IT FLOWS
IT TOOLS AND METHODS
FOR MANAGING
MIGRATION FLOWS

D2.5
REPORT ON THE LEGAL, SOCIETAL IMPACT AND ETHICAL MONITORING OF THE ITFLOWS PROJECT

1st version

Andrea Guillén & Emma Teodoro
Autonomous University of Barcelona
July 2022

Deliverable Factsheet

Title and number	*Report on the Legal, Societal Impact and Ethical Monitoring of the ITFLOWS (D2.5)*
Work Package	WP2
Submission date	30/07/2022
Authors	Andrea Guillén & Emma Teodoro (UAB)
Contributors	Thilo Gottschalk & Francesca Pichierri (FIZ), Mengia Tschalaer & Alexandra Xanthaki (BUL) Colleen Boland & Mario Macías (UAB)
Reviewers	Konstantinos Kalampokis (TRC) & Asli Okyay (IAI)
Dissemination level	PU (Public)
Deliverable type	R (Report)

Version Log

Issue Date	Version	Author	Change
02/05/2022	V0.1	Andrea Guillén & Emma Teodoro (UAB)	Initial deliverable structure and allocation of WP2 partners' initial contributions
15/07/2022	V0.2	Andrea Guillén & Emma Teodoro (UAB)	Introduction and Section 2
15/07/2022	V0.2	Mario Macías & Andrea Guillén (UAB)	Sections 5.1 and 5.2
15/07/2022	V0.2	Coleen Boland (UAB) & Mengia Tschalaer (BUL)	Sections 4 and 5.3
15/07/2022	V0.2	Thilo Gottschalk & Francesca Pichierri (FIZ)	Section 1.1
15/07/2022	V0.2	Alexandra Xanthaki (BUL)	Section 3
20/07/2022	V0.3	Emma Teodoro & Andrea Guillén (UAB)	Consolidated version sent to final WP2 review
26/07/2022	V0.4	Jonathan Andrew (ITFLOWS DPA)	Changes and feedback made by ITFLOWS DPA
27/07/2022	V0.5	Lilian Mitrou, Marta Poblet & Ruth Fee (ITFLOWS IEB)	Changes and feedback made by ITFLOWS IEB
29/07/2022	V0.6	Konstantinos Kalampokis (TRC) & Asli Okyay (IAI)	Changes and suggestions made by the internal reviewers
29/07/2021	V1.0	Andrea Guillén & Emma Teodoro (UAB)	Final version for submission

Disclaimer
This article reflects only the author's view and that the Agency is not responsible for any use that may be made of the information it contains. (art. 29.5 Grant Agreement)

EXECUTIVE SUMMARY

This Deliverable includes insights in the ad hoc and ongoing monitoring process of the project both from an internal and external perspective with the aim of ensuring the implementation of the ITFLOWS Regulatory model at all stages of the project research activities and particularly with regard to the EUMigraTool. For better oversight it distinguishes between data protection perspective (FIZ), the ethical perspective (UAB), the societal perspective otherwise known as the human rights considerations (BUL) and the gender perspective (UAB). It also includes the external and independent monitoring carried out by the Independent Ethics Board, the Data Protection Advisor, and the Independent Gender Committee. This report is the first version of the Report on the legal, societal impact and ethical monitoring of ITFLOWS. As the monitoring process is ongoing, its outcomes will be shared within the project over the course of the project on a regular basis. The overall results and an evaluation of the process will be included in the second and final version of the report in M36.

Table of Contents

List of Tables

List of Figures

Abbreviations

2FA: Two-factor authentication

API: Application Programming Interface

BUL: Brunel University London

CERTH: Ethniko Kentro Erevnas Kai Technologikis Anaptyxis

CMP: Compliance Monitor Platform

CNIL: Commission Nationale de l'Informatique et des Libertés (French Data Protection Authority)

CRI: Associazione della Croce Rossa Italiana

DPA: Data Protection Advisor

DPIA: Data Protection Impact Assessment

EC: European Commission

EMT: EUMigraTool

EtSIA: Ethical and Societal Impact Assessment

FIZ-IGR: FIZ-Karlsruhe – Legal Team

FIZ-ISE: FIZ-Karlsruhe – Information Service Engineering

FRONTEX: European Border and Coast Guard Agency

GAP: Gender Action Plan

GDPR: General Data Protection Regulation (Reg. EU 2016/679)

HLEG: High-Level Expert Group on AI

IAI: Istituto Affari Internazionali

IDT-UAB: Institute of Law and Technology - Universitat Autònoma de Barcelona

IEB: Independent Ethics Board

IGC: Independent Gender Committee

JWT: JSON Web Tokens

JSON: JavaScript Object Notation

MTU: Munster Technological University

NGO: Non-governmental organizations

OCC: Open Cultural Center

OIT: Oxfam Italia Onlus

ToU: Terms of Use

UAB: Universitat Autònoma de Barcelona

UUID: Universal Unique Identifier

WCAG: Web Content Accessibility Guidelines

WP: Work Package

INTRODUCTION

The following report depicts the current status of the legal, ethical and societal reporting in the ITFLOWS project. The legal, ethical and societal impacts of the project have successfully been identified in Deliverables D2.1, D2.2, D2.3 and D2.4 and the findings of these works are now subject to continuous review and implementation efforts on the technical and organisational level. This report provides a conclusive overview of the work and state of the legal, ethical and societal aspects in the ITFLOWS project, as well as a description of the short-, mid- and long-term steps that are envisioned to ensure legal, ethical and societal compliance over the course of, and with regard to the EUMigraTool (EMT), beyond the ITFLOWS project. Due to different implications for the project, the following sections will distinguish between the legal, ethical, societal and gender-related aspects.[1] Section 1 presents the legal monitoring, which covers both human rights and data protection monitoring. Section 2 describes the ethical monitoring of the ITFLOWS research activities. Section 3 covers the societal monitoring and Section 4 the gender monitoring of the ITFLOWS project. Lastly, Section 5 presents the legal, ethical and gender-related monitoring conducted by the external and independent monitoring bodies of ITFLOWS, namely, the Independent Ethics Board (IEB), the Data Protection Advisor (DPA) and the Independent Gender Committee (IGC).

---

[1] The submission date for this report was moved from M14 to M23 and hence depicts the technical developments and monitoring impacts at this point (i.e., M23).

1. LEGAL MONITORING

1.1. Data Protection monitoring

Within the ITFLOWS Project, the legal team of FIZ Karlsruhe (IGR) is focusing on data protection aspects of the project. This includes identification, guidance and monitoring of data protection impacts as well as corresponding mitigation measures. Previous deliverables mainly focused on the impact of the data processing conducted throughout the project, whereas this report aims to describe the monitoring approaches implemented in the project. The monitoring of the project's efforts pursues various objectives. Ostensibly, monitoring is a tool to keep track of the efforts of the various partners. From a project management perspective, monitoring can be used to guide the next steps to ensure a compliant development. At the same time, monitoring also puts some necessary pressure on the project's partners to fulfil the legal and ethical requirements by making their efforts – or the lack thereof – visible.

In putting the monitoring to use, ITFLOWS combines various approaches to keep track of the efforts. Furthermore, it puts some pressure on the responsible partners and provide guidance and tools for partners to facilitate their efforts to ensure lawful data processing in the context of the ITFLOWS project. At the same time, the implementation of such monitoring measures and surrounding activities also must be viewed from a research perspective, and all involved legal and ethical partners conduct research on their respective areas of expertise.

In this light, multiple measures have been implemented and enforced to ensure proper monitoring of the activities conducted in ITFLOWS. Such measures are described in the following sections. We thereby hope to provide guidance to similar projects and contexts as well general transparency with regard to the efforts made in the project.

1.2.1 Monitoring Approach

ITFLOWS presents the legal partners—as well as ethical and societal researchers focusing on compliance aspects of the project—with multiple challenges. These challenges need to be addressed in order to allow proper monitoring and steering of the relevant processes. This also holds true for data protection requirements

imposed on the project in relation to the safeguard of fundamental rights, specified by the GDPR. One of the main challenges in this regard is to keep track of the actions and processes that take place in a project with 14 partners across Europe. The bar was additionally raised due to the pandemic situation that cancelled any effort to meet and discuss challenges in depth and in person.

Independently of these general challenges, the following steps have been identified as relevant to ensure legal—and in particular data protection—compliance within the project.

In short, the legal team needs to:

a) see what is happening **(Overview)**,

b) understand what is happening **(Insight)**,

c) analyse the respective approaches **(Analysis)**,

d) identify the necessary measures **(Identification)**,

e) communicate the measures **(Communication)**,

f) review the implementation of the measures **(Review)**.

The following document will hence follow these steps to report on the implemented monitoring approaches, the progress made so far, the challenges that lie ahead, and how we aim to approach these challenges over the remainder of the project.

a) Overview

ITFLOWS combines fourteen partners across Europe in ten Work Packages (WP), two of which are particularly targeted at legal and ethical requirements. This report will focus on the work of the legal monitoring activity in WP2. From the ten Work Packages, four are particularly focused on data driven research and are hence the focus of data protection monitoring, as the involved data processing poses the most complex risks to individuals. The Work Packages—namely WP3, WP4, WP5 and WP6—each pursue different research approaches that are intended to be linked at a later stage in the project in the form of the EUMigraTool (EMT). The same, partly, is true within the Work Packages and for the individual partners. To this end, the nexus of Work Packages and partners within Work Packages depends on the compatibility of their respective research approaches.

For example, qualitative research approaches are often difficult to combine with quantitative approaches. At the same time, this combination may prove particularly

helpful and is hence also included in the ITLFOWS project. For legal monitoring it is necessary to keep track of the individual approaches as well as the envisioned or actual linking of the approaches. The variety of research approaches that are embedded in the project result in a high level of complexity not only from an organisational point of view, but also from the perspective of overseeing the legal implications of "the bigger picture". To properly evaluate the technical partners' work, the full picture needs to be taken into consideration.

At the time of writing, the connection between the Work Packages is clear overall; the EMT´s general architecture has been described in D6.2 (which complemented D6.1), providing an overview on how the Work Packages and the contained research approaches are intended to be combined and linked within the EMT. Given the complexity of the overall approach, a great effort was made to understand the bigger picture, technical interfaces, and the possible direction for the agile development of the project. To achieve this, within the first 12 months of the project, the legal monitoring team identified the envisioned and/or already used data sources in the project. To this end, bilateral and multilateral communication with the respective partners have been pursued. During the first months of the project, the initial level of communication towards the legal teams was initially slightly lower than expected, but increased over time due to additional efforts of the legal teams. Overall, we observed a broad variety of awareness and understanding of the impact, relevance and consequences of legal and ethical requirements imposed on the project. To build a foundation for the subsequent monitoring of approaches, this understanding is key. It was hence fostered by the legal and ethical team through additional communication efforts (i.e., frequent participation in technical calls, "questioning" approaches and solutions, bilateral discussions, etc.) to raise awareness of the relevant topics and align the project partners' understanding of the importance of lawful and ethically compelling approaches. To support communication and enable legal researchers to "see" what is happening in the various areas of the project, FIZ Karlsruhe further suggested multiple measures to ease the retrieval of information from the respective work packages. These measures included:

- standardized templates for "Action Items" in each Work Package;
- the proposal of a shared calendar;
- participation in other work package meetings;

- workshops for discussion;

- standardized forms to collect further information about data processing;

- introduction of a data driven monitoring platform.[2]

All of these measures foster the ability for the legal and ethical teams in the project to pursue the necessary legal and ethical research to support the project. Except for the shared calendar, all measures have been successfully implemented and used as a tool for communication and information retrieval. As a substitute for a shared calendar, it was decided to use MSTeams to communicate important meetings and insights from other work packages. It was initially discussed to set up an Etherpad as a collaborative tool for internal WP2 discussions, but most users preferred using MSTeams.

The substantial effort made by the legal and ethical team to obtain a sufficient overview of the project's approaches proved to be successful. At the time of writing this deliverable, the overall approaches and linkage between various technical tools and research outcomes is mostly clear. Equally important, the awareness of the project partners on the legal and ethical implications has been aligned and the necessity for legal compliance—and related consequences—is subject to a high degree of acceptance from all partners. This is particularly helpful as partners now also reach out on their own to the legal team to request support on specific topics (e.g. asking for opinion on the use of tools such as atlas.ti prior to starting work with them)—in contrast to opaque work "behind closed doors" and distrust regarding the legal and ethical teams that can regularly be observed at the intersection of law and technology. To this end, one of the key goals—namely to provide a foundation for continuous insight and oversight of the project's actions—has been achieved.

b) Insight

The legal monitoring eventually should result in tailored approaches for the different work packages and contexts in the project. To this end, the most important differentiation was between quantitative and qualitative research approaches. Qualitative research is mainly conducted in WP3 in the form of interviews with

---

[2] https://cm-p.eu

migrants. Beyond that, most Work Packages also conduct literature research. From a legal perspective the literature research does not constitute any obstacles. At the same time, quantitative (i.e., data driven) research is conducted in WP3, WP4, WP5 and WP6.

In the first phase, the goal was to get insights on the research approaches that are happening in the respective Work Packages. To achieve this, FIZ Karlsruhe reached out to the Work Package Leaders in ITFLOWS and asked 1) for initial technical information on the processing and 2) for articles that are related to the research approaches plan to apply in ITFLOWS. In line with the efforts detailed above, (Oversight a) this proved to be important not only to identify legal and ethical risks but also to be able to discuss challenges on par with technical partners properly and understand the technical challenges they are facing. For example, it would not be helpful to discuss anonymisation techniques if anonymisation would undermine the whole research approach—it is hence necessary to understand (on a technical level) why certain approaches are feasible and others not. For example, sometimes it can be more compelling to ensure data protection on the level of data security (e.g., processing on closed premises). The key papers that have been taken into consideration to generate insight on the underlying technical approaches in this regard are listed in Table 1. The selection is based on the recommendations collected from the technical partners, paired with individual research by team members of FIZ Karlsruhe (IGR).

Paper	Type
Böhme, M.H., Gröger, A., Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, Vol. 142.	Technical Description of Empirical Methodology for Google Trends Analysis.
Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. Science (New York, N.Y.), Vol. 359, No. 6373, pp. 325–329.	Comparable research approach but focused on US Migration.
Suleimenova, D., Groen, D. (2020). How Policy Decisions Affect Refugee Journeys in South Sudan: A Study Using Automated Ensemble Simulations. *Journal of Artificial Societies and Social Simulation*, Vol. 23, No. 1.	Description of the computational modelling of refugee movements in South Sudan based on Automated Ensemble Simulations.

Alam, M., Gesese, G.A., Sack, H. (2020). MigrAnalytics: Entity-based Analytics of Migration Tweets, available at: ceur-ws.org/Vol-2721/paper514.pdf.	Description of technical approach to generate knowledge graphs from Twitter Data.
Fafalios, P., Iosidis, V., Ntoutsi, E., Dietze, S. (2018). Tweetskb: A public and large-scale RDF corpus of annotated tweets. CoRR abs/1810.10308.	Description of TweetsKB

Table 1. List of technical papers (1).

In combination with the direct communication with the partners this approach provided a valuable first insight into the respective processing methods and helped the monitoring team to understand what is happening "under the hood" of ITFLOWS. The first results of this process are reflected in the prior deliverables of WP2. In this context, the initial examination of the technical papers proved helpful to bring legal and technical partners on the same page and enabled a valuable and continuous discussion regarding possible mitigation measures. During the course of the project and thanks to the interaction with the technical partners, new technical papers were recommended and hence examined to highlight the necessity for continuous collaboration between legal and technical partners. The list is shown in Table 2:

Paper	Type
Chen, Y., Gesese G.A., Sack H., Alam, M. (2021). Temporal Evolution of the Migration-related Topics on Social Media. Proceedings of the ISWC 2021 Posters and Demos Track, co-located with the 20th International Semantic Web Conference (ISWC 2021), CEUR Workshop Proceedings, Vol. 2980, CEUR-WS.org.	Description of how to capture the temporal evolution of migration-related topics on relevant tweets.
Chen, Y., Sack, H., Alam, M. (2021). MigrationsKB: A knowledge base of public attitudes towards migrations and their driving factors, available at https://arxiv.org/abs/2108.07593.	Focus on the use of Artificial Intelligence methods based on knowledge graphs and neural networks for analysing public attitudes toward migration on Twitter.
Dieng A.B., Ruiz F.J.R., and Blei D.M. (2020) Topic modeling in embedding spaces. *Transactions on Association of Computational Linguistics*, Vol. 8, pp. 439–453.	Explanation of topic modelling technique Embedded Topic Model (ETM)
Suleimenova, D., Bell, D., Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific Reports*, Vol. 7, No. 13377.	Explanation of the Simulation Development Approach (SDA) to predict the destinations of refugee movements in conflict regions.
Carammia, M., Iacus, S.M., Wilkin, T. (2022). Forecasting	Overview of an existing prediction approach

asylum-related migration flows with machine learning and data at scale. *Scientific Reports*, Vol. 12.	based on ML that integrates official statistics and non-traditional data sources.

Table 2. List of technical papers (2).

c) Analysis

Once information is gathered as is detailed above, the next natural step is the analysis of such information. The analysis that FIZ Karlsruhe (IGR) conducts in ITFLOWS entails legal as well as technical examination of the data processing activities conducted or planned to be conducted in the different project tasks. The legal impact of the data processing activities is evaluated in the light of risks posed to the fundamental rights and freedoms of the data subjects. The evaluation is mainly based on the current European data protection framework. Technical approaches are also assessed from a technical point of view in order to facilitate the identification of risks and provide tailored suggestions on technical and organisational measures helpful in mitigating the risks identified.

The main goal of analysing processing activities in ITFLOWS is to verify compliance with data protection law and identify data protection risks related to research activities and the EMT, therefore ensuring protection of privacy-related interests. This goal can only be successfully achieved with mutual cooperation between both parties, technical partners/controllers provide information concerning their activities and legal/ethical partners analyse it, as they commit themselves to understand each other's perspectives and expertise fields. This is why FIZ Karlsruhe, since the beginning of the project, has put emphasis on internal communication and technical preparation (see the subsections above **Overview** and **Insight**).

ITFLOWS requires an ongoing legal and technical examination and revision as the data processing within the project is subject to continuous development. At the core of the analysis task, there is the *Data Protection Impact Assessment* (DPIA) process (T2.2). The monitoring team has already extensively described the DPIA process and the DPIA methodology applied in the project in D2.3. While the overall context and goals of the processing is broadly clear, the underlying processing is subject to continuous changes. Equally, the reasons for such cases can be manifold and encompass, for example, performance tweaking or use cases that have been identified through the feedback of

the stakeholders on the ITFLOWS Users Board. With regard to data protection, a great deal of effort was put into the analysis activity over the course of the project to be able to adapt to changes in the processing and/or the purposes of the processing. As a first step, FIZ Karlsruhe conducted an in-depth analysis of *data sources* envisioned to be used in the project. The datasets were classified based on the risks they pose to natural persons when processed. Data sources which do not contain personal data were classified as low risk datasets; the ones partly or potentially containing personal data were classified as medium risk datasets; the ones containing personal data were classified as high-risk datasets. The analysis of data sources allowed the identification of high-risk datasets in need of specific consideration within the project. By extension, processing activities making use of such data sources were assessed as requiring a stricter evaluation and monitoring. This included not only the analysis (i.e., processing for research purposes) itself but also the subsequent use of data. Especially with regard to the latter, WP2 is analysing means and measures to make data available to interested parties without undermining fundamental rights of the data subjects. On this matter, it was suggested to include fine-grained access control not only on the EMT Frontend (i.e., the website) but also if data is provided through technical interfaces (i.e., APIs).

The focus of the analysis hence covers data processing methodologies and activities to be put in place by technical partners in the context of their research but also in a potential operational context of the EMT. FIZ researchers use a multi-tracked approach to fulfil such a task in the best way. As is detailed above, the approach includes bilateral discussions with technical partners, participation in technical calls, study of the technical aspects embedded in the project, preparation of questionnaires for technical partners and analysis of technical deliverables. The analysis of all information gathered through these channels is the main task for FIZ Karlsruhe (IGR) and WP2 members in general. This approach results in the identification of risks and related data protection requirements that are directed towards specific technical partners as described below.

d) Identification

The **analysis** task results in the determination of data protection requirements and

in the detection of risks concerning data processing. Once general requirements are pointed out and risks have been detected, WP2 aimed to define individual mitigation measures to fulfil such requirements and address in advance such risks that are identified within the project. The identification of such individual risks and potential mitigation measures proves to be particularly challenging in a project such as ITFLOWS. As described before, the multitude of aspects, partners and challenges can quickly go into the hundreds and is hardly depicted in the traditional deliverable. Historically, this often forced the legal and ethical teams to either shift to rather generic description of requirements that left technical partners uninformed as to the specific measures they can or need to implement. On the other side of the spectrum, detailed descriptions result in lengthy deliverables that are hardly accessible by non-legal personnel. To overcome this, ITFLOWS decided to pair the traditional approach with a more dynamic data-driven solution—namely the compliance-monitor platform (CMP). The CMP has been developed in the aftermath of a previous EU Project (TITANIUM) and is specifically targeting the needs of research and development projects such as ITFLOWS. It was hence agreed to make use of the platform and to provide input regarding the specific requirements to steer the external development of the platform in a way that addresses the needs of ITFLOWS and other projects. In consequence, the findings of the ITFLOWS project can also be transferred to other (EU) projects through the CMP.

The identification of adequate measures that fulfil legal requirements and address risks concern both legal partners and technical partners. Usually, technical partners draw on recommendations and suggestions coming from the legal team (e.g., in the case of Twitter analysis and interviews). In some circumstances, technical partners take the lead in proposing measures for a specific issue (e.g., in a recent meeting concerning the development of the EMT, CERTH pointed out that the institution has an already established set of measures that would mitigate the risks emerging from the use of a proprietary software). In both cases, it is a matter of teamwork as measures are then communicated, discussed, implemented and reviewed.

The CMP provides a centralized platform to collect and link legal and ethical requirements with the responsible partners in the project. It combines project management aspects (e.g., defining responsibilities) with a tool to collect and communicate (see below) the identified requirements.

In the identification phase, the platform provides a structured input form that allows legal personnel to collect requirements in bundles. It thereby enforces the collection of individual requirements in a consistent form and with all the necessary information in one place (see below, Figure 1 and2). The collected information includes:

- Title
- Responsible entity
- Related task in project
- Related topic
- Legal basis/source
- Risks analysis and a risk score
- Requirement description
- Solution proposal
- Relevant (research) documents
- Link to technical interface

Figure 1. CMP: Basic Info and Requirement Description.

Figure 2. CMP: Related Documents and Solution Proposal.

The structured collection of information aims to help technical partners to oversee what is actually required from them and the solution proposal helps them to implement proper solutions. As this is a new approach to handling legal and ethical requirements, ITFLOWS decided to go for a hybrid approach, combining platform driven requirement management with "traditional" deliverables.

e) Communication

Following the identification of the respective requirements it is necessary to communicate the requirements in a transparent and understandable form. To do so, various approaches can be used. Given the hybrid approach pursued in WP2, the communication of requirements in ITFLOWS takes place through multiple channels. On the one hand the means used to fulfil the aforementioned tasks in ITFLOWS include deliverables, internal meetings and workshops, email exchanges, and communication through MS Teams. In addition, the CMP is used to reflect the requirement of the agile environment the project works in. In consequence, the CMP is currently used in cases that require and allow for individual requirements whereas more general requirements have been collected and communicated through traditional means (i.e., deliverables). That being said, by May 2022, the monitoring team created more than 100 requirements on the platform that are directly communicated to the responsible partners. Currently ~40 requirements have been solved by the responsible partners. Most of the other requirements have a deadline by the end of the project and are hence not yet expected to be fulfilled/solved.

		Added	Requirement	Responsible Partner	Task	Topic	Source	<	⊘+	☑
⬆	⋮	Fri Sep 10 2021	Technical and Organizational Measures to ensure compliance in Twitter Analysis (FIZ)	FIZ	T3.3	REA Ethics Check	Art. 6 (1) (f) GDPR	☑	☑ 1	☑
⬆	⋮	Fri Sep 10 2021	Lawful Basis for Twitter Analysis - FIZ	FIZ	T3.3	REA Ethics Check	REA Ethics Check	☑	☑ 1	☑
⬆	⋮	Mon Feb 14 2022	Data Protection Training for NGOs (Interviews)	FIZ	T2.4	ITFLOWS Interviews	Various.	☑	☑ 1	☑
⬆	⋮	Mon Oct 04 2021	Art. 14 GDPR - Transparency Requirement	FIZ	T9.1	ITFLOWS Website	Art. 14 (1) GDPR	⊘	☑ 1	☑
⬆	⋮	Thu Feb 03 2022	Create a Privacy Policy for EMT	FIZ	T2.4	EMT Frontend	Art. 5 (1) lit. a GDPR	⊘	⊘	⊘

Rows per page: 5 ▾   1-5 of 8   < >

Figure 3. Exemplary list of requirements linked to a specific partner.

On the one hand, the platform enables legal and ethical partners to communicate requirements towards the responsible (technical) partners in a structured and transparent way (see Figures 3-5).

Specify collected data
R
Added: 2022-02-06T19:42:57.000Z
Requirement UUID: 2549dabf-0454-4e6f-87c8-1b3287a577eb

VIEW PARENT

Responsible Partner:

Terracom Ae (TRC)
79 Ethnikis Antistasis, Katsikas,
Ioannina
Greece

Contact: sgous@terracom.gr

Related Task:

T6.5

Visualisation, infrastructure integration
and deployment (M10-M28)

Info:

Project: ITFLOWS

Risk Rating: 4

Topic: EMT Frontend

Short description

Please specify what data is collected and processed through the EMT Frontend - i.e. data of website visitors/users. Please elaborate for what purposes the respective data is collected. Create a new solution for each data type/context (e.g. consent cookies, account data, youtube/googlead, Kibana, Elastic, Avada).

Full description

The EMT frontend collects a variety of data from the user through cookies, download tracking and scripts. Please specify which data is collected, why that collection is necessary and how the respectively collected information is handled (e.g. anonymization, retention dates, etc.)

Create a new solution for each data-point/context. E.g. data collection through Kibana, standard reporting in Avada-Wordpress Template, etc.

Identified transmissions include:

Admin-ajax.php

POST https://emt.itflows.eu/wp-admin/admin-ajax.php "action": "moove_activity_track_pageview", "post_id": "466", "is_single": "1", "is_page": "", "user_id": "3", "referrer": "https://emt.itflows.eu/privacy-policy/", "extras": "[]" { "data": { "post_id": 466, "user_id": 3, "status": "visited", "user_ip": "87.189.***.***", "display_name": "itflowsdemo", "city": "Hanover", "post_type": "post", "referer": "https://emt.itflows.eu/privacy-policy/", "month_year": "022022", "visit_date": "2022-02-02 13:38:53", "campaign_id": "1642681608466" }, "id": 1298

https://emt.itflows.eu:5601/api/ui/_counters/_report

{ "report": { "application_usage": { "dashboards-main": { "appId": "dashboards", "minutesOnScreen": 2.7506333333333335, "numberOfClicks": 0, "viewId": "main" }, "short_url_redirect-main": { "appId": "short_url_redirect", "minutesOnScreen": 0.013, "numberOfClicks": 0, "viewId": "main" } }, "reportVersion": 3, "uiCounter": { "DashboardPanelVersionInUrl-loaded-7.13.0": { "appName": "DashboardPanelVersionInUrl", "eventName": "7.13.0", "key": "DashboardPanelVersionInUrl-loaded-7.13.0", "total": 10, "type": "loaded" } }, "userAgent": { "kibana-user_agent": { "appName": "kibana", "key": "kibana-user_agent", "type": "user_agent", "userAgent": "Mozilla/5.0 (X11; Linux x86_64; rv:96.0) Gecko/20100101 Firefox/96.0" } } } }

Kibana/Elastic Maps

How are Map-tiles loaded for the visualisation. Does elastic track user data? Did TRC check ToU for the Elastic License?

Youtube/GoogleAd

The website uses google tracking cookies.

Consent Cookies

The website stores a variety of cookies on the end-users device. Please specify what cookies are used for what purpose.

Risk Assessment

Failure to inform users about data collection, transfers or analysis can result in fines from DP authorities.

Legal Sources & References

Art. 5 (1) GDPR - Transparency

Figure 4. Exemplary Requirement on the CMP.

Solution: Specification of EMT Frontend Data Collection

Added: Tue Apr 12 2022

Provided by: ggogolos@terracom.gr (TRC)   Score: Not accepted! (NaN)   Tracking Keyword: Tracking Url: https://emt.itflows.eu/

We collect data in various ways but only for security purposes.

Account Data we collect:

- At **Registration Form** (https://emt.itflows.eu/register/) we collect information (Username, Password, Name, Organization, Type of Organization, Country, Email Address, Website URL and Purpose of use) in order to successfully register to the site. All fields are mandatory.
- At **Contact Form** (https://emt.itflows.eu/contact/) we collect your Personal Information (Name, Email and Phone number) to respond to your inquiry or request and to send updates or notices about EMT.

Avada Theme

- The Avada Theme does not communicate with any external sites. All data needed to make the theme work properly are safely stored in the WordPress database.

Data collected for statistical analysis

We monitor and track EMT website activity. **What we track?**

- User activity (duration on the website - date and time of login and logout)
- IP address
- Pages & posts visited by each user
- Log suspicious admin activity.
- When a page/post was published, and who published it.
- If a plugin/theme was activated/deactivated.
- We store no cookies for statistical analysis purposes on users' computers and therefore requires no cookie opt-in from users.

Why is tracking necessary?

- Securing the site by tracking log of all user activity. If someone is trying to hack the site.

Additional Information

- **Consent Cookies** This website uses cookies to improve your experience while you navigate through the website. Out of these, the cookies that are categorized as necessary are stored on your browser as they are essential for the working of basic functionalities of the website. We also use third-party cookies that help us analyze and understand how you use this website. These cookies will be stored in your browser only with your consent. You also have the option to opt-out of these cookies. But opting out of some of these cookies may affect your browsing experience. You can at any time change or withdraw your consent from the Cookie Declaration on our website.

Cookie	Duration	Description
cookielawinfo-checkbox-analytics	11 months	This cookie is set by GDPR Cookie Consent plugin. The cookie is used to store the user consent for the cookies in the category "Analytics".
cookielawinfo-checkbox-functional	11 months	The cookie is set by GDPR cookie consent to record the user consent for the cookies in the category "Functional".
cookielawinfo-checkbox-necessary	11 months	This cookie is set by GDPR Cookie Consent plugin. The cookies is used to store the user consent for the cookies in the category "Necessary".
cookielawinfo-checkbox-others	11 months	This cookie is set by GDPR Cookie Consent plugin. The cookie is used to store the user consent for the cookies in the category "Other.
cookielawinfo-checkbox-performance	11 months	This cookie is set by GDPR Cookie Consent plugin. The cookie is used to store the user consent for the cookies in the category "Performance".
viewed_cookie_policy	11 months	The cookie is set by the GDPR Cookie Consent plugin and is used to store whether or not user has consented to the use of cookies. It does not store any personal data.

- **Necessary (Always Enabled)** Necessary cookies are absolutely essential for the website to function properly. These cookies ensure basic functionalities and security features of the website, anonymously.
- **Functional (optional)** Functional cookies help to perform certain functionalities like sharing the content of the website on social media platforms, collect feedbacks, and other third-party features.
- **Performance (optional)** Performance cookies are used to understand and analyze the key performance indexes of the website which helps in delivering a better user experience for the visitors.
- **Analytics (optional)** Analytical cookies are used to understand how visitors interact with the website. These cookies help provide information on metrics the number of visitors, bounce rate, traffic source, etc.
- **Advertisement (optional)** Advertisement cookies are used to provide visitors with relevant ads and marketing campaigns. These cookies track visitors across websites and collect information to provide customized ads.
- **Others (optional)** Other uncategorized cookies are those that are being analyzed and have not been classified into a category as yet.
- **admin-ajax.php** The admin-ajax. php file contains all the code for routing Ajax requests on WordPress. Its primary purpose is to establish a connection between the client and the server using Ajax. WordPress uses it to refresh the page's contents (for example search results) without reloading it, thus making it dynamic and interactive to the users.

Figure 5. Exemplary Solution on the CMP.

At the same time, the platform also provides partners with a simple way to communicate their efforts and solutions towards the other partners on a central platform in the form of structured solutions (see Figure 5).

In addition, the central collection of requirements and solutions also opens novel possibilities to communicate data protection efforts in a targeted way towards specific stakeholder groups. Such stakeholders can, for example, be the public or data protection authorities.
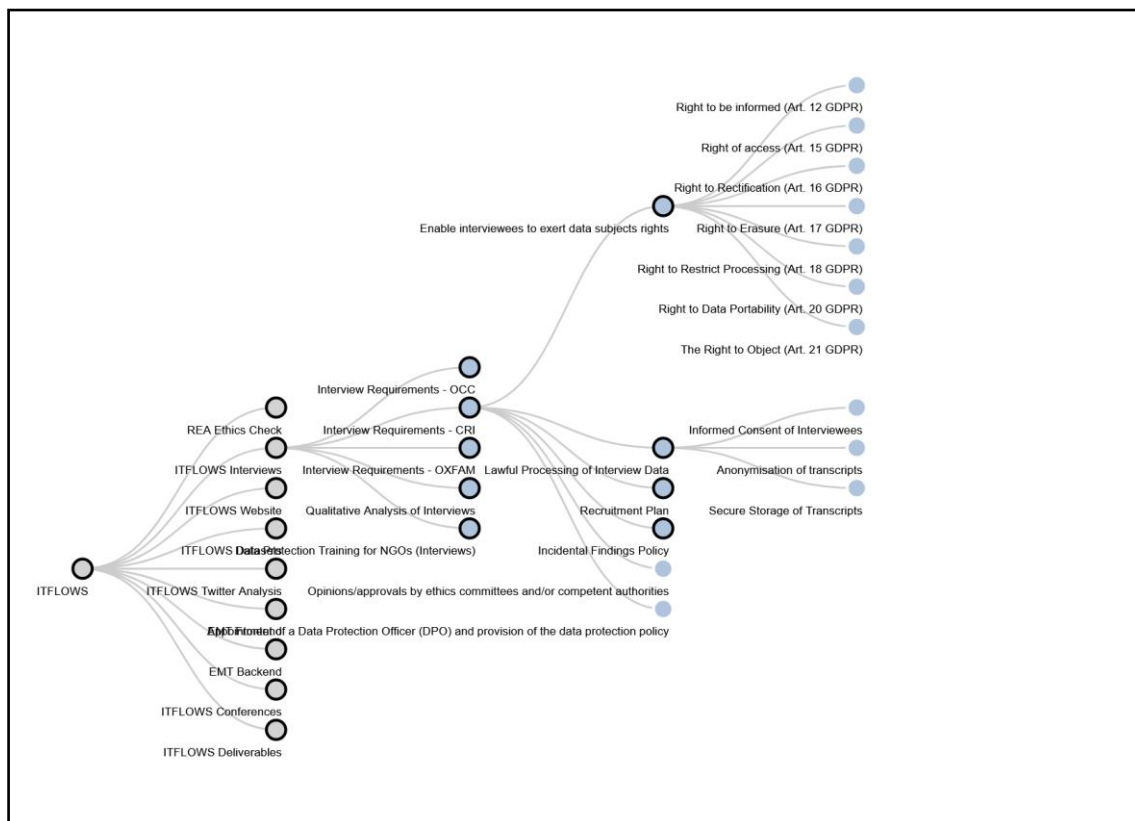
Figure 6. Example of a visual communication of ITFLOWS requirements.

To this end, the platform provides various degrees of detail in the publication of the requirements and their corresponding mitigation measures (i.e., solutions). For internal purposes the corresponding solution will be shown in full detail (see Figure 5 Exemplary Solution on the CMP) and provides the basis for the next step, namely the review of the solution. All related parties will also be informed about changes (e.g., new reviews, new solutions) on the respective requirement.

Regarding providing transparency to the public, visualisations can be used to emphasize and explain data protection (or any other) efforts in a straightforward

and compelling way.

The use of the CMP in the project in the context of data protection monitoring hence helps to comply with the transparency requirements laid down in Article 12 GDPR. In addition, the platform also enables ITFLOWS to provide relevant information to data protection authorities—if necessary—as every relevant step can be exported in detailed pdf-reports.

The communication and review process implemented in ITFLOWS hence effectively fosters compliance with data protection law and significantly limits the probability of negative outcomes in respect to data subjects. By following a 'data protection by design and by default' approach (Article 25 GDPR), the project aims at developing and providing a privacy compliant tool.

Requirements created by the monitoring team most often contain suggestions concerning adequate technical and organisational measures that can be put in place by technical partners (some of them already included in deliverables). Technical partners are asked to find solutions to the requirements and such solutions can be then reviewed directly by WP2 partners (and also by other technical partners involved in the project) in an agile manner. This makes the process of identification, communication, and review of measures more organized and efficient.

f) Review

Solutions should be subject to subsequent review by the relevant experts. The "traditional" approach opens limited possibilities to review solutions in a meaningful way especially due to lengthy intervals between deliverable publication and technical solutions. Participation in calls and workshops on the other hand provide short time-intervals between requirement, solution and review but lack the necessary transparency and trails of the process. To overcome this dilemma, the CMP comprises all three steps in a single and transparent way. Following the provision of a solution, the relevant legal and ethical partners can hence also provide reviews for the solution.

1.2.2 First results

The monitoring approach that is pursued in ITFLOWS so far has been proven to be successful. The legal monitoring pushed partners of ITFLOWS to a) think about

legally compliant processing approaches, b) adapt their approaches to legal requirements and c) conduct their research in legally compliant ways without losing on the quality of the research. In many cases, the underlying datasets consist of aggregated data (e.g., Eurostat), which pose limited to no risks to individuals on their own. A key discussion within the legal team was hence the intention to connect these datasets with data from other ends (e.g., Twitter data, interviews) of the project. This approach would leverage the data protection impact of the overall processing in the project. To minimize the risks of this connection before data is actually connected, FIZ Karlsruhe put relatively strong focus on both Twitter data and interview data.

a) General

With regard to the project in general, the legal monitoring was successful in terms of increasing overall awareness of data protection requirements. The deliverables D2.1, D2.3 and D2.4 provided guidance to all partners in the project regarding legal (as well as ethical and societal) matters and provided a solid legal foundation for the next steps in the project. Due to the complexity of the legal matters in the project, the outcomes of these deliverables have not only been explained and discussed with the relevant partners in the project, including the NGOs, but also in, individual training sessions and workshops with the partners.

Members of FIZ Karlsruhe also participated in multiple (web-)conferences to monitor progress in other projects and get insights on their research activities to identify possible synergies. It was identified that within the H2020 program no other projects pursue a similar track to ITFLOWS. Many projects appear to focus on surveillance measures and migration management, i.e., border control. As a consequence, it is expected that the end-users of other projects (e.g., Frontex and the national border control units) will have a great interest in ITFLOWS outcomes as well. From a legal perspective, this highlighted the necessity to include a proper user and rights management in the EMT, to avoid misuse of the tool (e.g., outputs of the EMT contributing to practices such as pushbacks).

The consequences of such findings are also reflected in the respective WP2 deliverables (in particular D2.3, D2.4) and also transferred to the specifications for

the EMT that are laid down in Section 3 and Section 4 of D6.1 and that have been discussed in several WP6 and EMT meetings. This again shows the importance of participation on the technical side of the project for successful monitoring and implementation of measures. The results of these efforts considerably strengthen the data protection compliance of the project and thereby also reduce risks of potential misuse of project outcomes. As an initial result, the measures that envisioned in the EMT to ensure legal compliance include, among others:

- provision of an informed consent form before collecting any data;
- summary of the content of data and why it is needed before collection in order to avoid collection of sensitive and unwanted data;
- implementation of privacy enhancing techniques such as anti-tracking, encryption of sensitive data and secure file sharing;
- anonymizing data whenever possible and applicable;
- separation of personal and sensitive data in different databases and separation from the rest of the EMT infrastructure.

In addition to the transfer of DP requirements into the respective WP and deliverables, the legal monitoring team at FIZ Karlsruhe also reviewed a series of technical deliverables (e.g., D6.1, D5.1, D6.2, D5.3), to evaluate if the identified requirements were sufficiently reflected in the document and if any other challenges were arising (e.g. proper anonymisation of Twitter data; implementation of data protection safeguards on the EMT frontend). Multiple workshops and discussions have been conducted with different partners. This included meetings with UAB, FIZ and MTU, CERTH and BUL.

b) Twitter data analysis (WP3, WP5)

The legal team has to date placed a particular focus on Twitter data analysis conducted in ITFLOWS, as data from Twitter has an initial relation to a natural person (high risk datasets). Due to the complexity of the technical aspects involved in such analytical work, FIZ-Karlsruhe (IGR team) examined the state-of-the-art methods and techniques in the field of Twitter data driven sentiment analysis. This research helped in the validation and analysis of research and processing approaches proposed by ITFLOWS partners, fostered the identification of data protection implications and risks as well as technical solutions. The work conducted

guides technical partners within the project towards the design of analytical models where technical and data protection requirements are aligned. In addition, the work also had impacts outside the project, resulting in an academic paper[3] describing data protection risks of Twitter analysis along the technical processing pipeline and proposing mitigation measures.

To this end, the legal team had a series of individual meetings with FIZ-ISE. In such occasions, many recommendations were made in order to ensure the privacy of Twitter users (e.g. the use of Batch Compliance, access restrictions to the MigrationsKB, anonymisation of User IDs). The implementation of such measures is ongoing and tracked on the CMP. Eventually, the implementation will result in an audit trail how data protection requirements have been addressed within the ITFLOWS project. The specific requirements and recommendations have been detailed in the aforementioned research paper as well as on the CMP. The research paper particularly discusses the findings from ITFLOWS and highlights risks and mitigation measures for all relevant processing steps. With this paper, WP2 outcomes and findings are spread beyond the internal borders of ITFLOWS and made available to the public. The paper pursues a structured approach following the technical processing pipeline of sentiment analysis on Twitter Data as implemented in ITFLOWS and identified during the legal research in the project. The structure can be broken down into a general description of the data protection obligations, a description of Sentiment Analysis approaches followed by the analysis of the technical processing pipeline that such approaches usually follow, namely:

- Source Identification,
- Data Collection,
- Data Cleansing,
- Analysis.

Each step contains specific technical measures and guidance on how to make the processing of Twitter data (more) compliant. The paper is hence directed towards technical as well as legal personnel to provide a foundation for further discussion between these two realms.

---

[3] Presented on the LREC Conference 2022 (https://lrec2022.lrec-conf.org/en/); currently in publication: Gottschalk, Thilo; Francesca, Pichierri; About Migration Flows and Sentiment Analysis on Twitter Data: Building the Bridge Between Technical and Legal approaches to data protection.

About Migration Flows and Sentiment Analysis on Twitter Data: Building the Bridge Between Technical and Legal approaches to data protection

Thilo Gottschalk, Francesca Pichierri

FIZ Karlsruhe - Leibniz-Institute for Information Infrastructure
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen

{francesca.pichierri, thilo.gottschalk }@fiz-karlsruhe.de

Abstract

Sentiment analysis has always been an important driver of political decisions and campaigns across all fields. Novel technologies allow automatizing analysis of sentiments on a big scale and hence provide allegedly more accurate outcomes. With user numbers in the billions and their increasingly important role in societal discussions, social media platforms become a glaring data source for these types of analysis. Due to its public availability, the relative ease of access and the sheer amount of available data, the Twitter API has become a particularly important source to researchers and data analysts alike. Despite the evident value of these data sources, the analysis of such data comes with legal, ethical and societal risks that should be taken into consideration when analysing data from Twitter. This paper describes these risks along the technical processing pipeline and proposes related mitigation measures.

Keywords: sentiment analysis, data protection, privacy

1. Introduction

Social media data are commonly processed for analysing and predicting social phenomena. They are relatively easy to obtain, cheap and contain a lot of valuable and diverse information - ranging from factual to subjective (Pereira-Kohatsu et al., 2019; Ligthart et al., 2021). Tweets are particularly popular among researchers due to their accessibility, actuality and ease of processing (Ligthart et al., 2021; Goritz et al., 2019). One particular field that shows strong interest in the use of such data is the field of migration studies and border security as can be observed by public funding directed towards research in this area[1], by research activities in general (Carammia et al., 2022), as well as by Frontex strategical-analysis documents[2] and public tenders (Frontex, 2019). In the field of migration studies, Twitter data analysis is considered very useful for a series of purposes such as measuring and predicting migration flows, providing necessary support to vulnerable groups/migrants/refugees, assessing the integration of migrants in destination countries or evaluating public opinion towards migration (Righi, 2019; Mijatović, 2021). The importance of such approaches was most recently highlighted in the context of the Ukraine war where social media intelligence (SOCINT) played an important role (Engelhaupt, 2022).

Despite the practical and analytical advantages, the processing of Twitter data can raise concerns regarding the right to data protection and privacy of Twitter users as well as affected third parties. Linkage of different datasets can produce a clearer picture of global migration flows but also raise risks for unwanted and inappropriate negative societal effects, e.g. for migrants and refugees.

Given these contrasting effects, it is crucial to design and implement analytical models and approaches in a way that balance technical and data protection requirements without undermining compliance with the legal framework, such as the General Data Protection Regulation (GDPR), nor the purposes of the data analysis. This alignment proves to be difficult especially for data scientists with no deeper understanding of the legal frameworks they conduct their work in. At the same time, legal experts often lack sufficient understanding of the technical approaches and the possible risks linked to them. This often results either in over-regulation or in non-compliance of the processing.

Where risks and potential negative consequences towards users are identified at an early stage, it is possible to adopt mitigation measures to address these risks and foster compliance with the data protection by design and data protection by default principle, enshrined in Article 25 GDPR. That being said, compelling approaches require interdisciplinary efforts involving legal experts as well as developers and data scientists to find a common language that is intelligible to all parties and to break down the knowledge barriers between different fields of expertise.

On these grounds, this paper aims to provide a foundation for structured approaches towards privacy preservation in the analysis of Twitter data and aims to build a bridge between technical and legal data protection approaches in Twitter data driven sentiment analysis. In-

---

[1]See e.g. the EU project ITFLOWS: https://www.itflows.eu/; the project METICOS:https://meticos-project.eu/; and EFFECTOR:https://www.effector-project.eu/

[2]https://frontex.europa.eu/we-know/situational-awareness-and-monitoring/strategic-analysis/

Figure 7. Research paper on data protection risks of Twitter analysis and mitigation measures.

c) Website and the EMT front-end

The data protection team recommended the implementation of an individual data protection statement (Privacy Policy) for the EMT. FIZ Karlsruhe developed a fitting privacy statement and tracked the implementation of it on the EMT front-end. In this context it was identified that the privacy policy was not correctly implemented on the first iteration. In this case the privacy policy was not publicly available for everyone but only visible for users that have been logged in before.[4] The required changes have been communicated and partly implemented. However, the monitoring identified that the links to the privacy policy are not yet active in all relevant places. Namely, the privacy policy is accessible through the footer of the website but not yet in the registration form itself. The necessity to update the form has been communicated to the responsible partner.

FIZ Karlsruhe further analysed and kept track of the cookies that are used by the EMT as well as the ITFLOWS website. For both, FIZ Karlsruhe regularly checked the website and aligned the cookie policy with the GDPR and with the most recent case-law.

Regarding, the EMT FIZ Karlsruhe provided guidance and recommendations on the implementation of security measures that prevent misuse of the tool. The recommendations, namely multiple levels of access that are driven by the respective backgrounds of the users and the intended usage of the tool are subject to ongoing implementation and will be tracked accordingly. To date, the recommendations already have been partly implemented through the implementation of an account management system and the corresponding registration mechanism. In particular, during the registration, the registrants will be asked their backgrounds and the intended purpose of usage of the EMT. Based on the context (e.g., municipalities), the access can be limited to reports, in contrast to broader access including predictions (e.g., for NGOs) reducing risks misuse of the EMT. In the course of the monitoring of the EMT it has been identified that various strategies are in place to track the use of the EMT. From the development side it has been specified that this kind of tracking is only used to avoid misuse of the EMT and for security purposes.

---

[4] https://emt.itflows.eu/privacy-policy/.

d) Interviews

The initial recommendations how to ensure compliance with data protection rights has been communicated on various levels. First, the relevant risks have been addressed within the initial deliverables as well as in bilateral discussions with the responsible partners. The solution proposals have also been addressed in D2.3 following the hybrid monitoring approach of ITFLOWS. With the integration of the CMP, the implementation of the necessary measures has been tracked on the platform. The NGOs provided solutions to the requirements based on the solution proposals provided by FIZ Karlsruhe. The CMP hence provides a full audit trail for the identified data protection requirements. Furthermore, as detailed in D2.3 a two-step anonymisation approach has been suggested and implemented by the project to ensure sensitive personal interview data cannot be accessed or misused by unforeseen parties. In short, only the NGOs conducting interviews and the two relevant research partners in the project can access the interview material, whereas third parties as well as other project members can only access aggregated or summarized data that does not contain personal information. Anonymisation of the interview data primarily takes place on the premise of the respective NGO. While multiple NGOs are involved in interviews, each one can only access their own interview data. The analysing research partners (IAI/UAB) can only access the data after anonymisation. Nevertheless, the research partners went through the data again to double-check the anonymisation. After these two steps, the actual analysis of the interviews is conducted.

In addition, the implementation has also been fostered through the provision of additional guiding documents and workshops with the NGOs where feasible anonymisation techniques have been discussed. As part of the monitoring efforts of FIZ Karlsruhe, the secondary processing of Interview data by IAI and UAB has been subject to individual guidance on the tool used to examine the transcripts (Atlas.ti). To this end, the high effort to strengthen communication between the legal and other teams of the project proved particularly valuable. The corresponding requirements are also laid down on the CMP to generate an audit trail for the compliant handling of interview data.

e) Next steps

As stated before, the monitoring approach has proven to be successful so far. For the coming months, FIZ-Karlsruhe will devote more effort towards a series of new objectives. FIZ Karlsruhe aims to further foster the usage of the compliance monitoring platform (CMP) among technical partners for keeping fast track of measures and solutions to requirements. The use of the CMP was so far mainly (and rightly) driven by the work of the legal and ethical partners which set requirements and proposed mitigation measures. However, as the technical specifications of the project achieve more clarity and robustness and the more the project develops, it would then be possible and beneficial to guide technical partners towards a more autonomous approach in the usage of the CMP, in particular when it comes to the description of solutions implemented. The CMP does not aim to substitute deliverables in terms of descriptive power. Instead, it aims to be a valid help in keeping track of the work conducted and generating evidence on the privacy preserving measures that have been implemented in the project. To this end, the CMP also results in a higher level of accountability for the individual partners as it is clearly visible what efforts have been taken. That being said, the CMP constitutes a novel approach in handling legal and ethical requirements. This novelty also comes with the downside of additional effort to adapt processes and communication towards the platform. FIZ Karlsruhe will further support the technical partner (CERTH) in reviewing the (partly proprietary) code of the large-scale model from a data protection perspective. Due to its internal policy, CERTH is not allowed to disclose and share the source code of the large-scale model with other technical partners. The use of proprietary software raises concerns among ITFLOWS researchers as it may compromise transparency, accountability and explainability in the project. As a consequence, CERTH was asked to implement measures to mitigate such risks (e.g., sharing of execute binary executables of the large-scale model with technical partners, proper documentation of algorithms in user documentation, etc.).

In addition to these goals, FIZ Karlsruhe expects an increase in the review work on technical deliverables. In order to increase the efficiency of the monitoring task and the promptness of legally compliant solutions, WP2 partners will be jointly reviewing additional technical deliverables in the coming months. The list includes:

D6.3, D3.5, D3.6, D3.7, D7.3, D6.4, D7.5.

In the following months the consortium will define the Terms of Use (ToU) that will accompany the privacy policy developed by FIZ Karlsruhe and that applies to the usage of the EMT-frontend; users of the frontend need to agree upon such ToU (i.e., by reading and clicking a checkbox) if they want to use the tool. A contractual framework needs to be implemented to ensure the end-users are aware of their obligations. As the content will be driven by various legal and ethical considerations, FIZ-Karlsruhe, as a member of WP2, will participate in and monitor the implementation of the ToU of the EMT particularly for what concerns data protection issues. The actual ToU should be sufficiently clear and specify the allowed use cases from an ethical and legal point of view. Relevant aspects that could be included are e.g.: prohibition to connect the EMT data to individual persons (e.g., during VISA applications); prohibition to trace back Twitter users; prohibition to base decisions (political/asylum, etc.) solely on EMT outputs; a statement pointing out the limited liability of ITFLOWS/EMT providers.

In addition to the aforementioned tasks and goals, FIZ Karlsruhe will continuously monitor the data processing activities in all tasks to ensure all changes and risks are properly assessed and – where necessary – mitigated.

2. ETHICAL MONITORING

As explained in D2.4 'Report on Human Rights, Ethical, Societal and Data protection risks assessments', the UAB performed an ethical risk analysis, the results of which are contained in D2.4. D2.4 identified and assessed the three most ethically challenging topics of ITFLOWS—according to the nature of the project. Namely, human participation, data protection and the technological development of the EUMigraTool.

During the first months of the project, the ethical monitoring focused on the interviews with migrants from M6 to M23 (T3.4). In this regard, multiple documents, guidelines and procedures were designed by the IDT-UAB together with T3.4 partners before the start of the interviews. These documents, guidelines and procedures cover both human participation and data protection. Most of them are listed in Table 3 (not an exhaustive list) (see also WP2 and WP10 Deliverables for further details):

ETHICAL DOCUMENTS/GUIDELINES/PROCEDURES FOR THE INTERVIEWS – HUMAN PARTICIPATION AND DATA PROTECTION (M1-M12)	DELIVERABLE	MONTH
Specific initial ethical guidelines for interviewing teams	D2.1	M5
The first draft of the conceptual paper (T3.1, Milestone 1), was reviewed by the IDT-UAB.	--	M5
- Informed Consent Forms and Information Sheets templates for research participants - Recruitment plans for research participants and interviewing teams. - Set of measures to protect migrants and minimise the risk of their stigmatisation - Incidental Finding Policy	D10.1	M6
Ethics approvals/positive opinions from the internal ethics committees/bodies of the NGOs in charge of conducting the interviews were requested and obtained before the starting of the interviews.	D10.2	M6
- Anonymisation procedure for the interviews - Technical and organisational measures to safeguard the rights and freedoms of research participants - Security measures to prevent unauthorized access to personal data	D10.3	M6
Identification and assessment of ethical risks and provision of mitigation measures to address them	D2.3	M6
Training session and training material for the NGOs and their respective interviewing teams involved in conducting the interviews with migrants, refugees and asylum seekers. In addition to the presentation, an Ethics Handbook was delivered as a guide for the interviewing teams, which is available on Teams and is continuously updated.	--	M7

- Ethics Handbook (v2.0) included in the Deliverable. - Two-step incidental findings transcription procedure	D2.4	M10

Table 3. Ethical documents, guidelines, and procedures for the interviews (T3.4) – human participation and data protection

The IDT-UAB has been continuously monitoring the development of the interviews and has been in contact with all NGOs to make sure that interviews were being conducted in an ethical way and that interviewing teams were not encountering any ethical issue. Only one ethical concern was raised by the NGOs during the interviews. OCC requested clarifications on how to proceed with the transcription of an interview where an incidental finding had been disclosed by the research participant. The IDT-UAB, together with WP2 partners, T3.4 partners, and the IEB and DPA agreed on a two-step procedure (see D2.4), which was included in the updated version of the Ethics Handbook for the interviewing teams (see the latest version of the Ethics Handbook in Annex 1).

The IDT-UAB also requested the NGOs to provide the Informed Consent Forms translated into all the languages in which the interviews were planned to be conducted.

In M18, the IDT-UAB contacted all T3.4 partners with the aim of:
1. Regarding NGOs: i) getting an update on the current status of the interviews; ii) checking if they had faced any ethical issues that needed to be tackled; iii) asking for feedback on the interviews and whether the ethical documents/guidelines/procedures designed were helpful for the interviewing teams.
2. Regarding the partners in charge of analysing the transcripts: checking if they were encountering data protection/ethical issues when analysing the transcripts.

The feedback obtained was very positive. NGOs informed us that they had not encountered any ethical issues and highlighted that the ethical documents, guidelines, and procedures were useful for the interviewing teams and that they had been followed at all times. The main difficulty faced by NGOs was the Informed

Consent Form. NGOs identified the Informed Consent Form as an obstacle to building trust with research participants. As explained in D3.5 'The 'making of' of real-time mixed migration journeys arriving in the EU: Formation and materialization of migration decisions', some research participants were hesitant, and sometimes reluctant, to sign the Informed Consent Form, while others felt embarrassed due to their illiteracy. Partners in charge of analysing the transcripts were already in the process of conducting the second step of the anonymisation procedure and were about to start using atlas.ti—after UAB and FIZ confirmation that it was an ethically and legally-compliant collaboration platform for the analysis of the transcripts.

The IDT-UAB updated the Ethics Handbook (v3.0) in M12 to include the anonymisation procedure, and informed the respective partners. This latest version of the Ethics Handbook (v3.0) can be found in Annex 1.

Regarding the topic of human participation and data protection on tasks other than the interviews, the IDT-UAB designed:

1. An Informed Consent Template Form to conduct T7.1 (an end-user board workshop to design visualisation mock-ups and indicative workflows to be implemented in the EUMigraTool, only attended by internal end-users). The aim of this informed consent template form was to obtain consent to record the live Zoom session and to take photographs.
2. Informed Consent Templates to conduct workshops with policy makers (T8.2). Ethical guidelines to conduct qualitative research activities were also provided by the ethical lead partner of the project to the Consortium (Section 5 of D.2.1).

As highlighted in this section from a human participation perspective, the IDT-UAB has also developed the data protection angle of the Informed Consent Form and Information Sheet and the Anonymisation procedure for the interviews, among others (see D10.3 for further details). As stated in the previous Section, the IDT-UAB and FIZ-IGR, in charge of the data protection issues that may arise in ITFLOWS, have established a close collaboration to monitor data protection compliance in

ITFLOWS. For instance, in the past months, the IDT-UAB spotted that the Privacy Policy link for the ITFLOWS Policy Conference 2022 led to the ITFLOWS Privacy Policy, this was communicated to FIZ-IGR, that included a Privacy Policy for the event.

2.1 Technological development of the EMT

2.1.1 Initial monitoring and preliminary AI Impact Assessment

During the first months of the project, the IDT-UAB and BUL conducted an Ethical and Societal Impact Assessment (EtSIA) as part of D2.3 'Report on Human Rights, Ethical, Societal and Data protection risks assessments' (M6). The ethical side of this impact assessment was carried out by the IDT-UAB and already included the identification and evaluation of several ethical risks that the technological development of the EMT could pose. In this regard, the ethical values at stake and the ethical risks were identified and mitigation measures were provided together with an initial overall assessment. As stated in D2.4 (M10), the IDT-UAB provided guidance on how to develop the user requirements in a way that clearly reflects the needs of end-users, and requested explanations on several aspects of the EMT design and development (e.g., types of models underlying the EMT, metadata extracted from tweets for Twitter analysis, among others). Lastly, the IDT-UAB reviewed D6.1 'Report on the specifications and architecture of the EMT platform', and included the ethical design principles to be observed by technical partners.

In M14, the IDT-UAB decided to conduct a preliminary AI Impact Assessment given the development stage of the EMT. The purpose of this preliminary AI Impact Assessment was to identify and assess, at that stage of the project (M14), the ethical risks posed by the EMT to ultimately minimise them. The results of the preliminary AI Impact Assessment, and the recommendations and mitigation measures provided to technical partners for their implementation were included in D6.2. 'Preliminary release of the EMT' and described subsequently.

The identification and assessment of ethical risks were conducted based on the Ethics Guidelines on Trustworthy Artificial Intelligence of the High-Level Expert

Group on Artificial Intelligence of the European Commission (HLEG)[5], the Assessment List for Trustworthy Artificial Intelligence for self-assessment of the HLEG[6], and the Ethically Aligned Design guidelines developed by the IEEE[7]. Following the methodological approach provided by such works – primarily the AI HLEG guidelines on trustworthy AI–, a set of ethical principles based on fundamental rights was identified as the backbone of the AI impact assessment to ensure that AI ethics is embedded in the EMT. According to the AI HLEG, these principles are: i) human autonomy, ii) prevention of harms, iii) fairness and, iv) transparency/explicability.

These principles were then turned into requirements for addressing the risks. These requirements are: i) human agency and oversight, ii) technical robustness and safety, iii) privacy and data governance, iv) transparency, v) diversity, non-discrimination, and fairness, vi) environmental and societal well-being and, vii) accountability.

Identification of the AI Ethical Principles

• **Human autonomy**[8]**:** "AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems."

This ethical principle is addressed in:

> • R1: Human agency and oversight

• **Prevention of harms:** "AI systems should neither cause nor exacerbate harm or

---

[5] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai
[6]https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
[7] https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf
[8] The following principles have been defined according to the Ethics Guidelines for Trustworthy AI: High-Level Expert Group on Artificial Intelligence of the European Union, 2019. "Ethics Guidelines for Trustworthy AI" https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure."

This ethical principle is addressed in:

- R2: Technical robustness and safety
- R3: Privacy and data governance
- R6: Societal and environmental well-being

• **Fairness**: "ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation."

This ethical principle is addressed in:

- R5: Diversity, non-discrimination and fairness
- R6: Societal and environmental well-being
- R7: Accountability

• **Transparency/Explicability:** "processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected."

This ethical principle is addressed in:

- R4: Transparency

Definition of the requirements for embedding AI ethical principles into the EMT and addressing potential risks

• **R1: Human agency and oversight**[9]: "AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight."

---

[9] The following requirements have been defined according to the Ethics Guidelines for Trustworthy AI: High-Level Expert Group on Artificial Intelligence of the European Union, 2019. "Ethics Guidelines for Trustworthy AI" https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

• **R2: Technical robustness and safety**: "A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured."

Technical robustness is also key for the system's accuracy, which "pertains to an AI system's ability to make correct judgements, or its ability to make correct predictions, recommendations, or decisions based on data or models. An explicit and well-formed development and evaluation process can support, mitigate and correct unintended risks from inaccurate predictions."

• **R3: Privacy and data governance**: "Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy."

• **R4: Transparency**: "This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business models." Decisions made by systems built on AI must be transparent, traceable and explainable.

• **R5: Diversity, non-discrimination and fairness**: "In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely

linked with the principle of fairness."

• **R6: Societal and environmental well-being:** "In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations."

• **R7: Accountability:** "The requirement of accountability complements the above requirements, and is closely linked to the principle of fairness. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use."

Preliminary ITFLOWS AI Impact Assessment questionnaire

As mentioned, the questionnaire sent to ITFLOWS technical partners (see Annex 2) was structured into seven requirements – R1: Human agency and oversight; R2: Technical robustness and safety; R3: Privacy and Data governance; R4: Transparency; R5: Diversity, non-discrimination, and fairness; R6: Environmental and societal well-being and, R7: Accountability. Each of these seven requirements were comprised of a set of questions which had to be answered internally by all ITFLOWS technical partners, with the lead of WP6. In this regard, WP6 oversaw provision of the consolidated answer to the AI impact assessment.

Technical partners were encouraged to answer to the questions to the greatest extent possible, refraining from providing yes/no answers. Technical partners were also duly informed on the AI ethical principles and the corresponding requirements, and on how to fill out the AI impact assessment questionnaire.

Specific AI ethical requirements for the EMT

The IDT-UAB identified a set of specific AI Ethical Requirements that must be embedded into the EMT, which are listed in Table 4. The granularity of these requirements enables the assessment of whether the AI ethical requirements are being implemented in the EMT and if so, to which extent. The questionnaire prepared by the IDT-UAB aims at disentangling whether these requirements have been implemented, are planned to be embedded or if (additional) technical and organisational measures are needed to comply with such requirements.

ID	AI Ethical Requirements
Human agency and oversight (HUM)	
HUM-Req1	Safeguards to prevent end-user's overconfidence in or overreliance on the EMT must be taken. Human-centric design principles must be implemented to leave meaningful opportunity for human choice.
HUM-Req2	Technical mechanisms should be implemented to ensure human control and oversight of the EMT.
HUM-Req3	End-users must be clearly informed about the functionalities, capabilities and limitations of the EMT, and the consequences of its use, through training sessions and materials.
Technical robustness and safety (TECH)	
TECH-Req4	Potential security risks and foreseeable uses of the EMT, including intended and unintended misuse, must be identified and addressed.
TECH-Req5	Technical measures to ensure the integrity and resilience of the EMT against potential attacks must be embedded.
TECH-Req6	An incremental back up of the EMT should be conducted on a daily basis and a full back up should be performed, at least, weekly.
TECH-Req7	The accuracy of the EMT must be assessed regularly. Technical partners should strive for the highest accuracy rates as is technically feasible. To this end, accuracy thresholds or benchmarks must be determined.
TECH-Req8	Technical measures to ensure that the data used in the EMT is accurate, comprehensive and up to date must be adopted.
TECH-Req9	Technical measures should be embedded to assess the need for additional data.
TECH-Req10	Technical measures should be implemented to measure the amount of inaccurate predictions of the EMT.

TECH-Req11	Technical measures must be adopted to monitor and test if the EMT is meeting its goals, purposes and intended applications.
Privacy and data governance (PRI)	
PRI-Req12	Regular assessments of the type and scope of data in the data sets used for the EMT, e.g., whether they contain personal data, must be conducted.
PRI-Req13	Privacy-preserving mechanisms, such as via encryption, anonymisation and aggregation, must be implemented.
PRI-Req14	Oversight mechanisms for data collection, storage, processing, and use should be embedded into the EMT.
PRI-Req15	Quality checks of the external data sources used for the EMT must be conducted.
PRI-Req16	Technical measures to ensure the quality and integrity of the data used for the EMT must be implemented.
Transparency (TRA)	
TRA-Req17	Technical measures to ensure traceability should be implemented. This may include documenting: i) the methods used for designing and developing the EMT; ii) the methods used to test and validate the EMT; and iii) the outcomes/results of the EMT.
TRA-Req18	The EMT must be designed to ensure its interpretability. This includes being able to analyse the training and testing data and to update and change it over time and having access to the internal workflow of the model.
TRA-Req19	The outcomes/results provided by the EMT should be made easily understandable to all end-users. Technical mechanisms to inform end-users on the reasons and criteria behind the EMT's outcomes/results should be implemented.
TRA-Req20	Technical measures and processes to consider end-users' feedback and use this to adapt the EMT should be implemented.
TRA-Req21	End-users must be made aware of the characteristics, limitations and potential shortcomings of the EMT.
Diversity, non-discrimination and fairness (DIV)	
DIV-Req22	The composition of the data sets used in the EMT must be assessed, with particular emphasis on its possible limitations.
DIV-Req23	Diversity and representativeness of users in the data must be ensured.
DIV-Req24	The use of technical tools to improve the understanding of the data, model

	and performance is highly encouraged.
DIV-Req25	Potential biases arising at the design, development, deployment and use phase of the EMT must be tested, monitored and addressed.
DIV-Req26	Engagement with different stakeholders in the EMT's design, development and use must be sought.
DIV-Req27	Accessibility and universal design of the EMT is highly encouraged to ensure that the EMT is usable by those with special needs or disabilities.
Societal and environmental well-being (SEW)	
SEW-Req28	The broader societal impact of the use of the EMT, both positive and negative, should be assessed and addressed accordingly.
SEW-Req29	The environmental impact of the EMT's design, development, deployment and use should be assessed and measures to reduce it should be implemented.
Accountability (ACC)	
ACC-Req30	Training sessions and materials must be delivered to end-users to help developing accountability practices, including the risk of misuse.
ACC-Req31	Technical measures must be embedded into the EMT to allow end-users to report potential vulnerabilities, risks or biases of the EMT.
ACC-Req32	Authentication and authorisation components must be embedded into the EMT.
ACC-Req33	Users' roles and privileges must be clearly defined for authorisation purposes.
ACC-Req34	Technical measures must be implemented to facilitate the EMT's auditability, such as ensuring traceability and logging of the EMT's processes and outcomes.
ACC-Req35	Oversight mechanisms must be implemented to log when, where, how, by whom and for what purpose data was accessed. These data logs must be reviewed regularly.

Table 4. Specific AI Ethical Requirements.

Analysis of the technical measures adopted at this stage and further ethical recommendations

The IDT-UAB analysed the answers provided by technical partners to evaluate whether the abovementioned requirements were met. In the following paragraphs the technical explanations and measures that have been identified by the IDT-UAB

from the responses to the questionnaire are presented. The identification of these explanations and measures was also elicited via a workshop organised by the IDT-UAB with all technical partners and with the participation of the IEB and DPA, in which some clarification on certain technical aspects were requested. Based on these explanations and measures, further mitigation measures to be embedded into the EMT are presented below.

Human agency and oversight

The following <u>technical explanations and measures</u> have been identified from the responses to the preliminary AI impact assessment:

1. The EMT is a decision-support system designed to aid/support humans in their decision-making processes.

2. The outputs of the EMT's AI modules will be accompanied by explainability features which will provide insights regarding how the outcomes of the tool has been produced.

3. Extensive documentation with example use-cases has been provided for the preliminary version of the EMT. Training webinars for end-users and training videos will be provided. The EMT will feature a helpdesk in order to assist users in their queries.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of human agency and oversight:

1. EMT explainability features must be provided by each EMT module avoiding technical jargon to ensure that end-users can comprehend **why** a certain EMT outcome has been produced.

2. Provide clarifications on the EMT helpdesk and its purpose. Consider expanding the features of the EMT helpdesk to include a reporting mechanism that allows users to flag errors, potential biases and systems' malfunctions.

Technical robustness and safety

The following <u>technical explanations and measures</u> have been identified from the

responses to the preliminary AI impact assessment:

1. Technical measures have been implemented to ensure the integrity and resilience of the EMT against potential attacks. Examples include SSL Certificates, secured servers, firewalls, and that the system is regularly backed up in an offsite location.

2. Regarding the likely impact of a failure of the EMT if it provides wrong results, or becomes unavailable, given that the EMT assists decision-making processes, but does not provide automated decisions, they have identified the following negative impacts: a) delay in the decision-making; b) miscalculation of resources at migrant receiving areas; c) misplacement of migrants in less-accepting areas. Users should treat EMT outputs as guidelines.

3. Information on the accuracy of the models will be provided on the EMT website.

4. Measures in place when there is a need for additional data: EMT has established a private data repository (CKAN) populated with data from public and trusted data sources and data is updated regularly. Data comprehensiveness is ensured by the EMT's front-end, where all the required legends and explanations are provided.

5. Technical partners evaluate accuracy and bias during the development phase of the EMT. Feedback from the end-users will be gathered in order to assess the need for additional data.

6. Potential harms caused by EMT inaccurate predictions have been identified by technical partners. Such inaccurate predictions could lead to miscalculations in resources or poor decisions related to the spatial distribution of migrants' allocation to reception centres. End-users should use the EMT outputs as guidelines or for consultative purposes, and not for making final decisions.

7. Users will be able to provide feedback on the quality of the results. This will allow the technical team to improve the models. The EMT will be constantly updated as new data becomes available in order to improve its accuracy.

8. Feedback from end-users will be gathered in the form of surveys. The technical team will have real-time information when such feedback is

received to react if needed. The feedback results will be publicly available in statistical form (fully anonymised).

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of technical robustness and safety:

1. Security measures to prevent security risks, in particular intended and unintended misuse must be embedded into the EMT.

2. Provide clarifications on the periodicity of the regular back-ups.

3. Provide clarifications on how accuracy and bias are evaluated during the development of the EMT.

4. Establish a threshold for the accuracy rates of the predictions. Below such a threshold, predictions cannot be shown to the end-user. Instead, users must be warned that a prediction could not be made due to a low accuracy rate.

5. In case of low/medium accuracy rates of a prediction, a warning must be implemented to alert users of the poor results of the prediction.

Privacy and data governance

The following <u>technical explanations and measures</u> have been identified from the responses to the preliminary AI impact assessment:

1. Type of data in the EMT data sets: ethical and legal use of data is ensured before using any data set.

2. EMT does not use personal data/identifiable data in its core. Potentially, identifiable data is used by individual components during the training phase, but this data is not passed to the EMT. All developers ensure full anonymisation of the data they use, and that no one besides them has access to this data.

3. All data stored in the EMT's repository (CKAN) is encrypted. The EMT only stores data needed for its functionality.

4. The EMT will ensure that no unauthorised access will be possible. However, the EMT or the ITFLOWS partners cannot control how data is used by end-users.

5. Data fed into the EMT comes from trusted sources. Thus, data quality checking is not needed. The EMT does not use Twitter data directly, but through a model that ensures that no bad-quality data will be used in order to minimise accuracy impacts on the model and in any case, it will be identified at the development case.

6. CKAN repository that the EMT uses for data storage has embedded mechanisms to ensure the quality and integrity of the data. Cybersecurity mechanisms have been put in place to ensure the security of the system.

7. Data governance: access rights policy controls have been implemented. The front-end logs all actions performed in the EMT.

According to these technical explanations and measures, the following ethical recommendations are provided by the ethical lead partner to comply with the ethical requirement of privacy and data governance:

1. Provide clarifications on the security measures adopted when managing and storing encrypted data in CKAN.

2. Provide clarifications on how to ensure non-authorised access in CKAN.

3. Provide further clarifications on the technical measures implemented to ensure the integrity of data in CKAN and the integrity of the EMT.

4. Provide clarifications on the technical measures taken to ensure the accuracy and quality of data in CKAN.

5. The collection of information from users must be limited following the data minimisation principle.

6. The datasets used to feed the EMT (including the so-called "trusted sources") must undergo a quality check before they are fed into the EMT.

7. Terms of Use of the EMT must be developed.

Transparency

The following technical explanations and measures have been identified from the responses to the preliminary AI impact assessment:

1. The EMT's modules are developed by research/academic partners. Their design, functionality and results have been published in scientific journal/conferences, and thus are publicly available for scrutiny.

2. Details regarding the EMT modules will also be provided within the EMT webpage to allow the users insight on the modules.

3. The EMT will include explainable features in its results, in a comprehensive manner avoiding technical language.

4. Explainability is one of the core design requirements of the EMT.

5. Information regarding the EMT's functionalities will be in the documentation pages on the website. Limitations and shortcomings will be listed as well.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of transparency:

1. The methods used for designing and developing, as well as for testing and validating the EMT, must be documented. A decision must be made regarding which design and development details will be publicly available in the EMT webpage. This information must be provided in clear and plain language, free from technical jargon.

2. EMT explainability features must be provided by each EMT module avoiding technical jargon to ensure that end-users can comprehend why a certain EMT outcome has been produced.

3. The limitations and shortcomings of the EMT must be included in the EMT webpage in an accessible manner and must be visible and clearly explained to end-users.

Diversity, non-discrimination and fairness

The following <u>technical explanations and measures</u> have been identified from the responses to the preliminary AI impact assessment:

1. Limitations stemming from the data sets: the supported functionalities and accuracy of each EMT module are defined with respect to the available data, and such acknowledgements will be made available on the EMT website.

2. Diversity and representativeness are core principles of the EMT data assessment process. Developers (design/development phase) make sure that the available data is representative of the whole population of interest and that no bias towards specific categories is produced.

3. All EMT modules are tested for bias (design/development phases). Any findings in this regard have been reported and mitigated accordingly to ensure the un-biased implementation of the EMT.

4. A core design principle is to avoid/mitigate potential disproportionate impacts of the EMT on persons/groups. The ITFLOWS team will assist in this by providing extensive and comprehensive training to potential users. However, the interpretation of the EMT results by the users cannot be controlled by technical partners.

5. Several workshops with end-users and stakeholders have taken place and more are planned to ensure that the EMT meets their requirements and standards.

According to these technical explanations and measures, the following ethical recommendations are provided by the ethical lead partner to comply with the ethical requirement of diversity, non-discrimination and fairness:

1. Provide clarifications on how diversity and representativeness of the datasets used for the EMT is ensured.

2. Provide clarifications on how the different EMT modules will be tested for bias at the design, development and implementation phase of the EMT.

3. Provide clarifications on how the reports made by end-users during the testing and use phase of the EMT will be technically addressed.

Societal and environmental well-being

The following technical explanations and measures have been identified from the responses to the preliminary AI impact assessment:

1. Societal impact assessment: WP2 monitoring tasks.

2. Measures to reduce the environmental impact of the EMT's life cycle: the EMT is fully compliant with the "Do no significant harm principle" (Arts 9 and 17 of the EU Sustainable Finance Taxonomy Regulation-six environmental objectives).

According to these technical explanations and measures, the following ethical recommendations are provided by the ethical lead partner to comply with the

ethical requirement of societal and environmental well-being:

1. Provide further clarifications on the measures implemented to assess the environmental impact of the EMT and to address it.

Accountability

The following <u>technical explanations and measures</u> have been identified from the responses to the preliminary AI impact assessment:

1. Training and education accountability practices: training for end-users will be provided in the forms of guides and examples that will be available within the EMT website. Workshops/webinars will be organised during the project.

2. Risk of misuse exists as the results/outputs could be misunderstood or miss-interpreted by the malicious or untrained users. In order to minimise the risk of misuse, the ITFLOWS Consortium monitors the access to the EMT and ensures that the training materials have been delivered. However, the actions of people cannot be monitored.

3. Extensive logging is built-in the EMT: all actions are logged, but the way in which EMT results are used cannot be monitored.

4. Processes for users to report potential vulnerabilities: feedback from end-users will be encouraged via built-in questionnaires and forms.

5. Authentication/Authorisation components embedded in the EMT: only authorised and authenticated users will have access to the EMT. The technical team will check users requesting access and ensure their access privileges.

6. Oversight logging mechanisms implemented (when/where/how/by whom/for what purposes) in the EMT: all actions in the EMT are logged in detail.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of accountability:

1. Technical measures to minimise the potential misuse of the EMT must be implemented.

2. Access rights must be clearly defined and differentiated based on the type of end-user (NGOs, municipalities, etc.).
3. Provide clarifications on the authentication and authorisation system.

As a result of the discussions held during the workshop, the IEB and DPA also provided their recommendations, which can be found in Section 5.1 and 5.2 respectively, and were also included in D6.2.

2.1.2 2nd AI Impact Assessment

The ITFLOWS AI Impact Assessment is meant to be conducted several times during the lifespan of the project to closely monitor the technical developments of the EMT to ensure that it is ethically compliant. Therefore, a second iteration of the AI Impact Assessment was conducted in M22 following the same methodological approach adopted in the preliminary AI impact assessment (See Section 2.1.1).

The updated questionnaire was sent to all ITFLOWS technical partners (see Annex 3) following the structure of seven requirements: R1: Human agency and oversight; R2: Technical robustness and safety; R3: Privacy and data governance; R4: Transparency; R5: Diversity, non-discrimination, and fairness; R6: Environmental and societal well-being and, R7: Accountability. A set of questions corresponded to each of these seven requirements, which had to be answered internally by all ITFLOWS technical partners, with the lead of WP6. In this regard, WP6 oversaw provision of the consolidated answer to the AI Impact Assessment.

This time, technical partners were also encouraged to answer questions to the greatest extent possible, including technical explanations, refraining from providing yes/no answers. Technical partners were again informed on the AI ethical principles and the corresponding requirements, and on how to fill out the AI Impact Assessment questionnaire.

For this second iteration of the AI Impact Assessment, another workshop with technical partners and the rest of the Consortium was held to clarify some technical aspects of the EMT. The specific AI ethical requirements identified by M18 again

serves to evaluate whether the technical and organisational measures explained by the technical partners in their answers satisfy the requirements or if further measures are needed. First, in order to make this assessment, the technical and organisational measures adopted thus far need to be identified. Secondly, further mitigation measures will be provided if necessary. Both the measures already embedded into the EMT and the mitigation measures that must be implemented by technical partners are presented below.

Human agency and oversight

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. The EMT website features a glossary, where relevant information and key definitions of the terms used in its various modules are available to end-users. The EMT's dashboards section provides detailed explanations on EMT visualizations and their intuition.

2. The EMT website provides, in its home page, information regarding: i) EMT's different functionalities (e.g., simulations and forecasting); ii) the technologies used; iii) the overall approach of the EMT modules.

3. EMT end-users have access to both historical data and predictions produced by the EMT. They can compare EMT predictions with ground truth data provided by the EMT's data sources.

4. Recommendations regarding EMT end-users' potential biases will be available on the EMT website.

5. Helpdesk information is available under the support tab on the EMT website, which is visible and easily accessible.

6. The EMT User Manual is also included under the support tab on the EMT website.

7. A ticketing system has been implemented in the EMT and can be found under the support tab on the EMT website. This allows EMT end-users to report bugs and/or errors and system malfunction to the EMT developers. The user is responsible to make the report as detailed as possible.

8. The EMT forum has been implemented in the EMT and can be found under the support tab on the EMT website. This allows EMT end-users to interact

amongst themselves, and also with the developers.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of human agency and oversight:

1. The information provided in the home page should be more comprehensive and comprehensible. More information should be included, and it should be written in plain and clear language, keeping in mind that the intended readers are end-users, who will possibly not have a technical background but still need to have a clear picture of the EMT, its functionalities, capabilities, and limitations. For instance, detailed explanations are needed regarding: i) Small-scale model; ii) Large-scale model; iii) Twitter analysis.

2. On the home page it should be clearly stated that the EMT predicts asylum seekers' and unrecognised refugees' flows, not migration flows. From a legal perspective this is a key distinction that must be made.

3. Graphs shall be displayed for the large-scale model (2.2. Predictions of Asylum Seekers/Unrecognised Refugees per destination and origin country) that compare EMT predictions with ground truth data provided by the EMT data sources, so that end-users can easily check the accuracy of the predictions. This has been already implemented for the small-scale model (1. Origin Countries with conflict locations and Asylum Seekers/Unrecognised refugee camps) but this is required for the large-scale model too.

4. The EMT website should also include information to prevent (or, at least, minimise) EMT end-users' potential biases, such as automation bias.

Technical robustness and safety

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. All SoA security measures are in place (SSL certificates, authentication and authorization of user access, 2FA, secured servers, firewalls, constantly updated software and more).

2. The EMT is regularly backed up in an offsite location to ensure that in case of an attack, it will be back online in a short period of time with minimal to no

data loss.

3. The backend and frontend of the EMT have implemented authentication and authorisation processes to ensure that only users approved for an account can use the tool. The EMT backend server uses token authentication techniques to ensure that only whitelisted servers can send requests. The EMT frontend uses various techniques for user authentication and authorisation including e-mail address validation, two factor authentication (2FA), and hierarchy roles to ensure that only specific groups of users have access to specific functions of the EMT.

4. Validation data from Eurostat have been used for the assessment of the EMT's forecasting accuracy.

5. Administrative bias of asylum applications has been mitigated using techniques found in the scientific literature.

6. The EMT will provide both: i) prediction points (exact number of estimated asylum applications); and, ii) prediction intervals (a range of possible values within which the number of asylum applications is expected to fall in with 95% statistical confidence). Prediction intervals are not available yet.

7. Warning legends will be assigned to the respective Dashboards of the EMT to alert end-users of the poor accuracy of a given performance.

8. The EMT website will display information for EMT end-users to inform them that the EMT is a decision-support system. This information is not available yet.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of technical robustness and safety:

1. Provide clarifications on the periodicity of the regular back-ups.

2. The authentication and authorisation components can help reduce the risk of misuse by third parties. However, technical mechanisms shall also be implemented to mitigate the risk of misuse by existing users, who are authenticated and authorised to use the EMT.

3. Clarifications of what administrative bias is, and how it has been addressed, are required.

4. Once the prediction intervals are included in the EMT, their level of statistical confidence should be mentioned.

5. It should be clearly stated in the EMT Home Page that the EMT is a decision-support system aimed at aiding end-users in their decision-making processes.

Privacy and data governance

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. Data are being processed on the EMT backend servers and computers of the partners (Step 1). Processed data are upload in the CKAN data repository and the EMT backend server only responds with results from CKAN (Step 2). Only specific partners have access to CKAN and they are allowed for the processing of such data only from their organisation. To that end, users' roles (admins, editors, simple users) with access to specific organisations and specific datasets within CKAN have been implemented.

2. Accounts with credentials have been implemented. Only selected partners have accounts with access to CKAN. 2FA is also used.

3. CKAN users have "editor" roles only to data from their organisation.

4. The data used for training and validation of the EMT models have been inspected for existence of personal data.

5. No identifiable data are used within the EMT, so there is a low risk of de-anonymisation. For Twitter data that could be identifiable, tweets are not stored anywhere within the EMT servers.

6. Sensitive information such as passwords are hashed and encrypted before stored in the EMT backend database.

7. Optional cookies related to user experience will be stored in a user's browser only if the end-user consent is provided. They will have the option to disable them at any time. User activity tracking is performed only for security reasons, and no cookies are required nor stored at all.

8. For Twitter semantic analysis, the only personal information collected from Twitter users are: ids, usernames, geo information (city and country names or city level coordinates). Original data is only accessible for the data

collector and can be processed only for research purposes. These data will be deleted after the project is completed.

9. Twitter users' identifications are anonymised with UUID version 4 objects as specified in RCF 4122 from Python library, to ensure that each user has a unique id that cannot be traced back and found on Twitter. Twitter usernames are omitted so they cannot be searched by their usernames.

10. Twitter users' geo information is mapped to the country or regional level using the city name, country name, or coordinates. The individual geo information is not exposed so the users cannot be linked.

11. Manual checks for all external data sources used in the EMT have been implemented. They include: i) inspection of data retrieval methodologies by the owners of data sources; and ii) assertion of consistent availability and proper formatting.

12. For Twitter analysis, the external data sources used included: Eurostat, UK Parliament Office for National Statistics, Statista and Asylum Information Database.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of privacy and data governance:

1. The Terms of Use of the EMT must be included in the EMT website.

Data protection recommendations have also been provided in Section 1.1.

Transparency

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. Methods and coding scripts used for designing and developing the underlying models of the EMT have been thoroughly documented. Testing and validation of the results have been documented in D6.2 and D6.3.

2. Information regarding the design and development is publicly available in the Home Page of the EMT website. This information explains the various approaches of the tool without including technical specifications.

3. Disclaimers informing the EMT users about the limitations of each dashboard

will be implemented on the EMT website.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of transparency:

1. The information provided in the Home Page should be more comprehensive and comprehensible. More information should be included, and it should be written in a plain and clear language keeping in mind that the intended readers are end-users, who will possibly not have a technical background but still need to have a clear picture of the EMT, its functionalities, capabilities, and limitations. For instance, detailed explanations are needed regarding: i) Small-scale model; ii) Large-scale model; iii) Twitter analysis.

2. In the Home Page it should be clearly stated that the EMT predicts asylum seekers and unrecognised refugees' flows, not migration flows. From a legal perspective this is a key distinction that must be made.

3. The limitations and shortcomings of the EMT must be included in the EMT webpage in an accessible manner and must be visible and clearly explained to end-users. The EMT website should also include detailed information about the limitations stemming from the datasets (e.g., gender is only divided into male/female/unknown).

One key issue regarding transparency that must be noted here is that at first CERTH was reluctant to disclose the source code of the large-scale model due to their internal policy. The use of proprietary software raises serious ethical concerns regarding not only transparency, but also accountability and explainability. These concerns are aggravated by the field in which the EMT operates, and the consequences its use may have for migrant individuals and groups. After several discussions between the Project Coordination, WP2 partners and WP6 partners, it was agreed with CERTH to share the source code only with relevant ITFLOWS partners that need access to the source code to conduct their activities. This agreement strikes a right balance between intellectual property rights and the ethical and legal implications of making decisions based on predictions produced by a black box. The names of the people designated to have access to the EMT have been

provided to CERTH, and NDAs between them and CERTH have been signed to ensure confidentiality.

Diversity, non-discrimination and fairness

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. Diversity and representativeness are a core principle of the tool's data assessment process. The developers during the design of the development phases make sure that the available data is representative of the whole population in interest, and not biased towards specific categories. If this is not possible due to lack of data, it is made available in the documentation.

2. Depending on the case, different methods have and will be implemented in the future to solve existing biases. Dataset biases have been mitigated through careful selection and processing of the data.

3. Both a ticketing system and forums have been implemented in the EMT so users can inform the developers and the teams working on the EMT of potential bugs/errors and make suggestions for future work in the EMT. The feedback from the users is very valuable to the developers, especially in the testing phase of the EMT. Users' feedback will be carefully analysed, and users' suggestions will be considered for future updates of the EMT.

4. The front-end has been developed in such a way that is accessible to all. More specifically, the guidelines derived from WCAG (Web Content Accessibility Guidelines) were followed in order to make the content accessible to people with disabilities. According to these guidelines, text alternatives are provided for any non-textual content and text content is readable and understandable with large and distinct font. Also, the colour contrast between background and foreground content is great enough to ensure legibility.

5. For the modules of forecasting and simulation, different features were used to capture as much of the variance of the dependent variable as possible. For example, in the case of forecasting using topic shares of the national press, intuitive visualizations were embedded into the models to provide partners with information on the unsupervised topics extracted by the LDA topic modeler. Similar approaches were used for all features tested, such as

correlation heatmaps.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of diversity, non-discrimination and fairness:

1. Provide clarifications on how the different EMT modules are tested for bias at the design, development and implementation phase of the EMT.

2. The EMT website should include detailed information about the limitations stemming from the datasets (e.g., gender is only divided into male/female/unknown).

Societal and environmental well-being

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. Extensive work has been conducted by WP2 partners.

2. Measures to reduce the environmental impact of the EMT's life cycle: the EMT is fully compliant with the "Do no significant harm principle" (Arts 9 and 17 of the EU Sustainable Finance Taxonomy Regulation-six environmental objectives).

According to these technical explanations and measures, recommendations have been provided in Section 3 to comply with the ethical requirement of societal and environmental well-being.

Accountability

The following <u>technical explanations and measures</u> have been identified from the responses to the 2nd AI Impact Assessment:

1. Users must complete and send a "Request an account" form that will be assessed by the project organisers for getting whitelisted. The user must, also, verify their email address for the account to be created. After the account is created, various authentication and authorisation processes have been implemented, like the two-factor authentication process, to ensure only

authorised users can login.

2. A strict and clear hierarchy has been developed for accessing the EMT. The general public only has access to historic data, while authorised users can also access predictions and advanced analytics.

3. Backend authentication and authorisation: Bearer tokens authentication systems have been implemented in the EMT backend servers. Requests made to the backend API must include authentication headers with bearer tokens based on JWT (JSON Web Tokens). Only a limited amount of the tokens will be created, and only very specific groups of users will be given access to the backend API.

   Frontend authentication and authorisation: i) email verification; ii) two-factor authentication (2FA); iii) all accounts are validated by the ITFLOWS team.

4. The activity logs will be reviewed periodically (period not defined yet) manually.

According to these technical explanations and measures, the following <u>ethical recommendations</u> are provided by the ethical lead partner to comply with the ethical requirement of accountability:

1. An auditing plan must be in place as soon as possible. Auditing the EMT is crucial to assess compliance and ensure accountability.

The IEB and DPA also provided their recommendations, which can be found in Section 5.1 and 5.2 respectively.

2.1.3 Other monitoring activities

The IDT-UAB, as WP2 leader, has been coordinating the development of WP2 tasks and deliverables. To this end, WP2 meetings have been scheduled on a bi-weekly basis from the beginning of the project ensuring the continuing monitoring of WP2 tasks and deliverables' developments as well as coordinating WP2 contributions to the ITFLOWS Consortium, particularly to WP3, WP5 and WP6. All WP2 partners (UAB, FIZ and BUL, including researchers of all three institutions working on human

rights, ethics and technology and on gender) and the Project Coordinator are participating in these WP2 meetings. The IDT-UAB also attends the bi-weekly technical meetings organised by WP6 in order to monitor the ethical and legal compliance of the EMT from the early stages of its design.

Additionally, the IDT-UAB has closely collaborated with partners participating in Task 3.4 (CRI, OIT, OCC, IAI and UAB). In this regard, several meetings were arranged by the IDT-UAB to discuss in detail issues related to the participation of migrants in the interviews and to prepare the documents, guidelines and procedures mentioned in Section 2.1.1.

Lastly, the IDT-UAB is in direct contact with both the IEB and DPA (external and independent monitoring bodies of ITFLOWS), to discuss particular ethical issues and also to organise the review of Deliverables, and their participation in specific meetings, including Plenary Meetings and the Project Review.

In Year 1, the IDT-UAB reviewed D1.1 'Data Management Plan' (M6) and D6.1 'Report on the specifications and architecture of the EMT' (M9). Regarding D6.1, the IDT-UAB provided specific recommendations to mitigate ethical risks related to the technical development of the EMT. In particular, it included those concerning user requirements, data sources, the EMT architecture and the design principles. In addition to the review, the IDT-UAB included in D6.1 a specific section on the ethical design principles to be observed by ITFLOWS technical partners.

In Year 2, the IDT-UAB has reviewed so far three other deliverables and provided the following input/recommendations:

- D6.2 'Preliminary release of the EMT''. The IDT-UAB provided recommendations and drafted Section 4 ('Ethics Section'), which describes the preliminary AI Impact Assessment, presents its results, the recommendations provided by the IDT-UAB, as well as the recommendations provided by the IEB and DPA.
- D6.3 'Report of migration modelling simulation'. Main recommendations:
  - Specify the sources used to identify the population distribution of conflict locations.
  - Information on the data processing activities conducted (e.g., cleaning

of data) before the data is fed into the model is required. This applies both to the small- and large-scale model.

- o The meaning of the graphs displayed regarding the first and second administrative level results (Figures 4, 6, 9, 11, 14, 16, 19, 22, and 24) should be explained.
- o The large-scale model performance (and its respective matrix with the Median Relative Error) should be explained in greater detail.
- o Regarding Twitter sentiment analysis, clarifications on when anonymisation takes place and how are required.

- D3.5 ' The 'making of' of real-life mixed migration journeys arriving in the EU: Formation and materialisation of migration decisions' (M23). Main recommendations:
  - o The brief paragraph that tackles the collaboration with WP2 should be more detailed and include the ethical issues that were raised regarding the interviews with migrants (T3.4) and how these were addressed. References to WP2 and WP10 deliverables that contain the ethical, human rights, and data protection procedures designed to ensure the ethical and legal compliance of the interviews with migrants, refugees, and asylum seekers should be referenced.
  - o Provide clarifications on how informed consent procedures were implemented during the interviews with migrants, refugees, and asylum seekers, how the obstacles to obtain it were overcome and if potential improvements can be implemented for future projects.

As explained in Section 1.2, the ITFLOWS Consortium is making use of the CMP as a tool to keep track of the actions taken by partners, which allow us to monitor ITFLOWS research activities. The IDT-UAB is actively adding ethical requirements and solutions, and also reviewing them. Figure 8 shows some of the ethical requirements that the IDT-UAB has included so far:

Requirements your organisation created

		Added	Requirement	Responsible Partner	Task	Topic	Source	<	⊘+	☰✓
⤢	⋮	Fri Feb 18 2022	Recruitment Plan for research participants	CRI	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.4	⊘	⊘	⊘
⤢	⋮	Thu Feb 24 2022	Incidental Findings Policy	OXFAM	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.6	⊘	⊘	⊘
⤢	⋮	Wed May 11 2022	Technical robustness and safety	CERTH	T2.2	EMT Backend		⊘	⊘	⊘
⤢	⋮	Thu Feb 24 2022	Likelihood of incidental findings	CRI	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.6	⊘	⊘	⊘
⤢	⋮	Fri Feb 18 2022	Recruitment Plan of the interviewing teams	OXFAM	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.4	⊘	⊘	⊘
⤢	⋮	Fri Feb 18 2022	Recruitment Plan for research participants	OXFAM	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.4	⊘	⊘	⊘
⤢	⋮	Thu Feb 24 2022	Appointment of a Data Protection Officer (DPO) and provision of the data protection policy	OXFAM	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.8	⊘	⊘	⊘
⤢	⋮	Thu Feb 24 2022	Incidental Findings Policy	CRI	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.6	⊘	⊘	⊘
⤢	⋮	Thu Feb 24 2022	Opinions/approvals by ethics committees and/or competent authorities	CRI	T3.4	ITFLOWS Interviews	REA Ethics Check Requirement No.7	⊘	⊘	⊘
⤢	⋮	Wed May 11 2022	EUMIGRATOOL: Preliminary AI Impact Assessment	UAB	T2.2	EMT Backend		⊘	☑ 1	⊘

Rows per page: 10 ▾   1-10 of 36   ‹ ›

Figure 8. CMP ethical requirements added by the IDT-UAB.

Some of these requirements include multiple sub-requirements. For instance, the requirement "Preliminary AI Impact Assessment" includes 7 sub-requirements that correspond to the ethical requirements mentioned. Likewise, the ethical requirements that the NGOs had to comply with are also included in the CMP as depicted in Figure 9:

Figure 9. Example CMP ethical sub-requirements for the interviews.

3. SOCIETAL (HUMAN RIGHTS) MONITORING

ITFLOWS has the possibility to positively influence the situation of migrants in the society in which they reside, as much as the societal understanding of migration. Equally, because it deals with a sensitive and highly politicised issue, that of migration, it also has the possibility to lead to opposite results. The effects of the Project on society are determined directly by its coherence with human rights standards, which of course also run through the legal and ethical monitoring. Since the inception of the project, human rights considerations have been identified as a core issue of the project. These considerations transcend the legal, ethical and social monitoring. Although the project falls within a cluster of an EC-funded security grant, human rights considerations are of utmost importance; hence the effects on society. Several risks have been identified since the start and as time went by, partners decided that human rights monitoring was strengthened.

Which rights are at risk

We have identified that the project may endanger the following rights; hence we are continuously monitoring whether activities and the development of the project may affect them:

- Right to asylum: We focus on the guarantees that the partners need to put in place so that the prediction of inflows does not lead to restriction or even denial of the right to asylum. We engage in discussions with all partners on the entities that should not have direct access to the data.

- Non-discrimination: We monitor that deliverables do not fall into populist assumptions and do not apply prejudicial statements on migrants, asylum seekers and refugees. We have been putting pressure to use data that will give some information on gender etc. We have repeatedly favoured an intersectional approach in the discussions. The deliverables on public attitudes should not be used to justify discrimination against migrants and asylum seekers.

- Non-refoulement/ no collective expulsions: Data from our project should not be used to lead to pushbacks nor collective expulsions. We are very concerned about allegations of pushbacks by FRONTEX and national enforcement agencies, and we think it is better not to give such bodies direct access to the data.

- Prohibition of torture and access to justice: The data from the project should not lead to abuse against migrants and refugees.

- Prohibition of arbitrary detention: Our project aims at allowing preparatory measures to be taken for the integration of inflows that the EMT will predict. Such measures must not include arbitrary detention, so we are monitoring the development of the project to ensure that our work includes any guarantees possible against such measures.

- Living conditions and socio-economic rights: Discussions on integration must not contravene the already existing standards on the adequate standard of living and socio-economic rights; neither should these rights be considered aspirations, but the minimum standards.

- Rights to privacy and family life: Integration measures should take into account these rights. Also, the interviews conducted and any data used should not violate such rights.

- Hate speech: Our data and conclusions on the public attitudes on migration should not be highjacked and used in populist debates that dehumanize migrants. Our work cannot be implied to justify such speech, as manifested on social media.

How was the monitoring conducted

The monitoring of societal impact coincides to a large degree with the legal and ethical monitoring, as discussed extensively in the previous sections. Specifically

regarding human rights, the basis for the monitoring has been our report on the legal EU and international frameworks (Deliverable D2.1). In order to facilitate that all partners are aware of the human rights 'reefs' of the project, the report includes a chart with the main human rights that are relevant to the project, their legal basis, the main case-law and the main sub-topics. This is, we felt, an important aid for the technical partners to have in mind while developing their technological tools.

The main bulk of monitoring has been conducted through regular, focused and in-depth discussions with our other partners and specifically with the technical partners. These were discussed at length in the previous sections; hence, we will only restrict our analysis to new elements. The discussions are facilitated by the review of other WP deliverables, which allowed us familiarize with the overall project. We have been meeting WP2 colleagues every two weeks discussing issues that have come to our attention during the monitoring. We have also been taking part in monthly discussions with WP6 colleagues to discuss issues that arise with respect to the human rights considerations.

The discussions with the technical partners have been invaluable and have revealed how difficult truly intra-disciplinary work is. The focus of our work is different and it has been proven challenging to understand each other and create a common pool of knowledge. Technical partners have found the legal discussions too theoretical and social scientists of WP2 have found the explanations of the technical partners very difficult to understand. We all needed a lot of patience and continuous efforts to understand each other's work. Invaluable in this respect have been the role of some partners who transcend the legal/ technical dichotomies and were able to act as intermediaries in offering 'translations' at certain points of the discussions.

Some special meetings were needed to work specifically on some issues. They have been conducted and have proven very valuable. In M16, for example, WP2 colleagues responsible for the societal and human rights impact have been attending the meeting convened between WP2 and WP6, "Discussion EMT: Ethical and legal aspects" to offer feedback to technical partners. In December 2021, we engaged in a discussion with the technical partners on sources that the partners are using and

looked at the definitions given in these sources.

The second year of the project we also started using the very helpful **ITFLOWS Compliance Monitoring Platform**: In M14 and M19, all WP2 members were trained on how to use the ITFLOWS Compliance Monitoring Platform. It has been populated to some degree since then.

Which issues have been revealed through our monitoring

Our monitoring revealed mainly the following concerns:

1. *Use of correct terminology: 'asylum seekers, refugees and migrants'*

During the first year of the project, our monitoring revealed that the partners faced some confusion in distinguishing among the legal terminologies 'asylum seekers', 'migrants' and 'refugees' on the project. This is very common: the difficulties and the blurring of these terminologies have been identified extensively in the literature.[10] Yet, in view of the simulations and the EMT, it was essential to be precise as to the individuals each part of the project deals with. Depending on the categorisation of the individual, the rights will be different,[11] especially when it comes to economic and social rights, important elements for integration[12] part of our project.

After identifying the issue, our continual monitoring revealed inconsistency in the sources used widely by our technical partners. We embarked on a study to monitor the use of words that can be legally defined in different ways than those that our technical partners would know. Our technical partners told us that they heavily rely on the statistics given by EUROSTAT, GDELT, FRONTEX, EMDAT, ACLED, the World Bank, the UNHCR AND HDX.

---

[10] For example, see N El Enany 'On Pragmatism and Legal Idolatry: Fortress Europe and the Desertion of the Refugee' (2015) 22 *International Journal on Minority and Group Rights* 7-38.
[11] UNHCR, 'UNHCR viewpoint: 'Refugee'or 'migrant- Which is right? (11 July 2016), https://www.unhcr.org/news/latest/2016/7/55df0e556/unhcr-viewpoint-refugee-migrant-right.html (assessed 19 July 2022).
[12] A Xanthaki 'Against integration, In favour of human rights' 20 (2016) 6 *International Journal of Human Rights* 815- 838.

In December 2021, we distributed the first paper on terminology to technical partners. We looked at each one of the sources used and highlighted the provisions that explained the different concepts in each of the source highlighting the main issues. For example, this is what our paper included on the term 'asylum seeker' and on FRONTEX:

ASYLUM SEEKER

UNHCR in https://www.unhcr.org/glossary/#a:

"A general term for any person who is seeking international protection. **In some countries, it is used as a legal term referring to a person who has applied for ~~refugee status~~ or a complementary international protection status and has not yet received a final decision on their claim. It can also refer to a person who has not yet submitted an application but may intend to do so, or may be in need of international protection**."

SO:

EUROSTAT *https://ec.europa.eu/eurostat/web/migration-asylum/asylum/database*

We use the following from Eurostat:

1. *Asylum applicants by type of applicant, citizenship, age and sex - monthly data (rounded) (migr_asyappctzm)*

Source: EUROSTAT, *Expert Group on Refugee and Internal Displaced Persons Statistics- International Recommendations on Refugee Statistics*, March 2018, at page 27 onwards makes the distinction between prospective asylum seekers and asylum seekers.

"1. Prospective asylum seekers: Persons with the intention of filing an application for asylum, but who have not yet done so for reasons which include being unable to file an application because of practical or administrative obstacles including

capacity constraints in the receiving country authorities. Asylum seekers in transit to another country are excluded from the scope of these recommendations.

2. Asylum seekers: Persons who have filed an application for asylum in a country other than their own and whose claims have not yet been determined. These include those filing primary applications or subsequent applications following an appeal. The date on which the application for asylum is filed marks their entry into the status of asylum seeker. They remain in the status of asylum seeker until their application is considered and adjudicated."

ACTION POINTS

⇒ The term 'asylum applications' is to be preferred to 'asylum seekers'

⇒ The term 'Asylum seekers' if used must be replaced with 'persons seeking asylum'

⇒ EUROSTAT seems to include as asylum seekers only those who have applied for refugee status formally:

⇒ We should have a statement/ footnote somewhere saying the above and identifying the discrepancy between EUROSTAT and UNHCR.

⇒ In this database, the term 'unaccompanied minors' is also included. This needs a footnote and definition.

We completed a similar exercise with every single source that was given to us by our technical partners. We compared the different terms and their meanings in every one of the sources given and reached specific conclusions. It was an invaluable exercise that helped us understand the source of confusion. After that, we engaged in further regular discussion with our technical partners on the use of these terms which further crystalised our favoured choice of terms in the project. We insisted that 'illegal migrants' is not used anywhere in the project, introduced the

terminologies of 'recognised' and 'unrecognised refugees' following the UNHCR terminology; we completed another note given to the technical partners and a glossary for the website and the database; and we continue to question all references to asylum seekers, migrants and refugees in all documents, presentations and discussions by our technical partners. It has been a very laborious process, but it did reveal that different terms are used by different international organisations which leads to confusion. This has informed our knowledge and our future recommendations.

We also engaged in monitoring the use of other terms. For example, we identified the issue of extreme poverty as used in the World Bank website, which is used by our partners. We asked our partners whether this was relevant, i.e., whether they had to define this in their searches. Concerning WP3, we questioned whether 'hate speech' or 'hatred' would be used when talking about public attitudes. Such references would have to comply with the human rights understanding of what hate speech is and would trigger human rights considerations around the limitations of freedom of expression. In discussions around integration, we challenged the emphasis on the economic and labour aspects of integration as international literature links integration with the respect for human rights. In general, we also favoured in our discussions with partners the use of intersectional approaches. To this end, we worked very closely with the gender experts and discussed whether and to which extent the data have some implied bias.

2. Human Rights, simulations and EMT

The human rights risks posed by technology have been at the forefront of our discussions and of our monitoring. Technology as a means to predict migration within a security context raises many concerns; and this is well documented in the area of migration.[13]

Who has direct access to the data we gather is a very sensitive matter that brings

---

[13] Niovi Vavoula, *Immigration and Privacy in the Law of the European Union – The Case of Information Systems* (Brill Nijhoff, 2022); also Valsamis Mitsilegas, Violeta Moreno-Lax and Niovi Vavoula, *Securitising Asylum Flows: Deflection, Criminalisation and the Challenges for Human Rights* (Brill Nijhoff, 2020)

serious human rights considerations: the purpose limitation principle must be the cornerstone in making the decision of who will have direct access to the data and the risk of this tool to curtailing migration must be taken into account. In this respect, we have spent a lot of time and effort in monitoring the development of technological tools and their access. Even from its inception, the project rejected direct access of the national states but local authorities are allowed to have the data of the project. This has caused us serious concerns. The possibility of direct access to the data of enforcement agencies also causes us great concern.

Direct access of enforcement agencies would be problematic. Allegations of pushbacks have been raised by several organisations and media both against Member States' enforcement agencies and FRONTEX. Allowing states and EU enforcement agencies to have direct access to the data incurs the risk of pushbacks, which constitutes a violation of non-refoulement. It also has the risk of violating the right to asylum as obstacles may be put to the application of asylum.

We have also been concerned about allowing local authorities to have access to the data on attitudes, as again, the data can be used by politicians for populist initiatives and speech that would lead to hate speech, stereotyping and ultimately denial of the right to asylum and discrimination against these individuals. We have been very eager to monitor how access is discussed and shaped in the next few months.

In addition, we have also been monitoring the kind of data that is being used both for the simulations and for the EMT. We have made clear to our technical partners that use of data that identifies any migrants would raise human rights concerns; and they have assured us that such data does not exist.

We have been very eager to ensure that no discrimination, direct or indirect, and no violation to the rights of privacy occurred in the gathering of the data in interviews and simulations. We had several discussions with the NGOs involved in the gathering of data from interviews. It was very clear that this was very much at the forefront of their work.

3. Identifying bias

Any project that involves migrants has to be monitored continuously for bias; in the way the partners discuss them, in the assumptions that they make, but also in the effects that the project will have on the local population.

We have been very aware that the work that ITFLOWS is doing on public attitudes may have such effects. The attitudes on migration may be influenced by bias and bring about further bias. Hence, we have been following the work quite closely and have been in communication with other WPs to ensure that the data gathered, the conclusions and the way the conclusions will be used does not lead to simplistic answers that maintain stereotypes that lead to negative societal impact and ultimately involve violations of human rights. This was discussed in depth in the Legal and Ethical Monitoring.

The work on integration has also raised our concern and we have been following it closely. Discussions on migrant women have been quite sensitive; so too have been assumptions about migrants' lifestyles. Arguments relating to migrants' family choices have been criticised as dehumanising migrants and adding to the populist panic. We continue to monitor these discussions so that we do take into account legitimate characteristics of migrants without adding to the prejudice and dehumanisation that leads to further violations of 'migrants as 'The Other.' We have been monitoring the development of the discussions so that no simplistic assumptions are made.

4. GENDER MONITORING

This section describes the involvement, monitoring activities and recommendations of the ITFLOWS Gender Committee over two years of ITFLOWS project activity. The activities carried out by the Internal Gender Committee are divided into Year 1 (M1-12, or September 2020 to August 2021) and Year 2 (M13-M23, or September 2021 to July 2022, through to the D2.5 deliverable deadline). The ITFLOWS Gender Committee is composed of both an internal body as well as an Independent Gender Committee (IGC). The monitoring activities of the IGC are reported in Section 5.3.

What follows briefly outlines the main tasks carried out by the Internal Gender Committee for each year. This report particularly highlights the activities that led to the successful acceptance of the ITFLOWS Gender Action Plan (GAP) and the ITFLOWS Gender Policy by the European Commission in May 2022. It overviews the key meetings and events, as well as monitoring and recommendations activities.

4.1 Internal Gender Committee Activities Year 1 (M1- M12)

4.1.1 The Gender Action Plan and the Gender Policy (D2.2)

One of the main tasks and a deliverable of the Gender Committee was the ITFLOWS Gender Action Plan (D2.2) and its annexed Gender Policy, drafted over the course of M1 to M5 of the project. D2.2 was submitted to the EC in M5 and accepted by the Commission in the project's Mid-Term Review.

The Internal Gender Committee conceptualised and devised a 91-page Gender Action Plan (GAP) (D2.2), which constitutes a living document, and clearly establishes how the project's gender commitment will be implemented, monitored, and fulfilled. It emphasises why this strategy is crucial in achieving the project's aim of providing the most accurate and effective solutions and policy recommendations for managing migration flows to the EU, from a human rights perspective. To this end, the GAP stipulates the actions to be taken in order to ensure the mainstreaming of gender and sexuality throughout the project and sets forth the monitoring process over the 3-year course of the project. Thus, the GAP is a tool to assist and encourage ITFLOWS researchers and project partners in achieving ITFLOWS' commitment to the Horizon 2020 'Promoting Gender Equality in Research and Innovation' policy.

Notably, the GAP's section on Compliance and Monitoring indicates how the Gender Committee will monitor the implementation of such recommendations throughout 3-year life cycle of project.

Annexed to the GAP is the ITFLOWS Gender Policy, a two-page document, signed by the Independent Gender Committee and by a representative from each ITFLOWS partner institution. The Gender Policy outlines the project's commitment to a) consider gender and sexuality as a priority in the project's design, methodology, analysis, and dissemination of outputs, underlining that gender and sexuality are central to an intersectional analysis of migration flows; and b) pursue gender equality and balance in project participation. The GAP and Gender Policy are available to all members of the Consortium. The work of the Gender Committee on these documents and their implementation was highlighted in the first ITFLOWS newsletter in M6.

Gender balance: In the GAP (D2.2) and Gender Policy, the ITFLOWS project commits to "gender balance," or equal participation of men and women in the project. As of M23 (this deliverable's submission) the ITFLOWS leadership teams and advisory boards, including the Independent Ethics Board, Expert Advisory Board, Independent Gender Committee and Steering Committee all reflect a gender composition of at least 50% female-identifying participants.

4.1.2 Publications

In M4, members of the Gender Committee and co-authors published an ITFLOWS Policy Brief: "Covid-19 Implications for Migrant Care Workers: A Gender Perspective"

4.1.3 Meetings and events

In order to ensure that the project remains fully committed to addressing gender and sexuality, the Internal Gender Committee attended regular meetings organized by WP2 leaders, as well as regular meetings of WP6 to monitor the project's EMT. In addition to discussions on deliverable monitoring, they also held a series of Internal Gender Committee meetings over the course of the year. Finally, they additionally organized: the Committee's corresponding presentation and workshop in the

ITFLOWS Kick-Off (Plenary) Meeting (M1); and a gender panel for the CEPS Ideas Lab "The role of essential migrant workers in the EU's recovery: Why does gender matter?" (M10).

4.1.4 Monitoring and Recommendations

To ensure that the project remains fully committed to an intersectional approach and the mainstreaming of gender-sensitive research, analysis and policy in all ITFLOWS work packages, the Internal Gender Committee engages in regular monitoring activities throughout the project and keeps the IGC advised.

Monitoring of human participation: As an example, the Internal Gender Committee reviewed the interview scripts for interviews procedures and the anonymisation procedures of the interviews as part of D10.3 submitted by ethical and legal partners in M6 (further described below). Moreover, the Gender Committee indicated to ITFLOWS partner NGOs which questions required extra caution and sensitivity, as well as pointed them to the relevant parts of the Gender Action Plan for conducting their work. In M7, as part of a WP2 training for of these NGOs, the Committee resent/reemphasized the gender and sexuality guidelines/gender policy for interviews to migrants. The internal Gender Committee oversaw that all NGO partners signed the Gender Policy by M8. It was noted that these procedures should be conducted with any project collaborator that participates in administering the interviews.

Monitoring of the EMT: Another example includes how in M6, the Internal Gender Committee met with researchers in WP3 and WP5, as well as a WP2 data protection expert, to discuss the possibility of including the gender dimension when working with Big Data/Twitter. It was explained that in order to arrive at an ITFLOWS system that that might attempt to determine the gender of the author of a tweet, this would involve working with a data set manually, using known tweets from persons of a specified gender to train a system to give better accuracy in terms of determining the gender of user who wrote a tweet. As the initial work plan only relates to

collecting tweets and not user data, this would have to have been already included in the methodology. Moreover, it would require extensive data on users that could have data protection implications. Ultimately, it was concluded that the literature points to primarily male use of social media among migrants in origin countries, and this would be something to acknowledge in the conclusions drawn from tasks using Twitter/Big Data. However, as it was not included in the WP3/WP5 work plan to begin with, it would be outside of the scope to detect and analyse gender in the scraped Twitter data (as it would require expanding the work and tasks); moreover, this kind of method could perhaps insert gender biases. As such, male use of Twitter/social media will be acknowledged, and the researchers and Internal Gender Committee will remain in contact for any possible incorporation of gender dimension in the future. Moreover, based on this discussion, the Internal Gender Committee will ensure that relevant deliverables remark on limitations involved in migration research relating to big data, as well as including gender mainstreaming from the start in research design, and perhaps would include this in the D8.1 (M34).

Deliverables: In addition to these examples, the Internal Gender Committee was both engaged in contributing to, reviewing or overseeing the work related to several deliverables to ensure the full compliance of all Work Packages to the mainstreaming of gender and sexualities in their outputs. The following deliverables were reviewed in or received input in Year 1, and are listed consecutively by project month:

- D9.1 Project Website and Project Branding (M3): Reviewed by Internal Gender Committee.
- D2.1 Report on the ITFLOWS Legal and Ethical framework (M3): the Gender Committee contributed a section on intersectionality, gender and sexuality to D2.1 and reviewed it.
- D2.2 Gender Action Plan (M5): Authored by Internal Gender Committee, as described above, with contributions by Independent Gender Committee (detailed in next section).

- D10.1 H-Requirement No. 1 (M6): As described above, the Internal Gender Committee ensured an adequate approach to gender and sexuality regarding human participation in ITFLOWS migrant interviews (T3.4).

- D10.2 H-Requirement No. 2 (M6) The Gender Policy specifically designed to be applied in the ITFLOWS Project is signed by relevant authorities.

- D1.1 Data Management Plan (M6): Reviewed by Internal Gender Committee

- D2.3 Report on Human Rights, Ethical, Societal and Data protection risks assessments (M6): Participation from Internal Gender Committee.

- D7.1 Report on Users Board Participatory Feedback (M6): The Internal Gender Committee assisted in preparing the M5 Workshop with the ITFLOWS Users Board, including revising the workshop survey, in addition to reviewing this deliverable.

- D6.1 Report on the specifications and architecture of the EMT (M9): Reviewed by Internal Gender Committee.

- D2.4 Report on the ITFLOWS Regulatory Model (M10): Input for gender and sexuality from Internal Gender Committee.

4.2 Internal Gender Committee Activities Year 2 (M13-M24)

4.2.1 Meetings and events

As noted above the section on Year 1, in order to ensure that the project remains fully committed to addressing gender and sexuality, the Internal Gender Committee attended regular meetings, in addition to other relevant meetings highlighted here. By Year 2, both WP2 monitoring and WP6 EMT meetings (where the internal Gender Committee attended in a monitoring capacity) were held and attended on a bi-weekly basis. Apart from these established regular meetings, other relevant ones during this period included:

- ITFLOWS Annual Meeting (M13): The ITFLOWS Annual Meeting consisted of a three-part series of events. The Internal Gender Committee participated in and/or presented at second and third events: the ITFLOWS Workshop on the EMT and ITFLOWS Expert Workshops.

- WP7 Users Board: In M17, the Internal Gender Committee attended the ITFLOWS Users Board Workshop to monitor.

- WP8 ITFLOWS Policy Meetings and Conference: In advance of the ITFLOWS Policy Conference (M22), since M18 the Internal Gender Committee attended and participated in WP8 meetings with WP8 and the Policy Working Group in organizing and preparing content for the conference, as well as ensuring involvement of invitees and participants with expertise in gender and sexuality, as well as those that engage in an intersectional approach.
- EMT Exploitation: In addition to attending the bi-weekly EMT meetings, the Internal Gender Committee was present for periodic meetings/discussions (roughly monthly) regarding future exploitation of the EMT, which began in M19 and will continue through to the end of the project.

4.2.2 Monitoring and Recommendations

Monitoring of EMT: In addition to monitoring the deliverables as described below, and attending the bi-weekly meetings, in M14, the Internal Gender Committee reviewed all WP6 datasets for gender and sexuality keywords, after having went back to the respective original databases for gender and sexuality keywords and indicators. It submitted recommendations as to how to better account for gender and sexuality based on this review. In M16, the Internal Gender Committee members attended a meeting between WP2 and WP6, "Discussion EMT: Ethical and legal aspects" to offer feedback to technical partners.

ITFLOWS Compliance Monitoring Platform: In M14 and M19, both members of the Internal Gender Committee were trained on how to use the ITFLOWS Compliance Monitoring Platform. While its implementation is still in development, it was populated to some extent by the Gender Committee in M19.

Deliverables: Again, as in Year 1, the Internal Gender Committee was engaged in reviewing several deliverables to ensure the full compliance of all WPs to the mainstreaming of gender and sexualities in their outputs. The following deliverables (and one milestone) were reviewed in or received input in year 1 by the Internal Gender Committee, and are listed consecutively by project month:

- D3.2 Analysis on migration drivers and trajectories along The Eastern Mediterranean Route (South-Central Asia & Middle East); Central and Western Mediterranean Routes (North, West, and the Horn of Africa); Atlantic Air Route (South America) (M15): contribution by the Internal Gender Committee.

- Milestone 3 Final conceptual paper on migration drivers and trajectories (M16): the Internal Gender Committee reviews Milestone 3. They also initially offered feedback on how to streamline this paper with the Gender Action Plan in M5.

- D4.3 Overview report on relevant socio-economic situation in EU member states: Dataset on economic situation as input in EMT and for other WPs (M15): Contributed to and reviewed by Internal Gender Committee.

- D4.2 Two reports: The impact of family migration and family reunification of refugees and other migrants; European Union Policies on Onward and Secondary Movements of Asylum-seekers and Refugees (M17): The Internal Gender Committee reviews and expresses concerns about the terminology of "fertility" and offers a solution in clarifying the ITFLOWS position on this terminology. This issue is signalled in the final version of the deliverable.

- D1.2 Interim Project Report (M18): Reviewed by Internal Gender Committee.

- D6.2 Preliminary Release of the EMT (M18): While most of reviewed deliverables receive comments in track changes, given the ethical implications and societal impact of this deliverable, a formal summary of the recommendations submitted by the Internal Gender Committee are provided here:

  o Always disaggregate when possible: Firstly, although it has been recognized in discussions with technical partners that limitations exist in the data sets, and it is recognized that there are various gaps in data on migration to the EU, it is again encouraged that all datasets that are disaggregated by sex to be incorporated accordingly into the tool's models. As of December 3, 2022, in the meeting between WP2 and WP6 "Discussion EMT: Ethical and legal aspects," the Internal Gender Committee communicated that it had gone through the datasets technical partners had indicated fed the EMT models and

output to examine whether they addressed gender and sexuality and followed up with those partners that had not provided information. In this discussion it was established that the agent-based model could possibly incorporate further gender-related indicators or data in its development, and it was stressed that this should be done if and when possible.

o Establish clear terminology: The Internal Gender Committee has also encouraged and supervised that an understanding of "Gender and sexuality" be noted in the glossary page of the EMT, which contextualizes the datasets that the tool uses. It was decided that further glossary terms outlined for the ITFLOWS project in its D2.2 Gender Action Plan would not be included in the EMT glossary, given that the datasets do not directly relate to gender and sexuality related indicators and terminology. However, these glossary terms would be made available on the ITFLOWS website.

o Maintain gender balance in EMT Users Board: Finally, the Internal Gender Committee encourages and remains conscious of the gender equal and balanced composition of the ITFLOWS Users Board, which continues to provide feedback on the development of the EMT.

o Refer to D2.2 Gender Action Plan: As always, with reference to the EMT, the Gender Committee reminds technical partners to refer to key sections of the D2.2 Gender Action Plan, including sections 3C "Gendering and actioning the EMT" and 3D "Gendering and actioning Big Data," which establish the expectations, guidelines and recommendations for approaching gender and sexuality in the creation, implementation and dissemination of the EMT.

- D7.2 EMT Strategy Training Report (M19): Reviewed by Internal Gender Committee.

- D4.1 Two reports: Gendered migrant integration policies in the EU (M20) Arrivals to Spain and obstacles versus opportunities from the migrant perspective (M20): Reviewed by Internal Gender Committee.

- D6.3 Report on migration modelling and simulation (M22): Upon review, the Internal Gender Committee again notes a need for further gender/intersectional disaggregation.

5. EXTERNAL MONITORING BODIES

5.1 ITFLOWS Independent Ethics Board

5.1.1 Independent Ethics Board activities Year 1 (M1-M12)

During the first year of the project, the IEB reviewed seven deliverables and provided recommendations that were duly implemented. The IEB also provided guidance at the request of the partners to clarify specific ethical and legal concerns. For more information, see D2.3 'Report on Human Rights, Ethical, Societal and Data protection risk assessments', the D2.4 'Report on the ITFLOWS Regulatory Model', and 'D10.4: GEN-Requirement No.5' (the report on the activities of the IEB during the first year of the project). The following table (Table 5) lists some of the main recommendations (not exhaustive):

IEB RECOMMENDATIONS (M1-M12)	DELIVERABLE	MONTH
Specific initial ethical guidelines for interviewing teams.	D2.1	M5
Economic compensation for research participants.	--	M5
- Informed Consent Forms and Information Sheets templates for research participants - Recruitment plans for research participants and interviewing teams. - Set of measures to protect migrants and minimise the risk of their stigmatisation - Incidental Finding Policy	D10.1	M6
Ethics approvals/positive opinions from the internal ethics committees/bodies of the NGOs in charge of conducting the interviews were requested and obtained before the starting of the interviews.	D10.2	M6
- Anonymisation procedure for the interviews. - Technical and organisational measures to safeguard the rights and freedoms of research participants - Security measures to prevent unauthorized access to personal data	D10.3	M6
Questionnaire for the interviews with migrants, refugees, and asylum seekers.	--	M6
Two-step incidental findings transcription procedure in relation to the interviews with refugees, migrants, and asylum seekers.	--	M9
Development of the EUMigraTool's architecture and specifications.	D6.2	M9

Table 5. IEB Recommendations (M1-M12).

The deliverables reviewed by the IEB in Year 1 are:

1. Deliverable 2.1 Report on the ITFLOWS Legal and Ethical framework (M5)

2. Deliverable 2.3 Report on Human Rights, Ethical, Societal and Data protection risks assessment (M6)

3. Deliverable 2.4 Report on the ITFLOWS Regulatory Model (M10)

4. Deliverable 6.1 Report on the specifications and architecture of the EMT platform (M9)

5. Deliverable 10.1 H - Requirement No. 1 (M6)

6. Deliverable 10.2 H - Requirement No. 2 (M6)

7. Deliverable 10.3 POPD - Requirement No. 3 (M6)

IEB members also produced D10.4 'GEN-Requirement No.5' in M12, and D10.5 'GEN-Requirement No.6' (M24) is currently being drafted.

5.1.2 IEB recommendations - Preliminary AI Impact Assessment (M14)

The IEB evaluated the results of the preliminary AI Impact Assessment and provided the following ethical and legal recommendations to be embedded into the EMT, which were included in D6.2:

Human agency and oversight

1. Provide further information about the EMT explainability features.

2. The main issue is "how" the EMT outcomes/results are produced. The outcomes produced are not self-justifiable and, consequently, "accountable" and "responsible". Additional difficulties arise due to the lack of transparency or explainability and comprehensibility of how these outcomes/results have been produced, as it is nearly impossible for an outsider to review such process and the basis of an output. Measures to improve transparency, explainability and comprehensibility need to be implemented.

3. Provide clarifications on the selection criteria followed for the datasets fed into the EMT.

4. EMT training must also cover appropriate decision-making based on outputs.

Technical robustness and safety

1. Negative societal impacts due to malfunctions of the EMT: Need to consider measures for end-users to be aware of: (i) how and when actions would be taken and (ii) by whom. This is to minimize negative societal impacts.

2. Provide clarifications on how bias and accuracy of the EMT at the design

phase is evaluated.

3. Provide clarifications on who/what warns technical partners about the need for additional data.

4. Provide clarifications on how EMT end-users will be provided with instructions specifying that the EMT results are only for guidelines and consulting.

Privacy and data governance

1. Potentially identifiable data will be used by individual EMT components during the training phase. Provide clarifications on what will be delivered during the training phase.

2. Provide clarifications on how data is going to be used and mitigation measures to reduce the potential misuse of data during the lifespan of the project.

3. Provide clarifications on how the planned 'indirect use of data via a model' will fix any issues relating to bad quality.

Transparency

1. Provide clarifications on the implementation of the explainability principle within the EMT.

Diversity, non-discrimination and fairness

1. Provide clarifications on how technical partners have assessed and acknowledged limitations related to the composition of the used data sets.

2. Provide clarifications on how diversity and representativeness are ensured during the data assessment process.

3. Provide clarifications on how end-users' feedback will be used to enhance/develop the EMT.

Societal and environmental well-being

1. Provide precise information on how the EMT is compliant with the "do no significant harm principle" (Articles 9 and 17 of the EU Sustainable Finance Taxonomy Regulation).

Accountability

1. Provide clarifications on the implementation of mitigation measures in the EMT to minimise the risk of misuse.

5.1.3 IEB recommendations - 2nd AI Impact Assessment (M22)

The IEB reviewed the answers provided by technical partners to the 2nd AI Impact Assessment of the EMT and provided the following recommendations to be implemented by technical partners:

Human agency and oversight

1. Provide clarifications on what is meant by "quality of the results" and how this quality is evaluated.
2. It is key to make the tool understandable to end-users. Concepts such as "system's malfunctions" should be explained to ensure that end-users are capable of flagging potential errors in the system.
3. There are currently no guidelines for end-users on how to avoid/report bias. A plan of the next steps to be undertaken to fill this gap is needed.

Technical robustness and safety

1. Provide the full list of mechanisms and measures implemented to ensure the protection of the EMT against cyberattacks.
2. Provide clarifications on how the system's malfunctions might lead to a negative social impact. Clarify whether the malfunction itself will produce this negative impact or whether it will be caused by the poor outputs/data resulting from the malfunction.
3. Provide clarifications on what is meant by "administrative bias" and the scientific literature used to mitigate "administrative bias". Provide further explanations on how administrative bias of asylum applications were mitigated. It must be clarified what positive steps to mitigate bias to the highest possible level were taken during the design process, as well as how these measures will be reviewed.

Privacy and data governance

1. Make it clear why all the original Twitter data is reserved only by the data collector and for research purpose only.
2. Provide clarifications on the accuracy of the geo-information obtained from Twitter analysis.

Transparency

1. The relationship between "transparency" and "traceability" should be made clear.
2. The information regarding the limitations and shortcomings of the EMT must be easily accessible, visible, and clearly explained to end-users on the website. A disclaimer cannot be considered an explanation.
3. Provide clarifications on the arrangements in place to audit the outputs of the EMT in terms of accuracy and potential risks.

Diversity, non-discrimination and fairness

1. Provide clarifications on the actions taken at the various stages of the EMT to resolve issues of bias. A disclaimer is insufficient and does not provide solutions.
2. Provide clarifications on the nature/contribution of information gained by the scientific literature to assess limitations in the specific/specified datasets. Provide clarifications on how scientific literature informed design and how this information is provided to end-users.
3. Further mechanisms to mitigate limitation for persons with disabilities must be put in place. It cannot be stated that the EMT has no limitations for people with disabilities. Consider, for instance, the effects of colour contrast for users with colour blindness, dyslexia or other conditions.
4. Provide more detail on the explainability features that have been embedded into the models underlying the EMT.

Societal and environmental well-being

1. Provide details on how the work undertaken in WP2 has been weaved into

all design phases of the EMT.

2. Provide clarifications on how compliance with the "Do no significant harm principle" has been assessed.

3. It must be considered that incorrect outputs that are taken as verbatim by a human who has no malicious intent can indeed do significant harm.

Accountability

1. Automated methods to review activity logs will require tracking and follow-up actions.

2. An auditing plan to assess compliance and ensure accountability must be put in place.

5.1.4 Other IEB monitoring activities Year 2 (M13-M24)

After reviewing Deliverable 6.2 'Preliminary release of the EMT', the IEB made the following recommendations regarding the architecture and design of the EMT:

1. Provide more information on the methods that will be used to detect anti-migration attitudes besides Twitter Sentiment Analysis.

2. Clarify what is meant by "various factors that may affect the attitude". Provide clarifications on where these "various factors" are drawn from.

3. Provide clarifications on the methodology followed to categorise emotions.

4. Provide clarifications on how the factors that have a negative/positive attitude towards migration are determined.

5. Use GDPR terminology for data protection issues.

6. Note that the use of data must be both legal and ethical.

7. Provide clarifications on who evaluates the accuracy and bias of the EMT.

8. Provide clarifications on how quality of the data gathered from social media is evaluated.

9. Provide clarifications on the EMT's cookies policy.

After reviewing Deliverable 6.3 'Report of migration modelling and simulation', the IEB made the following recommendations on the predictive models developed for the EMT:

1. Use of plain language to the extent possible for the technical description of the tool. Overuse of technical jargon would make the deliverable less accessible for non-technical partners.

2. Clarifications on the meaning of some ambiguous terms and expressions (e.g., 'validation file', 'algorithm assumptions', or 'forecasting').

3. Clarifications on the functionalities of the tool, especially on the use of data. Additional explanations on the methods of manual extraction of data and its posterior validation (e.g., to avoid the presence of 'bots') were requested.

4. The deliverable must always make clear that the EMT is designed to be compliant with the EU's highest ethical and legal standards.

5. Clarifications on the context, the sources, the precision of the terminology, the presentation, and the conclusions drawn from the use cases, to ensure that they were well-documented and presented to ensure the accuracy of the predictions.

5.2 ITFLOWS Data Protection Advisor

5.2.1 Data Protection Advisor activities Year 1 (M1-M12)

During the first year of the project, the DPA reviewed seven deliverables and provided recommendations that have been implemented. The DPA also provided guidance at the request of the partners to clarify specific ethical and legal concerns. For more information, see D2.3 'Report on Human Rights, Ethical, Societal and Data protection risk assessments' and D2.4 'Report on the ITFLOWS Regulatory Model'. The following table (Table 6) lists some of the main recommendations (not exhaustive):

DPA RECOMMENDATIONS (M1-M12)	DELIVERABLE	MONTH
Economic compensation for research participants.	--	M5
- Informed Consent Forms and Information Sheets templates for research participants. - Recruitment plans for research participants and interviewing teams. - Set of measures to protect migrants and minimise the risk of their stigmatisation. - Incidental Finding Policy.	D10.1	M6
- Ethics approvals/positive opinions from the internal ethics committees/bodies of the NGOs in charge of	D10.2	M6

conducting the interviews were requested and obtained before the starting of the interviews.		
- Anonymisation procedure for the interviews. - Technical and organisational measures to safeguard the rights and freedoms of research participants. - Security measures to prevent unauthorized access to personal data. - Use of publicly available data.	D10.3	M6
Interview grid for the interviews with migrants, refugees, and asylum seekers.	--	M6
Two-step incidental findings transcription procedure in relation to the interviews with refugees, migrants, and asylum seekers.	--	M9
Development of the EUMigraTool's architecture and specifications.	D6.2	M9

Table 6. DPA Recommendations (M1-M12).

In Year 1, the DPA reviewed the following deliverables:

1. Deliverable 2.1 Report on the ITFLOWS Legal and Ethical framework (M3)

2. Deliverable 2.3 Report on Human Rights, Ethical, Societal and Data protection risks assessment (M6)

3. Deliverable 2.4 Report on the ITFLOWS Regulatory Model (M10)

4. Deliverable 6.1 Report on the specifications and architecture of the EMT platform (M9)

5. Deliverable 10.1 H - Requirement No. 1 (M6)

6. Deliverable 10.2 H - Requirement No. 2 (M6)

7. Deliverable 10.3 POPD - Requirement No. 3 (M6)

5.2.2 DPA recommendations - Preliminary AI Impact Assessment (M14)

In the context of D6.2 'Preliminary release of the EMT', the DPA provided the following ethical and legal recommendations to be embedded into the EMT at this initial development stage:

Human agency and oversight

1. Provide clarifications on who provides guidance to support the decision-making process regarding unconscious bias.

2. Provide clarifications on explainability features of the EMT and its outputs/results.

Technical robustness and safety

1. Be more specific regarding how exposed the EMT is to cyberattacks. Provide explanations on the cybersecurity measures embedded into the EMT.

Privacy and data governance

1. Open-source data may still contain personal data. Additionally, inferences may be drawn from the processing of such open-source data. The notion that open-source publicly available dataset does not contain personal data must be challenged. Oversight mechanisms to ensure that the datasets used do not contain personal data must be implemented.

2. Quality checking of the data sources fed into the EMT (e.g., Eurostat, Frontex, UNHCR, etc.) is required.

Diversity, non-discrimination and fairness

1. The evaluation of the impact on persons or groups stemming from erroneous decision-making, should not only cover negative impacts but also disproportionate impacts (e.g., opportunity costs).

Societal and environmental well-being

1. Societal impacts of the EMT should also consider allocation of financial, human resources and opportunity costs.

Accountability

1. Provide clarifications on how the risk of misuse is mitigated.
2. Provide clarifications on how the EMT activity logs will be reviewed, whether it is ad hoc, or reviews will be conducted on a regular basis (e.g., every week).
3. Provide clarifications on how the EMT will be audited. An EMT Auditing Plan should be put in place.

5.2.3 DPA recommendations – 2nd AI Impact Assessment (M22)

In the context of the 2nd AI Impact Assessment of the EMT, the DPA provided the following recommendations to be observed by ITFLOWS technical partners:

Technical robustness and safety

1. Provide clarifications on what is meant by "warning legends" to alert users of the poor accuracy of a given prediction and how are they going to be implemented into the EMT.

Privacy and data governance

1. Provide clarifications on whether "accounts with secret credentials" is used a synonym to "secure accounts".
2. Provide clarifications on whether "no identifiable data" means "personally identifiable information".

5.2.4 Other DPA monitoring activities Year 2 (M13-M24)

After reviewing Deliverable 6.2 'Preliminary release of the EMT', the DPA made the following recommendations regarding the architecture and design of the EMT:

1. Consider CNIL ruling for the use of Google Analytics.
2. Justify why the range of proposed factors might have a negative/positive impact on the local attitudes towards migration.
3. Replace "potentially identifiable data" with "personally identifiable information".
4. Clarify whether the software updating is periodical or automated.
5. Provide clarifications on the EMT's cookies policy.

After reviewing Deliverable 6.3 'Report of migration modelling and simulation' the DPA made the following recommendations on the predicting models developed for the EMT:

1. Add further references related to the information provided for the countries selected as use cases.
2. Provide clarifications on the reasons to remove Afghanistan from the list of use cases.

3. Provide clarifications on the sociodemographic variable considered for conducting the analysis of the attitudes towards migration.

4. Consider the impact of CNIL's ruling, and possible subsequent rulings, on the lawfulness of the operation of Google Analytics vis-à-vis data transfers.

5. Clarifications on what is meant by "unrecognized refugees" are needed.

5.3 ITFLOWS Independent Gender Committee

As with the Internal Gender Committee, the activities carried out by the Independent Gender Committee are divided into Year 1 (M1-12, or September 2020 to August 2021) and Year 2 (M13-M23, or September 2021 to July 2022, through to the D2.5 deliverable deadline).

5.3.1 Independent Gender Committee Activities Year 1 (M1- M12)

The Gender Action Plan and the Gender Policy (D2.2)

Again, one of the principal tasks and deliverable of the Internal and Independent Gender Committees was the ITFLOWS Gender Action Plan (D2.2) and its annexed Gender Policy, drafted over the course of M1 to M5 of the project. The ITFLOWS GAP and Gender Policy, received input from and a final review by the Independent Gender Committee. Independent Gender Committee member Eleonore Kofman provided her section on integration and intersectionality on November 30, 2020. In a final review of the entire document on December 18, Eleonore Kofman suggested including sections on drivers and domestic violence, which were incorporated. Independent Gender Committee member Floya Anthias provided a section on intersectionality December 1, 2020. On her final review of the deliverable on December 17, she noted where to provide more inclusive language in terms of racialisation and the western vs. non-western binary; her changes and suggestions were incorporated.

Meetings and events

The Independent Gender Committee was convened twice in Gender Committee meetings, in addition to the times they participated in key meetings or events highlighted here:

- **ITFLOWS Kick-Off (Plenary) Meeting (M1):** The Independent Gender Committee presented on "Intersectionality" (Floya Anthias), and "Gendered Migration Flows" (Eleonore Kofman) to the ITFLOWS Consortium, as part of the Gender Committee workshop in the ITFLOWS annual Kick-Off meeting.

- **CEPS Ideas Lab** "The role of essential migrant workers in the EU's recovery: Why does gender matter?" (M10): The Internal Gender Committee and Independent Gender Committee (IGC) organized a gender panel for the CEPS Annual Ideas conference in June 2021. Independent Gender Committee member Eleonore Kofman was a panellist.

Monitoring and Recommendations

Monitoring of human participation: the Independent Gender Committee reviewed the interview scripts for interviews procedures and the anonymisation procedures of the interviews as part of D10.1 submitted by the IDT-UAB in M6 and offered feedback on the interview scripts.

Deliverables: In addition to these examples, the following deliverables were reviewed in or received input in year 1, and are listed consecutively by project month:

- D10.1 H-Requirement No. 1 (M6) As described above, the Independent Gender Committee ensured an adequate approach to gender and sexuality regarding human participation in ITFLOWS migrant interviews (T3.4).

- D10.2 H-Requirement No. 2 (M6) The Gender Policy specifically designed to be applied in the ITFLOWS Project is signed by the Independent Gender Committee.

5.3.2 Independent Gender Committee Activities Year 2 (M13-M24)

In Year 2, the Independent Gender Committees was again engaged in monitoring and reviewing activities in order to ensure the Gender Policy and the GAP were fully considered in al Work Packages.

Meetings and events

The Internal and Independent Gender Committees further held regular bi-monthly meetings (M13-M23). Apart from the established regular meetings, other relevant ones during this time period included:

ITFLOWS Annual Meeting (M13): The ITFLOWS Annual Meeting consisted of a three-part series of events. The Independent Gender Committee presented to the Consortium at the first of these three events among ITFLOWS project partners, reiterating their recommendations as to incorporating gender and sexuality into the Project.

Monitoring and Recommendations

Independent Gender Committee Monitoring of EMT: It should be noted here that the Independent Gender Committee met with the Internal Gender Committee to discuss the development of the EMT in M21. It made recommendations as to how to update the glossary to better reflect understandings of definitions of vulnerable and disabled, particularly with relation to the Qualification Directive (and how this is implemented or vulnerability assessments are carried out in varying Member States) as well as pointed to how the work of the Global Refugee Forum (as part of the Global Compact on Refugees) could also inform the glossary.

Deliverables: Again, as in Year 1, the Internal and Independent Gender Committees were both engaged in reviewing several deliverables to ensure the full compliance of all WPs to the mainstreaming of gender and sexualities in their outputs. The following deliverables (and one milestone) were reviewed in or received input in year 2, and are listed consecutively by project month:

- D3.2 Analysis on migration drivers and trajectories along The Eastern Mediterranean Route (South-Central Asia & Middle East); Central and Western Mediterranean Routes (North, West, and the Horn of Africa); Atlantic Air Route (South America) (M15): The Independent Gender Committee reviews and provides additional references on unaccompanied minors, single men, and the growing literature on who is vulnerable and in need of protection.

- D3.5 Qualitative report: Migration intentions, trajectories, and outcomes: The role of agency (M23): Independent Gender Committee reviews and asks for further expansion on the gender and intersectionality conclusions section. They acknowledge that more female interviews had been added to the various migratory routes, which highlights the gender composition of the routes.

CONCLUSIONS

This report has reviewed the monitoring strategy and activities conducted by the ethical and legal partners (WP2) of ITFLOWS in the first two years of the project. The aim of these activities was to ensure that the project complies with the identified legal framework (D2.3) and that it meets the highest European ethical standards. The report is divided in accordance with the main monitoring perspectives implemented in the project, namely: data protection (FIZ), ethical (IDT-UAB), societal (BUL), and gender (Gender Committee: UAB & BUL). In addition to the work of these partners, the monitoring process has been supported and supervised by three external bodies appointed at the beginning of the project: the Independent Ethics Board, the Data Protection Advisor, and the Independent Gender Committee.

Monitoring activities have included the drafting of deliverables setting the fundamental elements of the ethical and legal framework of the project (WP2 and WP10); providing guidance to the Consortium; organising meetings and seminars with the Consortium partners; identifying ethical and legal requirements, their level of implementation and the provision of mitigation measures (impact assessments); contributing to WP6 deliverables; and reviewing deliverables that may pose legal and ethical concerns.

Regarding data protection, during the first year FIZ focused on assessing the data sources for the development of the EMT (especially the use of Twitter), their associated risks, and how to mitigate them. Besides producing and reviewing deliverables, FIZ has been active in providing guidance to technical partners and has developed a compliance-monitor platform (CMP) to centralise and make more accessible the ethical, legal and data protection requirements. FIZ has also dealt with the data protection issues associated with the interviews with refugees, migrants and asylum seekers, and has collaborated in the design of the mechanisms implemented to make this particular research activity legally compliant. As already concluded in Section 1.1, the adopted data protection measures were successful since they contributed to raise awareness among partners and have led to the identification of the main data protection risks.

During the first 12 months of the project's lifespan, the ethical monitoring centred on the interviews with migrants, refugees, and asylum seekers. In this sense, the

IDT-UAB led the design of an integral ethical strategy to identify and mitigate the inherent risks associated with research activities involving vulnerable people. This strategy crystalised in the elaboration of informed consent procedures, an incidental findings policy, and an Ethics Handbook—among others. Likewise, the IDT-UAB has provided continuous guidance to the NGOs and other partners in charge of the interviews, reviewed the questionnaires, and supervised the whole process. From M12 to M24, the efforts of the IDT-UAB were redirected to monitoring the ethical aspects of the design of the EMT. In this sense, the two AI Impact Assessments conducted by the IDT-UAB have contributed to ensure the ethical compliance of the tool.

The participation of BUL has allowed partners to remain aware of the societal and human rights perspective at all times. They have attended and organised several meetings with partners to discuss the way ITFLOWS must be consistent with human rights in their research activities and in the design of the EMT. Their guidance and the bulk of recommendations they provided in D2.1 have been very valuable to clarify terminological misunderstandings, to properly address the existing legal framework in the field, and to identify bias. Their engagement played an important role in the monitoring process of the interviews with migrants, refugees, and asylum seekers, as well as in the preparation of the requirements to be embedded into the EMT.

As for the gender aspects, the involvement of the Gender Committee has been key to guarantee that all the ITFLOWS activities incorporate a gender and sexuality—as well as intersectional—perspective, and avoid gender and other forms of discrimination. The elaboration of the Gender Action Plan (D2.2), as well as the different informative activities and deliverables reviews, have achieved these goals.

Finally, the work of the external monitoring bodies (IEB, DPA, and IGC) reviewing deliverables and providing guidance to the ITFLOWS Consortium has added an additional layer of robustness to the ethical, legal, data protection, and gender aspects of the project.

ITFLOWS monitoring tasks are an ongoing process. WP2 partners must continue to work to ensure that all of the Project's research activities are legally and ethically

compliant. In particular, the next steps in the development of the EMT will be the major challenge to be addressed in Year 3. A second version of this report will be submitted in M36.

This Deliverable has been reviewed by the ITFLOWS Data Protection Advisor and by the Independent Ethics Board.

REFERENCES

Legal Sources

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR). https://eur-lex.europa.eu/eli/reg/2016/679/oj

Ethical guidelines (Chronological Order)

- UNHCR, 2016. *UNHCR viewpoint: 'Refugee'or 'migrant- Which is right?* https://www.unhcr.org/news/latest/2016/7/55df0e556/unhcr-viewpoint-refugee-migrant-right.html

- IEEE, 2019. *Ethically Aligned Design.* https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf

- High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019. *Ethics guidelines for trustworthy AI* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

- High-Level Expert Group on Artificial Intelligence (AI HLEG), 2020. *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment.* https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Secondary Sources (Alphabetical Order)

- Alam, M., Gesese, G.A., Sack, H., 2020. 'MigrAnalytics: Entity-based Analytics of Migration Tweets', available at: ceur-ws.org/Vol-2721/paper514.pdf.

- Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., Weinstein, J., 2018. 'Improving refugee integration through data-driven algorithmic assignment'. *Science*, 359 (6373): 325–329.

- Böhme, M.H., Gröger, A., Stöhr, T., 2020. 'Searching for a better life: Predicting international migration with online search keywords'. *Journal of Development Economics,* 142.

- Carammia, M., Iacus, S.M., Wilkin, T., 2022. 'Forecasting asylum-related migration flows with machine learning and data at scale'. *Scientific Reports*, 12.

- Chen, Y., Gesese G.A., Sack H., Alam M., 2021. 'Temporal Evolution of the Migration-related Topics on Social Media. Proceedings of the ISWC 2021 Posters and Demos Track, co-located with the 20th International Semantic Web Conference (ISWC 2021)'. *CEUR Workshop Proceedings,* 2980, CEUR-WS.org.

- Chen, Y., Sack, H., Alam, M., 2021. 'MigrationsKB: A knowledge base of public attitudes towards migrations and their driving factors', available at https://arxiv.org/abs/2108.07593.

- Dieng A.B., Ruiz F.J.R., and Blei D.M., 2020. 'Topic modeling in embedding spaces'. *Transactions on Association of Computational Linguistics*, 8: 439–453.

- El Enany, N., 2015. 'On Pragmatism and Legal Idolatry: Fortress Europe and the Desertion of the Refugee'. *International Journal on Minority and Group Rights*, 22: 7-38.

- Fafalios, P., Iosidis, V., Ntoutsi, E., Dietze, S., 2018. 'Tweetskb: A public and large-scale RDF corpus of annotated tweets'. CoRR abs/1810.10308.

- Gottschalk, T., Pichierri, F., 'About Migration Flows and Sentiment Analysis on Twitter Data: Building the Bridge Between Technical and Legal approaches to data protection'. [Presented on the LREC Conference 2022 (https://lrec2022.lrec-conf.org/en/); currently in publication]

- Mitsilegas, V., Moreno-Lax, V., and Vavoula, N., 2020. *Securitising Asylum Flows: Deflection, Criminalisation and the Challenges for Human Rights*. Nijhoff: Brill.

- Suleimenova, D., Bell, D., Groen, D., 2017. 'A generalized simulation development approach for predicting refugee destinations'. *Scientific Reports,* 7 (13377).

- Suleimenova, D., Groen, D., 2020. 'How Policy Decisions Affect Refugee Journeys in South Sudan: A Study Using Automated Ensemble Simulations'. *Journal of Artificial Societies and Social Simulation,* 23 (1).

- Vavoula, N., 2022. *Immigration and Privacy in the Law of the European Union – The Case of Information Systems.* Nijhoff: Brill.

- Xanthaki, A., 2016. 'Against Integration, In Favour of Human Rights'. *International Journal of Human Rights,* 20 (6): 815- 838.

ANNEX 1 – Ethics Handbook for the interviewing team (v3.0)

ITFLOWS TRAINING WORKSHOP

24th March 2021

(v3.0 - updated 24th September 2021)

itflows

IT TOOLS AND METHODS
FOR MANAGING
MIGRATION FLOWS

Ethics Handbook for the
interviewing team

The aim of this ethics handbook is to highlight the most relevant ethics issues that must be taken into account before conducting the interviews with migrants, refugees and asylum-seekers in ITFLOWS. This handbook also provides general ethical tips to address ethical concerns that may arise during the interviews. The ultimate goal is to ensure the ethical compliance of this research activity.

This handbook only contains basic information regarding some relevant ethical requirements previously identified in the context of ITFLOWS interviews. This means that you must be familiar with the content of the documents referenced in this handbook: Deliverable 2.1 (Section 5), Recruitment plan, Incidental Findings Policy, Gender Action Plan, Informed consent procedures and template and the Anonymisation techniques document.

Andrea Guillén and Emma Teodoro (IDT-UAB)

Ethics Handbook
Ethical principles governing
the interview research activity

1. Integrity
2. Reliability
3. Honesty
4. Respect
5. Accountability

1. Autonomy
2. Doing no harm
3. Equity
4. Diversity
5. Competence
6. Voluntary
   participation
7. Confidentiality &
   privacy
8. Transparency &
   accountability

General ethical principles

• All research activities foreseen within the Project should be conducted in strict compliance with the **general principle of integrity**.

• In addition to the principle of integrity, the following principles must be applied: **reliability, honesty, respect and accountability**.

(See Deliverable 2.1, Section 5, p. 157)

Ethical principles governing the interviews

ITFLOWS researchers involved in conducting qualitative research activities should also adhere to specific ethical principles. These principles have been identified and translated into ethical guidelines to be implemented by ITFLOWS interviewing teams in order to ensure that ITFLOWS interviews are conducted ethically.

• **Autonomy:** You should ensure the right of people to make their own decisions concerning their lives and particularly their participation in the Project.

• **Doing no harm:** You should prioritise the dignity, safety and well-being of participants as well as that of all members of the team. Individual or collective actions that may increase racism, discrimination, the criminalisation of migration or retraumatise migrants, refugees and asylum seekers should be avoided.

• **Equity:** You should take proactive actions to minimise potential negative impacts that may occur due to unbalanced power relationships.

• **Diversity:** You should respect cultural, ethnic, gender and sexual orientation differences. Ethnocentric research perspectives and behaviour must be avoided.

• **Competence:** Adequate training should be provided to the interviewing team. Profiles with expertise and empirical background in the field of migration should be prioritised.

• **Voluntary participation:** You must obtain participants' informed consent before their involvement in the interviews.

• **Confidentiality and privacy:** Processing of personal data must be compliant with the European and national data protection legal framework.

• **Transparency and accountability:** You should present the interview in a clear and accurate manner, avoiding biased and misleading information that makes it excessively attractive for the participants.

(See Deliverable 2.1, Section 5, p. 160-161)

Ethics Handbook
Human participation

Ethical guidelines to conduct the interviews

• You should **comply with the Guidance note on Research of refugees, asylum seekers and migrants** of the EU Directorate-General for Research and Innovation.[1]

• Do not create **unjustified expectations** in research participants about their future residence in the EU Member States, their status as asylum seekers or reward for their participation. Nevertheless, all research participants in the interviews will be provided with a small compensation with the aim of recognizing their time, effort and valuable participation.

• You should present the interview in an **unbiased manner** and free of misleading emphasis that makes it excessively appealing.

• Interviews must be conducted in a **comfortable and private setting** to favour the well-being of the research participant. Questions that could potentially cause distress, discomfort or fear should be carefully managed. If the described feelings are detected, the principles of autonomy and dignity should guide the actions to be taken by the interviewing team, e.g., to take a break or to avoid insisting on specific questions.

• Authorisations from national/local/reception centres authorities must be obtained before conducting the interviews.

(See Deliverable 2.1, Section 5, p. 162-164 and 'The roadmap of Human Rights and your Interview')

Recruitment plan: interviewing team

• NGOs will be responsible for ensuring the **adequate background and expertise** of the interviewing team. It is recommended to prioritise people with a refugee or migrant background, or from the same culture as the research participants, or with experience in communicating with migrants and refugees. The quality of translators is crucial to ensure fluid communication, fully informed consent, accurate and precise interpretations, and that research participants' voice is heard.

• **Under no circumstances external organisations will conduct interviews.** However, in case of necessity, NGOs can rely on their contact networks with other associations in search of support. If support from external organisations was needed, prior communication to the ITFLOWS Ethical Lead Partner (UAB) and its corresponding approval will be required to ensure ethical and legal compliance.

• You can resort to **trustworthy contacts** in the field, such as community leaders, to enable the recruitment of participants and to set fluid communication channels with them.

• Anyone involved in any way in the interviews must sign a **confidentiality agreement**.

• When the research participant is a woman, the presence of **female interviewers** is strongly recommended. Nevertheless, you must take into consideration the opinion and preferences of the research participant.

• **Public health directions** will be followed at all times when conducting interviews. This may imply avoiding quarantine areas or any place that could entail any sort of risk for the interviewing teams and research participants. All interviews will be conducted following all **sanitary measures** to protect interviewing teams' and research participants' health. Particular emphasis will be placed on COVID-19 given the current global pandemic.

• Interviewing teams will fully adhere to the **internal guidelines and best-practices codes** of their organisations.

(See Recruitment plan)

---

[1] European Commission (2020). Guidance Note – Research on Refugees, Asylum Seekers & Migrants. Retrieved from https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/guide_research-refugees-migrants_en.pdf

Ethics Handbook
Human participation

Recruitment plan: research participants

• **No minors will be interviewed.** You will rely on the personal documentation of the candidates (for example, C3 for asylum seekers in Italy) or any other reliable mean to assure the legal age of the research participants. You can resort to trustworthy contacts in the field and community leaders to verify the age of the research participants. In doubtful cases in which the verification of the age is impossible, the candidate will necessarily be excluded from participation in the interview.

• People who are already identified as **especially vulnerable groups** will also be automatically excluded from participation in the interview. Especially vulnerable people includes people with evident health concerns and psychological traumas, as well as victims of human trafficking and genital mutilation, among others. See the Incidental Findings Policy (see below), which establishes action protocols for you to react to unexpected findings of this type. (See Incidental Findings Policy)

• Ensure **gender representativeness** in the interview sample. Gender parity will be guaranteed by ensuring a proportion of 1 woman per every three or four men. The objective is to ensure a proportion as close as possible to 1/3. (See Gender Action Plan)

• Participation must be **voluntary** and based on **free and informed consent.** You will inform research participants about the nature of the project and the interview, and about their rights. They will be able to withdraw their consent at any time or to momentarily interrupt the interview without any detriment. (See Informed consent procedures and template)

• Personal data from interviews will be **anonymised.** (See Anonymisation techniques document)

• Research participants in the interviews will be provided with a **small compensation.** In this regard, an acknowledgement of receipt proving the transmission of the small compensation must be signed. Each NGO partner in charge of conducting the interviews must decide and specify the type of compensation (monetary compensation, voucher or some gift) that will be offered to all research participants. This compensation will be subject to the approval of the Ethics Lead Partner (UAB). **As a general rule monetary compensation is highly discouraged.** In this respect, if one of the NGOs decides to consider monetary compensation, a specific request in this respect - justifying the suitability of this compensation - must be sent to the Ethics Lead Partner of the project (UAB). This request will be subject to the specific approval of the Independent Ethics Board (IEB) and the Data Protection Advisor (DPA).

(See Recruitment plan)

Incidental findings policy

• Incidental Findings refer to risks that may emerge in any research activity involving human participants, which are unrelated to the purpose of the research activity. In the context of the interviews, incidental findings may comprise indications of criminal activity and human rights violations. NGOs have evaluated **sexual and gender-based violence and trafficking in human beings** as **highly likely** to be incidental findings in the context of the interviews to be conducted in ITFLOWS.

• You must be familiarised with the **ITFLOWS Incidental Findings Policy** and strictly follow the **applicable national referral system.**

• You must **respect the autonomy of the research participant** to freely decide on the next steps to take, if any.

• You must comply with the ethical principles that govern the Incidental Findings Policy and its procedure: i) **protection of migrant's best interests;** ii) **do no harm;** iii) **zero-tolerance approach;** iv) **procedural fairness: accountability and transparency;** v) **fair benefit-sharing;** and, vi) **shared responsibility.**

(See Incidental Findings Policy)

Ethics Handbook
Human participation

Incidental findings policy: procedure

• You must follow the **ITFLOWS Incidental Findings Procedure**:

○ <u>Stage 1</u>: You must **immediately report any disclosure or suspicion** of an incidental finding, or at least within the first 24 hours to the line manager. Such information can be provided verbally or in writing.

○ <u>Stage 2</u>: You will comply with the principles of respect **for dignity, participation, privacy and security** and continue with **active listening**. Safeguarding procedures must be applied and the action/decision to be taken in protecting the person is achieved within the team, considering the specific circumstances of the research participant and the <u>referral system of each country</u> where the interview is taking place.

○ <u>Stage 3</u>: You will provide **detailed information** to the research participant on the **specific services available** in the territory (following the applicable national referral system) and will maintain open communication with the research participant. Additionally, the **relevant phone numbers** within the applicable National Referral System must be provided to the research participant. **As a general rule, you must respect the autonomy of the research participant** to freely decide on the next steps to take, if any. The purpose of this approach is to avoid putting the research participant in danger and allow them to exercise their personal judgement as to the best course of action. Nevertheless, if the interviewing team considers that the **research participant is in concrete danger**, emergency services will be contacted to ensure their immediate protection and wellbeing.
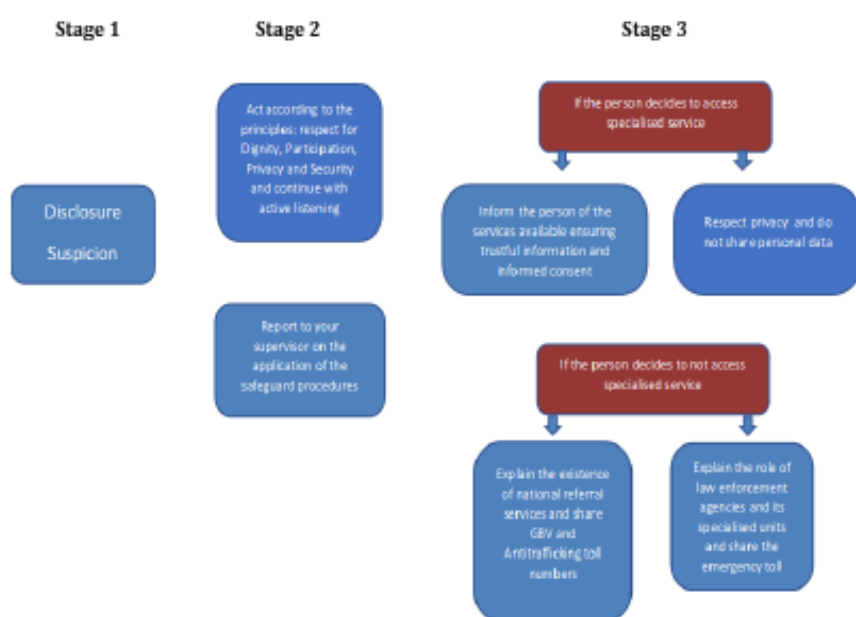
(See Incidental Findings Policy)

Ethics Handbook
Protection of personal data

Anonymisation techniques to be applied to interviews

• NGOs and interviewing teams must **respect and preserve** the **anonymity** of all research participants.

• Interviewing teams will record the **audio** of the interviews **to facilitate its transcription**, and for this sole purpose.

• The recording will be **deleted as soon as the transcript is finished**, and in any case, no later than **24 hours after the transcription.**

• The transcript of the interview will not contain research participants' **direct identifiers**, such as names. Additionally, **other types of personal data** that clearly enables identification must be erased.

• Transcribers will receive training on what types of data must be erased.

• Interviews will be associated with a **unique code** and **indirect markers** (date, place, gender, age group, educational/professional background, and country of origin), whose combination must not lead to the identification of the participant.

• Informed consent forms gathered from research participants will in no way be associated with the interview's unique code and will be kept on the NGOs premises in an **encrypted form.**

• NGOs will evaluate if any of these interview markers combined with the information disclosed during the interview itself can lead to the identification of the research participant. In such a case, interview markers will be **generalised.**

• These anonymised transcripts will be safely stored on the NGOs premises using **encryption**, will not be shared with any third party, and will be destroyed at the end of the project. The **anonymised transcripts** will only be shared with IAI and UAB in a format that is **encrypted**. NGOs must also send, in an **encrypted** way, the **table compiling all interviews** with the corresponding indirect markers.

(See Anonymisation techniques document)

Two-step incidental findings transcription procedure (I)

This general procedure applies to all interviews in which information regarding an incidental finding emerges. The purpose of this procedure is to establish clear guidelines for NGOs, interviewers and the partners involved in the analysis of the interview transcripts on **how to proceed when information regarding an incidental finding is disclosed during the interviews.**

This procedure does not revolve around the Incidental Findings Policy but **whether information disclosed during the interview related to an incidental finding should be included in or removed from the transcript**, and in case of including it how to proceed from an ethical and data protection perspective. The procedure consists of two steps.

Ethics Handbook
Protection of personal data

Two-step incidental findings transcription procedure (II)

<u>**Step 1**</u>: Determining whether the information about the incidental finding is relevant for analysis purposes within ITFLOWS. This step is divided into two stages:

1. <u>Before the transcription of the interview</u>, you must produce a **general summary** that should not contain any personal data and must send it to IAI and UAB. Under no circumstances will the original transcript be sent to IAI and UAB.

   The following recommendations must be observed:

   a) All interviewers must be properly trained in:
      • Ethics Handbook for the interviewing team and the ethical documents referred to in the handbook.
      • The ITFLOWS Incidental Finding Policy designed for the interviews.
      • The applicable referral system to deal with incidental findings in the country where interviews are conducted.
      • The anonymisation procedure to ensure that the general summary (if an incidental finding emerges) and the transcript does not contain personal data.

   b) You must draw up the general summary **following the anonymisation procedure** in place to minimise the risk of including personal data.
   c) You must submit jointly with the general summary **your personal opinion** on the relevance/irrelevance of the information related to an incidental finding disclosed during the interview.

2. Based on the general summary, **IAI and UAB will determine the relevance/irrelevance** of the information and will inform the NGO and the interviewer about their decision.

   IAI and UAB assessment on the relevance of such information should be properly documented. The evaluation must take into consideration the opinion of the interviewer that conducted the interview. This decision-making process should be performed following a cautionary approach. This approach entails that if the relevance of the disclosed information cannot be determined, the information should be deemed irrelevant and should thus not be included in the transcript.

   In case there is no unanimous assessment on the significance of the information disclosed in terms of relevance and necessity for the research purpose of the interviews: IAI decision, as leader of this task, will prevail. Dissenting opinions and the rationale behind the final decision must also be properly documented.

   If the "do no harm" ethical principle is compromised in a particular case, the relevance/irrelevance check does not longer apply, and the information should not be included in the transcript.

<u>**Step 2**</u>: There are two possible outcomes:

1. If IAI and UAB consider that the information is **relevant** → The two-step anonymisation approach applicable to all transcripts will be followed, i.e., you must anonymise the information related to the incidental finding, and IAI/UAB will double-check it.
2. If the information is deemed **irrelevant** → You should not include this information in the transcript.

Ethics Handbook
Gender

Gender policy

ITFLOWS is committed to considering gender and sexuality as a major priority in project design, methodology, analysis, and dissemination of outputs, acknowledging that gender and sexuality are central to an intersectional analysis of migration flows. In addition, gender equality and balance among its participants forms part of this commitment.

• **Intersectionality**: Pursue an intersectional approach to migration research and analysis in order to understand the manner in which the EU migration and asylum systems create dynamics of in- and exclusions along lines of gender, race, ethnicity, social group, or sexuality – taking into consideration colonial legacies.

• **Gender-based violence and gender stereotyping**: Remain sensitive to migrants' experiences with gender-based and sexual violence (trafficking, female genital mutilation (FGM), rape or sexual assault, forced marriages etc.) and how gender stereotyping contributes to the marginalisation and stigmatising of women, girls as well as men, boys and LGBTQI+ persons.

• **Gender discrimination within the EU**: Be aware of gender and sexuality discrimination within the EU in the area of migration policies, laws, employment, education, health and with respect to COVID-19.

• **Gender and Technology**: Recognise that technology can intensify gender inequalities because women and girls may have less access to technological literacy due to socioeconomic disadvantages. The development of the EMT must be taken to actively mitigate bias, ensuring that (a) gender equality is considered a key principle in the research, (b) gender is included as a variable in the study design, and (c) data are presented in disaggregated fashion at all levels of intersectional analysis.

Useful resources:

- OECD: Addressing Emerging Human Trafficking Trends and Consequences of the COVID-19 Pandemic. https://www.osce.org/odihr/human-trafficking-COVID-19-report
- Key types of sexual and gender violence from UN Women. https://www.unwomen.org/en/what-we-do/ending-violence-against-women/faqs/types-of-violence
- Council of Europe Fact Sheet, Protecting the rights of migrant, refugee and asylum-seeking women and girls. https://edoc.coe.int/en/refugees/8053-protecting-the-rights-of-migrant-refugee-and-asylum-seeking-women-and-girls.html

(See Gender Action Plan)

ANNEX 2 – Preliminary AI Impact Assessment Questionnaire

November 2021

EUMIGRATOOL: Preliminary AI Impact Assessment

QUESTIONNAIRE OF THE EMT AI IMPACT ASSESSMENT FOR ITFLOWS TECHNICAL PARTNERS (Internal Working Document)
Andrea Guillén, Emma Teodoro
(UAB-IDT)

1.   Methodology

The purpose of this preliminary AI Impact Assessment is to identify and assess, at this stage of the project, the risks posed by the EMT in order to minimise them.

The assessment will be conducted on the basis of recent relevant works published by the High-Level Expert Group on Artificial Intelligence of the European Commission (AI HLEG)[14] and by the IEEE[15].

Following the methodological approach provided by such works—primarily the AI HLEG guidelines—a set of ethical principles based on fundamental rights has been particularly identified as the backbone of the AI Impact Assessment to ensure that AI ethics are embedded in the EMT. According to the AI HLEG, these principles are: i) **human autonomy,** ii) **prevention of harms,** iii) **fairness and,** iv) **transparency/explicability.**

These principles are turned into requirements for addressing the risks. These requirements are: i) **human agency and oversight**, ii) **technical robustness and safety**, iii) **privacy and data governance**, iv) **transparency**, v) **diversity, non-discrimination, and fairness,** vi) **environmental and societal well-being** and, vii) **accountability.**

Each requirement is comprised of a set of questions which must be answered internally by all **ITFLOWS technical partners, with the lead of WP6. WP6 should provide a consolidated AI impact assessment.**

---

[14] High-Level Expert Group on Artificial Intelligence of the European Union, "Ethics Guidelines for Trustworthy AI" [2019] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai and "Assessment List for Trustworthy Artificial Intelligence" [2020] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

[15] https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf

2. Practical guidelines to fill out the questionnaire

The questionnaire is structured into the seven requirements identified by the AI HLEG in their 'Ethics Guidelines for Trustworthy AI' and the 'Assessment List for Trustworthy AI'. As mentioned, the list of requirements that will be evaluated is the following:

- R1: Human agency and oversight
- R2: Technical robustness and safety
- R3: Privacy and Data governance
- R4: Transparency
- R5: Diversity, non-discrimination, and fairness
- R6: Environmental and societal well-being
- R7: Accountability

Each requirement is comprised of a set of questions specifically targeted at addressing, in practice, the level of implementation of each ethical principle in the EMT. Each question has an alphanumeric identification (ID) with two components: the number of the requirement (R1, R2, R3, …) and the number of the question (Q1, Q2, Q3 …). For example, the ID 'R4-Q3' refers to the third question (Q3) of the requirement related to transparency (R4).

A table, which must be filled out by WP6 leader (CERTH) on behalf of ITFLOWS Technical Partners <u>to the greatest extent possible</u>, can be found below the questionnaire of each requirement. <u>**Do not provide yes/no answers, provide as many details as possible**</u>. The tables are divided into two columns. The first contains the ID of each requirement's question and the second column must be completed with the corresponding answer.

Before filling out the questionnaire tables, the subsequent sections must be read. Section 3 identifies the four ethical principles that are at stake in the EMT and how these are addressed in the requirements. Section 4 briefly explains the meaning and scope of each requirement. The content of Section 3 and 4 is based on the 'Ethics Guidelines for Trustworthy AI'.

3. Identification of the ethical principles addressed by the AI requirements

3.1. Human autonomy
"AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems."

This ethical principle is addressed in:

- R1: Human agency and oversight

3.2. Prevention of harms

"AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure."

This ethical principle is addressed in:

- R2: Technical robustness and safety
- R3: Privacy and data governance
- R6: Societal and environmental well-being

3.3. Fairness

"ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation."

This ethical principle is addressed in:

- R5: Diversity, non-discrimination and fairness
- R6: Societal and environmental well-being
- R7: Accountability

3.4. Transparency/Explicability

"processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected."

This ethical principle is addressed in:

- R4: Transparency

4. Definition of the requirements for ensuring ethical principles and addressing potential risks

- **R1: Human agency and oversight**

"AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight."

- **R2: Technical robustness and Safety**

"A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence

of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured."

Technical robustness is also key for the system's accuracy, which "*pertains to an AI system's ability to make correct judgements, or its ability to make correct predictions, recommendations, or decisions based on data or models. An explicit and well-formed development and evaluation process can support, mitigate and correct unintended risks from inaccurate predictions*."

- **R3: Privacy and data governance**

"Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy."

- **R4: Transparency**

"This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business models."

Decisions made by systems built on AI must be transparent, traceable and explainable.

- **R5: Diversity, non-discrimination and fairness**

"In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness."

- **R6: Societal and environmental well-being**

"In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations."

- **R7: Accountability**

"The requirement of accountability complements the above requirements, and is closely linked to the principle of fairness. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use."

5. Questionnaire and answers

Please, fill out the following questionnaire according to the guidelines provided in Section 2 of this document.

R1: Human agency and oversight

- R1-Q1. Can you describe the level of human control or involvement?
- R1-Q2. Do you plan to put in place mechanisms and measures to ensure human control or oversight?
- R1-Q3. What training will/do users have?

R1: Human agency and oversight	
ID	*Answer*
R1-Q1	
R1-Q2	
R1-Q3	

R2: Technical robustness and safety

- R2-Q1. How exposed is the EMT to cyberattacks? What measures do you plan to implement to ensure the integrity and resilience of the EMT against potential attacks?
- R2-Q2. What can be the likely impact of a failure of the EMT if it provides wrong results, becomes unavailable, or provides societally unacceptable results (e.g. discrimination)?
- R2-Q3 Do you plan to assess how accuracy is measured and assured in the EMT?
- R2-Q4. Do you plan to put in place measures to ensure that the data used in the EMT is comprehensive and up to date?
- R2-Q5. Do you plan to put measures in place to assess whether there is a need for additional data, for example to improve accuracy or eliminate bias?
- R2-Q6. What harm would be caused if the EMT makes inaccurate predictions?
- R2-Q7. Do you plan to put in place ways to measure whether the EMT is making an unacceptable amount of inaccurate predictions?
- R2-Q8. How do you plan to monitor and test if the EMT is meeting its goals, purposes and intended applications?

R2: Technical robustness and safety	
ID	*Answer*
R2-Q1	
R2-Q2	
R2-Q3	
R2-Q4	
R2-Q5	
R2-Q6	
R2-Q7	
R2-Q8	

R3: Privacy and data governance

- R3-Q1. Do you plan to assess the type and scope of data in your data sets (for example whether they contain personal data)?
- R3-Q2. Do you plan to take measures to enhance privacy, such as via encryption, anonymisation and aggregation?
- R3-Q3. Do you plan to establish oversight mechanisms for data collection, storage, processing, and use?
- R3-Q4. Do you plan to assess the extent to which you are in control of the quality of the external data sources used?
- R3-Q5. Do you plan to you put in place processes to ensure the quality and integrity of your data? How are you verifying that your data sets have not been compromised or hacked?
- R3-Q6. What protocols, processes and procedures are you following to manage and ensure proper data governance? Do you assess who can access users' data, and under what circumstances? Do you plan to ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?

R3: Privacy and data governance	
ID	*Answer*
R3-Q1	
R3-Q2	
R3-Q3	
R3-Q4	
R3-Q5	
R3-Q6	

R4: Transparency

- R4-Q1. Do you plan to establish measures that can ensure traceability? This could entail documenting the following:
    - Methods used for designing and developing the EMT:
        - Rule-based systems: the method of programming or how the model was built;
        - Learning-based systems: the method of training the algorithm, including which input data was gathered and selected, and how this occurred.
    - Methods used to test and validate the EMT:
        - Rule-based systems: the scenarios or cases used in order to test and validate;
        - Learning-based systems: information about the data used to test and validate.
- R4-Q2. To what extent can the decisions and hence the outcome made by the EMT be understood by users?
- R4-Q3. Are you designing the EMT with interpretability/explainability in mind from the start?
- R4-Q4. Do you plan to communicate to users the characteristics, limitations and potential shortcomings of the EMT?

R4: Transparency	
ID	*Answer*
R4-Q1	
R4-Q2	
R4-Q3	
R4-Q4	

R5: Diversity, non-discrimination and fairness

- R5-Q1. Do you plan to assess and acknowledge the possible limitations stemming from the composition of the used data sets?
- R5-Q2. Are you considering diversity and representativeness of users in the data?
- R5-Q3. Do you plan to put in place processes to test and monitor for potential biases during the design, development, deployment and use phase of the EMT?
- R5-Q4. Do you plan to assess whether there could be persons or groups who might be disproportionately affected by negative implications due to the use of the EMT?
- R5-Q5. Are you including/engaging with different stakeholders in the EMT's design, development and use?

R5: Diversity, non-discrimination and fairness	
ID	*Answer*
R5-Q1	
R5-Q2	
R5-Q3	
R5-Q4	
R5-Q5	

R6: Societal and environmental well-being

- R6-Q1. Have you assessed the broader societal impact (both positive and negative) of the EMT?
- R6-Q2. Do you plan to put in place measures to reduce the environmental impact of the EMT's life cycle? (sustainable and environmentally friendly AI)

R6: Societal and environmental well-being	
ID	*Answer*
R6-Q1	
R6-Q2	

R7: Accountability

- R7-Q1. Do you plan to provide training and education to help developing accountability practices? Does its content include the risk of misuse?
- R7-Q2. Do you plan to establish mechanisms that facilitate the EMT's auditability, such as ensuring traceability and logging of the EMT's processes and outcomes?
- R7-Q3. Do you plan to establish processes for users to report potential vulnerabilities, risks or biases in the EMT?
- R7-Q4. Are authentication and authorisation components embedded into the EMT?
- R7-Q5. Are oversight mechanisms implemented to log when, where, how, by whom and for what purpose data was accessed?

R7: Accountability	
ID	*Answer*
R7-Q1	
R7-Q2	
R7-Q3	
R7-Q4	
R7-Q5	

ANNEX 3 – 2nd AI Impact Assessment Questionnaire

itflows
IT TOOLS AND METHODS
FOR MANAGING
MIGRATION FLOWS

June 2022

EUMIGRATOOL: AI Impact Assessment

QUESTIONNAIRE OF THE EMT AI IMPACT ASSESSMENT FOR ITFLOWS TECHNICAL
PARTNERS (Internal Working Document)
Andrea Guillén, Emma Teodoro
(UAB-IDT)

6. Methodology

The purpose of this AI Impact Assessment is to identify and assess, at this stage of the project (M22), the risks posed by the EMT in order to minimise them. This AI impact assessment is based on the results of the Preliminary AI Impact Assessment conducted last year (M14).

As with the preliminary AI impact assessment, this assessment will also be conducted on the basis of the work published by the High-Level Expert Group on Artificial Intelligence of the European Commission (AI HLEG)[16] and by the IEEE[17], with a particular focus on the field of deployment of the EMT and its purpose – migration management in the humanitarian context.

Following the methodological approach provided by such works—primarily the AI HLEG guidelines—a set of ethical principles based on fundamental rights has been particularly identified as the backbone of the AI Impact Assessment to ensure that AI ethics are embedded in the EMT. According to the AI HLEG, these principles are: i) **human autonomy,** ii) **prevention of harms,** iii) **fairness and,** iv) **transparency/explicability.**

These principles are turned into requirements for addressing the risks. These requirements are: i) **human agency and oversight**, ii) **technical robustness and safety**, iii) **privacy and data governance**, iv) **transparency**, v) **diversity, non-discrimination, and fairness,** vi) **environmental and societal well-being** and, vii) **accountability.**

Each requirement is comprised of a set of questions which must be answered internally by all **ITFLOWS technical partners, with the lead of WP6. WP6 should provide a consolidated AI impact assessment.**

---

[16] High-Level Expert Group on Artificial Intelligence of the European Union, "Ethics Guidelines for Trustworthy AI" [2019] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai and "Assessment List for Trustworthy Artificial Intelligence" [2020] https://ec.europa.eu/digital-single-market/en/news/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
[17] https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf

7. Practical guidelines to fill out the questionnaire

The questionnaire is structured into the seven requirements identified by the AI HLEG in their 'Ethics Guidelines for Trustworthy AI' and the 'Assessment List for Trustworthy AI'. As mentioned, the list of requirements that will be evaluated is the following:

- R1: Human agency and oversight
- R2: Technical robustness and safety
- R3: Privacy and Data governance
- R4: Transparency
- R5: Diversity, non-discrimination, and fairness
- R6: Environmental and societal well-being
- R7: Accountability

Each requirement is comprised of a set of questions designed by WP2 partners specifically targeted at addressing, in practice, the level of implementation of each ethical principle in the EMT. Each question has an alphanumeric identification (ID) with two components: the number of the requirement (R1, R2, R3, …) and the number of the question (Q1, Q2, Q3 …). For example, the ID 'R4-Q3' refers to the third question (Q3) of the requirement related to transparency (R4).

A table, which must be filled out by WP6 leader (CERTH) on behalf of ITFLOWS Technical Partners <u>to the greatest extent possible</u>, can be found below the questionnaire of each requirement. <u>**Do not provide yes/no answers, provide as many details as possible**</u>. The tables are divided into two columns, the first contains the ID of each requirement's question and the second column must be completed with the corresponding answer.

Before filling out the questionnaire tables, the subsequent sections must be read. Section 3 identifies the four ethical principles that are at stake in the EMT and how these are addressed in the requirements. Section 4 briefly explains the meaning and scope of each requirement. The content of Section 3 and 4 is based on the 'Ethics Guidelines for Trustworthy AI'.

8. Identification of the ethical principles addressed by the AI requirements

8.1. Human autonomy

"AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems."

This ethical principle is addressed in:

- R1: Human agency and oversight

8.2. Prevention of harms

"AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings. This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure."

This ethical principle is addressed in:

- R2: Technical robustness and safety
- R3: Privacy and data governance
- R6: Societal and environmental well-being

8.3. Fairness

"ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation."

This ethical principle is addressed in:

- R5: Diversity, non-discrimination and fairness
- R6: Societal and environmental well-being
- R7: Accountability

8.4. Transparency/Explicability

"processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected."

This ethical principle is addressed in:

- R4: Transparency

9. Definition of the requirements for ensuring ethical principles and addressing potential risks

- **R1: Human agency and oversight**

"AI systems should support human autonomy and decision-making, as prescribed by the principle of respect for human autonomy. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight."

- **R2: Technical robustness and Safety**

"A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the principle of prevention of harm. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured."

Technical robustness is also key for the system's accuracy, which "*pertains to an AI system's ability to make correct judgements, or its ability to make correct predictions, recommendations,*

or decisions based on data or models. An explicit and well-formed development and evaluation process can support, mitigate and correct unintended risks from inaccurate predictions."

- **R3: Privacy and data governance**

"Closely linked to the principle of prevention of harm is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy."

- **R4: Transparency**

"This requirement is closely linked with the principle of explicability and encompasses transparency of elements relevant to an AI system: the data, the system and the business models."

Decisions made by systems built on AI must be transparent, traceable and explainable.

- **R5: Diversity, non-discrimination and fairness**

"In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with the principle of fairness."

- **R6: Societal and environmental well-being**

"In line with the principles of fairness and prevention of harm, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations."

- **R7: Accountability**

"The requirement of accountability complements the above requirements, and is closely linked to the principle of fairness. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use."

10. Questionnaire and answers

Please, fill out the following questionnaire according to the guidelines provided in Section 2 of this document.

R1: Human agency and oversight

- R1-Q1: What information about each EMT module -*that aims to assist the decision-making process of end-users of the EMT*- is provided to end-users? Please, indicate where this information can be found and if it is clearly visible and accessible.
- R1-Q2: What information referred to the EMT outputs/results, in particular on how the EMT has produced a specific output/result, is provided to end-users?
- R1-Q3: How would an end-user assess quality of results?
- R1-Q4: What recommendations/guidance are provided to end-users to avoid unconscious bias? Please, indicate where this information can be found.
- R1-Q5: Can you confirm that information related to the EMT helpdesk purpose and functionalities is provided to end-users? Please, detail the information provided to end-users.
- R1-Q6: Have you implemented, as part of the EMT helpdesk, a reporting mechanism to allow end-users to flag errors, potential biases and systems' malfunctions?

R1: Human agency and oversight	
ID	*Answer*
R1-Q1	
R1-Q2	
R1-Q3	
R1-Q4	
R1-Q5	
R1-Q6	

R2: Technical robustness and safety

- R2-Q1: Have you assessed the level of exposure of the EMT to cyberattacks? Please, provide explanations regarding: i) how cybersecurity risks have been assessed; ii) results of such assessment; iii) and measures embedded into the EMT against potential attacks.
- R2-Q2: Have you assessed the potential misuse (intended and unintended) of the EMT? What technical measures have been embedded into the EMT to prevent its misuse?
- R3-Q3: How have you assessed bias and accuracy of the EMT at the design and at the development phases of the EMT?
- R2-Q4: Has a threshold been established for the accuracy rates of the EMT predictions? Please, provide technical explanations about how this accuracy threshold has been determined.
- R2-Q5: Have you implemented technical measures into the EMT to ensure that users cannot have access to those EMT predictions that are below the accuracy threshold established as a minimum by technical partners? Have you implemented warning mechanisms to alert users of the poor accuracy of a given prediction?
- R2-Q6: What instructions are provided to EMT end-users to inform them that the EMT results serve only as a tool that aid their decision-making? Where is this information located into the EMT?
- R2-Q7: What awareness-raising measures addressed to EMT end-users have been implemented to minimise negative societal impacts due to malfunctions of the EMT?

R2: Technical robustness and safety	
ID	*Answer*
R2-Q1	
R2-Q2	
R2-Q3	
R2-Q4	
R2-Q5	
R2-Q6	
R2-Q7	

R3: Privacy and data governance

- R3-Q1: Please, provide detailed information regarding:
  - Security measures implemented for the processing of data in CKAN.
  - Mechanisms implemented to ensure non-authorised access to CKAN.
  - Technical measures implemented to ensure the integrity of data in CKAN.
  - Technical measures implemented to ensure the accuracy and quality of data in CKAN.
  - How the risk of de-anonymisation has been assessed. Mitigation measures implemented to address the potential misuse of data during and beyond the lifespan of the project.
  - Mitigation measures implemented to address the privacy of website users (EMT front-end).
  - How indirect use of data via a model will fix bad quality.
- R3-Q2: Is the collection of information from twitter users compliant with the data protection principle of data minimisation? How have you ensured compliance with this principle?
- R3-Q3: Is the processing of information from twitter users compliant with the data protection principle of data accuracy? How have you ensured compliance with this principle?
- R3-Q4: Concerning Twitter analysis, what measures did you take to avoid any linkage to personal data? Please provide detailed information regarding the technical measures implemented to ensure privacy of Twitter users.
- R3-Q5: Have you implemented quality checks to be conducted to external data sources before processing them into the EMT? Please, provide detailed information on what do these quality checks consist of.
- R3-Q6: Have you developed the Terms of Use of the EMT? Where can this information be found? Is it visible and easily accessible?

R3: Privacy and data governance	
ID	*Answer*
R3-Q1	
R3-Q2	
R3-Q3	
R3-Q4	
R3-Q5	
R3-Q6	

R4: Transparency

- R4-Q1: Regarding technical measures implemented to ensure traceability, can you confirm that you have documented methods used for designing and developing the EMT; methods used to test and validate the EMT; and the results/outcomes of the EMT?
- R4-Q2: What information regarding the design and development of the EMT is publicly available on the EMT Webpage? Is this information provided in a clear and plain language free from technical jargon? Where can this information be found? Is this information easily accessible?
- R4-Q3: What information on the limitations and shortcomings of the EMT has been made available on the EMT website? Is this information easily accessible, visible, and clearly explained to EMT end-users?
- R4-Q4: Have oversight mechanisms been implemented to ensure both, that the datasets fed into the EMT do not contain personal data, and to control the impact of potential inferences due to the processing of open-source data?
- R4-Q5: What mechanisms have been foreseen to audit the EMT?

R4: Transparency	
ID	*Answer*
R4-Q1	
R4-Q2	
R4-Q3	
R4-Q4	
R4-Q5	

R5: Diversity, non-discrimination and fairness

- R5-Q1: How have diversity and representativeness of the datasets used for the EMT been ensured? What criteria has been followed to establish the variables to be taken into account?
- R5-Q2: How have the different modules of the EMT been tested or will be tested for bias at the design, development, and implementation phases of the EMT? How do you plan to solve potential biases raised at the design, development, deployment and use phases of the EMT?
- R5-Q3: How have technical partners assessed and acknowledged limitations related to the composition of the used data sets? What information regarding such assessment will be provided to end-users? How will this information be provided?
- R5-Q4: What reporting mechanisms have been implemented to gather end-users' feedback during the testing and use phase of the EMT? How will end-users' feedback be used to enhance/develop the EMT?
- R5-Q5: Is the EMT usable by people with special needs or disabilities?
- R5-Q6: Describe in detail the explainability features that have been embedded into the models underlying the EMT.

R5: Diversity, non-discrimination and fairness	
ID	*Answer*
R5-Q1	
R5-Q2	
R5-Q3	
R5-Q4	
R5-Q5	
R5-Q6	

R6: Societal and environmental well-being

- R6-Q1: How has the societal impact of the EMT been assessed? Which measures have been implemented to reduce negative societal impacts that the design, development, deployment, and use of the EMT may cause?
- R6-Q2: How has the environmental impact of the EMT been assessed? Which measures have been implemented to reduce negative environmental impacts that the design, development, deployment, and use of the EMT may cause?
- R6-Q3: Can you provide detailed information on how the EMT is compliant with the "Do no significant harm principle" (Articles 9 and 17 of the EU Sustainable Finance Taxonomy Regulation)?

R6: Societal and environmental well-being	
ID	*Answer*
R6-Q1	
R6-Q2	
R6-Q3	

R7: Accountability

- R7-Q1: Which technical measures have been implemented or will be implemented to minimise the risk related to the potential misuse of the EMT?
- R7-Q2: Are access rights to the EMT clearly defined and differentiated based on the type of end-user? Can you provide explanations on which and how different access rights are implemented in the EMT?
- R7-Q3: Can you provide technical explanations regarding the EMT authentication and authorisation system?
- R7-Q4: How will the EMT activity logs be reviewed (ad hoc, on a regular basis…)?
- R7-Q5: Is there an EMT Auditing Plan in place? Can you describe how the EMT will be audited and by whom?

R7: Accountability	
ID	*Answer*
R7-Q1	
R7-Q2	
R7-Q3	
R7-Q4	
R7-Q5	