cover them

Besides personnel costs of the researchers involved in data collection analysis, and the hosting for the ITFLOWS Website, in principle, there will not be additional costs associated in making the database FAIR and ensuring long term preservation. At the same time, *Gold Open Access* fees for publishing in indexed scientific journals have been allocated in the budget of some partners. In particular, the ITFLOWS partners UAB, CEPS and IfW have a specific budget to cover Open Access expenses for publishing in scientific journals, as part of *Other Direct Costs/Other Goods and Services*.

Despite such expenses, and as noted earlier, all scientific publications produced within ITFLOWS will be available as open access and freely available via both “gold” and “green” open access models, where the standard approach will be to follow the “green” route to open access.

As mentioned in Section 2, the ITFLOWS Coordinator, the Universitat Autònoma de Barcelona (UAB), has its own institutional repository: The *Dipòsit Digital de Documents*⁶ (or Digital Repository of Documents). The DDD is the tool that collects, manages, disseminates and preserves the scientific, educational and institutional production of the UAB. It also gathers digital documents that are part of, or complement, UAB libraries’ collections. The DDD repository provides an organised, open access and interoperable collection, has hosted datasets since 2017, and supports all University-affiliated researchers, enabling their published works to be made freely available via the “green” open access route. In this sense, the DDD will help the project to fulfil requirements in terms of open access, as all publications produced by ITFLOWS will be stored there.

In some particular cases, the “gold” route will be followed provided the Consortium identifies a strategic rationale for doing so (e.g. significantly greater impact is to be achieved through open access publication by the publisher).

⁶ <https://ddd.uab.cat/?ln=en>

3.2) Responsibility for data management in the project

As project coordinator, the UAB holds ultimate responsibility for the data management in the project (except in particular cases described below). However, each ITFLOWS partner shall create, manage, analyse, store and share data and/or datasets with respect to the applicable national and international legislation on data protection. Also, the quality control of these data falls under responsibility of the institution leading the respective Work Package.

Prior to the start of the ITFLOWS project, all partners signed a Consortium Agreement, which defined the roles and responsibilities of all partners and formalised the organisational structure, progress monitoring, and decision-making procedures.

The ITFLOWS Data Protection Advisor (DPA) and the ITFLOWS Ethical Board (composed of experts from UAB, FIZ and BUL, together with an Independent Ethics Board) will provide specific advice and guidance on how to tackle potential issues and risks related to the use of data during the lifetime of the project.

The UAB, coordinator of ITFLOWS, will be responsible for the data management, except for the following cases:

- * For interruptions, losses or any damage on data stored in the DDD that are due to force majeure;
- * For any misuse of data/dataset created or collected by a Partner;
- * For any issue, damage, legal violation or privacy breach that occurs during interviews or during the handling of interview data among the respective NGOs and Partners;
- * For any breach of internal obligations and/or international/national legislation committed by a Partner;
- * For content (data and/or documents) not duly reviewed under ITFLOWS internal mechanisms;
- * For any excessive costs incurred by a partner when publishing via Gold Open Access.

3.3) Resources for long term preservation

Regarding the question of long term data preservation, to date, no specific arrangements have been made by the ITFLOWS Consortium. However, all ITFLOWS documents deposited in the DDD will have a permanent link that can be viewed in the record.

4. Data security

4.1) Provisions in place for data security and long term preservation and curation measures

Apart from what is described in Part 2 and Part 5 regarding the Project's compliance with the EU GDPR, it is important to highlight that—in order to ensure data security—we will implement technical and organisational measures (c.f. Article 32 GDPR) appropriate to the risks identified through the legal and ethical guidance implemented in the project (e.g. WP2, IEB, DPA).

Specifically, the DDD uses the *Invenio* software, which is open source software under an MIT license that was created at the European Council for Nuclear Research (CERN). The hardware system builds in redundancy models to guarantee reliability, and is subject to the usual renewal and security protocols of the University Information Service (*Servei d'Informàtica de la UAB*). The UAB applies the same digital information security and preservation procedures for backups, including those stored outside the university's facilities.

The Open Archives Initiative model allows interoperability with the OAI-PMH metadata transmission protocol (Open Archive Initiative - Protocol for Metadata Harvesting). This protocol allows the visibility of the documents from different platforms and collectors: Google Scholar, BASE, CORE, etc.

In order to fulfil the European mandates, the DDD is OpenAIRE compliant as it meets all the requirements of metadata required by the European Union, to be able to revise projects. Persistent IDs are provided for each document and author identifiers (ORCID) are included in the metadata. The Library Service will verify publishing rights to guarantee copyright policies for deposited documents.

The UAB Governing Council (*Consell de Govern*) has approved the policy for preservation of documents available in the DDD, which states the institutional sustainability of its infrastructure, as well as the distribution of the different responsibilities within the services and governing bodies of the University.

In addition, regarding the data repository for the EMT, we will use CKAN. CKAN is a tool for creating open data websites, helping to manage and publish collections of data. It is widely used by national and local governments, research institutions, and other organisations collecting data. CKAN is open source software, with an active

community of contributors who develop and maintain its core technology. In addition, it is modified and extended by an even larger community of developers who contribute to a growing library of CKAN extensions.

CKAN is built with Python on the backend and Javascript on the frontend, and uses the Pylons web framework and SQLAlchemy as its ORM. Its database engine is PostgreSQL and its search is powered by SOLR. It has a modular architecture that allows extensions to be developed to provide additional features such as harvesting or data upload.

CKAN uses its internal model to store metadata about different records, and presents it on a web interface that allows users to browse and search this metadata. It also offers a powerful API that enables third party applications and services to be built around it.

In terms of security measures for the EMT, we plan to strictly put into practice all the standard measures for data security:

- Encrypted storage/transmission;
- Regular offline backups;
- Transmission only over secured/encrypted channels;
- Use of firewalls for securing the network;
- Access control with personal user accounts and permissions depending on need-to-access policy;
- Access to the data/tool only via VPN (Virtual Private Network) if deemed necessary;
- Use two-factor authentication and public/private key encryption.

Currently, the CKAN repository is only available to Project partners. After discussion among all Project partners, the Project's WP2 Ethics partner, IEB and DPA will evaluate whether to make the CKAN repository available to other entities, and will offer an explanation as to this by M25 of the project. This M25 Consortium meeting and decision-making will include discussion of access to CKAN and information as to the backend API that holds all the endpoints that serve the output files to the frontend, or files that exist in CKAN. It will also include discussion of EMT users being able to download raw data files below EMT visualisations. In the meantime, as

stated earlier, EMT Reports offering analysis by technical and research partners as to the latest data outputs of the EMT are being delivered to all audiences in M25, M30 and M36.

5. Ethical aspects

Ethical or legal implications that impact data sharing when conducting ITFLOWS research activities are addressed in strict compliance with the ethical and legal framework specifically identified for the project, particularly the EU GDPR, as well as the national legal frameworks in force.

In large part, the ethical aspects of the ITFLOWS project are outlined in the Grant Agreement, and Article 34 of the same regarding Ethics and Research Integrity.

Work Package 2 (WP2) is specifically designated for ethics, with task T2.2 identifying data protection and privacy risks, mandating two continual assessments throughout the project that will affect data sharing: a Privacy and Data Protection Impact Assessment in parallel with the Ethical and Societal Impact Assessment; and an ITFLOWS Regulatory Model (T2.3), with both ethical and legal hard and soft law applicable instruments. These provisions reflect how the ITFLOWS project has committed to privacy by design and privacy by default principles, as well as to the technical methodologies to efficiently implement these principles.

Moreover, a new Work Package (WP10) has been included for the monitoring of ethical issues within the project as a result of the ethics check to which the project was subject before the signature of the Grand Agreement. Data Protection aspects are specifically addressed in the context of the added deliverables imposed in WP10 such as, for instance: the need to provide clarification on the deployment of anonymisation techniques, on tracking and observation of participants and on the appointment/tasks/responsibilities of the ITFLOWS Data Protection Advisor; the need to submit a Data Management Plan; and the need to provide an Informed Consent Template including information regarding destruction of data. These aspects were addressed in Deliverables 10.1, 10.2 and 10.3, submitted in January 2021. These are the specific deliverables imposed by the Commission as a result of the ethics check conducted by the European Commission, as mentioned above.

The protection of personal data within ITFLOWS is conducted in strict compliance with the ethical and legal requirements of: the EU GRPR; National and local

provisions applicable to the ITFLOWS Consortium partners in the field of privacy and data protection; Guidelines issued by the European Data Protection Supervisor and the European Data Protection Board; and the ITFLOWS DPA and IEB decisions and recommendations.

It is important to note that WP leaders are responsible for their WP activities, and must consult or notify the Ethics lead partner and the IEB whenever there is a doubt, risk or issue in accordance with the internal and external ethics strategy specifically designed by the Ethics lead partner in order to address general ethical issues. Data protection issues that may entail risks in terms of jeopardising an individual's data protection and privacy rights have been addressed in particular in the context of the ethics strategy.

The Ethics lead partner, with the support of the IEB and the DPA, monitors that data processing activities comply with the principles of EU data protection law: purpose limitation, data minimisation, data accuracy, and proportionate data retention period and data security.

All data collected in ITFLOWS is initially made available for analysis by all project partners. As described in Part 2, all data stored follows an anonymisation or pseudo-anonymisation process. Formats such as images, videos and audio data are avoided as these are not easy to efficiently anonymise. The project only stores extracted, meaningful information that meets the anonymisation requirements.

The project's ethics monitoring instruments and bodies described above address how ITFLOWS involves human subjects, both in data generated through interviews with migrants, refugees and asylum seekers (vulnerable populations), as well as through participatory workshops and pilot tests of the project EMT with Users Board organisations from the EU Member States.

The transfer of data on human subjects to the ITFLOWS repository takes place only when accompanied with informed consent, the project's ethics approval and again in conformance with international, EU, national and local law. As previously stated, all data transferred to the project repository is either pseudonymised or

anonymised. The partner generating the data is responsible for the anonymisation or pseudonymisation process, and for ensuring that identifiable variables are not transferred to the ITFLOWS repository. Directly identifiable variables include (but are not limited to) official identification numbers, name, phone number, postal code, e-mail address, address, geographical coordinates (that would risk identification). Any variables are provided at the lowest possible resolution that is necessary for analysis to mitigate risk (for example, year of birth or age instead of birth date).

Finally, the Ethics lead partners design the informed consent forms for the project's qualitative research activities foreseen in the project, in close cooperation with the partners involved in developing such research activities.

6. Other issues

DMP parts 2.2 and 5 outline the UAB Institutional policies, Spanish national legislation, as well as European Union policies and regulation that inform the project's procedures for data management. Again, as noted in Part 1, UAB holds ultimate responsibility for managing the project's data, although Consortium partners do have their own institutional policies and procedures that they must adhere to. These include the following:

European University Institute (EUI)

- EUI President Decision No. 10/2019 of 18 February 2019 Regarding Data Protection at the European University Institute (EUI): <https://www.eui.eu/Documents/AboutEUI/Organization/PresidentDecisionOnPrivacy.pdf>
- EUI Guide on Good Data Protection Practice in Research (April 2019 edition): <https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf>

Ethniko Kentro Erewnas Kai Technologikis Anaptyxis (CERTH)

- Πολιτική Προστασίας Δεδομένων Προσωπικού Χαρακτήρα: <https://www.certh.gr/01BF01A8.el.aspx>

Center for European Policy Studies (CEPS)

- CEPS Data Privacy Policy: <https://www.ceps.eu/about-ceps/data-privacy-policy/>

Institut für Weltwirtschaft (IfW)

- IfW Privacy Policy: <https://www.ifw-kiel.de/privacy/>

Istituto Affari Internazionali (IAI)

- IAI Privacy Policy: <https://www.iai.it/it/note-legali>

FIZ Karlsruhe (FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur)

- FIZ Privacy Policy: <https://www.fiz-karlsruhe.de/en/ueberuns/datenschutzerklaerung>
- German Federal Data Protection Act: https://www.gesetze-im-internet.de/englisch_bdsch/

Cork Institute of Technology (CIT)

- CIT Data Protection Policy (2018 edition) https://www.cit.ie/contentfiles/gdpr/180528_CIT_Data_Protection_Policy_Final_V1.0_GB_Approved_5_July_2018.pdf

Associazione della Croce Rossa Italiana (CRI)

- CRI Privacy Policy: <https://www.cri.it/privacy>

Oxfam Italia Onlus (OXFAM ITALIA)

- OXFAM ITALIA Policy on the Treatment of Personal Data: <https://www.oxfamitalia.org/informativa-privacy-oxfam-italia-intercultural/>

Center for the Study of Democracy (CSD)

- CSD Personal Data Protection and Privacy Policy: <https://csd.bg/footer-menu/privacy/>

Open Cultural Center (OCC)

- OCC Privacy Policy: <https://openculturalcenter.org/privacy-policy/>

Terracom AE (TRC)

- TRC Privacy Policy: <https://www.terracom.gr/en/privacy-policy>

Brunel University London (BUL)

- BUL Information Compliance Data Protection Policy (2018 edition): <https://www.brunel.ac.uk/about/documents/pdf/DP-policy.pdf>
- BUL Guidance note: use of personal data in research: <https://www.brunel.ac.uk/about/documents/pdf/persdatares.pdf>

7. Further support in developing the DMP

This DMP has been created using the tool *EinaDMP*⁷. Specific online training from UAB Library personnel was received by the authors of this DMP, in order to access and interact with the online application. In addition, during the process of developing this document, the authors received continuous feedback and assistance. Each of the partners will follow their national and institutional procedures for data management, in addition to this DMP.

Main references:

- Data Management in the context of Horizon 2020 (Available at: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm);
- The Open Research Data (ORD) Pilot in H2020 (Available at: <https://www.openaire.eu/what-is-the-open-research-data-pilot>);
- OpenAIRE Guidelines for Literature Repositories, Data Archives, and CRIS Managers based on CERIF-XML (Available at: <https://guidelines.openaire.eu/en/latest>);
- DMP for H2020. Zenodo repository (Available at: <https://zenodo.org/record/2635768#.X3pTXWhKjIU>);
- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 DOI: 10.1038/sdata.2016.18 (2016). (Available at: <https://www.nature.com/articles/sdata201618>)

⁷ <https://dmp.csuc.cat>