IT TOOLS AND METHODS FOR MANAGING MIGRATION FLOWS

Preliminary release of the EMT

Georgios Stavropoulos
Nikolaos Gkevrekis
Ilias Iliopoulos
Garyfallia Papadopoulou

The Centre for Research & Technology

February 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 882986.

Deliverable Factsheet	
Title and number	*Preliminary release of the EMT* (D6.2)
Work Package	WP6
Submission date	28/02/2022
Authors	Georgios Stavropoulos, Nikolaos Gkevrekis, Ilias Iliopoulos, Garyfallia Papadopoulou (CERTH)
Contributors	Thilo Gottschalk, Francesca Pichierri, Mehwish Alam & Yiyi Chen (FIZ), Cristina Blasi, Colleen Boland, Emma Teodoro & Andrea Guillén (UAB), Lenka Drazanova (EUI), Mengia Tschalaer, Derek Groen & Diana Suleimenova (BUL), Andre Groeger & Konstantin Boss (UAB), Stelios Gkouskos, & Konstantinos Kalampokis (TRC), Tobias Heidland (IfW).
Reviewers	Thilo Gottschalk (FIZ) Emma Teodoro (UAB) Independent Ethics Board (IEB)
Dissemination level	PU (Public)
Deliverable type	DEM (Demonstrator)

Version Log			
Issue Date	**Version**	**Author**	**Change**
20/12/2021	V0.1	Georgios Stavropoulos	ToC & structure
08/02/2022	V0.2	WP2 Ethical & Legal Contribution	Section 4
08/02/2022	V0.3	Georgios Stavropoulos, Nikolaos Gkevrekis, Ilias Iliopoulos, Garyfallia Papadopoulou (CERTH)	Initial version sent to reviewers
11/02/2022	V1.0	Thilo Gottschalk (FIZ) Emma Teodoro (UAB)	Internal Review
16/02/2022	V1.1	Lilian Mitrou, Marta Poblet Ruth Fee Jonathan Andrew	ITFLOWS IEB and DPA Review
18/02/2022	V2.0	Georgios Stavropoulos, Nikolaos Gkevrekis, Ilias Iliopoulos, Garyfallia Papadopoulou (CERTH)	Addressed initial review comments
25/02/2022	V2.1	Cristina Blasi (UAB)	Coordinator final review and layout adjustments. Version ready to be submitted.

Executive Summary

This deliverable aims to release the first version of the EUMigraTool (EMT) application that was developed within the ITFLOWS project.

The EMT is a software platform that brings together all of the knowledge generated by the ITFLOWS project. It provides stakeholders with a collection of tools that will allow them to conduct simulations and forecasts on different facets of migration, ranging from the number of people projected to leave from a specific region within specified origin countries towards a country within EU, to potential tensions that may arise when migrant populations enter EU territory.

The EMT consists of two parts:

(1) the backend, which is responsible for downloading the relative data, process these data through various models developed within the ITFLOWS project and provide the frontend with outputs for simulations and predictions.

(2) the frontend, which provides the user with the visualizations of the aforementioned outputs in a useful and understandable way.

In this document, which accompanies the EMT application, we analyse the data gathering and data updating methods that were chosen. Next, there is a general overview of the models that were developed as part of the ITFLOWS project alongside an overview of the available routes in the backend server. Following, there are sections regarding the ethical and legal obligations of the EMT. We finalise this document by providing a manual for the first release of the EMT, that is relevant for the version that is live by the end month 18 of the ITFLOWS project (February 2022), a section with all the data sources EMT is utilizing and the future work needed to be done based on the feedback we received during the Users Board meeting that took place on January 31, 2022.

***Keywords**: EUMigraTool, simulation, prediction, requirements, architecture.*

Index of Figures

Figure 1: *Dataset from CKAN repository UI* 11
Figure 2: *Data pipeline.* 13
Figure 3: *Pie chart indicating the qualitative presence of each component in the LSM.* 16
Figure 4: *Topic based prediction pipeline.* 17
Figure 5: *'Inside' a topic extracted by the LDA model. Example of one LDA topic. share.* 18
Figure 6: *Small scale model construction.* 20
Figure 7: *Visualisation of conflict and camp locations for the Nigeria and Syria countries.* 21
Figure 8: *Visualisation of conflict and camp locations for the Nigeria and Syria countries.* 22
Figure 9: *Overview of the enlarged country map of Nigeria, which contains conflict zones (red circles), camps in neighbouring countries (blue circles), and other major settlements (yellow circles).* 22
Figure 10: *Overall Framework* 23
Figure 11: *Backend infrastructure* 26
Figure 12: *Preview of the Map Element that displays the Conflict Zones and Refugee Camps.* 33
Figure 13: *Layer style configuration* 34
Figure 14: *Tooltip Fields* 34
Figure 15: *Preview of the Dashboard "Asylum Seekers / Unrecognised Refugees* 35
Figure 16: *Options dashlet where the user can select the countries of interest* 36
Figure 17: *Preview of the Predictions Dashboard for February 2022* 36
Figure 18: *Sample of the csv file created for the visualization* 37
Figure 19: *Preview of the "Attitudes based on Twitter analysis" Dashboard* 39
Figure 20: *Layer Settings and Term Joins for the Map Dashlet* 40
Figure 21: *Diverging stacked bar chart displaying the negative and positive factors* 41
Figure 22: *Sample of the Vega Code used to create the diverging stacked bar* 42
Figure 23: *EMT architecture* 43
Figure 24: *Cookies collection* 46
Figure 25: *Cookies list* 48
Figure 26: *Information on the website about cookies.* 48
Figure 27: *Php code block* 49
Figure 28: *Php code block* 49
Figure 29: *Mail notification for user account application* 51
Figure 30: *User's profile page* 52
Figure 31: *Options available for user registration application* 52
Figure 32: *User Roles* 53
Figure 33: *Role form* 54
Figure 34: *EMT login page.* 71
Figure 35: *Registration form* 72
Figure 36: Confirmation message or the registration process 73
Figure 37: *Mail notification of account review* 74
Figure 38: *Mail notification account approved* 74
Figure 39: *Filling the credentials while on the Login page* 75
Figure 40: *Suggested authenticators* 76
Figure 41: *Set up 2FA* 76
Figure 42: *Download of the preferable 2FA* 77
Figure 43: *Scan the QR code using the 2FA mobile app* 77
Figure 44: *Enter the 6-digt token* 78
Figure 45: *Enter the 6-digit token* 79
Figure 46: *Logout* 79
Figure 47: *Home page* 80
Figure 48: *Dashboards page.* 81
Figure 49: *Nigeria case and Syria displacement map.* 83
Figure 50: *Selection box options with Nigeria as example.* 83
Figure 51: *Clicking the arrows button will zoom into the selected country* 84
Figure 52: *Hover information* 85

Figure 53: *Hover information with conflict data* 85
Figure 54*: Visualisations on Asylum Seekers* 86
Figure 55*: Filter application* 87
Figure 56*: Result of filter application* 87
Figure 57*: Information about the asylum seekers/ non-recognised refugees arriving to Italy* 88
Figure 58*: Example of asylum applications from Mali to France* 89
Figure 59*: Temporary prediction dashboard* 90
Figure 60: *Filter selection process on Attitudes dashboard* 91
Figure 61*: Apply changes by pressing the button in the red circle* 91
Figure 62*: Influential Attitudes Diverging Stacked Bars* 92
Figure 63: *Data sources* 93
Figure 64*: Glossary* 94
Figure 65*: Partners' page* 95
Figure 66*: Account page* 96

Index of Tables

Table 1: List of countries of origin and destination Member States, listed in alphabetical order with no relation between countries in the same row. 9
Table 2: Examples of request links 13
Table 3: List of most prominent factors influencing attitudes towards immigration 19
Table 4: List of most necessary cookies collected 47
Table 5: AI Ethical Requirements 60
Table 6**:** DPA legal and ethical recommendations 67
Table 7: IEB legal and ethical recommendations 69

Table of Contents

Executive Summary ... 3
Index of Figures ... 4
Index of Tables ... 5
Table of Contents ... 6
Abbreviations ... 8
1 Introduction ... 9
2 Architecture ... 11
2.1 Data gathering ... 11
2.2 Data Repository ... 13
2.3 Backend-services ... 14
2.3.1 Large-scale model ... 14
2.3.2 Small-scale model ... 20
2.3.3 Twitter Analysis ... 23
2.3.4 Google analytics ... 24
2.3.5 EMT server infrastructure ... 25
2.4 Frontend ... 28
2.4.1 Visualisations ... 31
2.4.2 Predictions ... 32
2.4.3 Attitudes ... 38
3 Design Approach ... 43
3.1 Data protection, privacy, and security by design ... 43
3.1.1 Backend-Processing/Migration Flow Analytics ... 44
3.1.2 User Data ... 45
3.2 User Administration System ... 50
3.2.1 Registration Process ... 50
3.2.2 Account mail notifications ... 50
3.2.3 Approve Registration ... 51
3.2.4 User Roles ... 53
4 Ethics Section ... 55
4.1 Purpose and methodology of the ITFLOW preliminary AI Impact Assessment ... 55
4.2 Specific AI ethical requirement for the EMT ... 58
4.3 Analysis of the technical measures adopted at this stage and further ethical recommendations ... 60
4.4 DPA and IEB legal and ethical recommendations ... 67
4.5 Additional legal and ethical concerns identified by WP2 ... 69
5 User Manual ... 71
5.1 User registration ... 71
5.2 Account mail notifications ... 73
5.3 First-time Login ... 75
5.4 Two factor authentication setup ... 75
5.5 User Login ... 78
5.6 Logout ... 79
5.7 Home ... 80

5.8 Dashboards 80
5.8.1 Origin Countries with conflict locations and Asylum Seekers / Unrecognised refugee camps 82
5.8.2 Asylum seekers / non-recognised refugees' arrivals in Europe Historic Data 85
5.8.3 Asylum seekers / non-recognised refugees' arrivals in Europe Predictions 89
5.8.4 Tension Function of the EMT: Attitudes and sentiment of European countries towards migration from Twitter analysis 90
5.8.5 Influential Attitudes 92
5.9 Data sources 92
5.10 Glossary 93
5.11 Partners 94
5.12 Account 95
6 Data sources 97
7 Conclusions and Future Work 99
7.1 Towards more accurate predictions 99
7.2 Ethical recommendations 99
7.3 EMT Platform development 99
7.4 Users Board Feedback 100
8 Appendixes 101
8.1 Appendix 1 – AI Impact Assessment Questionnaire 101
8.2 Appendix 2 – Eurostat datasets 110

Abbreviations

ACLED: Armed Conflict Location & Event Data Project
AIDA: Asylum Information Database
API: Application Programming Interface
CAMEO: Conflict and Meditation Event Observations
CESSDA: Catalogue of the Consortium of European Social Science Data Archives
COO: Countries of Origin
CRI: Associazione della Croce Rossa Italiana
CSA: Coordination and Support Action
EASO: European Asylum Support Office
EC: European Commission
EMT: EUMigraTool
ESS: European Social Survey
EU: European Union
EVS: European Values Study
FAO: Food and Agriculture Organization
FRONTEX: European Border and Coast Guard Agency
GA: Grant Agreement
GDP: Gross Domestic Product
GKG: Global Knowledge Graph
GQG: Global Quotation Graph
GRG: Global Relationship Graph
HDX: The Humanitarian Data Exchange
HTTP: Hypertext Transfer Protocol
ILO: International Labour Organization
IMF: International Monetary Fund
IOM: International Organization for Migration
IPC: Integrated Food Security Phase Classification
IPR: Intellectual Property Rights
ISSP: International Social Survey Programme
MVC: Model, View and Controller
NGO: Non-Governmental Organisation
OSM: OpenStreetMap
REIGN: Rulers, Elections, and Irregular Governance (
SPI: Standard Precipitation Index ()
TB: Terabyte
ToR: Terms of References
UB: Users Board
UN: United Nations
UNDP: United Nations Development Programme
UNHCR: United Nations High Commissioner for Refugees
UNODC: United Nations Office on Drugs and Crime
URL: Uniform Resource Locator
WDI: World Development Indicators
WP: Work Package
XML: Extensible Markup Language

1 Introduction

This deliverable's goal is threefold: (1) to accompany the preliminary release of the EUMigraTool (EMT) application, which can be accessed by selected end-users on the page https://www.emt.itlfows.eu/, and (2) to describe the EMT's general architecture (complementing Deliverable 6.1) and (3) to offer a roadmap for the complete system as it is up and working.

In essence, the main objective of the tool is:

- to provide accurate migration flow predictions,
- to provide adequate evidence to practitioners and policymakers involved in various stages of migration management, and
- to propose solutions for reducing potential tensions between migrants and EU citizens, by considering a wide range of human factors and utilizing multiple sources of data.

At this stage, the first version of EMT is released a fully functional by the end of project month 18 (28/02/2022). Such early version is already accessible by selected members of the Users Board. The first version of the EMT focuses on the two categories: (a) prediction and (b) attitudes identification. Whereas the prediction tool offers two interactive dashboards, the attitudes tool offers two main functions for risk of tension identification, with different graphs and charts.

As predefined in previous deliverables, EMT is focusing on 8 countries of destination and 8 countries of origin as shown in Table 1 .

Countries of Origin	Destination Countries
Afghanistan	France
Eritrea	Germany
Iraq	Greece
Mali	Italy
Morocco	Netherlands
Nigeria	Poland
Syria	Spain
Venezuela	Sweden

Table 1: *List of countries of origin and destination Member States, listed in alphabetical order with no relation between countries in the same row.*

The deliverable is structured and organised as follows:

Section 1: is the report's introduction, defining the document's scope and relationship to other deliverables.

Section 2: provides an overview of the EMT's architecture and components from the data gathering and storing to the AI models responsible for analysing the data and all the backend and frontend infrastructure.

Section 3: presents the design principles and approach including data privacy information and the administration system.

Section 4: includes a full manual of the EMT platform and how users are expected to operate it.

Section 5: describes the data sources EMT's AI modules are using to produce predictions and tension identification.

Section 6: provides a brief discussion upon the entire document and the conclusions that were drawn. It also includes what the future work for EMT is and what users can expect from the developer's team.

The ITFLOWS Independent Ethics Board and the Data Protection Advisor have reviewed this Deliverable and provided some recommendations. Such recommendations regarding the development of the EMT can be found in Section 4 of this document.

2 Architecture

The overall architecture of the EMT prediction tool - a software intensive structure - is the composition of several subordinate system modules, which comprise software elements, their externally visible attributes and properties of those elements, and the relationships among them. These **modules** are data gathering, data repository, backend and frontend services.

2.1 Data gathering

This section describes the automated functions for downloading data from data sources. Ideally, data should be updated **automatically**. To solve this problem various scripts have been created to handle the process of downloading data and pushing these data to our data repository which is CKAN. Up to the moment this document is being created, this is applicable only for Eurostat and GDELT. In the future, automated data gathering processes will be implemented for all data sources.

Python scripts collect data from GDELT, Eurostat and all the other data sources.[1] After filtering the data to our needs, translation and cleaning of certain words, data gathering uses **CKAN** as a data repository. Given these datasets, EMT web app will be able to pull the data which it needs and analyse them to make accurate predictions.

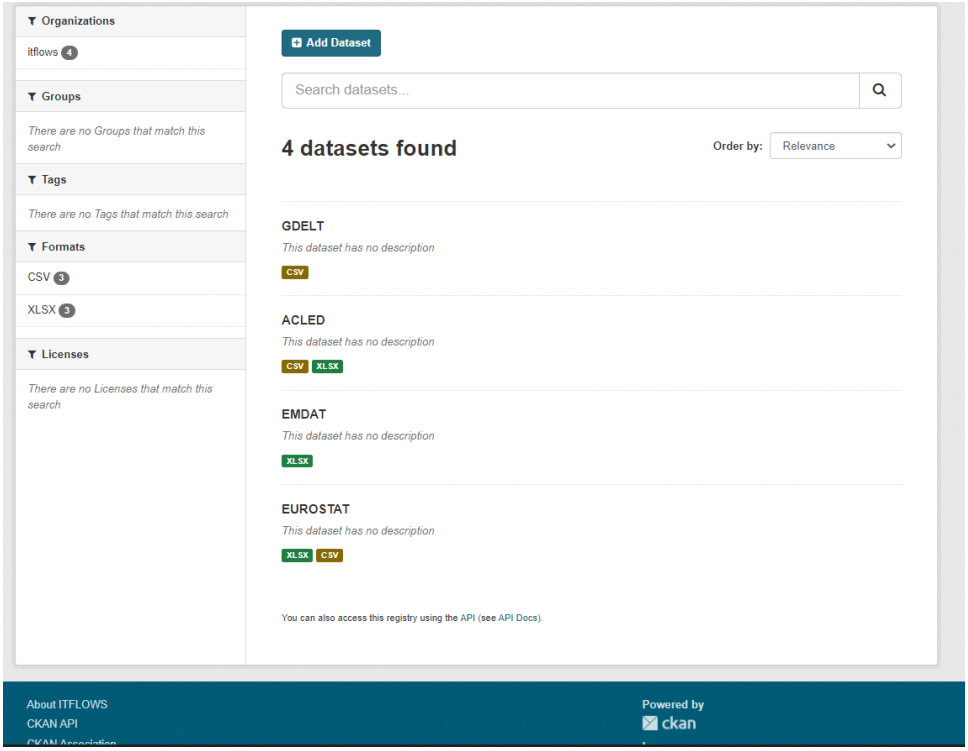


Figure 1: *Dataset from CKAN repository UI*

[1] More details about specific data sources can be found in *Section 5.9* of this deliverable (Data sources).

This script will automatically run every first day of each month, to collect the data sets from the previous months.

This python programme creates a specific URL, to collect specific data which depends on the needs of the **Large-Scale Model**. The script for GDELT, for example, encapsulates and apposes in URL data for every country, viz source country and source language to gather information for a specific country. For example, a URL from GDELT is:

https://gdelt.github.io/#api=doc&query=&sourcecountry=FR&sourcelang=fra&contentmode=ArtList&maxrecords=75&timespan=1d

The URL apart includes the source country (e.g., "Fr" is for France), source language (e.g., "fra" is for French), maximum records (75), as well as the start datetime and end datetime (1 day).

The programme consists of two parts: (1) the data gathering and (2) the data update.

- **Data gathering:**

  This part is responsible for downloading all datasets that are needed for the Large-Scale Model and potentially other models as well. It downloads the datasets from the sources in the requested format.

- **Data update**:

  This part responsible for updating the already existing datasets from data gathering every month and storing them to CKAN. That way EMT backend can pull and push datasets when needed. The Large-Scale Model will only use datasets that already reside in CKAN. The reason of this practice is twofold:

  a. It is easier and faster to monitor exactly which datasets are fed into the model, than downloading directly from the specific source.

  b. AI models sometimes require refined datasets. By firstly downloading the data, pre-process them and then upload them to CKAN, it is ensured that the Large-Scale Model only uses datasets tailored for its needs.

URL, MobileURL, Date, Title
https://www.afghanistansun.com/news/272221742/top-taliban-delegation-heads-for-oslo-talks,"","2022-01-23 00:45:00","Top Taliban delegation heads for Oslo talks"
https://www.afghanistannews.net/news/272222628/humanitarian-aid-tops-agenda-as-taliban-meet-western-officials,"","2022-01-23 07:00:00","Humanitarian Aid Tops Agenda as Taliban Meet Western Officials"

http://www.rawa.org/temp/runews/2022/01/20/taliban-launch-raids-on-homes-of-afghan-womens-rights-activists.html,"","2022-01-23 17:30:00","Taliban launch raids on homes of Afghan women rights activists « RAWA News"
https://tolonews.com/afghanistan-176425,"","2022-01-23 20:45:00","Ministry Says Schools to Open for All in Spring"
https://www.afghanistannews.net/news/272223582/talks-with-taliban-begin-in-norway,"","2022-01-23 16:45:00","Talks with Taliban Begin in Norway"

***Table 2**: Examples of request links*

Table 2 illustrates some examples of the links that are created from the "data gathering" programme, as described above, to request data that is needed from GDELT database. This dataset is then being fed into the AI modules to produce predictions for future months.

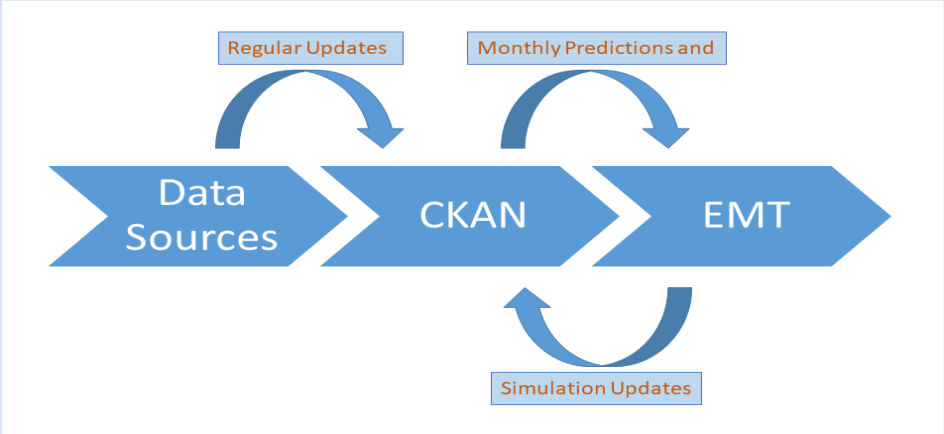


***Figure 2:** Data pipeline.*

2.2 Data Repository

The EMT uses CKAN[2] as data repository. CKAN is a robust Open-Source data portal platform that makes data more accessible by providing tools for publishing, sharing, locating, and using data more efficiently. It can be extremely beneficial to data publishers and application developers, including in the development of the EMT. The Large-Scale Model uses thus CKAN for pulling data, and it automatically updates its content via automated mechanisms implemented in EMT's servers.

More details regarding CKAN and its repository have already been provided in previous Deliverable D6.1.

[2] http://emt.itflows.eu:5000/ EMT's data repository page.

2.3 Backend-services

2.3.1 Large-scale model

The Large-Scale Model (LSM) has two main functions:

1. Producing European asylum application predictions monthly for a variety of bilateral cases, using state of the art machine learning approaches (neural network architectures and time series analysis). This is the process of incorporating techniques for the correlation analysis between raw data sources and simulation, as mentioned in the ITFLOWS Grant Agreement (GA) description.

2. Providing intuition on immigration and migration attitudes for all European destination countries using Twitter Sentiment Analysis data (FIZ) and most influential determinants of attitude towards immigration indicated by Dražanová et al. (also presented in Deliverable 5.2).[3]

PREDICTIONS

The first function includes not only several machine-learning algorithms (Long Short-Term Memory, Multi-Layer Perceptron neural networks as well as Kalman Filtering) performing predictions on asylum applicants in Europe but also various data transformation processes (e.g., headline news data downloaded in Arabic and French need to be translated into English, country names need to be removed from the headlines, daily topic shares need to be concatenated to monthly ones, etc.). The machine-learning prediction algorithms use different components, shown also in Figure 3 below, as input features:

- Topic shares based on topic modelling of national press by a Latent Dirichlet Allocation (LDA) model. (CERTH)
- Agent-based simulations of migrants' concentration in camps by Flee. (BUL)
- Asylum application forecasts based on Google Trends Analytics. (UAB)
- Asylum applicants including first-time asylum applicants and subsequent asylum applicants. (EUROSTAT)

All algorithms used for prediction are evaluated using the **R-squared (R^2) metric** in combination with the **Root Mean Squared Error (RMSE)** to decide the best performing set of

[3] Dražanová, L. (2022). Sometimes it is the little things: A meta-analysis of individual and contextual determinants of attitudes toward immigration (2009–2019). *International Journal of Intercultural Relations*, *87*, 85-97.

algorithm-component for every bilateral case (origin-destination countries pair). Nonetheless, only R-squared is taken into consideration when comparing the general performance of an algorithm across cases, since RMSE is a case dependent metric.

R-squared is a statistical measure that measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 1 scale.

- 0 represents an algorithm that does not explain any of the variation in the dependent variable around its mean.
- 1 represents an algorithm that explains all the variation in the dependent variable around its mean.

Upon comparing the LSTM and MLP networks to both traditional machine-learning algorithms, such as Linear Regression, Random Forest Regression and Support Vector Regression, and time series analysis techniques, such as Kalman Filtering and autoregressive models, the results indicate that neural networks are the best performing approach. Furthermore, a comparison between all combinations of algorithm-component led to the conclusion that the LSTM network combined with the national topic shares performs generally better. Specifically, the LSTM network had a superior performance in 42 out of 64 (8 origin countries times 8 destination countries) bilateral cases.

However, depending on the bilateral case, different components and approaches performed optimally. For example, the LSTM algorithm using monthly topic shares of the origin country's national press as input features performed very well (R^2 =0.5-0.8) in most high-traffic bilateral cases (e.g., Afghanistan-Italy, Iraq-Germany etc.) while poorly (R^2 =0.1-0.5) in all low traffic cases (e.g., Mali-Greece, Morocco-Sweden etc.) where time series analysis on historical data performed drastically better (R^2 =0.4-0.7). Moreover, in the case of Eritrea, since no available headlines were found in the GDELT database (making it impossible to produce predictions using topic shares from the LDA model) the predictions will be based only on Google Trends Analytics. In other words, since every component's performance varies across cases there is an urgent need for case-specific solutions. For that reason, all components were used in the bilateral cases they provided optimal performance and as shown in Figure 3, in 4 cases, out of the 64, the agent-based simulations data (in combination with the topic shares) improved the model's accuracy and in 10 cases time series analysis on historical data performed optimally. It is important to mention, though, that optimizing for every bilateral

case minimises the predictions' error but raises an important limitation: there is a high probability of overfitting, and a lot of time is required for every new case to be added in the EMT. Therefore, a significant improvement in the future would be to develop one generic model which is robust across cases.

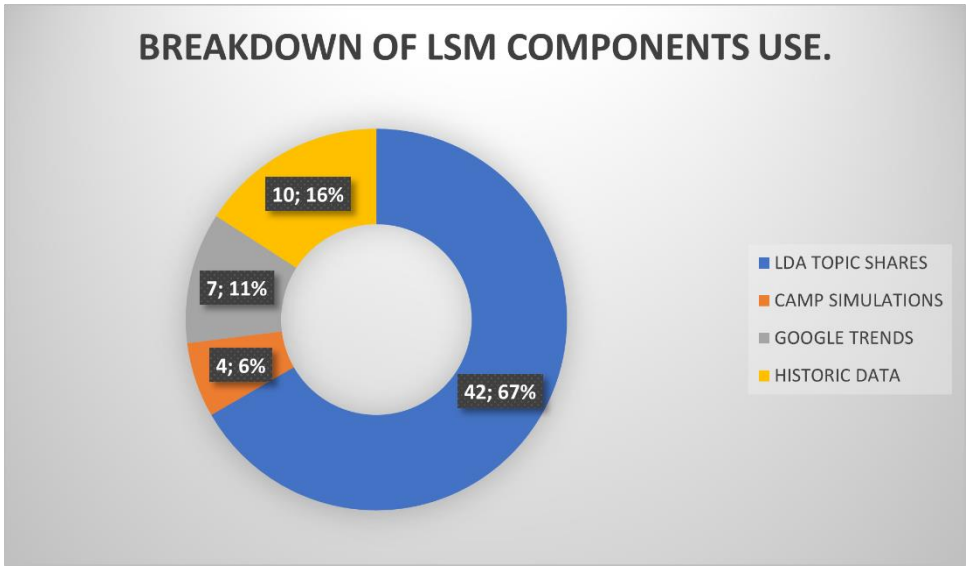


***Figure 3**: Pie chart indicating the qualitative presence of each component in the LSM.*

All data presented in the pie chart of Figure 3 are expected to slightly deviate in the future since the EMT is still under development.

National Press Topic Modelling

Topic modelling is a statistical method for discovering the abstract "topics" that occur in a collection of documents and has been used by the scientific community across a variety of different disciplines.[4] However, there are remarkably few papers utilising the method for migration prediction. Ahmed et al. presented a system combining several information sources, including the GDELT (Global Database of Events, Language and Tone) project, for mass-migration forecasting using the 2015 European refugee crisis as a case study. A very promising effort was made by the Conflict Forecast Project (Christopher et al.) where data on conflict histories together with a corpus of over 4 million newspaper articles were used in a combination of supervised and unsupervised machine learning methods to predict conflict at the monthly level in over 190 countries.

The LSM's topic classifier consists of a LDA model trained using the ALL-NEWS dataset

[4] Eickhoff, M., & Neuss, N. (2017). Topic modelling methodology: its use in information systems and other managerial disciplines.

(publicly available from Components) which features 2.7 million news articles and essays from 27 American publications. The dataset includes date, title, publication, article text, publication name, year, month, and URL (for some). Articles mostly span from 2013 to early 2020. Although, this dataset is more than sufficient for training, in terms of size, a significant limitation of it is that it contains articles from only American publications leading to slightly polarized classification topics. LDA is an unsupervised method, which means we do not, prior to the model being executed, know the relationship between the news headlines and their topics. A key aspect of LDA is to group headlines into a fixed number of topics, which must be given as a parameter when executing the algorithm. An essential process is therefore to estimate the optimal number of topics. To estimate the number of topics, a cross-validation method was used to calculate the coherence, as used in Röder et al.,[5] and it is a metric used to evaluate language models. Increasing the coherence score is identical to maximising the interpretability of the output topics. Upon comparing the coherence score of different number of topics ranging from 20 to 140, number 30 was decided to be the best performing one.

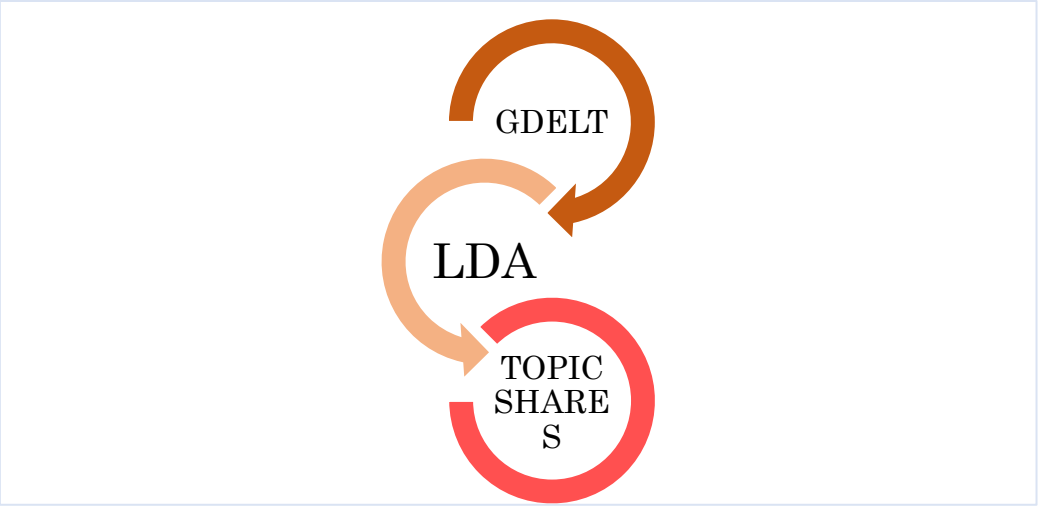


Figure 4: *Topic based prediction pipeline.*

Once training the LDA modeler is completed, the model is ready to classify text to one or more of its topics. A key process of the topic-based prediction pipeline, shown in Figure 4, is the data gathering and processing (GDELT). Approximately 250 news headlines per day are downloaded from GDELT's API 2.0 (but only from articles that include the origin-country's name at least 3-4 times, so as to narrow the noise of irrelevant articles and use the ones concerning national matters) and are classified by the LDA modeler to topic shares which are

[5] Röder, M., Both, A., & Hinneburg, A. (2015, February). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408).

then used as input to the aforementioned machine learning algorithms (LSTM, MLP). Since most of the Middle East and Africa origin countries do not use English as their native language, headlines in Arabic and French were selected for download (depending on the origin country) and they were translated into English using a Google Translator API.

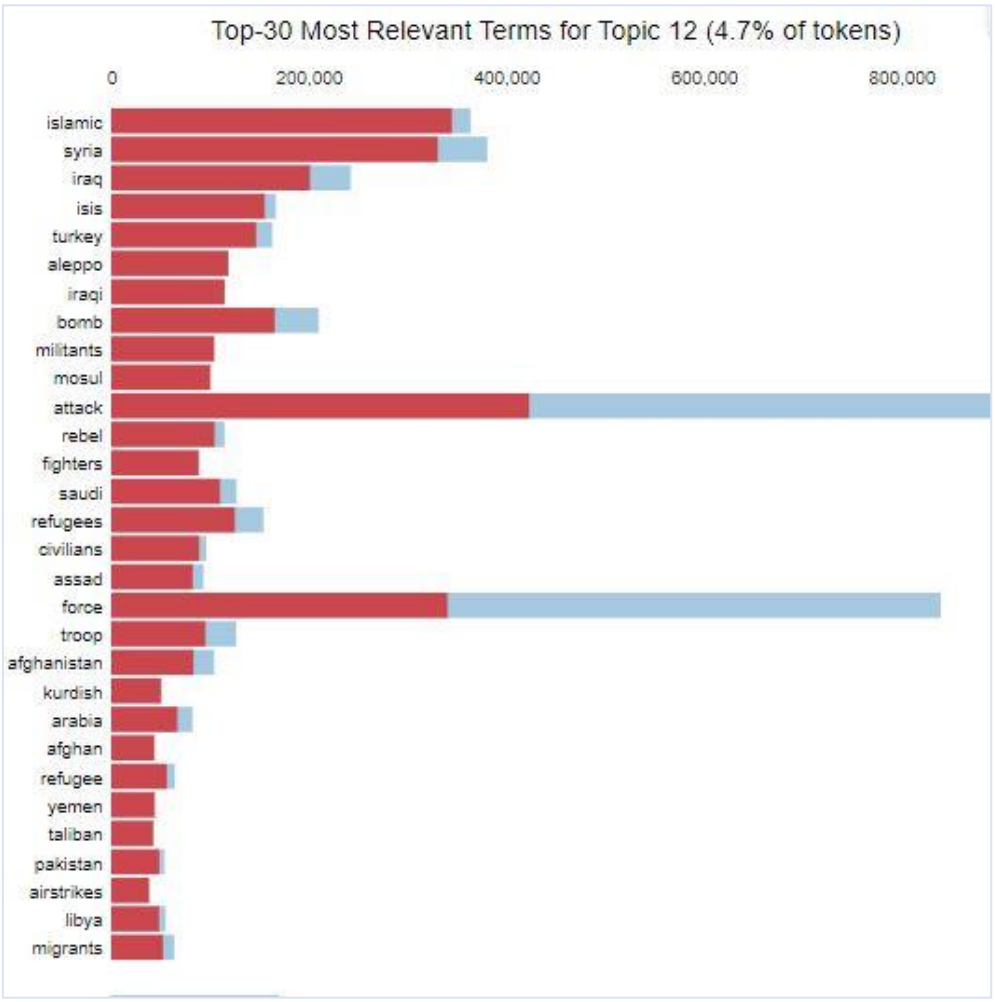


Figure 5: *'Inside' a topic extracted by the LDA model. Example of one LDA topic. share.*

Since unsupervised training is used for the LDA topic modeler, all the topics extracted are not humanly annotated. However, exploring the 30 most relevant terms of a sample topic (Figure 5 shows the example of topic number 12) reveals the human context of it. For example, topic 12 seems to be a topic of violence related to terrorism and it constitutes a 4.7% of all words(tokens) used in the training dictionary.

ATTITUDES

The second function of the LSM regarding attitudes includes the following components:

- Twitter Sentiment Analysis features (more details regarding the analysis can be found at section 2.3.3)
- Eurostat data on factors that social studies literature (Dražanová et al)[6] considers to significantly affect attitudes towards migration (See Deliverable 5.2 of ITFLOWS)

The factors that are highlighted and analysed as the most prominent in influencing attitudes to migration are shown in Table 3 below. At individual level, we concentrate on a series of sociodemographic variables such as age, education, urban share of population and being a minority, on a series of economic factors such as unemployment, GDP per capita and other individual characteristics such as the level of religiosity.

FACTOR	EFFECT ON ANTI-MIGRATION ATTITUDE
AGE	PROPORTIONAL
EDUCATION	INVERSELY PROPORTIONAL
URBAN SHARE OF POPULATION	INVERSELY PROPORTIONAL
RELIGIOSITY	PROPORTIONAL
GDP	INVERSELY PROPORTIONAL
UNEMPLOYMENT	PROPORTIONAL
MINORITY SHARE	INVERSELY PROPORTIONAL

Table 3: *List of most prominent factors influencing attitudes towards immigration.*

For all the aforementioned factors, correspondent data (EUROSTAT) are gathered and processed in the following way:

1. The median value of all European countries is subtracted from every individual country's value and the remaining values are adjusted on a scale from -1 to 1 (normalization). That way, negative values correspond to lower magnitude of the desired property while positive to higher. For example, Italy has the highest average population age (+1) while Ireland has the lowest (-1).
2. Afterwards, these values are visualized in a variety of ways (found to be intuitive for the end user) depending on the direction of the effect they have on the attitudes towards migration (positive or negative).

[6]Dražanová, L. (2022). Sometimes it is the little things: A meta-analysis of individual and contextual determinants of attitudes toward immigration (2009–2019). *International Journal of Intercultural Relations*, *87*, 85-97.

From the Twitter Knowledge Base created by FIZ (as mentioned in section 2.3.3) the large-scale model calculates the percentage of tweets expressing sentiment (hate, positive, negative, neutral sentiments) per country. At the time of writing this deliverable, period selection of this percentage is not supported but in the future a monthly period will be defined by users.

2.3.2 Small-scale model

For the small-scale model, we use the Flee agent-based modelling code (https://flee.readthedocs.io). Flee can forecast asylum-seekers/non-recognised refugees escaping violent conflicts and is able to do so even when little or no training data is available. The code focuses on migration from the country of conflict to adjacent countries; its outputs then serve to inform the input conditions for the large-scale model.

Flee is developed using Python3 and can be run either locally as a single process, or in parallel using a remote resource such as a supercomputer. It is open-source and distributed under a permissive BSD 3-clause license.

The construction and application of our small-scale simulations follows our in-house Simulation Development Approach (SDA), which is explained in detail in Suleimenova, D. et al. (2017),[7] and which is graphically shown in Figure 6:

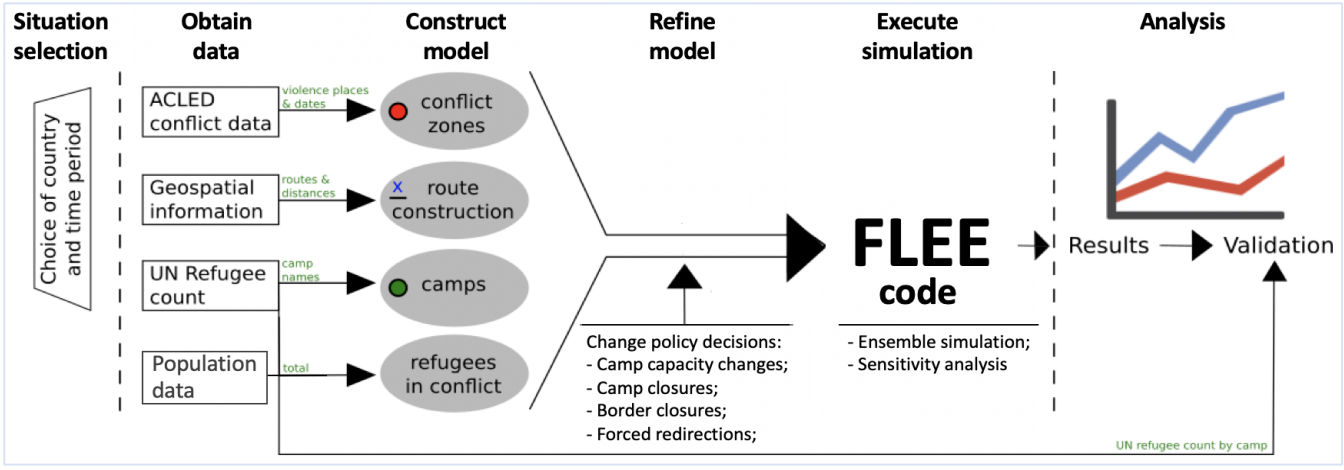


Figure 6: *Small scale model construction.*

[7] Suleimenova, D. et al. (2017) "A Generalized Simulation Development Approach for Predicting Refugee Destinations", Scientific Reports, 7:13377]

Two conflict situations are constructed and simulated using SDA and Flee for the preliminary release of EMT. The first conflict is in **Nigeria** and is modelled from 1 March 2016 - 20 April 2021 for 1887 simulation days. The second conflict is in **Syria** and is modelled from 1 January 2017 - 20 April 2021 for 1571 simulation days. The model identified conflict locations in these countries from ACLED and camp locations in neighbouring countries, which are demonstrated in Figure 7, Figure 8 and Figure 9. Currently, the simulation results are visualised using a static indication of these locations, which is going to be improved with a timeline bar. This visualisation will allow users to easily observe the predicted or forecasted distribution of asylum-seekers/non-recognised refugees arriving at camp locations.

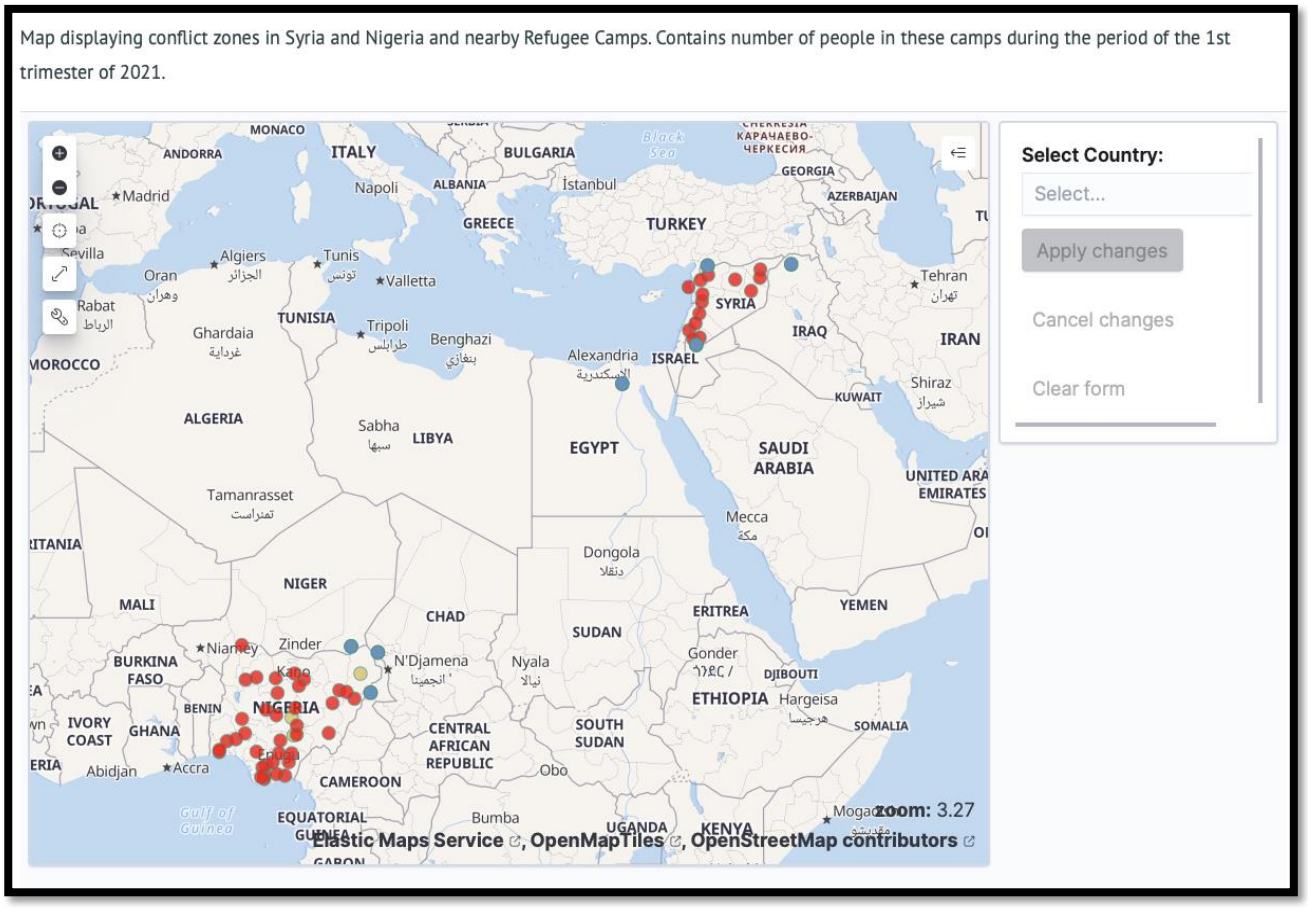


Figure 7: *Visualisation of conflict and camp locations for the Nigeria and Syria countries.*

.

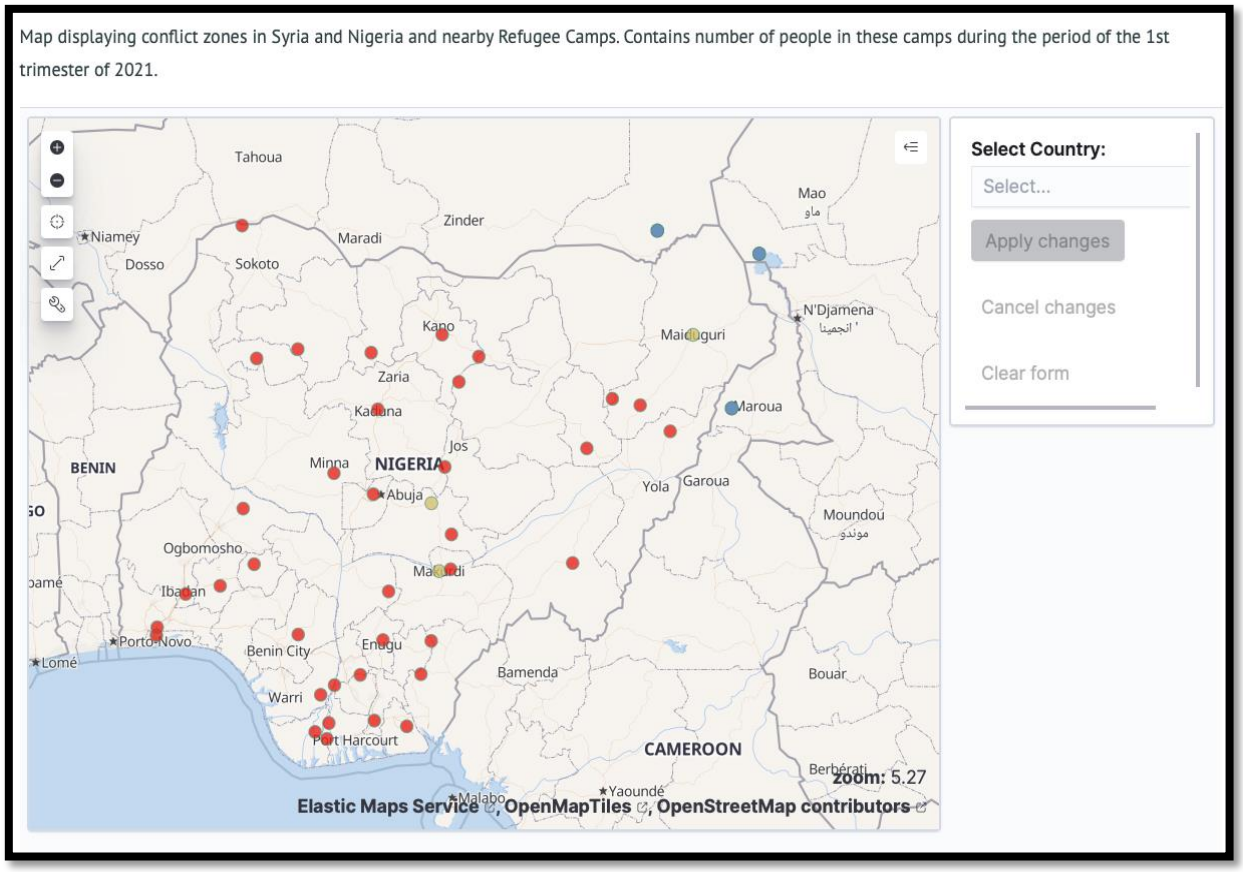


Figure 8: *Visualisation of conflict and camp locations for the Nigeria and Syria countries.*

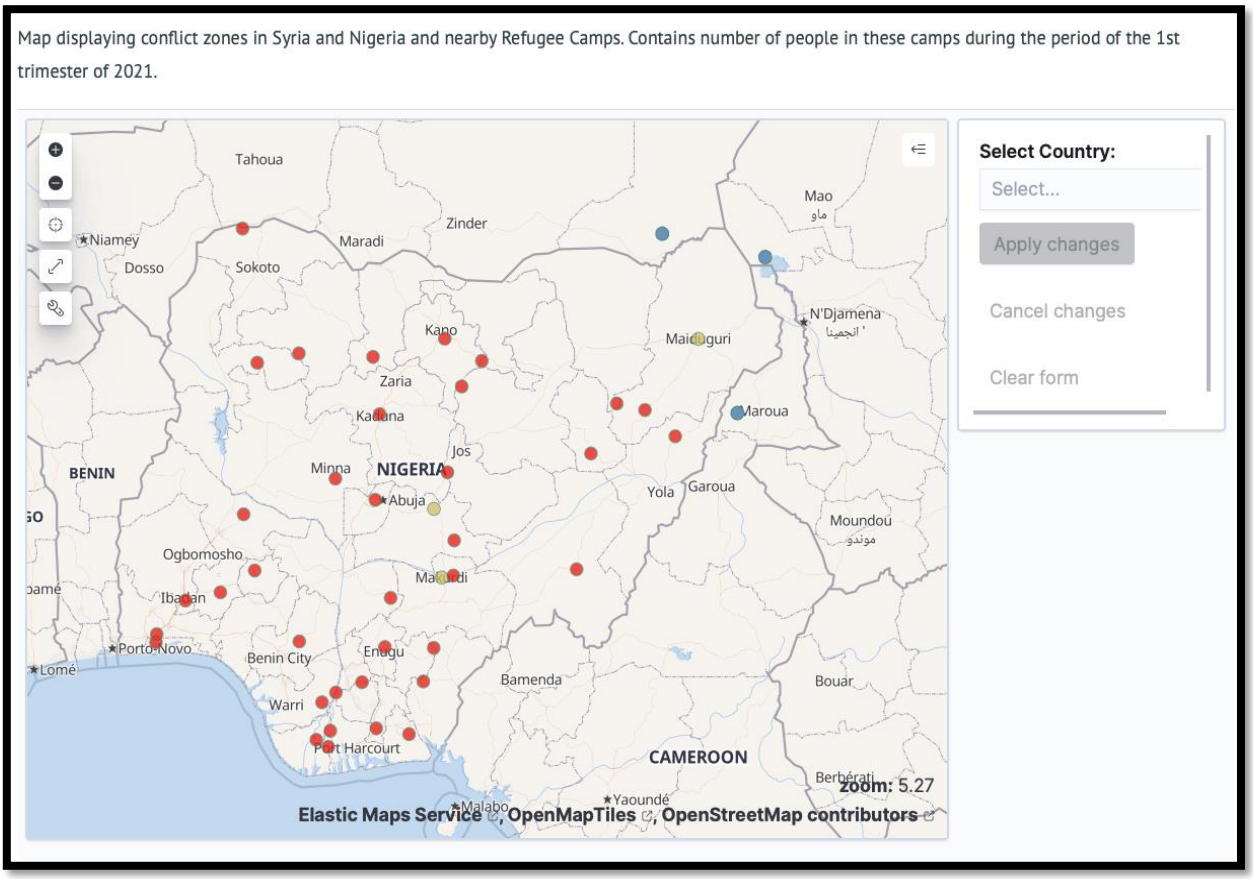


Figure 9: *Overview of the enlarged country map of Nigeria, which contains conflict zones (red circles), camps in neighbouring countries (blue circles), and other major settlements (yellow circles).*

2.3.3 Twitter Analysis

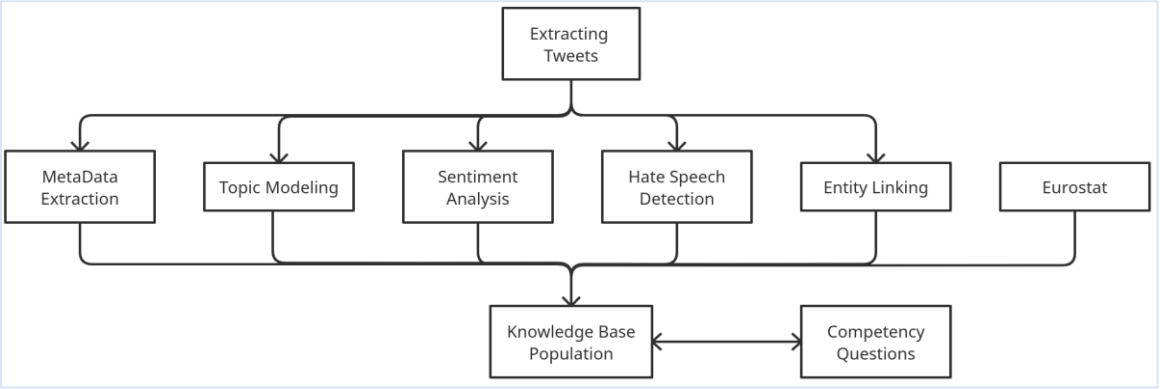


***Figure 10**: Overall Framework*

To analyse the public attitudes towards migration in the destination countries in Europe, the geotagged tweets are extracted using migration-related keywords. The irrelevant tweets are then filtered out by using state-of-the-art neural network-based topic modelling technique Embedded Topic Model.[8] It further utilises contextualised word embeddings (BERT) and transfer learning for sentiment analysis and attention-based neural networks for hate speech detection. Temporal and geographical dimensions are then explored for measuring the public attitudes towards migrations in a certain period in a certain region. To identify the potential social and economic factors driving the migration flows, external databases from sources such as Eurostat and Statista, are used to analyse the correlation between the public attitudes and the established economic indicators in a specific region in a certain period.

The overall framework for constructing a migration-related knowledge base from tweets is shown in the image above. The first step is to define migration-related keywords and perform keyword-based extraction of geo-tagged tweets along with their metadata. Furthermore, topic modelling is performed for refining the tweets by removing relevant tweets crawled in the tweet extraction phase. Contextual embeddings are then used for performing sentiment analysis. To further analyse the negative sentiments in terms of hate speech against the immigrants/refugees, tweets are further classified into three classes, i.e., hate, offensive, and normal. Finally, an analysis of factors causing negative sentiment or the hatred against immigrants/refugees is performed. To make this information queryable with the help of SPARQL queries, the knowledge base is constructed and populated with information extracted using the previously described framework.[9]

[8] (Dieng et al. 2020) Dieng Adji Bousso, Ruiz Francisco J. R., and Blei David M (2020) "Topic modelling in embedding spaces" Transactions on Association of Computational Linguistics, vol 8, P 439–453

[9] For a more detailed analysis of this task, see Deliverables 5.3 and 5.4 of ITFLOWS.

2.3.4 Google analytics

The integration of Google analytics into the EMT (particularly, in the Large-Scale model) is to follow Böhme et al.'s (2020) approach of using Google Trends search indices for migration-relevant keywords in origin countries to (i) create a real-time measure of bilateral migration intentions to EU destinations, and (ii) practically implement short-term k-step ahead forecasting of refugee arrivals to the EU using Google Trends as a leading indicator.

The data used in this task can be categorised in terms of the dependent variable (to be predicted), "classical" migration predictors (push- and pull-factors), and the Google Trends Indices. The resulting dataset features monthly frequency at the bilateral migration flow level between origin and destination countries. The dependent variable used is provided by Eurostat and captures asylum applicants as well as first-time asylum applicants, respectively.

To proxy for origin country push-factors of different dimensions, we rely on violent conflict indicators from ACLED, political events from REIGN and ELVI, agricultural production from FAO, natural disaster events from EM-DAT, labour market conditions from ILO, and macroeconomic performance from IMF. In terms of EU destination country pull-factors, we include the same data sources as for origins plus a set of detailed economic production indicators from Eurostat.

Additionally, we extract Google Trends time series for 196 different topical migration-related keywords in combination with specific EU destination country names (e.g., "visa Germany"). We cover a diverse set of more than 150 origin countries worldwide using 10 official local languages in which we extract Google Trends (i.e., Arabic, Dari, English, Farsi, French, Hausa, Pashto, Portuguese, Spanish, Turkish). The Google Trends based on the list of keywords are downloaded using the official Google API implemented in Python.
Since asylum applications are recorded as "x asylum seekers from country O filed for asylum in country D", we construct a monthly bilateral panel with economic, social, environmental and Google Trends data for each corridor (i.e., a corridor is between countries O and D). The panel includes more than 1000 variables available for forecasting. For aggregate migration flows we repeat the same procedure to construct a bilateral panel with yearly frequency.

The forecasting procedure is done corridor-by-corridor to best account for the relative idiosyncrasies that may exist between certain origins and destinations. We use AR(p) models,

ADL (autoregressive distributed lag) and the random walk as the simple univariate benchmark. We compare the performances to the Random Forest, Elastic Net and LSTM networks and preliminary results show the Random Forest model to deliver the most promising results this far.[10]

2.3.5 EMT server infrastructure

EMT's backend server consists of a variety of modules. It is mainly developed using JavaScript and Python. The backend server is running on NodeJS and ExpressJS and then there are numerous smaller programs and scripts assisting the server. Most of them are written in JavaScript and are responsible for processing and adjusting the output data files to serve the needs of the frontend server.

Wordpress is handling the frontend which is running on a nginx server and the visualizations are handled by Kibana/Elastic. Nginx[11] is a web server that can serve web content, such as Wordpress. More details can be found in *Section 2.5* of this deliverable, *Front-end*.

Data gathering features are, also, part of the backend infrastructure of the EMT. These scripts are responsible for downloading data from various sources and then pulling and pushing these data to EMT's data repository, CKAN, as explained in *Section 2.2* of this deliverable, *Data Gathering*.

All these modules are running on the same server in Docker containers. Docker creates containers suitable for each separate service offering high adjustability and performance. . It streamlines and accelerates the workflow while allowing developers to freely choose any tool, application stack, and deployment environment appropriate for the needs of the project. In terms of system resources, this implies they are far more efficient than hypervisors. Containers run on top of a single Linux instance rather than virtualizing hardware and that makes it a better choice than VMs. Last but not least, it makes it simple for developers to package, ship, and execute any application as a lightweight, portable, self-contained container that can operate almost anywhere.

[10] A more detailed analysis of Google Trends task will be provided in D3.6 on migration forecasting with high-dimensional Digital trace data.
[11] https://www.nginx.com/

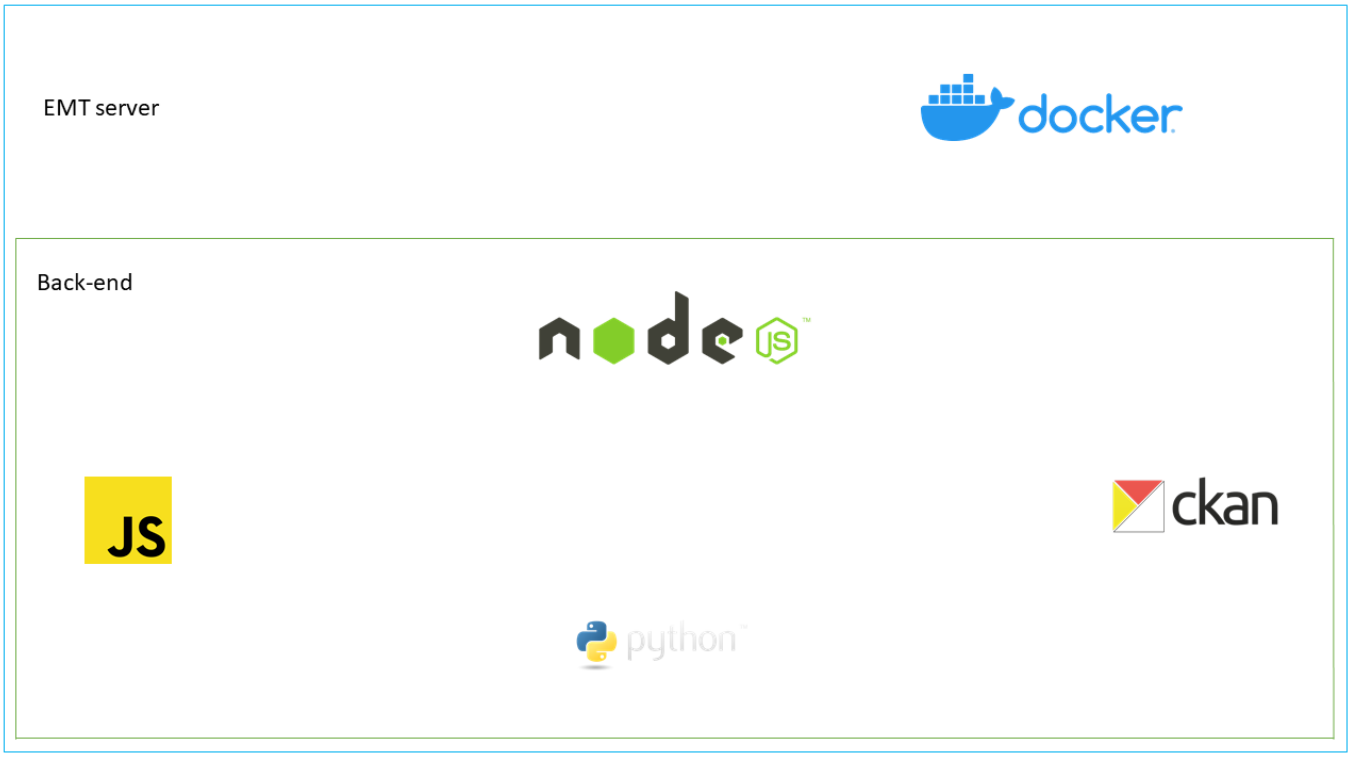


***Figure 11**: Backend infrastructure*

More information about EMT's backend infrastructure and architecture can be found in deliverable D6.1, in sections 4, 5, 6, 7 and 8.

The backend server offers all the required routes and endpoints that are needed for the frontend. The routes are categorised based on the service they offer, usually based on the data analysis model output files which is offering to the frontend. These files are output from all the models that are created for this project, including the Large-Scale Model, the Small-Scale Model, Google Trends, Twitter analysis and Influential Factors analysis for attitude towards migration. Not all described routes are currently in use from the frontend as EMT is still under development.

- **Attitude routes**

  Attitude routes include endpoints regarding the attitudes towards migration based on twitter analysis and other influential factors outside of twitter. More information regarding the Twitter analysis can be found in Section 2.3.3 of this deliverable as well as in Deliverable D6.1 of ITFLOWS. Details regarding the influential factors can be found in deliverable D5.2.

Twitter analysis attitudes based on MigrationsKB and deliverables D5.3 and D5.4:[12]

- getSentiments: returns the sentiments of European countries towards migration. Includes positive, negative, neutral and hate sentiments.
- getGdprHateSpeech: returns the GDPR indicator values and average indicator values for hate speech sorted by year and country.
- getTweetsMigrants: returns the number of tweets which contain the hashtag "migrant" or "immigration" sorted by year and number of tweets.
- getTweetsRefugees: returns the number of tweets which contain the hashtag "refugee" sorted by year and number of tweets.

Influence of various factors on attitudes towards migration. This analysis is described in Deliverable D6.1 and with more details in deliverable D5.2 :

- getInfluentialAttitudes: returns all the factors and their influence towards migration. The data are preprocessed to serve the needs of a bar diagram.
- getInfluentialSpectrogram: returns the same factors and their influence as the above endpoint but the data are pre-processed to serve the needs of a spectrogram (spider diagram).

❖ **CKAN routes**

CKAN routes include all the endpoints regarding information and details from EMT's CKAN data repository.

- getCKAN: returns all the details of the CKAN repository.
- getResourceMetadata: returns all the metadata of a specific resource/dataset that exists on the CKAN repository.
- getIndicators: returns a simple file with all the indicators of all the datasets on CKAN. This controller depends on the output of a python script responsible for getting and saving all these indicators.

[12] (Chen et al. 2021) Chen Yiyi, Sack Harald, Alam Mehwish (2021), "MigrationsKB: A Knowledge Base of Public Attitudes towards Migrations and their Driving Factors", arXiv: 2108.07593. See also Deliverables D5.3 and D5.4 of ITFLOWS.

❖ **Large-Scale Model routes**

Large-Scale Model endpoints offer prediction output files from the Large-Scale Model and some static historic data.

- getLDA: returns the output of an LDA (Latent Dirichlet Allocation) model.
- getMigrantsData: historic data of asylum seekers from Eurostat.
- getASPredictions: returns the output file with all the asylum seekers prediction from the Large-Scale Model.

❖ **Simulation routes**

Simulation routes include all the endpoints regarding the simulation on camps on neighbouring EU countries.

- getNigeriaInputLocation: returns the input location files from the simulation model for Nigeria.
- getNigeriaOutput: returns the output file from the simulation model for Nigeria.
- getSyriaInputLocation: returns the input location files from the simulation model for Syria.
- getSyriaOutput: returns the output file from the simulation model for Syria.

New routes will be implemented as the need arises. The more information EMT offers, the more routes will be needed.

2.4 Frontend

In general, when the terms Frontend and Backend are mentioned in computer science, they refer to the differences between two major parts, the presentation layer (Frontend), and the data access layer (Backend) of the software or the hardware.[13] More specifically, the Front-end of a website is the part that the users can interact with. Everything that the users see when they are navigating around the web including fonts, colours, dropdown menus etc. constitute the Front-end.

Regarding EMT's Frontend, it was designed taking into consideration the features, modules,

[13] As retrieved from I AM ON IT https://www.iamonit.de/frontend-backend/

plugins and requirements offered by WordPress, in order to create a user-friendly structure that the tool should have in the end. More specifically, these include:[14]

- **Automated Content Management System**, the web site content must be user intuitive, managed dynamically by non-technical department staff in the simplest possible way. Appropriately trained and authorised end-users shall be able to add, change and delete site content without having to manipulate any HTML or scripting code. As such, it was chosen to develop the website in WordPress CMS.
- **WordPress Advantages**, WordPress is a free and open-source Content Management System (CMS) written in PHP and paired with a MySQL or MariaDB database. Features include plugin architecture and a template system, referred to within WordPress as Themes. WordPress comes with great standard features, like easy content authoring, reliable performance, and excellent security. Below are some of the advantages that led to its selection for the EMT web site:
    - Customisable Designs.
    - SEO Friendly.
    - Responsive Mobile Sites.
    - High Performance.
    - Manage on the Go.
    - High Security.
    - Powerful Media Management.
    - Easy and Accessible.
- **WordPress Requirements**
    - PHP 7.4 or greater.
    - MySQL 5.7 or greater OR MariaDB 10.2 or greater.
    - Nginx or Apache with mod_rewrite module.
    - HTTPS support.
- **Modules and plugins.** The plugins we used are free, compatible with the WordPress version used, compatible with other plugins we used and stable with fewer bugs. Below follows the list with a brief description of modules and plugins that were taken advantage off to implement the web site:
    - **Advanced iFrame**, which includes any webpage as short code in an advanced iframe or embeds the content directly.

[14] As retrieved from WordPress Plugins | WordPress.org

- **Advanced iFrame custom folder**, which constitutes the folder of advanced iframe where custom files are stored. Advanced iframe uses the internal plugin editor to modify custom files and since WordPress 4.9.3 version, this is only possible if the files are in a "real" plugin folder.
- **2FAS Prime — Two Factor Authentication**, simple and secure token-based authentication.
- **Avada Builder**, the advanced, premium drag & drop Avada Website Builder entails the streamline of work, something that helps in order to save time for other tasks.
- **Avada Core**, Avada Core Plugin for the advanced, premium drag & drop Avada Website Builder.
- **GDPR Cookie Consent**, a simple way to show that the EMT website complies with the E-Privacy Directive as well as with the GDPR.[15]
- **Google Analytics for WordPress by MonsterInsights**, which offers the ability to see how visitors find and interact with the website. So, it helps in order to increase visibility and number of revisits. In regard with Google Analytics and GDPR the data sharing, ID function and sharing data were disabled. When it comes to MonsterInsights Google Analytics Plugin specifically, the Anonymise IP Addresses (requesting Google Analytics to anonymize the information sent by the tracker objects by removing the las octet of the IP address prior its storage) were enabled and the Demographics and Interests Reports for Remarketing and Advertising and Usage Tracking were disabled. Finally, analytics cookie is disabled by default and the user is being asked for his/her consent for whether it is enabled or not.
- **Ultimate Member**, it constitutes the easiest way to create powerful online communities and beautiful user profiles with WordPress.
- **Widget Context**, which is very helpful in order to show or hide widgets depending on the section of the site that is being viewed. It entails the configuration of the widget visibility rules under the individual widget settings.
- **WP Mail SMTP**, reconfigures the wp_mail() function to use Gmail/Mailgun/SendGrid/SMTP instead of the default mail() and creates an “options” page to manage the settings.

[15] As retrieved from European Commission Data Protection https://ec.europa.eu/info/law/law-topic/data-protection_en

- **WP Super Cache**, very fast caching plugin for WordPress.
- **WPForms Lite**, beginner friendly WordPress contact form plugin. Using the Drag & Drop form builder, specialized and unique WordPress forms are created.

Regarding user access and management, the access and use of EMT are permitted only to registered users, namely, selected NGOs' and municipalities. Other users are not allowed to access the system. Any interested user needs to go through a registration and approval process in order to be accepted and acquire access to the system. It is described elaborately on Chapter 5, sections 5.1 and 5.3 of the current deliverable.

Also, it is worth noting that during the design and development of the EMT, the guidelines derived from Web Content Accessibility Guidelines (WCAG)[16] were taken into consideration. These guidelines are developed through the W3C process in cooperation with individuals and organizations around the world, with the upper goal of providing a single shared standard for web content accessibility that meets the needs of individuals, organisations, and governments internationally. The WCAG documents explain how to make web content more accessible to people with disabilities. More specifically, concerning the implementation:

- The Avada Theme used is fully WCAG compatible.
- Text alternatives are provided for any non-textual content.
- Text content is readable and understandable with large and distinct font.
- The colour contrast between background and foreground content is great enough to ensure legibility. White background & black font on any text page, dark background & white font on text in the footer.

At this stage of the project the main idea behind EMT's design approach is to allow end-users to test the developed functionalities as well as the interface to provide to the ITFLOWS consortium, feedback and suggestions on it.

2.4.1 Visualisations

Data visualisations give the user a clear idea of what the information means by providing visual context through maps and/or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets. Therefore, visual interfaces were designed for the EMT to provide users with the best experience possible.

[16] W3C https://www.w3.org/TR/WCAG20/

To deliver the most interactive and most useful visualisations, Elasticsearch - a distributed, free, open search and analytics engine for all types of data - has been incorporated including textual, numerical, geospatial, structured, and unstructured data. Elasticsearch is the central component of the Elastic Stack, a set of free and open tools for data ingestion, enrichment, storage, analysis, and visualisation. For the visualisation of the data, Kibana was used, a free and open Frontend application that sits on top of the Elastic Stack, providing search and data visualisation capabilities for data indexed in Elasticsearch.[17] We use the features offered by Kibana and the relative plugins by enhancing them and configuring them in a way that best addresses in the identified needs and user requirements documented in the User Board meetings. The selection of this open source was due of its great visualisation activities. Thus, increasing the chances of a successful exploitation at the end of the project.

When the user visits https://emt.itflows.eu there are various Dashboards available to explore. More specifically, regarding the term dashboard, is used to refer to a data visualisation tool that tracks, analyses, and displays data, metrics, and critical data points. Dashboards can feature visualised data via charts, tables, and gauges.[18] The user can use Dashboards in order to see the information of his/her interest with graphical ways. Dashboards that are currently available (with more to come at the near future) are divided in two major categories:

- Predictions
- Attitudes

The respective visualisations of these categories are described analytically in the following sections of this deliverable.

2.4.2 Predictions

The category of "Predictions" includes the predictions made considering (1) historic data from conflict zones and data from the ongoing conflict zones and (2) the flows of migrants applying for asylum in Europe. More specifically, the Predictions category, consists of the following:

- **Origin Countries with conflict locations and Asylum Seekers / Unrecognised refugee camps**
- **Asylum Seekers / Unrecognised Refugees per destination and origin country.**

Based on the Grant Agreement (GA) description of this deliverable, one of the key functionalities of the EMT is the visual analytic tools that were developed for interactively exploring

[17] Elastic, 2022, https://www.elastic.co/
[18] Sigma https://www.sigmacomputing.com/blog/how-to-build-data-visualization-dashboards/

simulation results and collected data. The following sections describe elaborately these visualisations concerning the Predictions category.

2.4.2.1 Conflict Zones and Refugee Camps

This dashboard displays Conflict Zones, Towns and Refugee Camps for various countries in a world map. The red dots represent the Conflict Zones the yellow dots represent the Towns near those zones whereas the green dots represent the Refugee Camps on nearby countries. It also provides information about population on those Camps and the location of each point of interest.

***Figure 12**: Preview of the Map Element that displays the Conflict Zones and Refugee Camps.*

Figure 12 shows a preview of the Map Element that displays the Conflict Zones and Refugee Camps for the countries of Syria and Nigeria. This visualisation displays thus the major areas with conflict zones in Syria and Nigeria. It also displays refugee camps in nearby countries like Chad, Cameroon, Niger, Egypt etc. By selecting the country of interest, users can see the conflict zones and refugee camps on the map for this specific country. Also, by hovering on each of the circles, they can get further information like name of the region, location and population in refugee camps.

The Options Component of the dashboard is designed in Kibana by selecting the index of the map component and selecting the “Country” as method of filtering the map data so as it will display only the data of the selected country. To create this kind of visualisation Elasticsearch is used to collect the data and Kibana to create the map component. The data is imported by various endpoints that were provided by the Back-end server, including the following list:

1) http://emt.itflows.eu:8080/simulation/getNigeriaInputLocation.
2) http://emt.itflows.eu:8080/simulation/getNigeriaOutput.
3) http://emt.itflows.eu:8080/simulation/getSyriaInputLocation.
4) http://emt.itflows.eu:8080/simulation/getSyriaOutput.

The layer Style is adjusted to be the location type, so the colour varies according to the three types of points (conflict zones, towns, and refugee camps). The tooltip fields were chosen to display the Region, Name of the Location and Population (this point contains data only for refugee camps). The visibility is 0 –> 24 on Zoom levels and the opacity was set to 81%.

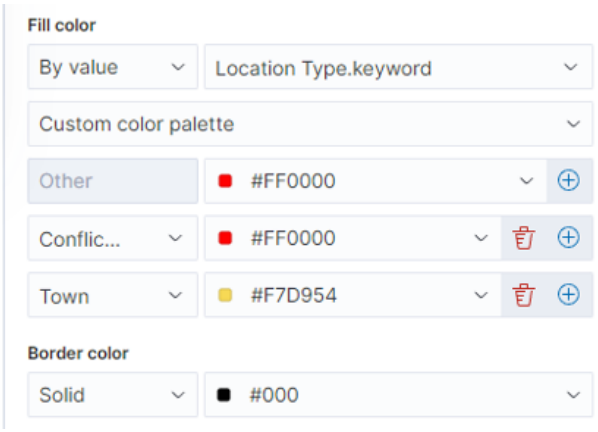


***Figure 13**: Layer style configuration*

Figure 13, shows the Layer Style and Fill Colour configuration, in order to distinct the type of locations into conflict zones (red), towns (yellow) and refugee camps (green). The border colour is black for all the points.

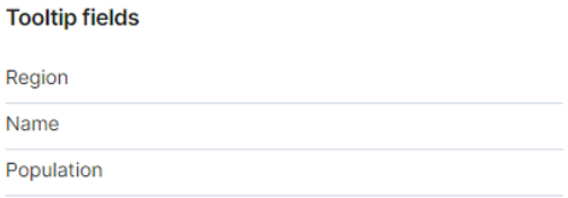


***Figure 14**: Tooltip Fields*

Figure 14 shows the Tooltip Fields to display further information (Region, Location, Population) when the user hovers onto the circles. Users can keep the information displayed for as much as they wish by right clicking on the point of interest to compare and gather information about more than one points.

2.4.2.2 Asylum Seekers / Unrecognised Refugees per destination and origin country

This dashboard displays information about asylum seekers and unrecognised refugee arrivals in the EU from other countries. It displays the number of people for each country of origin and destination along with pie charts with information about gender and age group statistics for the selected countries of interest.

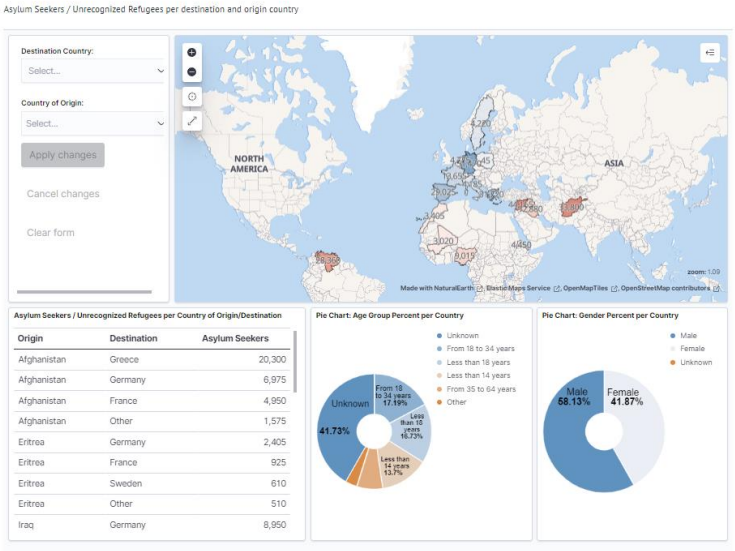


***Figure 15**: Preview of the Dashboard "Asylum Seekers / Unrecognised Refugees per destination and origin country"*

The Dashboard of the Asylum Seekers / Unrecognised Refugees per destination and origin country displays an option element to filter all the rest elements where the user can select the countries of interest. Depending on the countries the user selected, the rest of the elements are adjusted accordingly.

This visualisation displays the arrivals in European Countries. By selecting a Country of Arrival and/or Destination Country on the left side of the screen, the user can see the total flow of Asylum Seekers and Unrecognised refugees between those countries on the map (as well as the numbers) and a table element displays in detail the flow. There are also two pie charts, the first one is located on the bottom left side of the screen and displays the age groups that consist of the population of the selected country and on the bottom right of the screen the percentage of gender per selected country.

Firstly, a Filter Option was created, which selects a single country of arrival and destination country so the user can display the information for only the selected countries. After the selection, all the other reports correspond to that and display the data of the selected country, too.

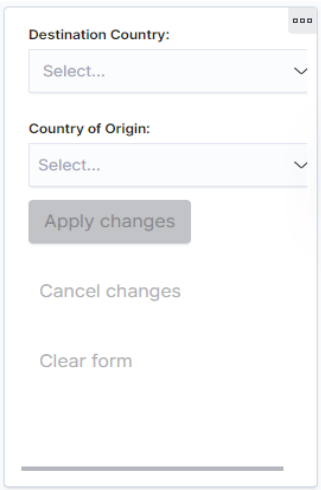


Figure 16: *Options dashlet where the user can select the countries of interest*

Figure 16 shows the Options dashlet where the user can select the countries of interest. The dashboard elements will change accordingly and only display information that refers to the select countries. At the centre of the dashboard, the user can find a map of Europe that displays the position of the selected country as well as the number of all the asylum seekers. On the bottom left of the screen there is a table report that displays the names of the countries of origin as well as the number of people that moved from each country to the selected country of interest. Lastly there are two pie charts that display the age group and age percentage of the people of the selected countries.

2.4.2.3 Predictions of Asylum Seekers / Unrecognised Refugees per destination and origin country

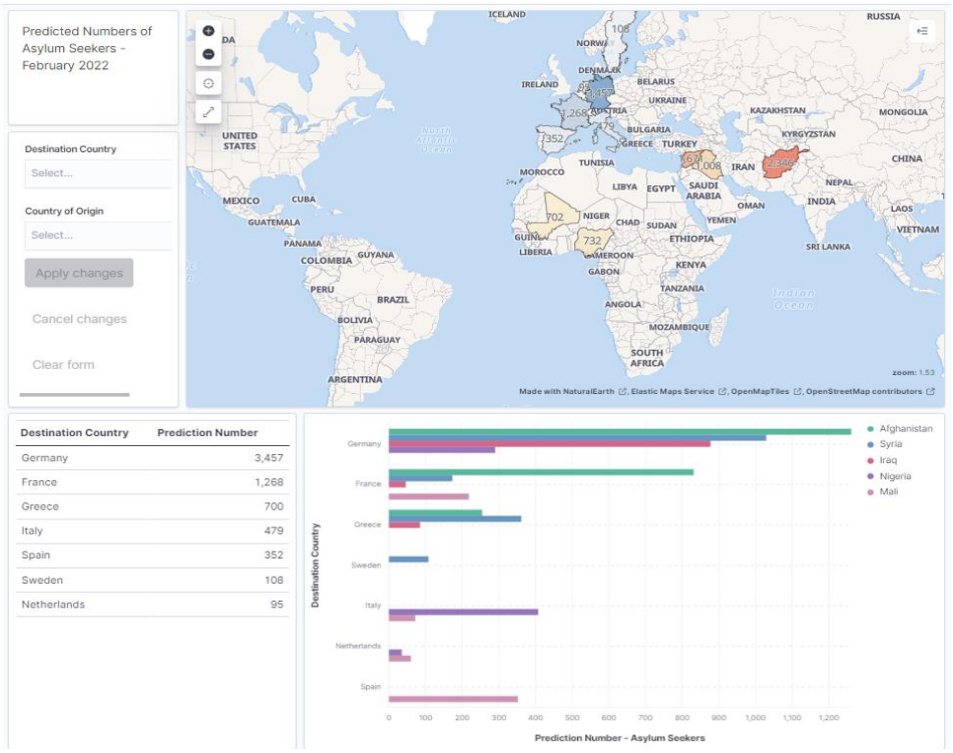


Figure 17*: Preview of the Predictions Dashboard for February 2022*

Users can access such page by clicking on ***Dashboards > Predictions > Asylum Seekers / Unrecognised Refugees per destination and origin country > Predictions.*** *This dashboard contains four elements. An option element where the user can select the destination and origin country, a map element displaying the countries, a horizontal bar chart that displays the predicted numbers of asylum seekers / country of origin and a table element that displays the numbers in a more compact manner.*

This dashboard displays information about predictions made for the number of Asylum Seekers in countries of the EU. By selecting the country of interest on the filters' element, the dashboard will adjust and display information only for the selected country. In the middle of the screen there is a map element displaying the countries – the orange graded countries are the countries of origin whereas the blue graded countries are the countries of destination. There is also a horizontal bar chart at the bottom right that displays the predicted numbers of Asylum Seekers / Country of Origin. The chart will change according to which destination country the user selects. Finally, on the left side of the dashboard there is a table element that displays the numerical values of each prediction.

To create this dashboard, we used the data provided by the backend server at the http://emt.itflows.eu:8080/largeScale/getASPredictions endpoint and by filtering them and keeping only the useful data for the visualization we created the index on Elasticsearch and proceeded with the visualisation at Kibana.

Country of Origin	Top values of Dest	Predicted Number of Asylum Seekrs
Syria	France	161
Syria	Germany	6,143
Syria	Greece	556
Syria	Sweden	86
Afghanistan	Germany	1,358
Afghanistan	Greece	405
Nigeria	Germany	283
Nigeria	Italy	509
Nigeria	Netherlands	283
Morocco	France	33
Morocco	Germany	952
Morocco	Netherlands	74
Mali	Italy	51
Mali	Spain	383
Iraq	France	31
Iraq	Greece	115

Figure 18*: Sample of the csv file created for the visualization*

As it is shown on the Figure 18, the values that were extracted and used are the country of origin, country of destination and predicted number of Asylum Seekers.

2.4.3 Attitudes

The Dashboard "Attitudes" includes attitudes towards migration by analysing Twitter data and various factors that may affect the attitude based on the analysis in Deliverable D5.2. More specifically, such Attitudes category, consists of the following sub-sections:

- **Attitudes based on Twitter analysis**
- **Influential Attitudes Chart**

The following sections describe elaborately the above visualisations concerning the Attitudes category.

2.4.3.1 Attitudes based on Twitter analysis

This dashboard displays information about how tweets that express sentiment about migration flows affect sentiment (hate, neutral, negative and positive) within the various European countries. It provides filters for the country and the sentiment, and displays a map with the tweet percentage, a bar chart with the number of tweets per year (2013-2020) and a metric element displaying the number of total tweets.

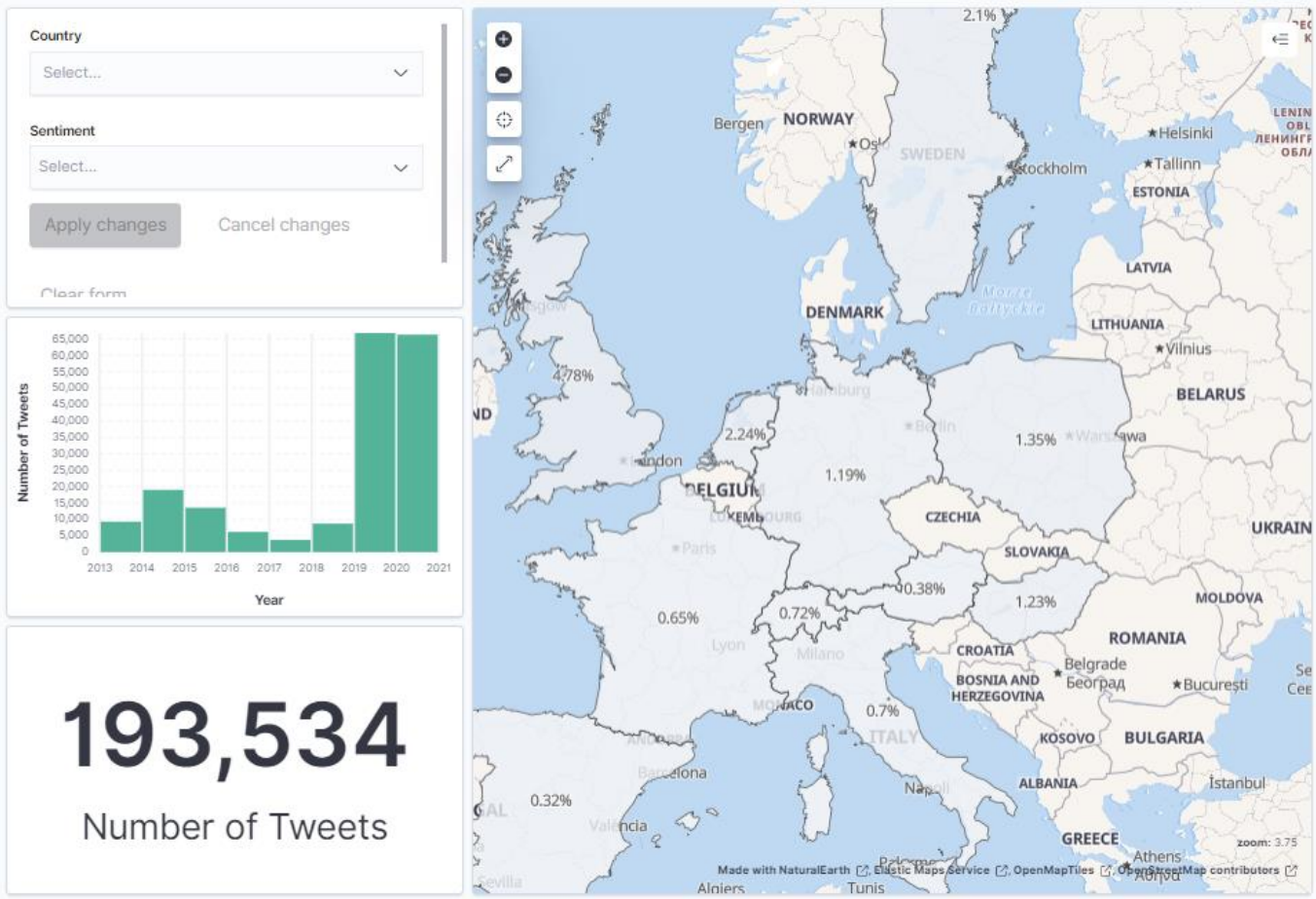


Figure 19: *Preview of the "Attitudes based on Twitter analysis" Dashboard*

Figure 19 shows a preview of the "Attitudes based on Twitter analysis" Dashboard. It consists of an Options' element that filters the dashboard by Country of Interest and Sentiment of Interest. The bar chart displays the number of tweets per year, the metric report displays the total number of tweets and the map element the percentile of each emotion (%).

This dashboard displays data and information about European Countries and the percentage of tweets expressing sentiment or emotion like hate, positive sentiment, negative sentiment and neutral sentiment. It consists of four dashlets. A Filter option dashlet where the user can select country of interest and the sentiment. Then the rest of the dashlets will correspond to those filters and display the correct data. Underneath there are two more dashlets. The first one displays a bar chart with the number of tweets of the selected sentiment and the selected country per year (2013 – 2021) – the higher the bar chart the bigger the number of tweets. On the bottom of the dashboard, there is a metric report that counts all the number of tweets after the selected filters are applied. When the user changes country and/or sentiment, the number on the metric dashlet changes accordingly.

Finally, the map dashlet displays the Percentile of each sentiment. The percentile value is 1,

the zoom levels are 0 to 24, the opacity is set to 80% and the fill colour is calculated by value (Percentage) with a grade of white to light blue as the percentage goes higher.

The data is imported by endpoints provided by the backend (/attitude/getSentiments) and filtered to keep only the data useful for the creation of the dashboard.

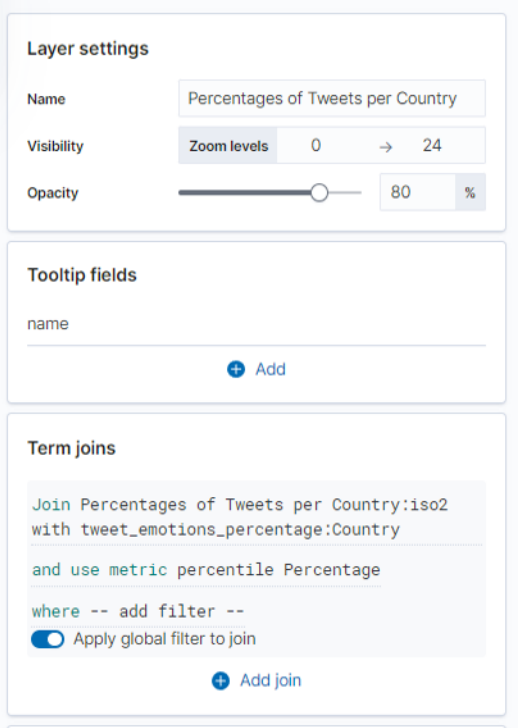


Figure 20: *Layer Settings and Term Joins for the Map Dashlet*

Figure 20, shows the Layer Settings and Term Joins for the Map dashlet. The term joins, uses a metric percentile percentage with no further filters.

2.4.3.2 Influential Attitudes Chart

This dashboard displays a diverging stacked bar with the negative and positive factors towards migration for the countries of the EU. The factors on the left side of the chart are considered negative whereas the factors on the right side of the chart are considered positive factors for a country.

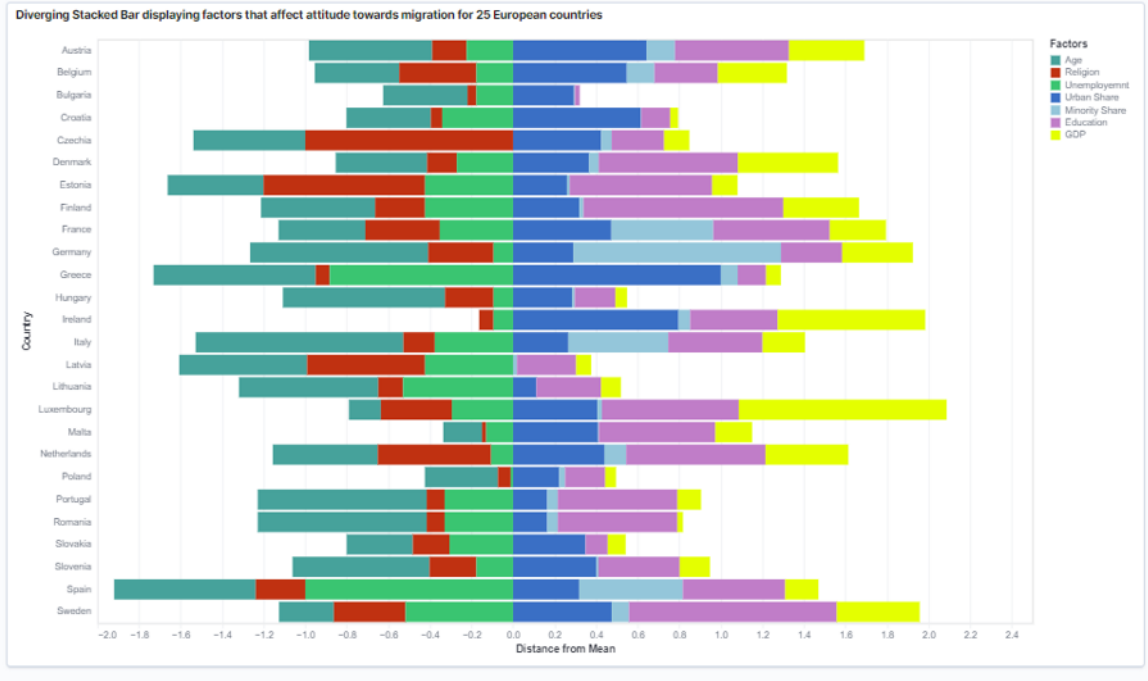


***Figure 21**: Diverging stacked bar chart displaying the negative and positive factors*

Figure 21 shows the diverging stacked bar chart displaying the negative and positive factors in the EU. The factors that are on the left side of the 0 scale of Distance from Mean are the negative factors whereas the right side displays the positive factors that affect the attitude of countries of EU.

A diverging bar chart is a bar chart that has the marks for some dimension members pointing up or right, and the marks for other dimension members pointing in the opposite direction (down or left, respectively). This diverging stacked bar chart displays the negative and positive factors for the countries of the EU. The factors that are displayed on the left side (Age, Religion and Unemployment) affect the attitudes in a negative manner whereas the factors that are displayed on the right side (Urban Share, Minority Share, Education and GDP per Capital) affect the attitudes in a positive manner. Details for this conclusion and the analysis can be found in Deliverable D5.2 of ITFLOWS. The width of each bar is equal to the distance of mean of every country. As in the case of Influential Attitudes Spectrogram the factors that have negative and positive influence are exactly the same.

The data is pulled from the endpoint provided by the Backend (/attitude/getInfluentialSpectrogram) and filtered to keep only the data useful for the creation of the diverging stacked bar.

```
{"question": "Sweden", "type":"Urban Share","percentage": 0
  .4747663551401869 }
{"question": "Sweden", "type":"Minority Share","percentage": 0
  .08342412552289187 }
{"question": "Sweden", "type":"Education","percentage": 1.0 }
{"question": "Sweden", "type": "GDP", "percentage": 0.3989453293155403}
    ]
  },
  "transform": [
    {
      "calculate": "if(datum.type === 'Age',-3,0) + if(datum.type ===
        'Religion',-2,0) + if(datum.type==='Unemployemnt',-1,0) + if
        (datum.type =='Urban',0,0) + if(datum.type ==='Minorities',1,0) +
        if(datum.type ==='Education',2,0)",
      "as": "q_order"
    },
    {
      "calculate": "if(datum.type === 'Unemployemnt' || datum.type ===
        'Religion', datum.percentage,0) + if(datum.type === 'Age', datum
        .percentage,0)",
      "as": "signed_percentage"
    },
    {"stack": "percentage", "as": ["v1", "v2"], "groupby": ["question"]},
    {
      "joinaggregate": [
        {
          "field": "signed_percentage",
          "op": "sum",
          "as": "offset"
```

Figure 22: *Sample of the Vega Code used to create the diverging stacked bar*

Figure 22 shows a sample of the Vega Code used to create the diverging stacked bar. Vega syntax is used to design the chart with every detail and import the data to fill out the diverging stacked bar chart.

3 Design Approach

The software architecture of the EMT system was determined using a hierarchical method. This technique begins with the ITFLOWS notion and analyses its basic concepts in an abstract conceptual manner. The high-level architecture is then defined by combining the conceptual modules with the system's available models and external data saved in EMT's data repository. Diving deeper into the hierarchy, the static structure of the EMT system is determined analytically by presenting the key functional modules after carefully analysing all the Use Cases and their stated end-user functional needs. This method is depicted in Figure 23 More analytical details have already been presented in Deliverable D6.1.

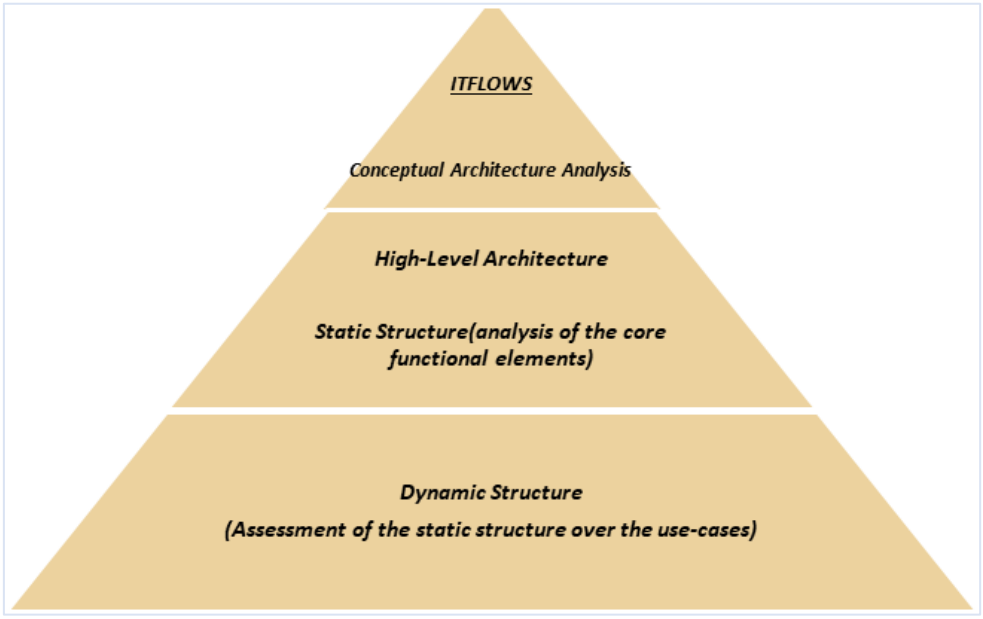


***Figure 23**: EMT architecture*

3.1 Data protection, privacy, and security by design

The development of the EMT adopts a "data protection-by-design" and a "data protection by default" approach, in line with Article 25 of the EU General Data Protection Regulation (GDPR). Information and guidelines on how to best put in place such approach have been provided by WP2 partners, in D2.1, D2.3 and D2.4 as well as during internal meetings and reviews. Technical and organizational measures are being currently designed to effectively implement the data protection principles provided by the GDRP such as transparency, accuracy, data minimization, integrity, accountability (to name a few), and ensure protection of data subjects rights.

Data protection requirements related to the EMT concern two general data flows that may contain personal data. First, the underlying processing (e.g., models, Twitter analysis) and second the user data (e.g., login data, usage collection) that is related to the individuals accessing the platform.

3.1.1 Backend-Processing/Migration Flow Analytics

First, the EMT will get input data from various sources (WP3, WP4, WP5 and WP7 tasks). Lawful and fair use of data is being ensured before using any data set. As stated in the last AI impact assessment questionnaire, *individual components may process personally identifiable information during the training phase, but this information is not passed to the EMT*. Therefore, the EMT does not use personal data in its core. *In addition, all developers ensure full anonymization of the data they use, and that no one besides them has access to these data*.

After getting the input data, it is stored in the EMT data repository, which is CKAN. When data is managed in a central repository such as CKAN the connected databases as well as the connection to this data need to be sufficiently secure. Furthermore, only data that is needed for the functionality of the EMT is extracted from the datasets and store as stated in D6.1 (Section 2.3.1).

*Direct **access** to the data is only allowed for the system administrators and maintainers.* At the time this deliverable is submitted, standard access control managed by CERTH ensures that only necessary ITFLOWS members can access the data. Technical partners please describe in depth a) how you ensure that no unauthorised access will be possible in CKAN in the future. As all online systems, EMT (and its components) is exposed to cyberattacks. To protect it, besides allowing only authorised users access to it, SoA security measures are in place, starting with SSL certificates, secured servers, firewalls and constantly updated. In addition, the system is regularly backed up in an offsite location, ensuring that in case of an attack, it will be back online very quickly with minimal to no data loss.

Mechanisms to automatically update the data as they become available are currently being developed to ensure **data accuracy**. These mechanisms will monitor the pre-defined data sources and if new data become available, they will download it and update the CKAN repository.

Data quality is ensured by using only data from verified and trusted sources (see data

sources in section 5). No anonymous or unverified data is allowed in the CKAN repository. Accuracy and bias are evaluated during the development phase of the EMT by evaluating various components based on the data currently available. Despite that, models are constantly under development and improvement. Users will be able to provide feedback on the quality of the results in the form of surveys. This will allow the technical team to improve the models. The technical team will have real-time information when such feedback is received, to act upon if needed. The feedback results will be publicly available in statistics form (fully anonymised) through the EMT website. Morover, predictions and simulations accuracy is being evaluated from the development stages of the EMT's various modules. Most of EMT data comes from trusted sources (e.g., Eurostat, UNHCR, etc.). Besides this, EMT also uses data from social media (Twitter) where quality could fluctuate.

Concerning **transparency**, *most of the EMT's modules are developed by research and academic partners, and their design, functionality and results have been published in scientific journals or conferences, thus are publicly available for scrutiny. Moreover, design details will also be provided within the EMT webpage to allow the users insight on the modules. In some cases, the source code of the modules could be available for more detailed monitoring.*

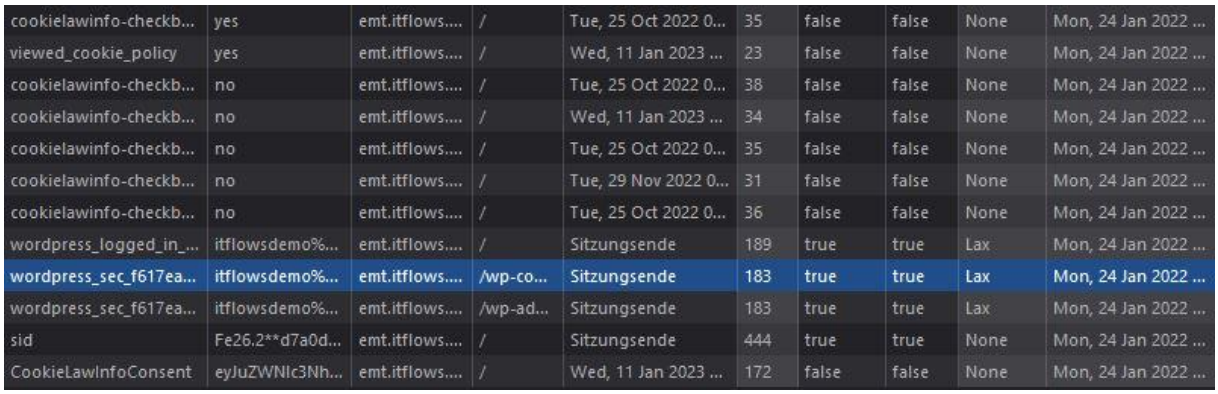
The EMT does not make decisions, but it provides suggestions and insights for the human users that make decisions. *The EMT will include explainable features in its results (i.e., explanations as to why this result was produced) in a comprehensive and not too technical language. All the details of the EMT's functionalities and models will be in the documentation pages on the website. Limitations and possible shortcomings will also be listed.*

3.1.2 User Data

The EMT is designed as a web platform that enables users to access the information generated in the underlying data analysis (see above). To this end, the EMT processes user information to provide users with a compelling and understandable frontend. To do so, various "out-of-the-box" visualisation approaches are implemented on the frontend making use of third-party software/libraries from Elastic.[19] The frontend is based on a WordPress-instance.[20] Both approaches are well established, subject to regular maintenance and fulfil state-of-the-art security requirements. Nevertheless, the EMT frontend should be designed in a privacy

[19] https://www.elastic.co/de/
[20] https://wordpress.org/

preserving way. This requires that only necessary data are collected/stored by the responsible party. The use of "out of the box solutions" often comes with a downside that the "default" settings result in the broad collection of information from the user. For example, Kibana collects usage data by default.[21] Unnecessary data collection needs to be disabled. During the research and development phase, the usage collection is likely to be a bit broader. Currently, the EMT collects the following information for the purposes described below



cookielawinfo-checkb...	yes	emt.itflows....	/	Tue, 25 Oct 2022 0...	35	false	false	None	Mon, 24 Jan 2022 ...
viewed_cookie_policy	yes	emt.itflows....	/	Wed, 11 Jan 2023 ...	23	false	false	None	Mon, 24 Jan 2022 ...
cookielawinfo-checkb...	no	emt.itflows....	/	Tue, 25 Oct 2022 0...	38	false	false	None	Mon, 24 Jan 2022 ...
cookielawinfo-checkb...	no	emt.itflows....	/	Wed, 11 Jan 2023 ...	34	false	false	None	Mon, 24 Jan 2022 ...
cookielawinfo-checkb...	no	emt.itflows....	/	Tue, 25 Oct 2022 0...	35	false	false	None	Mon, 24 Jan 2022 ...
cookielawinfo-checkb...	no	emt.itflows....	/	Tue, 29 Nov 2022 0...	31	false	false	None	Mon, 24 Jan 2022 ...
cookielawinfo-checkb...	no	emt.itflows....	/	Tue, 25 Oct 2022 0...	36	false	false	None	Mon, 24 Jan 2022 ...
wordpress_logged_in_...	itflowsdemo%...	emt.itflows....	/	Sitzungsende	189	true	true	Lax	Mon, 24 Jan 2022 ...
wordpress_sec_f617ea...	itflowsdemo%...	emt.itflows....	/wp-co...	Sitzungsende	183	true	true	Lax	Mon, 24 Jan 2022 ...
wordpress_sec_f617ea...	itflowsdemo%...	emt.itflows....	/wp-ad...	Sitzungsende	183	true	true	Lax	Mon, 24 Jan 2022 ...
sid	Fe26.2**d7a0d...	emt.itflows....	/	Sitzungsende	444	true	true	None	Mon, 24 Jan 2022 ...
CookieLawInfoConsent	eyJuZWNlc3Nh...	emt.itflows....	/	Wed, 11 Jan 2023 ...	172	false	false	None	Mon, 24 Jan 2022 ...

***Figure 24**: Cookies collection*

Cookies Privacy Overview

The website uses cookies in order to improve user experience while he/she navigates through the website. Out of these, the cookies that are categorized as necessary are stored on user's browser as they are essential for the working of basic functionalities of the website. Also, are used third-party cookies that help in order to analyse and understand how the users use the website. These cookies will be stored in user's browser only with his/her consent. The user will also have the option to opt-out of these cookies. But opting out of some of these cookies may affect his/her browsing experience. We used the plugin of the company webtoffee[22] in order to be cookie consent.

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.

Cookie	Duration	Description

[21] https://www.elastic.co/guide/en/kibana/master/telemetry-settings-kbn.html
[22] webtoffee, 2022 https://www.webtoffee.com/product/gdpr-cookie-consent/

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 are 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.

***Table 4**: List of most necessary cookies collected*

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 analyse 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 customised ads.

Others (optional)

Other uncategorized cookies are those that are being analysed and have not been classified into a category as yet.

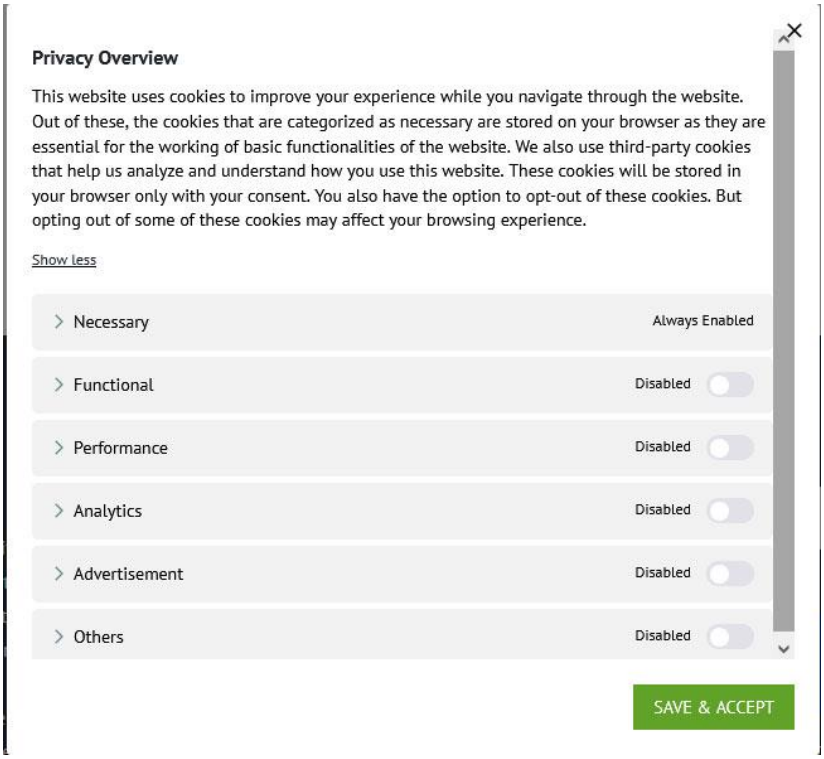


***Figure 25**: Cookies list*

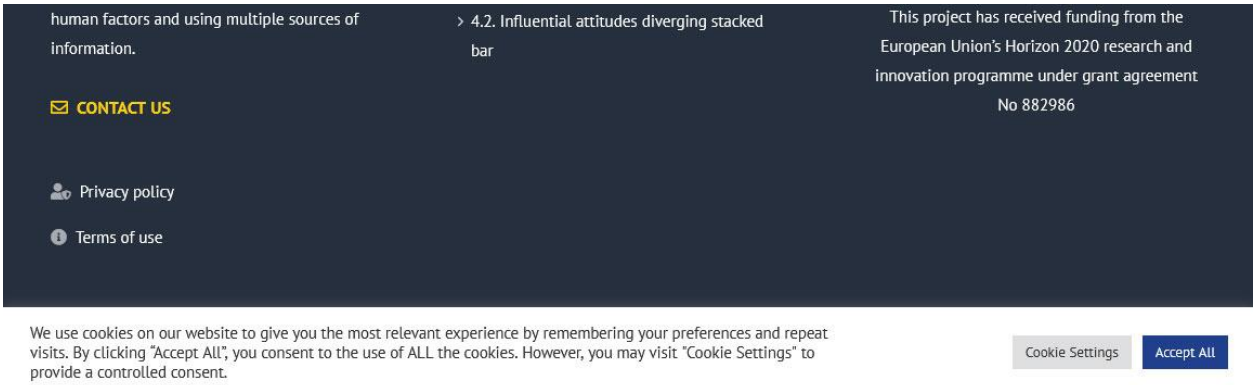


***Figure 26:** Information on the website about cookies.*

Figure 26 shows the information of the website about cookies. The user can choose the ''Cookie Settings'' button in order to see the Cookie list of Figure 25.

```
        "data": {
                "post_id": 409,
                "user_id": 3,
                "status": "visited",
                "user_ip": "87.189.***.***",
                "display_name": "itflow****",
                "city": "Han***",
                "post_type": "post",
                "referer": "https://emt.itflowsitflows-ITFLOWS.eu/",
                "month_year": "012022",
                "visit_date": "2022-01-24 10:46:21",
                "campaign_id": "16426709*****"
        },
        "id": 306
}
```


***Figure 27**: Php code block*

The admin-ajax. php file (Figure 27) 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 without reloading it, thus making it dynamic and interactive to the users.[23]


```
{
        "report": {
                "application_usage": {
                        "short_url_redirect-main": {
                                "appId": "short_url_redirect",
                                "minutesOnScreen": 0.009066666666666667,
                                "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": 2,
                                "type": "loaded"
                        }
                },
                "userAgent": {
                        "kibana-user_agent": {
                                "appName": "kibana",
                                "key": "kibana-user_agent",
                                "type": "user_agent",
                                "userAgent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64;
rv:96.0) Gecko/20100101 Firefox/96.0"
                        }
                }
        }
}
```


***Figure 28**: Php code block*

[23] (Kinsta, 2022)

Two plugins are used for reporting «**User Activity Tracking and Log**». These plugins are the following:

1. Activity Log

   Monitor and track EMT website activity.

   • If someone is trying to hack the site.

   • When a post was published, and who published it.

   • If a plugin/theme was activated/deactivated.

   • Log suspicious admin activity.

   • Securing the site by tracking log of all user activity

   Export to CSV: Export Activity Log data records to CSV.

   Data Privacy and GDPR Compliance: Tools to help the site be GDPR compliant, including Export/Erasure of data via the WordPress Privacy Tools.

2. User Activity Tracking and Log

   Track user activity & duration on the website. This plugin is especially useful for tracking users on membership sites. **Page** & **posts** visits are tracked. The plugin stores no cookies on users' computers and therefore requires no cookie opt-in from users.

3.2 User Administration System

3.2.1 Registration Process

As mentioned before, during the registration process of the user, the request of their registration is sent through the system to the administrator. The administrator will be notified via email that there is a pending request and in turn will activate the account after reviewing that all the user details meet the predefined requirements.

3.2.2 Account mail notifications

At the same time, an email is received from the system administrator. This email will inform them that a user has just applied for an account and is waiting to be reviewed. The email also contains all the details and information of the application.

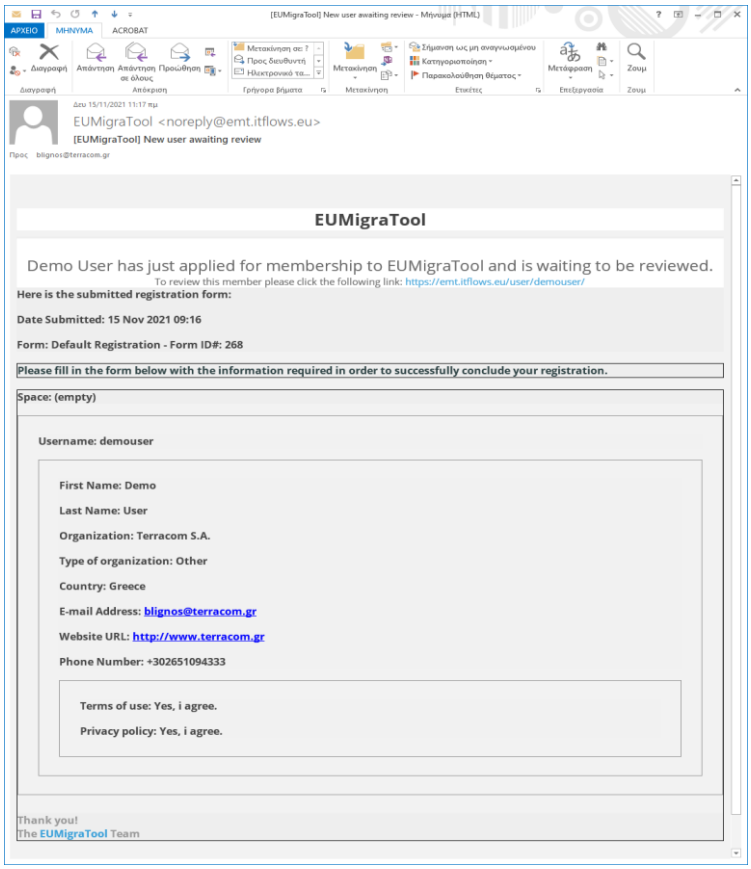


***Figure 29**: Mail notification for user account application*

Figure 29 shows the notification sent to the system administrator, notifying them about a user account application.

3.2.3 Approve Registration

By clicking on the "**review link**" the system administrator will be transferred to the users' profile page where they will be able to approve or not the registration request.

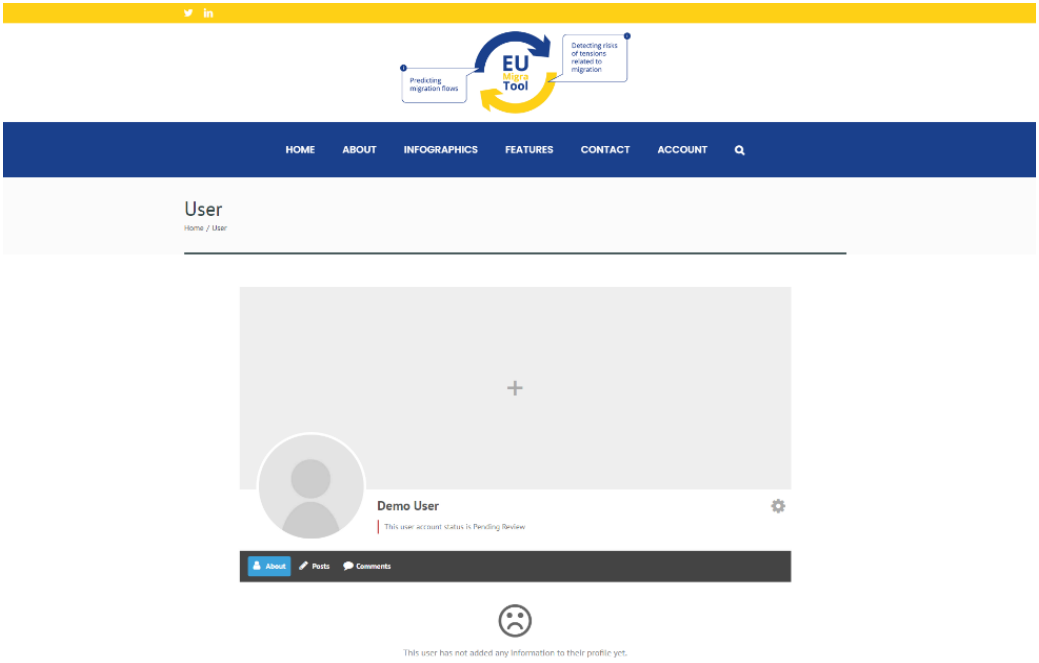


***Figure 30**: User's profile page*

Figure 30 shows the User's profile page inside the environment of the EMT.

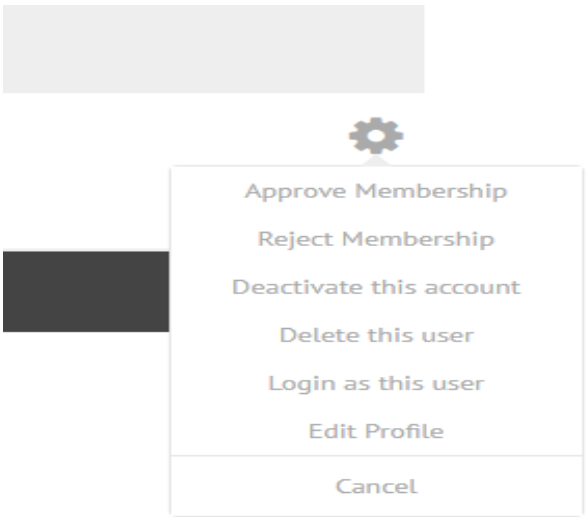


***Figure 31:** Options available for user registration application*

As it is depicted above, the system administrator can choose between other options too when it comes to user registration application (Figure 31). More specifically, there are also the following options:

- **Reject registration.**
- **Deactivate account.**
- **Delete account.**

- **Edit profile.**
- **Login as this user.**
- **Cancel.**

After the approval of the registration request, the administrator gives the user permission to login to EMT respectively.

3.2.4 User Roles

Although users might be approved to be registered to EMT, this does not mean that they have the role assigned automatically. More specifically, from the WP dashboard, in the User section, the system administrator can choose or change the role of the user accordingly.

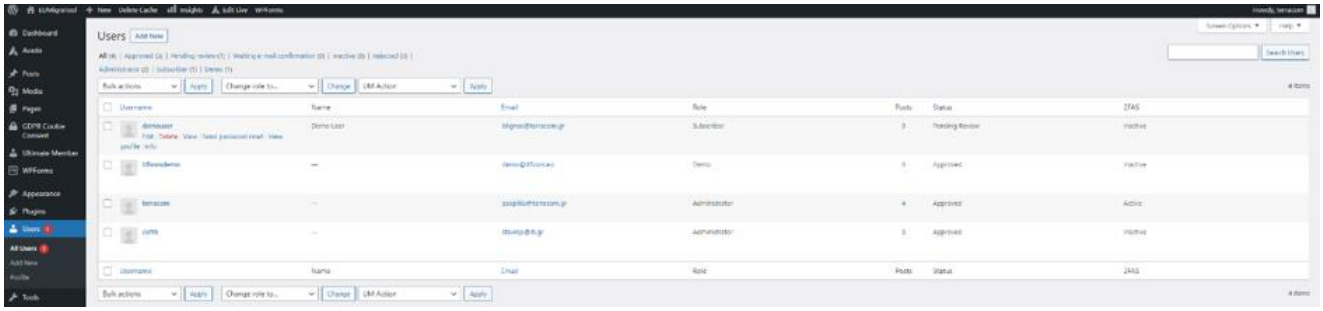


***Figure 32**: User Roles*

Figure 32 presents the process of changing the user role by the system administrator through the WP dashboard.

EMT has 3 main Role Groups for the users. These roles are the following:

- ITFLOWS users,
- NGOs,
- Municipalities.

It is worth noting that every role has different access permissions in the EMT tool.

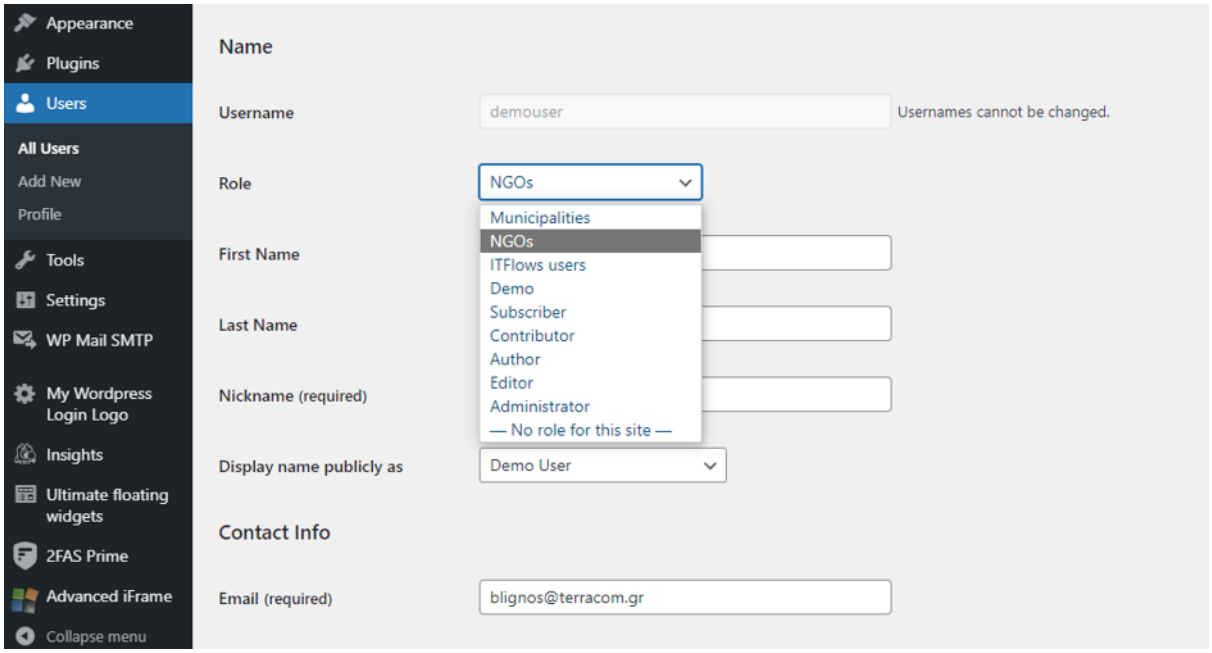


***Figure 33**: Role form*

Figure 33 shows the user form, in which the user administrator can edit user's information. After the system administrator approves the registration, an email will be sent to the user informing them that their account is approved and has access to the EMT as mentioned before.

4 Ethics Section

4.1 Purpose and methodology of the ITFLOW preliminary AI Impact Assessment

The purpose of the preliminary AI Impact Assessment was to identify and assess at this stage of the project (M14), the ethical risks posed by the EMT to ultimately minimise them. The identification and assessment of ethical risks were conducted on the basis of the Ethics Guidelines on Trustworthy Artificial Intelligence of the High-Level Expert Group on Artificial Intelligence of the European Commission (HLEG)[24], the Assessment List for Trustworthy Artificial Intelligence for self-assessment of the HLEG[25], and the Ethically Aligned Design guidelines developed by the IEEE[26].

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:**[27] "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

[24] https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai
[25] https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
[26] https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf
[27] 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

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 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:**[28] "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

[28] 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

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 those 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 above, the questionnaire sent to ITFLOWS technical partners (see **Appendix 1** below) 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.

4.2 Specific AI ethical requirement for the EMT

ID	AI Ethical Requirements
Human agency and oversight	
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-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 poten-

	tial 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 it 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 number 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-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-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-Req22	The composition of the data sets used in 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-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-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 5: *AI Ethical Requirements*

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

Human agency and oversight

The following technical explanations and measures 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 ethical recommendations 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 technical explanations and measures 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. For instance, SSL Certificates, secured servers, firewalls, 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 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 in migrants' placements. End-users should use the EMT outputs as guidelines and consulting 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 statistics form (fully anonymised).

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 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 technical explanations and measures 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 fully 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 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 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 collections 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 checking 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. EMT's modules are developed by research/academic partners. Their design, functionality and results have been published in scientific journal/conferences, 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 feature 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 ethical recommendations 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 a 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 technical explanations and measures 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 towards 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 technical explanations and measures 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 misinterpreted 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 ethical recommendations 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.

4.4 DPA and IEB legal and ethical recommendations

ID	Recommendations - Data Protection Advisor
Human agency and oversight	
DPA-HUM-Rec1	Provide clarifications on who provides guidance to support the decision-making process regarding unconscious bias.
DPA-HUM-Rec2	Provide clarifications on explainability features of the EMT and its outputs/results.
Technical robustness and safety	
DPA-TECH-Rec3	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	
DPA-PRI-Rec4	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.
DPA-PRI-Rec5	Quality checking of the data sources fed into the EMT (e.g., Eurostat, Frontex, UNHCR, etc.) is required.
Diversity, non-discrimination and fairness	
DPA-DIV-Rec6	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	
DPA-SEW-Rec7	Societal impacts of the EMT should also consider allocation of financial, human resources and opportunity costs.
Accountability	
DPA-ACC-Rec8	Provide clarifications on how the risk of misuse is mitigated.
DPA-ACC-Rec9	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).
DPA-ACC-Rec10	Provide clarifications on how the EMT will be audited. An EMT Auditing Plan should be put in place.

Table 6: *DPA legal and ethical recommendations*

ID	Recommendations – Independent Ethics Board
Human agency and oversight	

IEB-HUM-Rec1	Provide further information about the EMT explainability features. How will/is EMT designed to provide explainability
IEB-HUM-Rec2	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.
IEB-HUM-Rec3	Provide clarifications on the selection criteria followed for the datasets fed into the EMT. Provide clarifications on how “no-bad quality” data is used.
IEB-HUM-Rec4	EMT training must also cover appropriate (fair/ transparent/explainable) decision-making based on outputs.
Technical robustness and safety	
IEB-TECH-Rec5	Negative societal impacts due to malfunctions of the EMT: end-users’ awareness-raising measures about how and when actions would be taken and by whom to minimise such negative societal impacts.
IEB-TECH-Rec6	Provide clarifications on how bias and accuracy of the EMT at the design phase is evaluated.
IEB-TECH-Rec7	Provide clarifications on who/what warns technical partners about the need for additional data.
IEB-TECH-Rec8	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	
IEB-PRI-Rec9	Potential data that may identify natural persons will be used by individual EMT components during the training phase. Provide clarifications on what is meant with training phase.
IEB-PRI-Rec10	Provide clarifications on how data is going to be used/processed and mitigation measures to reduce the potential misuse of data during and beyond the lifespan of the project.
IEB-PRI-Rec11	Provide clarifications on how indirect use of data via a model will fix bad quality.
Transparency	
IEB-TRA-Rec12	Provide clarifications on the implementation of the explainability principle within the EMT.
Diversity, non-discrimination and fairness	

IEB-DIV-Rec13	Provide clarifications on how technical partners have assessed and acknowledged limitations related to the composition of the used data sets should be provided.
IEB-DIV-Rec14	Provide clarifications on how diversity and representativeness are ensured during the data assessment process.
IEB-DIV-Rec15	Provide clarifications on how end-users' feedback will be used to enhance/develop the EMT.
Societal and environmental well-being	
IEB-SEW-Rec16	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	
IEB-ACC-Rec17	Provide clarifications on the implementation of mitigation measures in the EMT to minimise the risk of misuse.

Table 7: IEB legal and ethical recommendations

4.5 Additional legal and ethical concerns identified by WP2

Legal

A glossary of common terms is a necessary component of the EMT for a number of reasons. Most importantly, a glossary ensures that used terms reflect their legal and/or other scientific meaning, according to the discipline they stem from. That way, the risk of using a term in the wrong way which could jeopardise the accuracy of the outputs of the EMT is reduced. Moreover, the use of a glossary ensures consistency across consortium, interdisciplinary, mutual communication and learning. This glossary is currently under development, and can be found here: https://emt.itflows.eu/data/glossary/

Gender

As part of its monitoring structure, the Gender Committee has kept informed of EMT development and explained its concerns and recommendations regarding the EMT in bi-weekly meetings with technical partners, with certain recent contributions of particular note and reemphasised in recommendations here:

1. **Always disaggregate when possible**: Firstly, although it has been recognised in discussions with technical partners that limitations exist in the data sets, and it is recognised that there are various gaps in data on migration to the EU, it **is again encour-**

aged 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 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.

2. **Establish clear terminology:** The Gender Committee has also **encouraged and supervised that an understanding of "Gender and sexuality" be noted in the glossary page** of the EMT, which contextualises 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.

3. **Maintain gender balance in EMT Users Board**: Finally, the 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.

4. **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.

5 User Manual

Please note that the EMT is still under development and will continue to be updated even after the preliminary release. That being said, visualizations and dashboards may change and there may be inconsistencies with this manual. This manual is submitted as part of deliverable D6.2 accompanying the EMT application, and it is accurate for the version released on the end of month 18 (February 2022).

5.1 User registration

As explained above, in order to be registered to the EMT the user must follow a specific pathway. First of all, users request their registration through the system. Then, the administrator will be notified via email that there is a pending request and will activate the account after reviewing that all the user details and information meet the requirements. The registration pathway includes:

- The user must visit the EMT website: https://emt.itflows.eu.
- The user must open the Registration Form by clicking on the "**Register**" button (Figure 34).

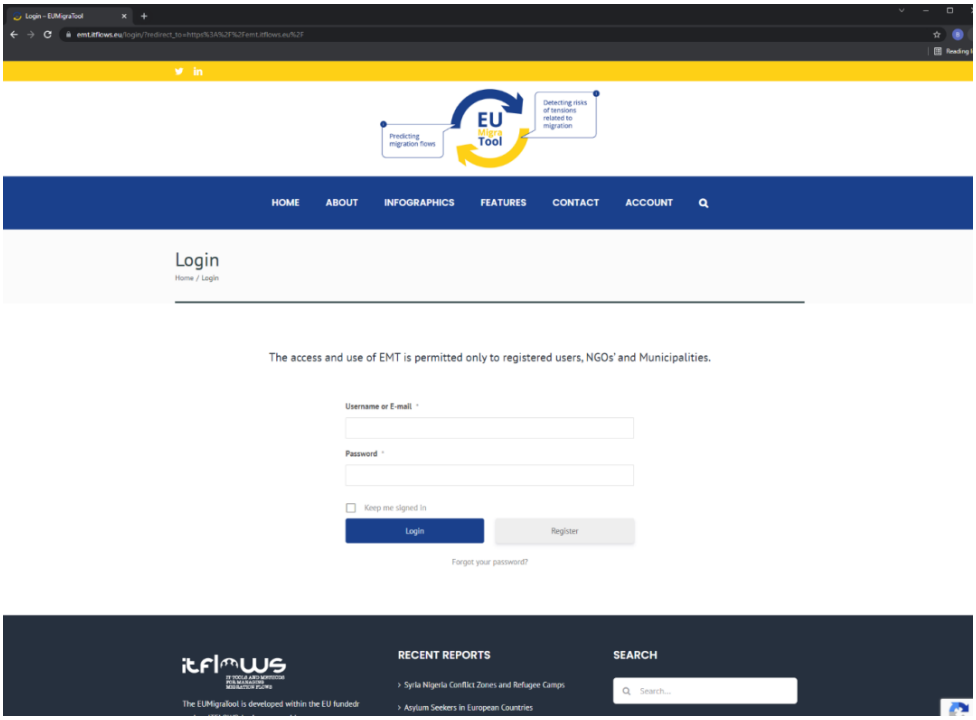


Figure 34: EMT login page.

- The next step includes the filling of the form with the information required in order to successfully register to the system.

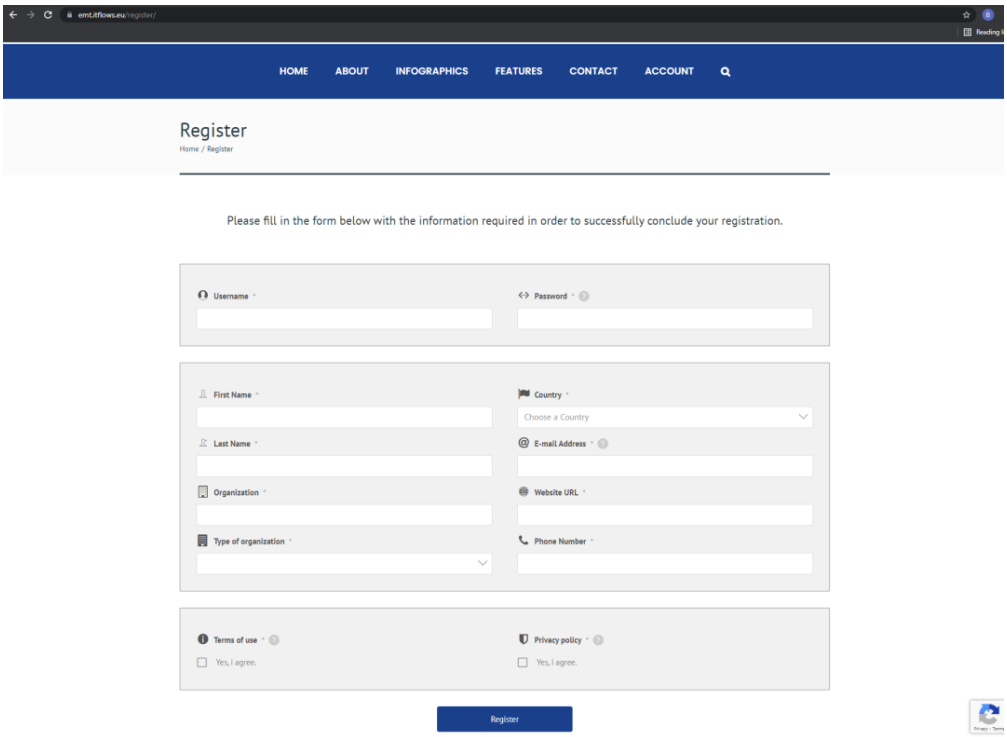


***Figure 35**: Registration form*

Figure 35 shows the registration form. More specifically, while on the registration form, the user is required to fill in the fields with the following information:

- **Username**: Using a username of their choice.
- **Password**: Using the password of their choice. The password must contain at least 8 characters, one lowercase letter, one capital letter and one number.
- **First Name**: User's first name.
- **Last Name**: User's last name.
- **Organization**: User's organisation.
- **Type of Organisation**: There is a dropdown list which includes the three available types of organization that the user can choose from:
  - NGO.
  - Municipality.
  - Other.
- **Country:** Users can choose from the dropdown list their country.
- **Email Address**: Email address of the user (Note: They should use their business email address).
- **Website URL**: User's website.
- **Phone Number**: User's phone number.
- **Terms of use**: Users need to confirm that they agree to the terms of use.
- **Privacy Policy**: Users need to confirm that they agree to the privacy policy.

All of the above fields are mandatory and in order to pass to the next step, the user must fill every field accordingly.

- After filling in all the fields the user should click on the "**Register**" button.
- Then, the following message will appear on the screen for confirmation of the registration request:

  ''Thank you for applying for membership to our site. We will review your details and send you an email letting you know whether your application has been successful or not''.

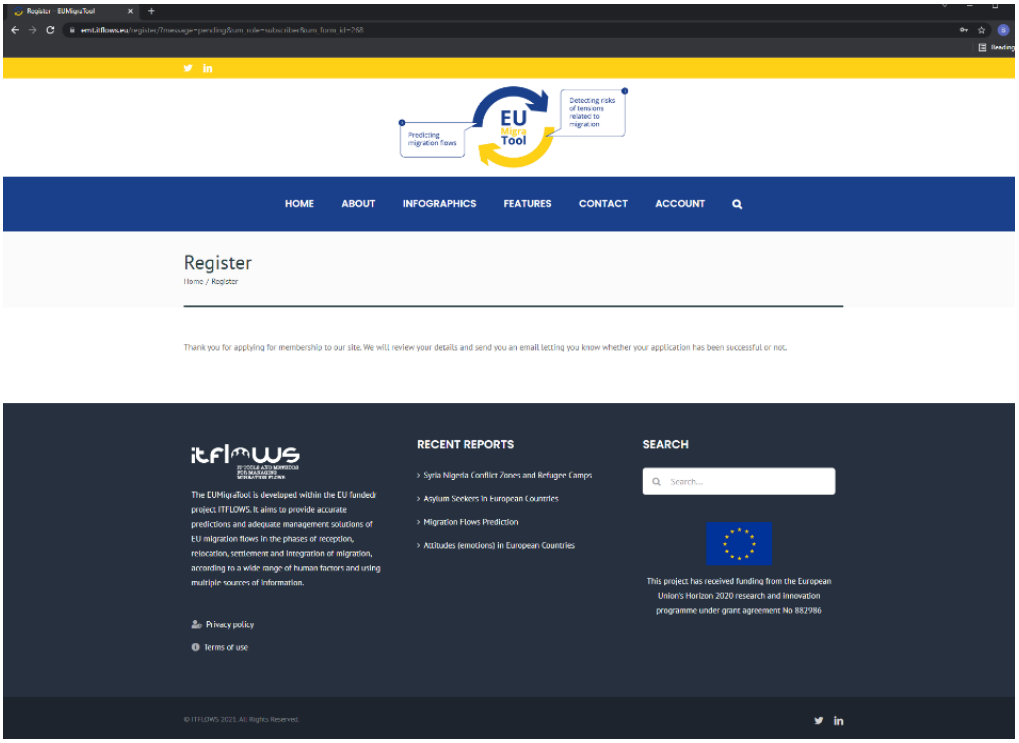


Figure 36: Confirmation message or the registration process

- In addition, an email will be sent to both the user and the system administrator informing them about the registration process.

5.2 Account mail notifications

The system is designed to inform the user about the registration process. The account mail notifications component of the system is responsible for sending an email to the user. This email will inform them that the account request has to be manually reviewed.

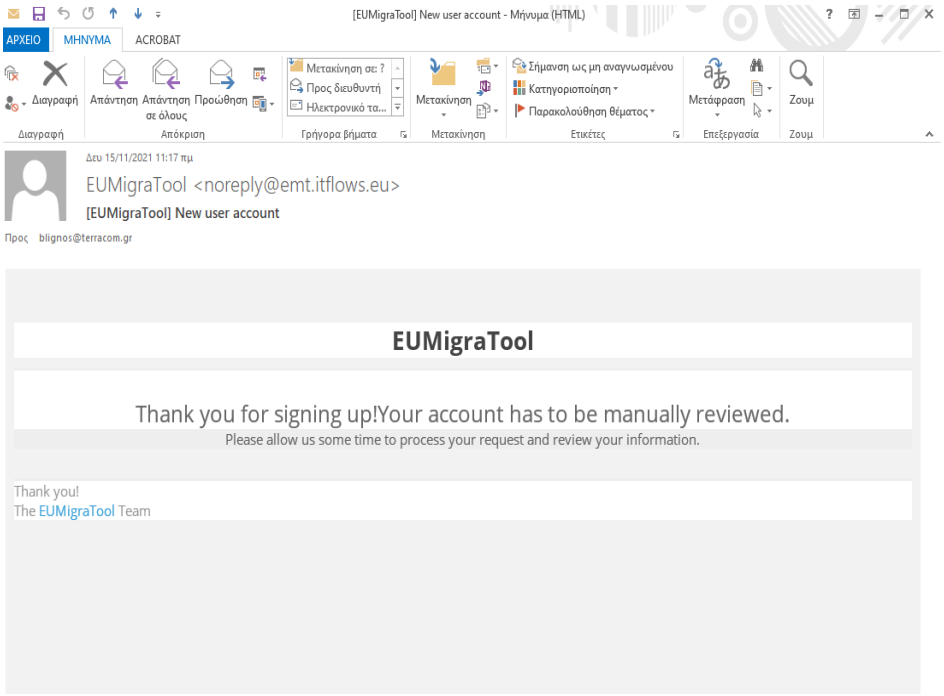


***Figure 37**: Mail notification of account review*

Figure 37 shows the mail notification received by the user in order to be informed about the manually review of their application by the system administrator. As seen already in section 3.2 above, after the system administrator approves the registration, an email will be sent to the user informing them that the account is approved.

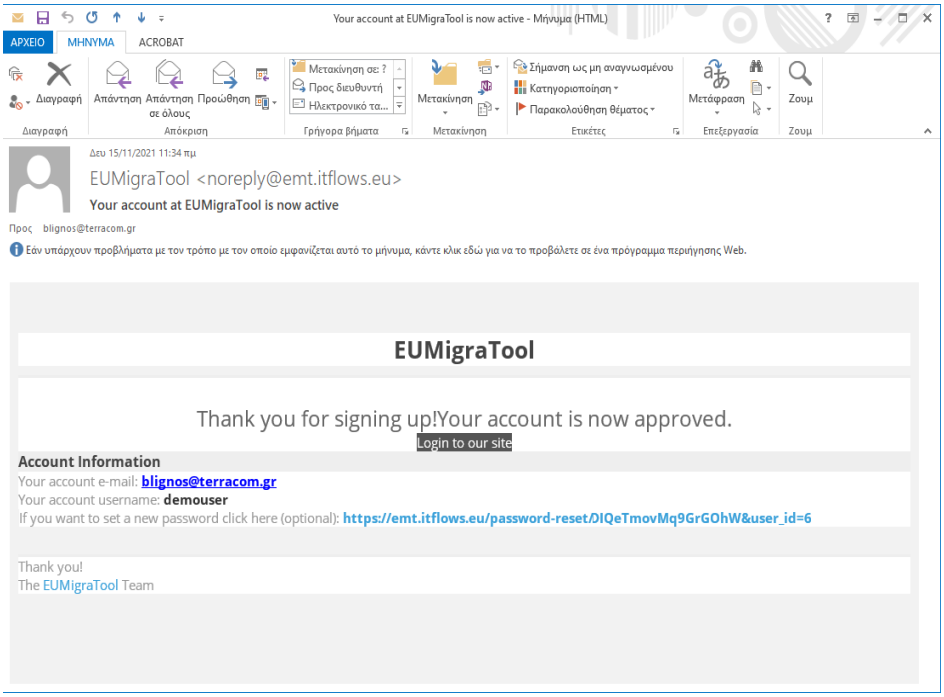


***Figure 38:** Mail notification account approved*

Figure 38 depicts the mail notification sent to the user with the approval of the account registration application.

5.3 First-time Login

After concluding the registration and activation process of the account, the user is in place to login to EMT. In order to do so, he/she is called to follow the next steps:

- Visit EMT website: https://emt.itflows.eu.
- Fill in the credentials.
- Click the "**Login**" button.

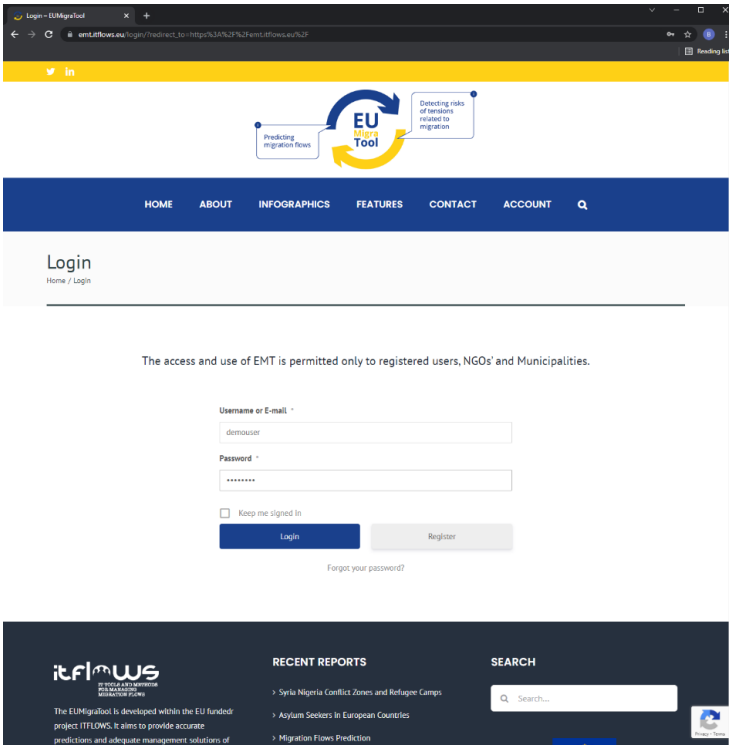


Figure 39: Filling the credentials while on the Login page

Figure 39 shows the insertion of the credentials in order to login to EMT now that the user has created and activated his/her own account.

5.4 Two factor authentication setup

The first time users login to the EMT, they will be asked to ensure the security of their account by enabling the Two Factor Authentication (2FA). To setup his/her 2FA, it is required to follow the next five steps:

1. **Download a 2FA Application.** Users need to download on their smartphone a 2FA Application. The suggested apps to acquire such a tool, are the following:

- Google Authenticator.
- Microsoft Authenticator.
- 2FAS Authenticator.

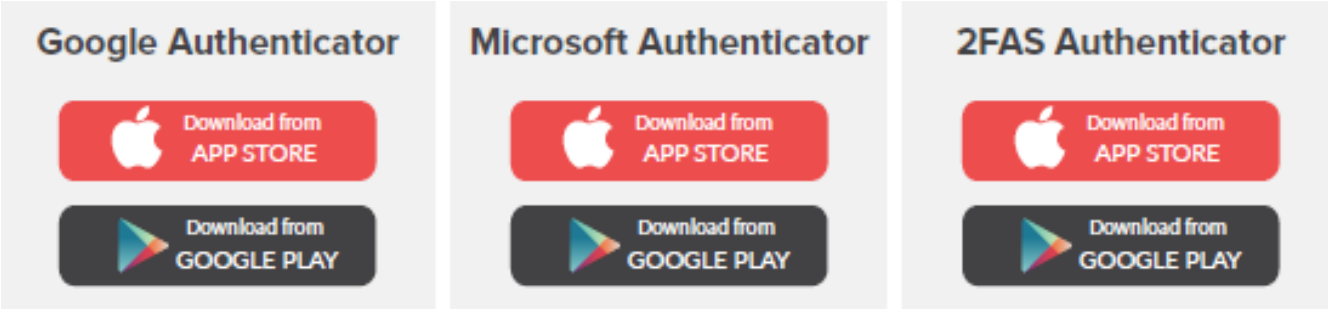


***Figure 40**: Suggested authenticators*

Figure 40 depicts the icons with the respective links included in the EMT website with the suggested authenticators for 2FA authentication. The user has to choose one of them in order to proceed with the 2FA authentication setup.

2. **Click "OK"**. After choosing the 2FA application of their preference, a message appears on the screen informing users about the 2FA.

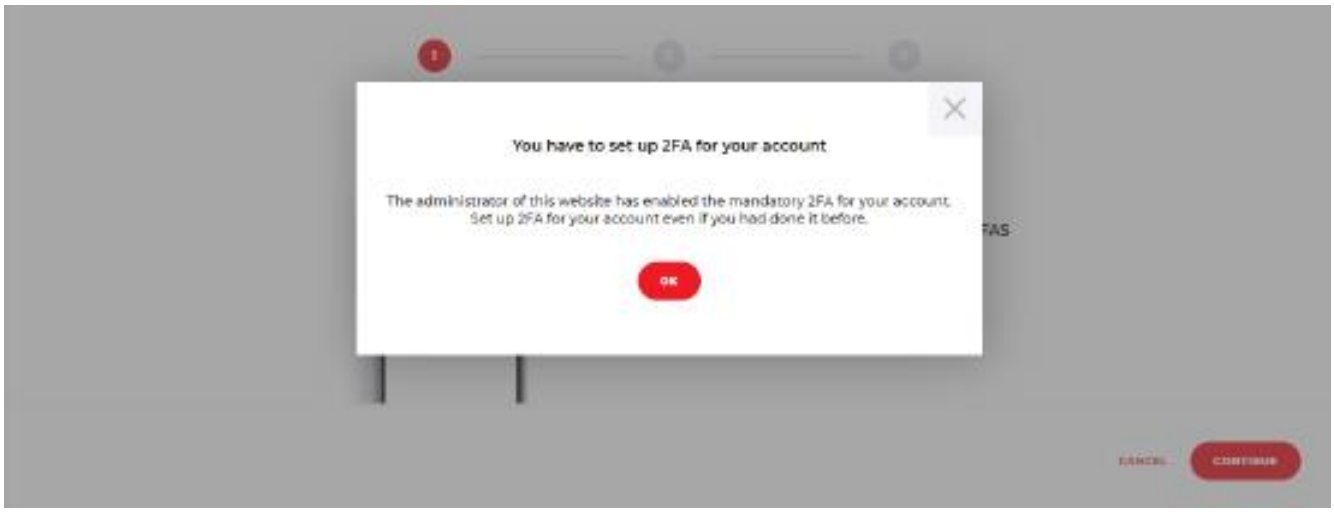


***Figure 41**: Set up 2FA*

As it is depicted on Figure 41, the system administrator has enabled the mandatory 2FA for user's account. Now the user must click ''**OK**'' to proceed.

3. **Click ''Continue''.** Then, the user must choose the red icon ''Continue'' in order to proceed to the next step.

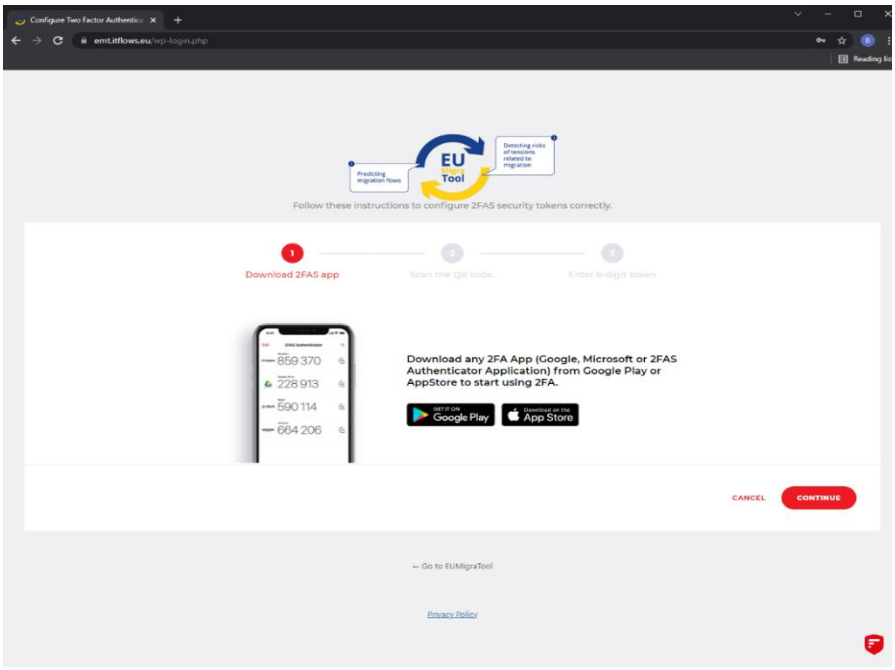


***Figure 42**: Download of the preferable 2FA*

4. **Scan the QR code using the 2FA mobile app.** The user has to open the Authentication App on the smartphone and scan the QR-Code that appears on the screen. Then the user has to click on the red icon ''**Continue**''.

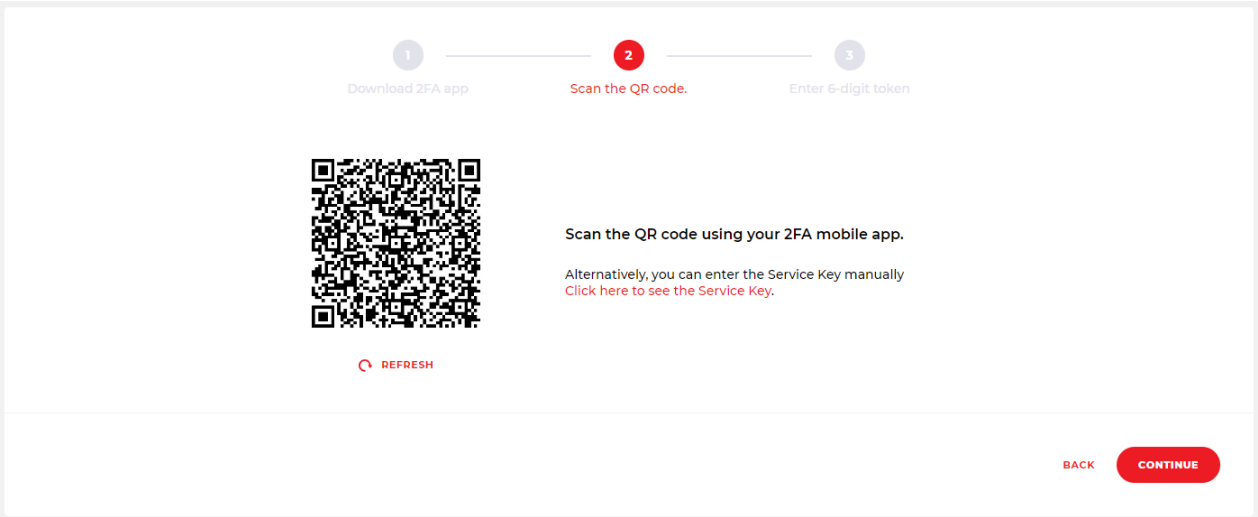


***Figure 43**: Scan the QR code using the 2FA mobile app*

Figure 43 depicts the step of scanning the QR code using the 2FA mobile app.

5. **Enter the 6-digit token.** During this step a 6-digit token has generated by the 2FA application, and the user has to enter it and click the "**Finish configuration**" button.

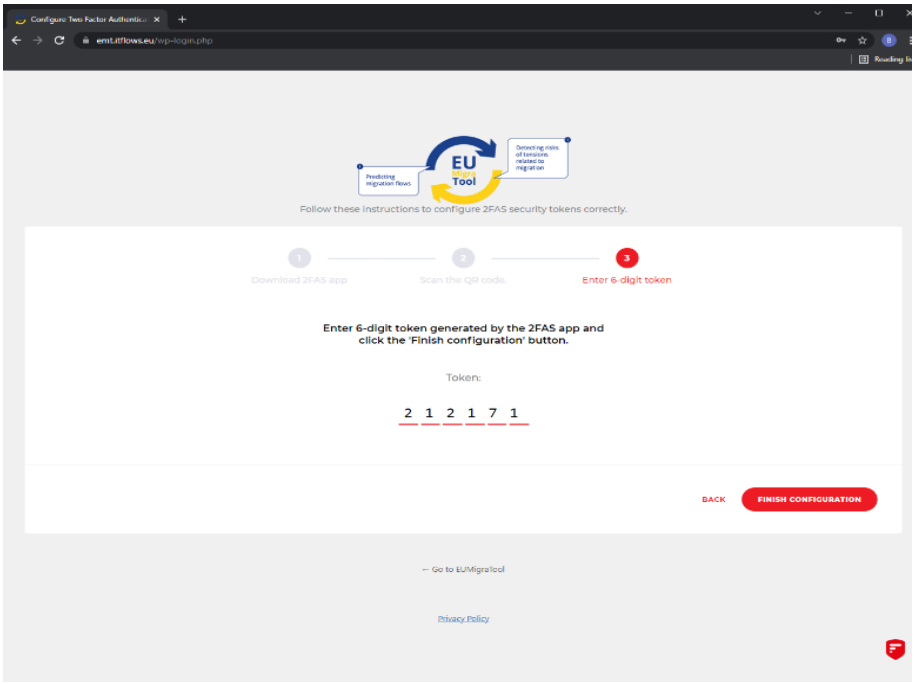


***Figure 44**: Enter the 6-digt token*

Figure 44 depicts the final step of entering the 6-digit token generated by the 2FA app. At this stage the account is ready and secure. The user can use the EMT without concern.

5.5 User Login

When the account of the user is activated and protected, they can login to the EMT by following the next five steps:

- Visit EMT website: https://emt.itflows.eu.
- Fill in the credentials.
- Click the "**Login**" button.
- Enter the 6-digit token generated by the 2FA app.
- Click the "**login**" button.

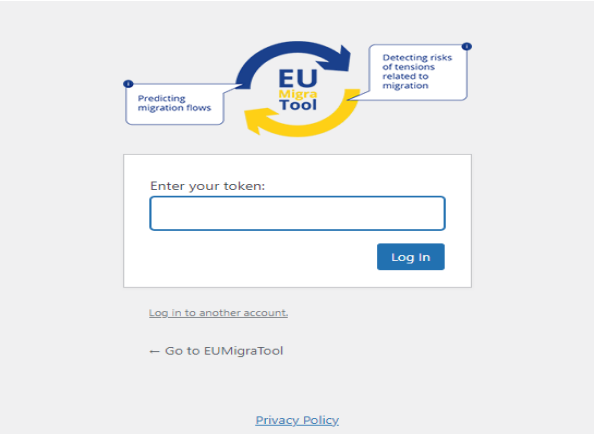


***Figure 45**: Enter the 6-digit token*

Figure 45 depicts the step of entering the generated 6-digit token, in order to successfully login to the system.

5.6 Logout

Each time users complete their navigation through the EMT, it is strongly recommended to logout from the system. Also, after 60 minutes of inactivity the system is logging out the user and a re-login is required. The user has the ability to easily logout by selecting at the main menu the tab Account and click on the Logout link. Then they return to the login page of the system.

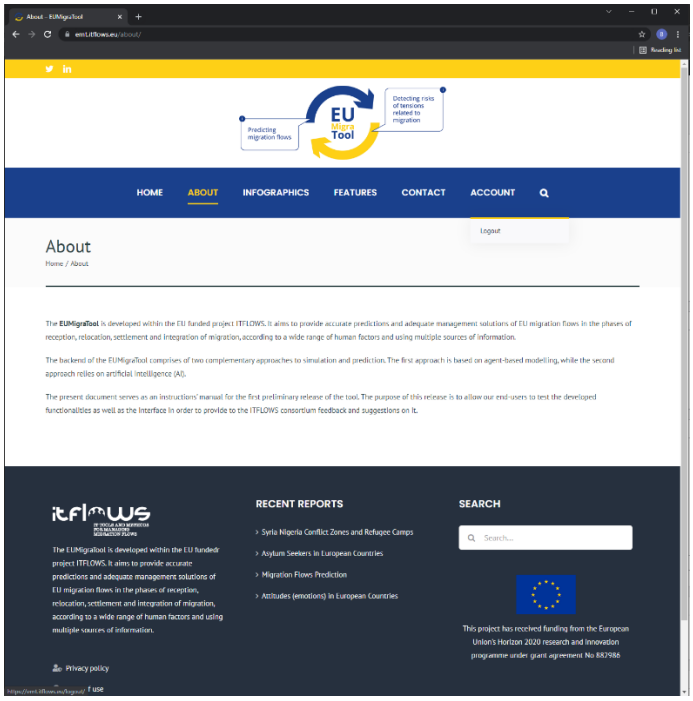


***Figure 46**: Logout*

Figure 46 depicts the logout icon in case of quitting the EMT and log out by the user.

5.7 Home

After the login of the user to the EMT, the first page presented is the "Home" page. Here, the user can be informed about prediction and management of migration in general and learn about the purpose of the EMT and contribution on this subject.

More specifically, as seen on the EMT website, the purpose of EMT is to provide accurate predictions and adequate management solutions of migration flows in the European Union in the phases of reception, and integration of migration, using multiple sources of information. These insights will be provided by models.

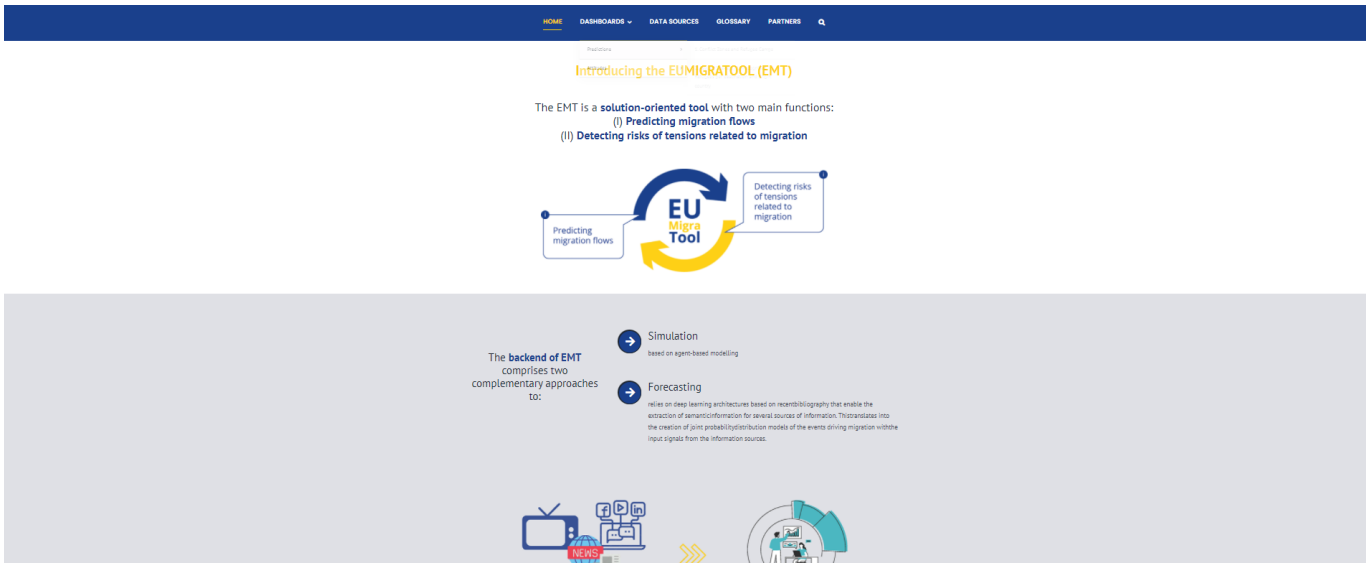


***Figure 47**: Home page*

Figure 47 presents the "Home" page of the EMT website. While on the "Home" page, the user can watch a short video presenting and describing the ITFLOWS project purpose and contribution on prediction and management of migration. At the same time, the user can identify the consortium partners of the project.

5.8 Dashboards

The next page that the user can navigate is the Dashboards page. As already explained above, this page includes the dashboards with the visualisations of the EMT. More specifically, the user can choose between the following options:

1. **Origin Countries with conflict locations and Asylum Seekers / Unrecognised refugee camps**. This map displays conflict zones in Syria and Nigeria and nearby refugee camps and contains the number of people in these camps during the period of the 1st trimester of 2021.
2. **Asylum seekers / unrecognised refugees per destination and origin country**

a. **Historic Data.**

This map displays an option element to filter all the rest elements where the user can select the countries of interest. Depending on the countries the user selected the rest elements are adjusted accordingly.

b. **Predictions.**

This graphic displayed the prediction for February 2022, and it is implemented for testing purposes. Major changes on this dashboard may be implemented in the future by combining it with the previous dashboard (Historic Data).

3. **Attitudes based on Twitter analysis.** It includes a map of Europe which displays the percentage of tweets expressing sentiments (hate, positive, negative, neutral sentiments) per country for a period (2013 – 2020).
4. **Influential Attitudes Chart.** It includes a diverging stacked bar displaying several factors (e.g., age, religion) and how these factors affect the attitudes towards migration.

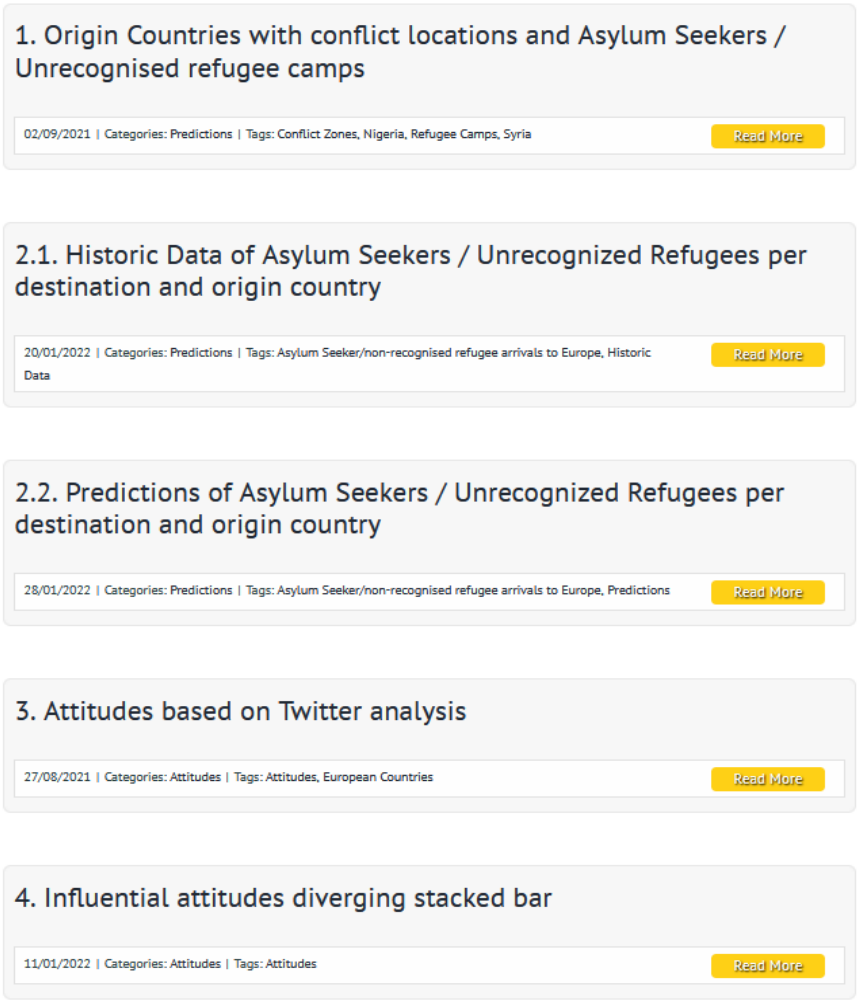


Figure 48: *Dashboards page.*

Figure 48 shows the dashboards included to the Dashboard page of the EMT:

5.8.1 Origin Countries with conflict locations and Asylum Seekers / Unrecognised refugee camps

As explained already above, the agent-based simulations seek to predict the number of migrants displaced from a specific country due to conflict situations in specific areas within such country. The simulation module estimates the cumulative number of migrants who will arrive in refugee camps in neighbouring countries over a period of time (the simulations run on a daily basis). This information is essential to further identify how many of such migrants might continue their journey towards the European Union seeking for protection.

It can be accessed directly from the following link:
Syria-Nigeria Conflict Zones – EUMigraTool (ITFLOWS.eu)

Initially, the map displays the case of Nigeria, at the end of the simulation period. The simulation for Nigeria runs from 01-03-2016 to 30-04-2021 and for Syria from 01-01-2017 to 20-04-2021.

As we have already explained above, green dots represent camps, yellow dots represent cities/towns and red dots represent conflict areas. Please note that the colours of the dots may change and are not necessarily accurate with the following figures. The legend of the map will always be updated, and it is definitely more accurate than the colours of this manual.

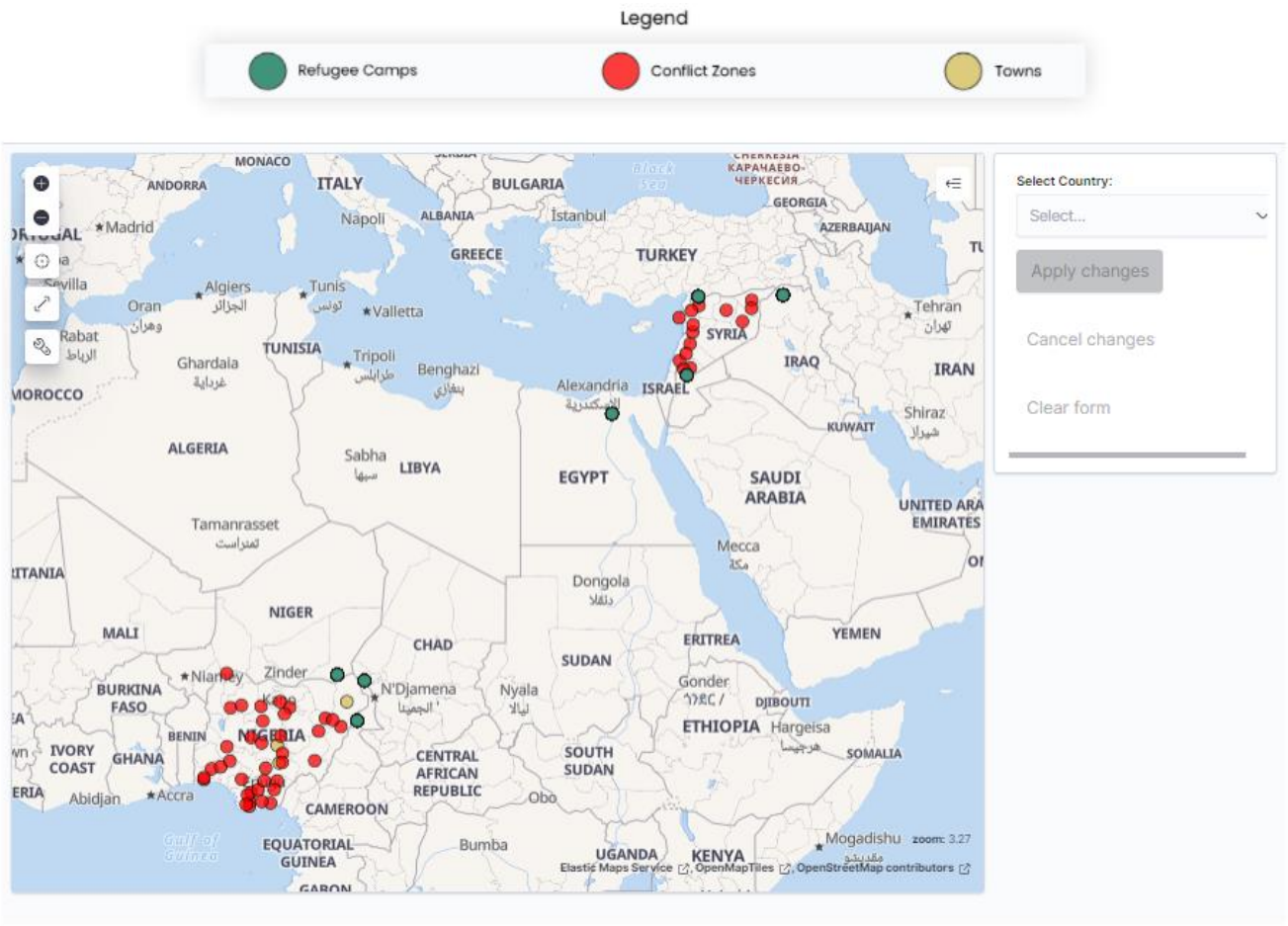


***Figure 49**: Nigeria case and Syria displacement map.*

The user can select one of the two countries in the top right selection box and by applying the changes, the map only displays the conflict zone of the selected country.

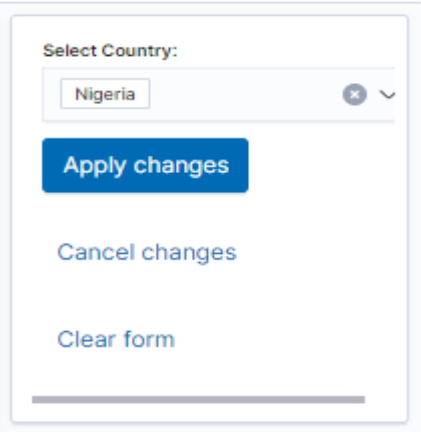


***Figure 50**: Selection box options with Nigeria as example.*

The user can now see only the selected country dots and by clicking on the arrows that can be found in the options button on the left side of the map, the visualisation is zooming towards the country that was selected.

***Figure 51**: Clicking the arrows button will zoom into the selected country*

Once the map is zoomed in the selected country, the user can check all the information regarding the camps, some major cities and the conflict zones. By clicking on any of the green dots (refugee camps), the user can see information on how many migrants have arrived at this location, along with other additional information like the region, the name of the camp and the population, as shown in Figure 52.

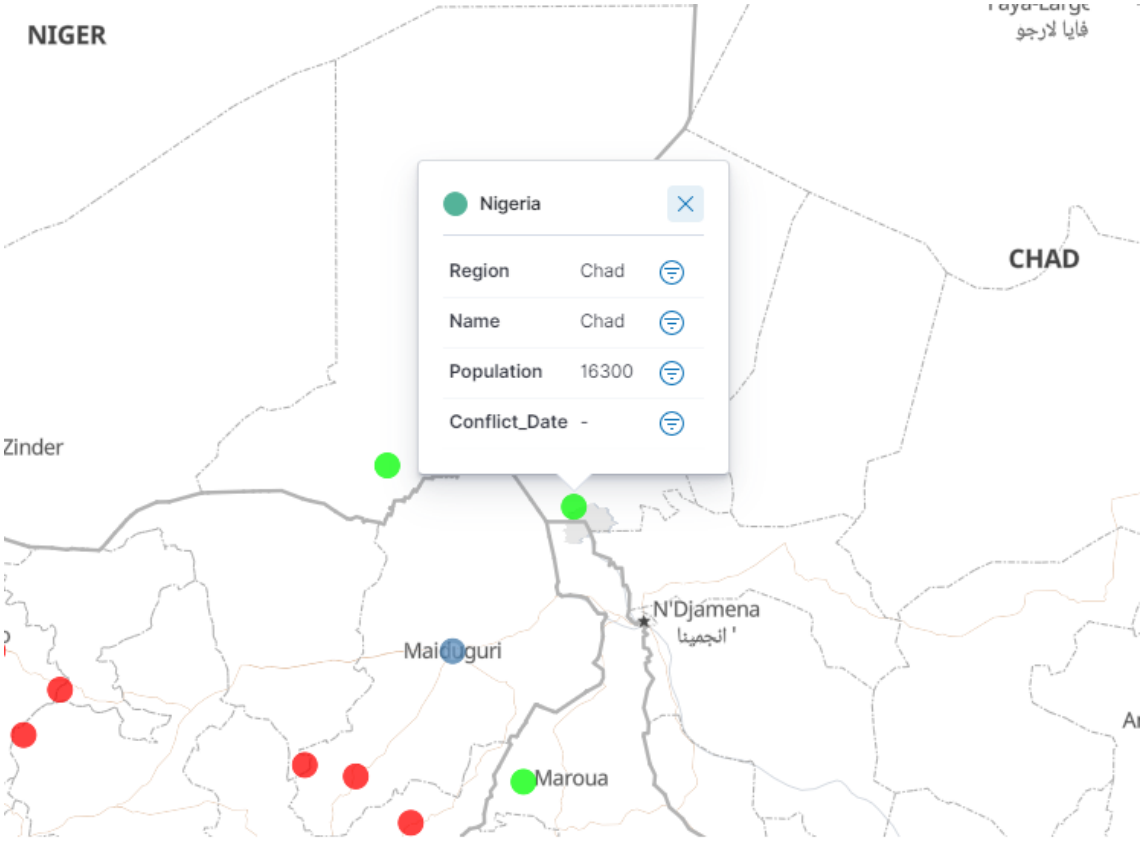


***Figure 52**: Hover information*

The red dots on the map denote the various conflict areas. By clicking on them, the user can view more information, such as the name, population and the date the conflict appeared in the specific location.

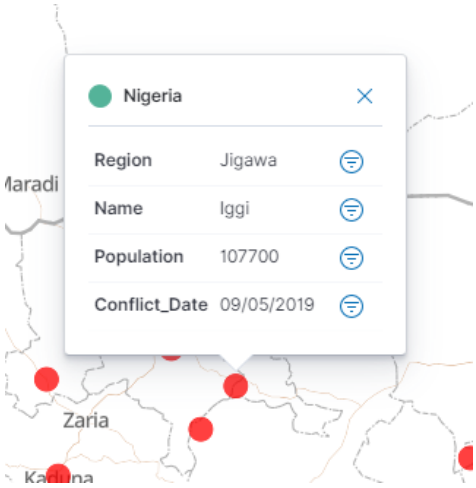


***Figure 53:** Hover information with conflict data*

5.8.2 Asylum seekers / non-recognised refugees' arrivals in Europe Historic Data

This functionality visualises the number of migrants seeking for asylum that are expected in various EU countries from a specific country of origin over a period.

Dashboard 2 can be directly accessed under:

[Asylum Seekers in European Countries – EUMigraTool (ITFLOWS.eu)]

Initially the visualisation shows arrivals from all the available origin countries and destination countries selected for EMT.

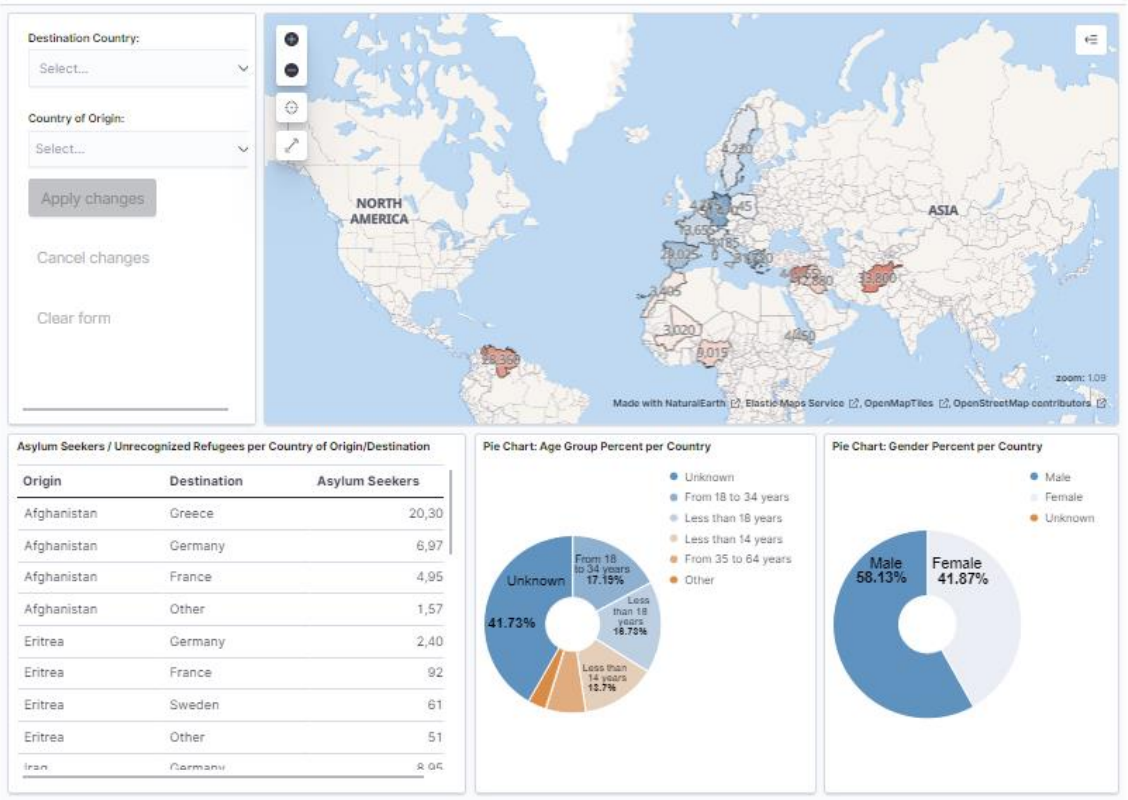


***Figure 54**: Visualisations on Asylum Seekers*

At this stage, the available origin countries are:

- Afghanistan
- Eritrea
- Iraq
- Mali
- Morocco
- Nigeria
- Syria

And the destination countries are:

- France
- Germany
- Greece
- Italia

- The Netherlands
- Poland
- Spain

The user can apply filters regarding the destination and origin countries. The filters can be applied using the menu on the left of the map, as shown in Figure 55.

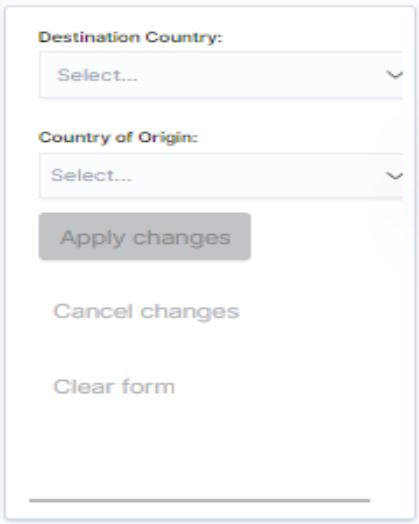


***Figure 55**: Filter application*

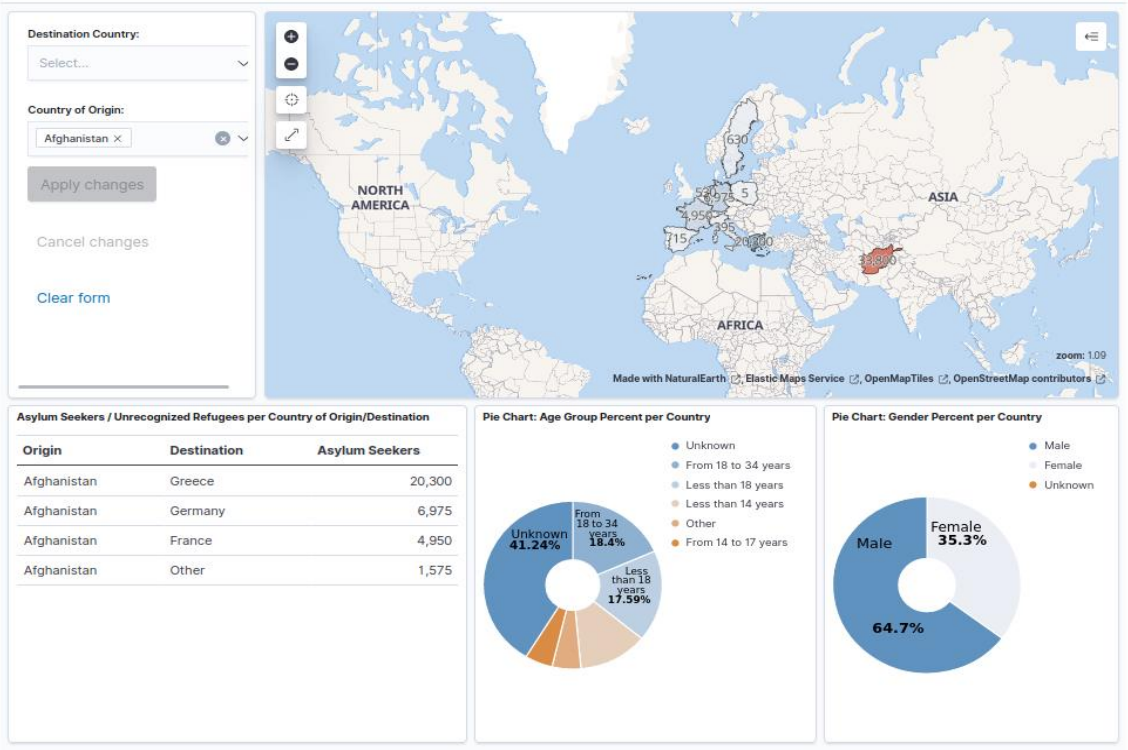


***Figure 56**: Result of filter application*

In Figure 56 the tool illustrates the numbers and statistics from asylum seekers / non-recognised refugees arriving to Europe from Afghanistan. In a similar way, a destination country within EU can be selected.

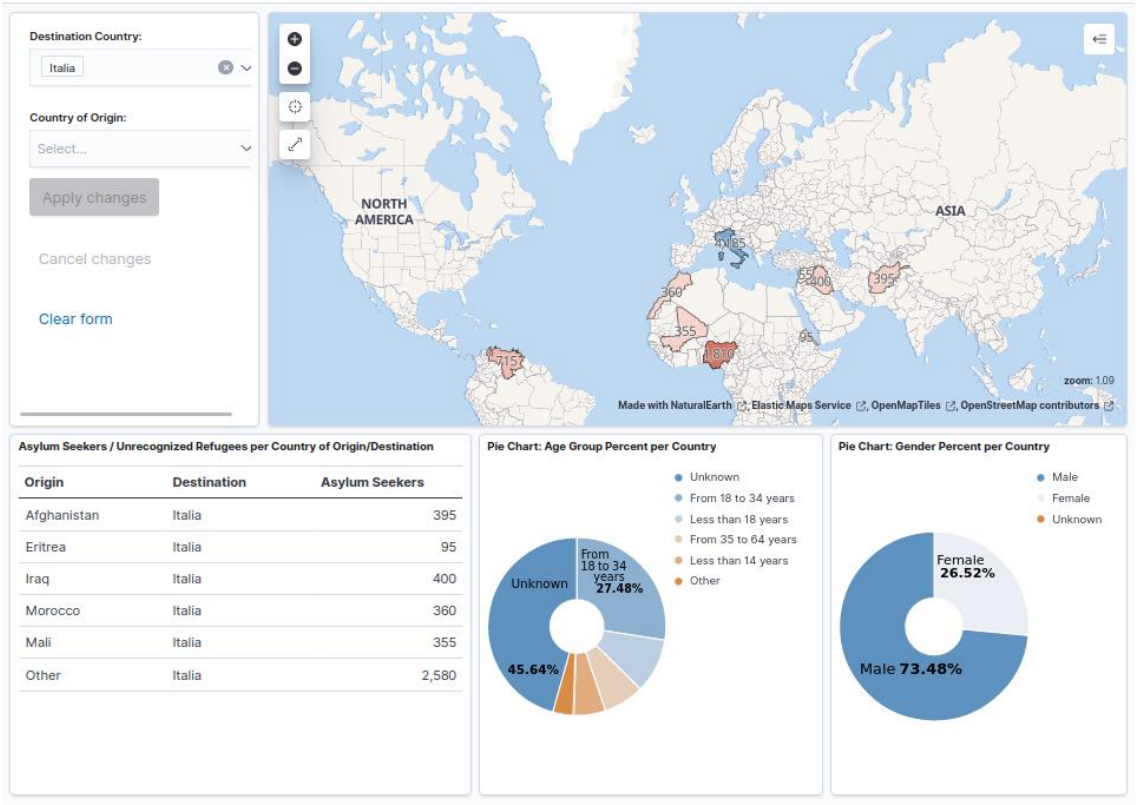


***Figure 56**: Information about the asylum seekers/ non-recognised refugees arriving to Italy*

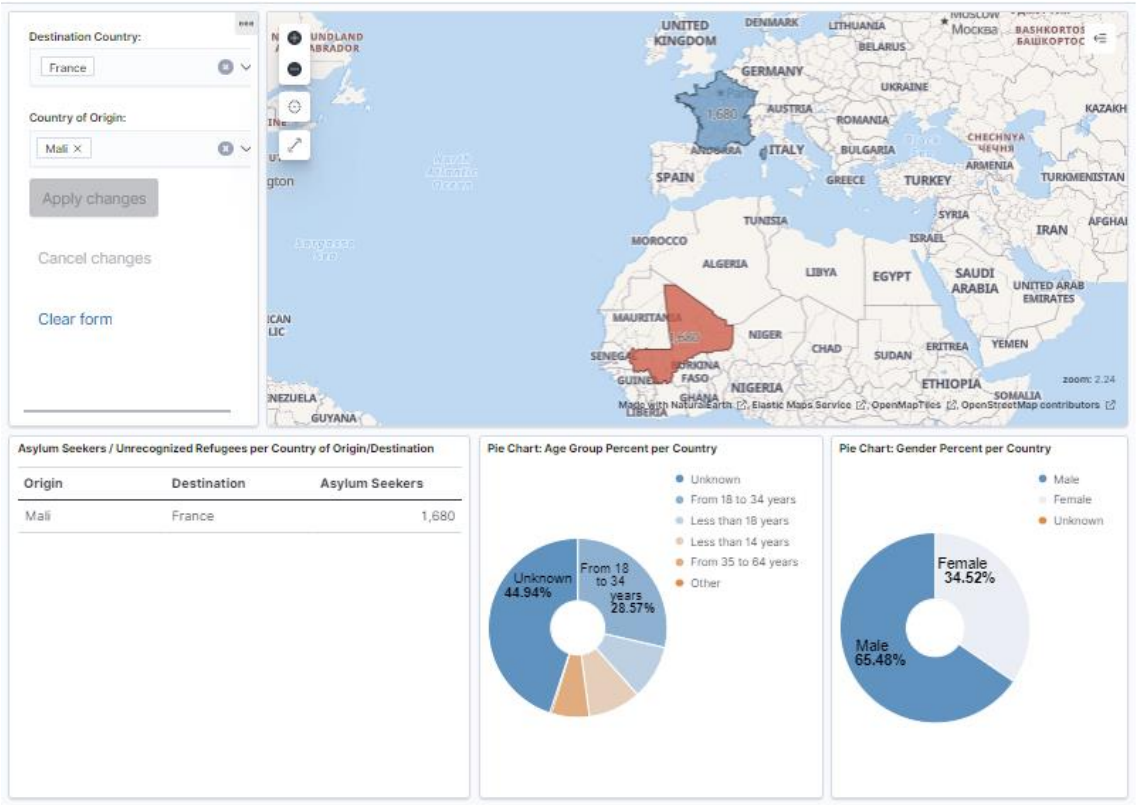


***Figure 57**: Example of asylum applications from Mali to France*

There is also the ability to specifically select both a destination and origin country as shown in the last example. By clicking the "Clear form" button from the side panel, all the options are cleared and by applying the changes the user is shown the initial world map with all the destination and origin countries.

5.8.3 Asylum seekers / non-recognised refugees' arrivals in Europe Predictions

This functionality visualises the number of migrants seeking for asylum that are expected in various EU countries from various origin countries for the month of February 2022 as these numbers were predicted on 25/01/2022. It is a testing page, and the prediction graphs and information are expected to be combined with Dashboard 2.1 on Historic Data.

It can be accessed directly be following this link:
Asylum seekers / non-recognised refugees' arrivals in Europe Predictions

This temporary dashboard offers the same options as the Historic Data dashboard, and it is working on a similar way.

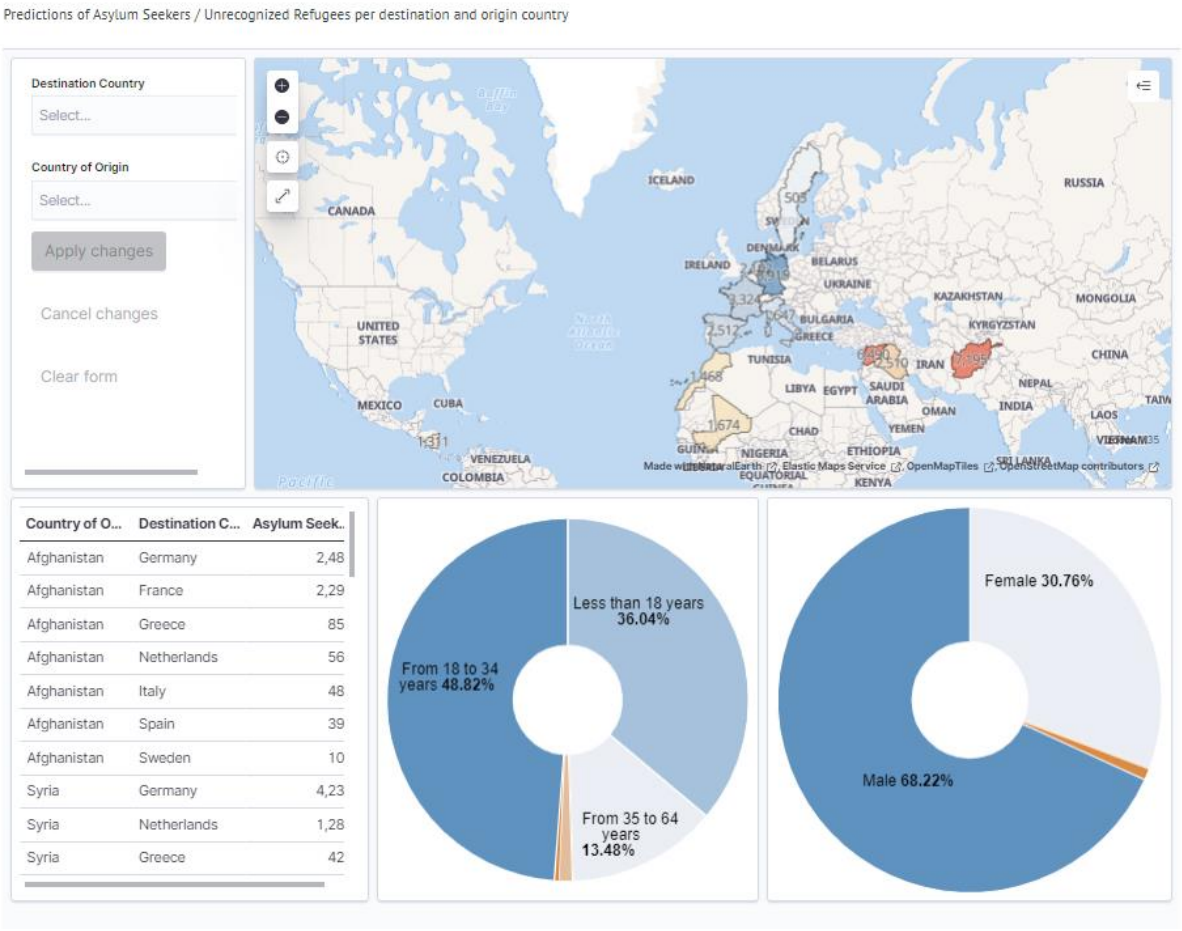


***Figure 58**: Temporary prediction dashboard*

The user can select either the destination country, the origin country or both. By applying the changes, the tables, graphs and generally the infographics of this dashboard will change accordingly.

5.8.4 Tension Function of the EMT: Attitudes and sentiment of European countries towards migration from Twitter analysis

This part of the EMT utilises Twitter data in order to detect hate speech, positive, negative and neutral sentiment related to migration (e.g., discussions on refugees, camps, etc.).

It can be accessed directly from here:
Attitudes based on Twitter Analysis – EUMigraTool (ITFLOWS.eu)

To create a filter, the user needs to select the country and the type of sentiment from the options menu on the left side of the map. For the indicator, the possible values are:

- Country Name: select one or more countries to focus on.

- **Sentiment:** select specific sentiments

Here, Austria is being selected, with the "Sentiment" of "negative sentiment"

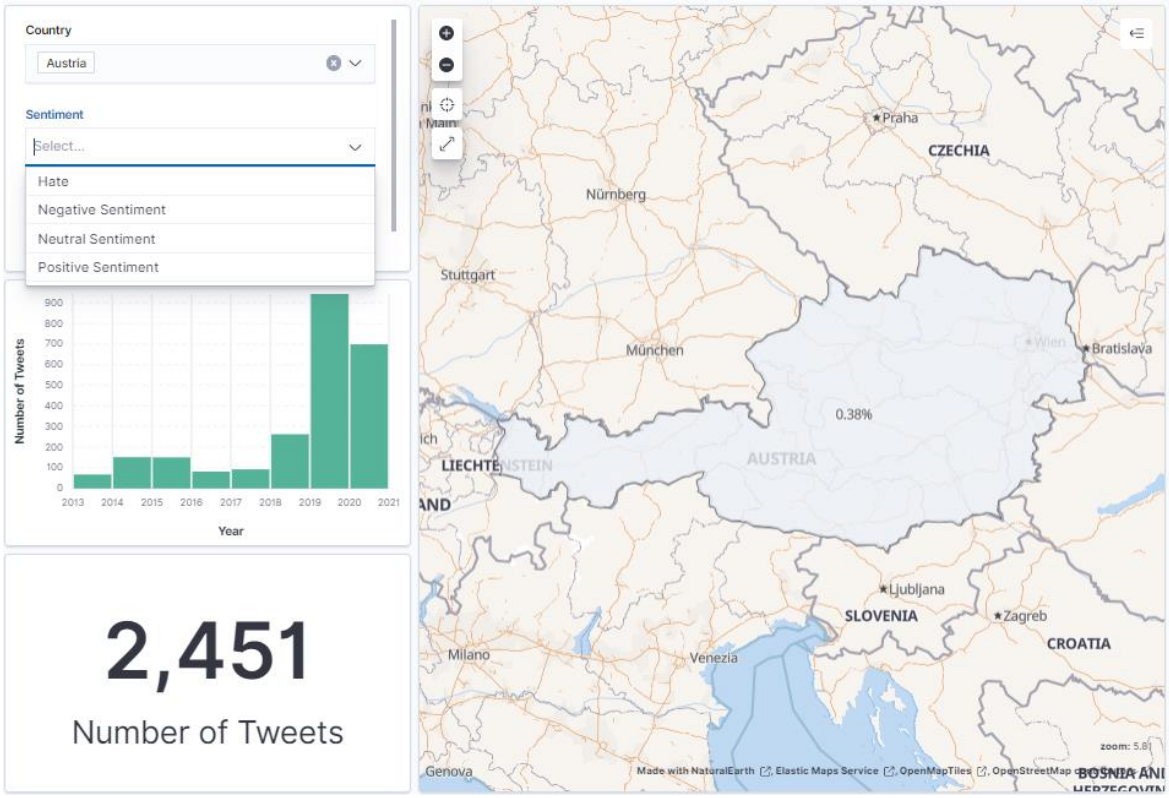


Figure 60: *Filter selection process on Attitudes dashboard*

Click apply changes:

Figure 61*: Apply changes by pressing the button in the red circle*

Below you will see that on the left side of the page, the page shows the number of tweets per EU country (in this case Austria), for the period of 2013 to 2020, that have been identified as containing "negative sentiment." The numbers shown on each country represent the percentage of tweets detected satisfying the defined filters (e.g., negative tweets, so here 0.38% negative tweets in Austria).

5.8.5 Influential Attitudes

The Influential Attitudes Diverging Stacked Bars consists of a bar graph showing the distance from the mean of each attitude for each country specifically.

It can be accessed directly from here:

Influential Attitudes Spectrogram – EUMigraTool (ITFLOWS.eu)

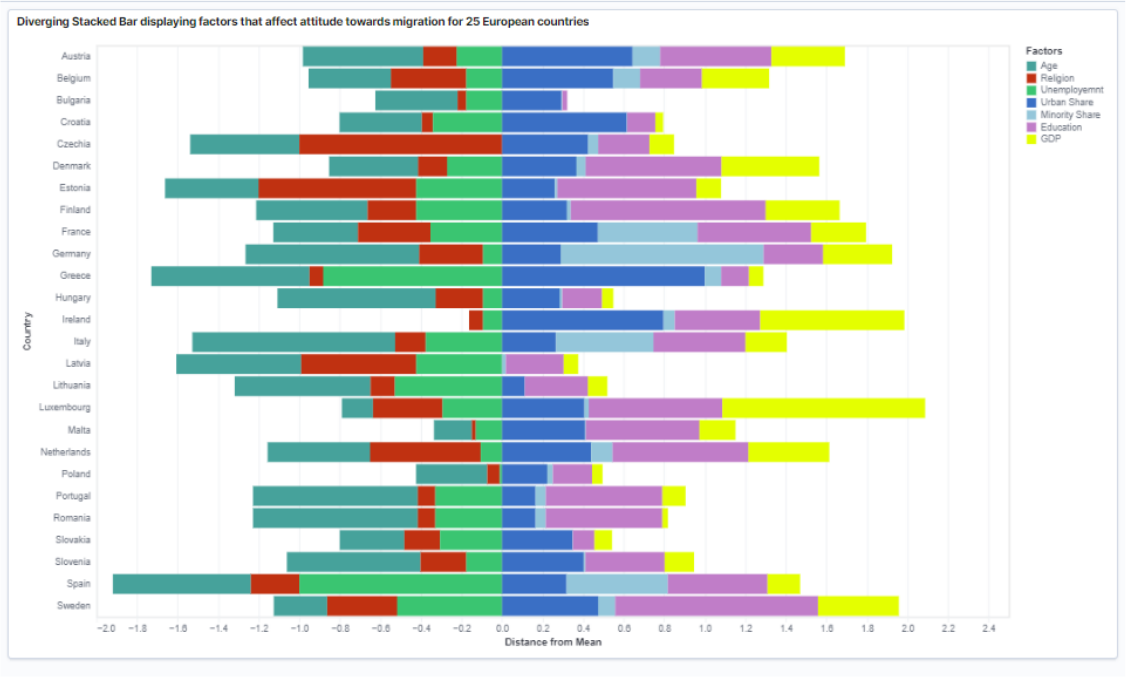
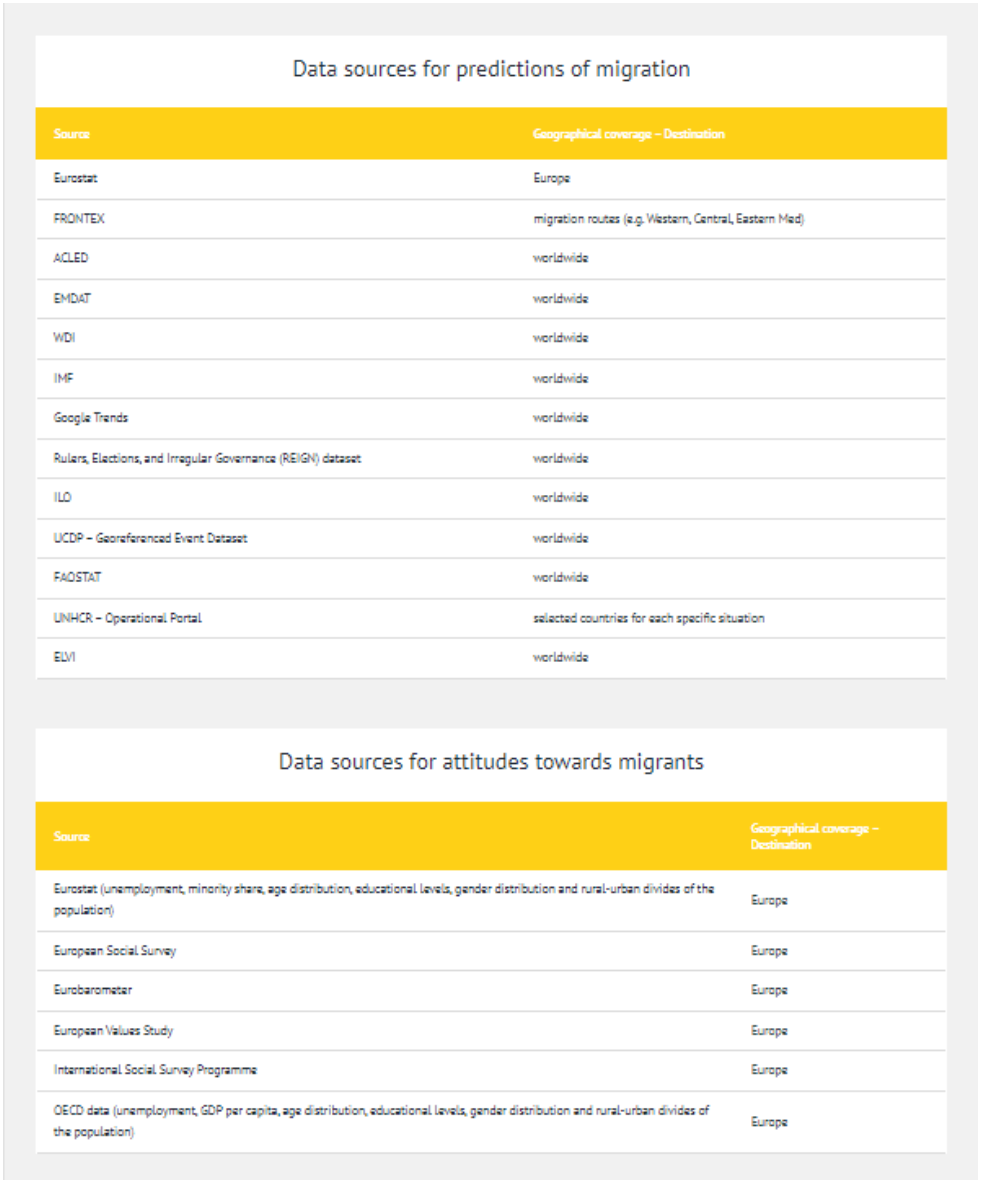


***Figure 592**: Influential Attitudes Diverging Stacked Bars*

There are no options in this bar. The user is able to visualise the general attitude of each country towards migration based on the analysis of different factors. Details of the analysis can be found on deliverable D5.2 of ITFLOWS.

5.9 Data sources

In order to result to the most complete visualisations to the Dashboards included in the EMT, the use of appropriate data sources is required. These data sources include data sources for predictions of migration and data sources for attitudes towards migrants.



Data sources for predictions of migration

Source	Geographical coverage – Destination
Eurostat	Europe
FRONTEX	migration routes (e.g. Western, Central, Eastern Med)
ACLED	worldwide
EMDAT	worldwide
WDI	worldwide
IMF	worldwide
Google Trends	worldwide
Rulers, Elections, and Irregular Governance (REIGN) dataset	worldwide
ILO	worldwide
UCDP – Georeferenced Event Dataset	worldwide
FAOSTAT	worldwide
UNHCR – Operational Portal	selected countries for each specific situation
ELVI	worldwide

Data sources for attitudes towards migrants

Source	Geographical coverage – Destination
Eurostat (unemployment, minority share, age distribution, educational levels, gender distribution and rural-urban divides of the population)	Europe
European Social Survey	Europe
Eurobarometer	Europe
European Values Study	Europe
International Social Survey Programme	Europe
OECD data (unemployment, GDP per capita, age distribution, educational levels, gender distribution and rural-urban divides of the population)	Europe

Figure 603: *Data sources*

Figure 60 presents the Data sources page of the EMT. If users press on the icon of their interest, then it will redirect them to the respective site.

5.10 Glossary

Sometimes it is difficult to possess the whole knowledge of the migration subject or to deal with the law and policy terms. In order to facilitate and ameliorate the use of EMT, ITFLOWS project has created a glossary with the purpose to clarify the terms used by the system as a whole.

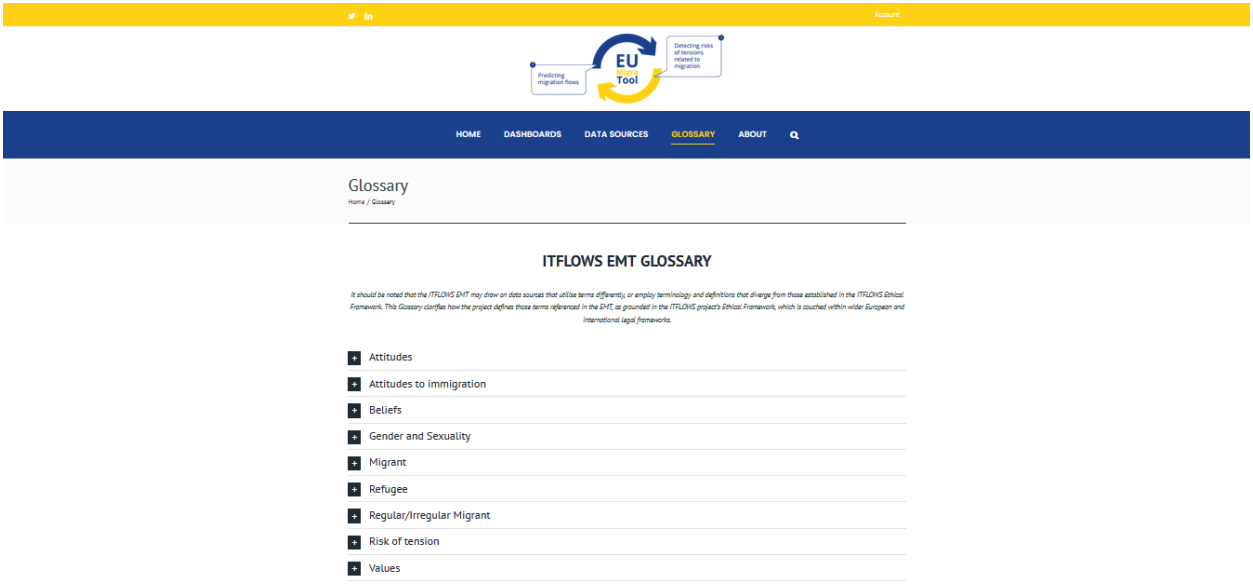


***Figure 614**: Glossary*

As it is written on the Glossary page too, it should be noted that the ITFLOWS EMT may draw on data sources that utilise terms differently or employ terminology and definitions that diverge from those established in the ITFLOWS Legal Framework. This Glossary clarifies how the project defines those terms referenced in the EMT, as grounded in the ITFLOWS project's Legal Framework, which is couched within wider European and international legal concepts and instruments.

5.11 Partners

The "Partners" page is the final one of the EMT and includes information about the consortium and the participating partners, and funding of the project. Also, links to interesting sources are provided.

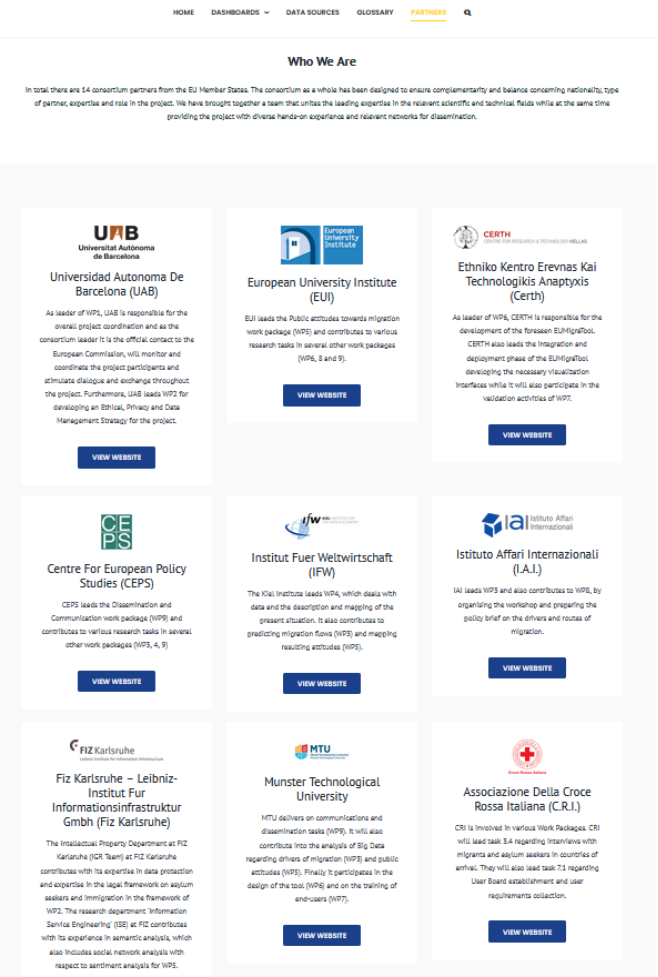


***Figure 625**: Partners' page*

Figure 62 shows the page *Partners*. At the bottom of the page, there is also written information about the project and recent reports. Also, the user can find and read the Privacy Policy and the Terms of Use of the proposed solution. Finally, a logo of the EU is depicted and escorted by the information that the project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 882986.

5.12 Account

While navigating through the EMT, users can show and edit their Account Page. In such page, they can proceed to edit and modifications regarding the account. The form consists of:

- **Username**
- **First name**
- **Last name**
- **Email address**

Also, if users wish to change their password, they could do so by choosing the field ''**Change Password**''.

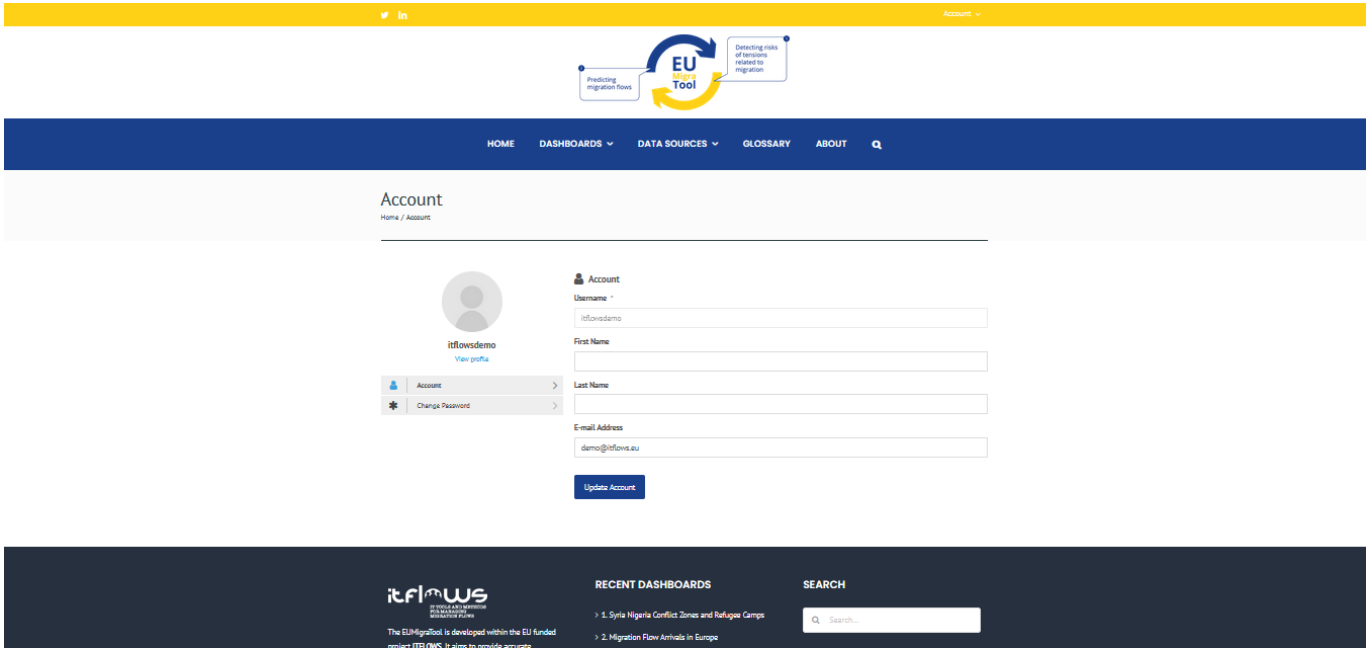


***Figure 636**: Account page*

Figure 63 shows the account page of the user. Last but not least, if the user wishes to, he/she can connect with the social media ITFLOWS accounts. Current social media channels are LinkedIn and Twitter.

6 Data sources

As seen above, EMT is using numerous datasets from various data sources for its Large-Scale model. Following is a list of these sources.

- **EUROSTAT**

  Eurostat is one of the main data sources EMT is using. Some of the categories include crime and criminal justice, various demographics, gender equality, social status data, health and living situation, education, and economic indicators. More information regarding the datasets EMT is using from this source are shown in **Appendix 2** below.

- **GDELT**

  GDELT is the largest, and highest resolution open platform that monitors the world's news media from nearly every country in print, broadcast, and web formats, in over 100 languages, in almost real time and that stretches back to January 1, 1979, through present day, with daily updates.
  EMT is using the GDELT 2.0 DOC API for monitoring the national press of origin countries. This API makes it possible to search the full text of all monitored coverage from January 1st, 2017, to this day and return a list of matching articles sorted by relevance, date, or even sentiment. The extraordinary thing about this database it that text searching is not limited to only English news coverage but includes the English translations of coverage from 65 languages.
  The main component the EMT is utilizing from this database is: headlines of news articles referencing the origin-country's name 3-4 times in the main text. The sampling frequency is once again daily, and it includes maximum 250 articles per day.

- **EM-DAT**

  The Emergency Events Database (EM-DAT) was launched in 1988 by the Centre for Research on the Epidemiology of Disasters (CRED) and contains data on the occurrence and effects of over 22,000 mass disasters in the world from 1900 to the present day. The database is compiled from various sources, including UN agencies, non-governmental organizations etc. EM-DAT contributes to EMT with datasets of records of technological and natural disaster events, including fatalities, in the whole world. Each data point consists of the disaster event, the country the incident happened and the source of information, providing very useful input for conflicts detection.

- **ACLED**

  The Armed Conflict Location & Event Data Project (ACLED) is a registered non-profit organization in the United States that collects real-time data on the locations, dates, actors, fatalities, and types of all reported political violence and protest events around the world.

  EMT is using the ACLED's records of violence events that include battles, protests, riots, and fatalities to capture any possible correlations and causations between violence events and the creation of migration flows to Europe. ACLED also includes datasets regarding disorder involving the media, peacekeepers and health workers, political violence targeting women, coronavirus related disorder events, indirect killing of civilians and sexual violence.

- **COMPONENTS**

  Components is a publication and research group that assembles, investigates, and editorialises large datasets. Most of the data they use in their research are freely available for use.

  It provides EMT with the ALL-NEWS Dataset which contains 2.7 million news articles and essays from 27 American publications. The data include date, title, publication, article text, publication name, year, month, and a URL for most of these articles. Articles mostly span from 2013 to early 2020 and are used for the unsupervised training of the LDA topic modeler.

7 Conclusions and Future Work

7.1 Towards more accurate predictions

In the future, the origin countries list (for which we perform asylum application predictions) will also include Eritrea, Venezuela, Colombia, and Honduras.

As for the predictions, the goal is to minimise the error (Root Mean Squared Error) between predictions and ground truth as well as make the resulting model more robust across different bilateral cases of migration. This might be possible by:

- Using embedded methods like *hierarchical cross-section forecast reconciliation* to combine forecasts from different models with different intuition.
- Experimenting with bigger time step forecasts (N months ahead instead of 1 and 3 as it is for now)
- Further experimenting with different input features and models.
- Improving LSM's backbone element which is the LDA Topic Modeler. At time of writing this deliverable the modeler is trained using 2.7 million articles from all American publications. It would be very interesting to train different modelers for every specific origin case (e.g., a modeler trained on Arabic text from articles of Middle East publications, capable of producing precise topic shares of Iraq's national press.). That way the Large-Scale model can become more sensitive to critical nuances of the origin country's native language.

7.2 Ethical recommendations

The developers' team and Work Package 6 will work closely with the responsible ITFLOWS boards and the ethical teams in order to ensure all the ethical guidelines are followed thoroughly. This is a high priority for the developers and the EMT will comply with the ethical recommendations described in Section 4 of this deliverable.

7.3 EMT Platform development

The EMT platform will be updated continuously according to the received feedback. There will be an open pipeline between the end users and the developers. The developers' team will monitor the feedback and keep the platform up to date with their needs. The focus will be on upgrading the user experience and EMT's functionality.

7.4 Users Board Feedback

Following the Users Board meeting held online on January 31 2022, the following points were established as fertile ground for future work and improvement of the EMT. Moreover, a plan on how these proposals will be addressed will be provided.

Changes on the frontend:

- Clarify the term "unrecognised refugees" in the glossary. (PENDING)
  The glossary will be updated to include the definition of this team by the legal team of the project.
- Update the visualisation of the Small-Scale model, the current version it is not the latest.
  Once the data is ready, the visualisations on the small-scale model will be updated. (PENDING)
- Remove the spectrogram visualisation, as stacked bar was more successful.
  The page showing the spectrogram visualisation has been removed from the web application. (DONE)
- Use both a map and a graph to show prediction numbers (some thought the predictions graph is fine, some thought it would be better in a map).
  The Prediction dashboard does now include a map similar to the Historic Data dashboard alongside the bar graph. (DONE)

Changes on the backend:

- Possibility to identify entry points and regions of arrival within the different MS. Data sources and all existing literature will be explored to identify feasibility of this addition. (PENDING)
- Possibility to add new countries of origin such as Somalia, Bangladesh, and Pakistan. GDELT headlines will be downloaded and classified to topic shares for all countries. Then an optimal solution for all new bilateral cases will be developed. (PENDING)
- Add weather data triggering migration flows. Different climate and weather indicators will be examined for correlation with the dependent variable and the best performing ones will be included as new component in the LSM. (PENDING)

8 Appendixes

8.1 Appendix 1 – AI Impact Assessment Questionnaire

IT TOOLS AND METHODS FOR MANAGING MIGRATION FLOWS

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)[29] and by the IEEE[30].

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 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 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.**

[29] 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

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

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.**

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 those 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 number 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	

8.2 Appendix 2 – Eurostat datasets

Category:	Crime and criminal justice	**Data code:**	CRIM_OFF_CAT
Indicator:	Recorded offences by offence category		
Time coverage:	2008-2018	**Frequency:**	Annual
Unit of measure:	Number of offences	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Crime and criminal justice		
Link	https://ec.europa.eu/eurostat/databrowser/view/CRIM_OFF_CAT__custom_511226/default/table?lang=en		
Indicator description and rationale: Police-recorded offences by crime such as: homicide, assault, kidnapping, sexual violence, rape, sexual assault, robbery, burglary, burglary of private residential premises, theft, theft of a motorized land vehicle and drug crimes. The need to provide information on the development of crime in the European Union was recognized in the Hague Programme adopted by the European Council in 2004.			
Category:	Crime and criminal justice	**Data code:**	CRIM_GEN
Indicator:	Crimes recorded by the police by offence category		
Time coverage:	1993-2007	**Frequency:**	Annual
Unit of measure:	Number of crimes	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Crime and criminal justice; Crime - historical data		
Link	https://ec.europa.eu/eurostat/databrowser/view/CRIM_GEN__custom_511357/default/table?lang=en		
Indicator description and rationale: The type of crimes recorded by the police according to their offence category, mainly: (i) Total crime: These figures include offences against the penal code or criminal code. Less serious crimes (misdemeanours) are generally excluded. (ii) Homicide (country and city): This is defined as intentional killing of a person, including murder, manslaughter, euthanasia and infanticide. Causing death by dangerous driving is excluded, as are abortion and help with suicide. Attempted (uncompleted) homicide is also excluded. The counting unit for homicide is normally the victim (rather than the case).			

(iii) Violent crime: This includes violence against the person (such as physical assault), robbery (stealing by force or by threat of force), and sexual offences (including rape and sexual assault).
(iv) Robbery: a sub-set of violent crime (see above). It is defined as stealing from a person with force or threat of force, including muggings (bag-snatching) and theft with violence. Pickpocketing, extortion and blackmailing are generally not included.

(v) Domestic burglary: s gaining access to a dwelling by the use of force to steal goods.
(vi) Theft of a motor vehicle: all land vehicles with an engine that run on the road which are used to carry people (including cars, motorcycles, buses, lorries, construction and agricultural vehicles, etc.).
(vii) Drug trafficking: illegal possession, cultivation, production, supplying, transportation, importing, exporting, financing etc. of drug operations which are not solely in connection with personal use.

The need to provide information on the development of crime in the European Union was recognized in the Hague Programme adopted by the European Council in 2004.

Category:	Demography	**Data code:**	DEMO_PJAN
Indicator:	Population by sex		
Time coverage:	1960-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Demography and migration; Population		
Link	https://ec.europa.eu/eurostat/databrowser/view/DEMO_PJAN__custom_500791/default/table?lang=en		

Indicator description and rationale: The recommended definition is the 'usually resident population' and represents the number of inhabitants of a given area on 1 January of the year in question (or, in some cases, on 31st December of the previous year). However, the population transmitted by the countries can also be either based on data from the most recent census adjusted by the components of population change produced since the last census, either based on population registers.

Usual residence means the place where a person normally spends the daily period of rest, regardless of temporary absences for purposes of recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage.

Population statistics are widely used for planning actions, monitoring and evaluating programmes in a number of policy areas in the social and economic fields.

Category:	Demography	**Data code:**	DEMO_MMONTH
Indicator:	Deaths by month		
Time coverage:	1960-2019	**Frequency:**	Monthly
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Demography and migration; Mortality		
Link	https://ec.europa.eu/eurostat/databrowser/view/demo_mmonth/default/table?lang=en		
Indicator description and rationale: Means the permanent disappearance of all evidence of life at any time after life birth has taken place (postnatal cessation of vital functions without capability of resuscitation). Statistics on deaths and demographic statistics in general are widely used for planning actions and for monitoring and evaluating programmes in a number of social and economic policy areas.			

Category:	Demography	**Data code:**	DEMO_MLEXPEC
Indicator:	Life expectancy by sex		
Time coverage:	1960-2018	**Frequency:**	Annual
Unit of measure:	Years	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Demography and migration; Mortality		
Link	https://ec.europa.eu/eurostat/databrowser/view/DEMO_MLEXPEC__custom_501071/default/table?lang=en		
Indicator description and rationale: Life expectancy at certain ages represents the mean number of years still to be lived by a person who has reached a certain exact age, if subjected throughout the rest of his or her life to the current mortality conditions (age-specific probabilities of dying). Statistics on deaths and demographic statistics in general are widely used for planning actions and for monitoring and evaluating programmes in a number of social and economic policy areas.			
Category:	Demography	**Data code:**	DEMO_NIND

Indicator:	Marriages and crude marriage rate		
Time coverage:	1960-2018	**Frequency:**	Annual
Unit of measure:	Individuals and rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Demography and migration; Marriages and divorces		
Link	https://ec.europa.eu/eurostat/databrowser/view/DEMO_NIND__custom_501269/default/table?lang=en		
Indicator description and rationale: The ratio of the number of marriages during the year to the average population in that year. The value is expressed per 1000 inhabitants. Demographic trend and developments have a strong impact on the societies of the countries and of the EU. Demographic statistics on the size and structure of Europe's population are used in a wide range of economic, social and environmental policy areas.			
Category:	Demography	**Data code:**	YTH_DEMO_010
Indicator:	Child and youth population by sex		
Time coverage:	2008-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Youth; Youth population		
Link	https://ec.europa.eu/eurostat/databrowser/view/YTH_DEMO_010__custom_510810/default/table?lang=en		
Indicator description and rationale: People under 15 years of age and between and between 15 and 29 years of age. The 'usually resident population' and represents the number of inhabitants of a given area on 1 January of the year in question (or, in some cases, on 31st December of the previous year). However, the population transmitted by the countries can also be either based on data from the most recent census adjusted by the components of population change produced since the last census, either based on population registers. Usual residence means the place where a person normally spends the daily period of rest, regardless of temporary absences for purposes of			

recreation, holidays, visits to friends and relatives, business, medical treatment or religious pilgrimage. Population statistics are widely used for planning actions, monitoring and evaluating programmes in a number of policy areas in the social and economic fields			
Category:	Gender equality	**Data code:**	SDG_05_20
Indicator:	Gender pay gap in unadjusted form		
Time coverage:	2002-2018	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 5 - Gender equality		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_05_20/default/table?lang=en		
Indicator description and rationale: The indicator measures the difference between average gross hourly earnings of male paid employees and of female paid employees as a percentage of average gross hourly earnings of male paid employees. The indicator has been defined as unadjusted because it gives an overall picture of gender inequalities in terms of pay and measures a concept which is broader than the concept of equal pay for equal work. All employees working in firms with ten or more employees, without restrictions for age and hours worked, are included. The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 5 on gender equality. In order to promote women's social and economic empowerment, SDG 5 calls for, among other things, recognition and value of unpaid care and domestic work, equal rights and access to economic and natural resources, technology, basic and financial services and property. Equal pay for equal work is one of the European Union's founding principles, embedded in the Treaties since 1957. Article 157 of the Treaty on the Functioning of the European Union provides that each Member State shall ensure that the principle is applied. Directive 2006/54/EC enshrines the principle of equal pay. Closing the gender pay gap is also one of the objectives of the Commission's Strategic engagement for gender equality as well as the European Pillar of Social Rights.			
Category:	Gender equality	**Data code:**	SDG_05_30
Indicator:	Gender employment gap from 20 to 64 by percentage of total population		

Time coverage:	2000-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 5 - Gender equality		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_05_30/default/table?lang=en		

Indicator description and rationale: The indicator measures the difference between the employment rates of men and women aged 20 to 64. The employment rate is calculated by dividing the number of persons aged 20 to 64 in employment by the total population of the same age group. The indicator is based on the EU Labour Force Survey.

The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 5 on gender equality.

In order to promote women's social and economic empowerment, SDG 5 calls for, among other things, recognition and value of unpaid care and domestic work, equal rights and access to economic and natural resources, technology, basic and financial services and property.

Indicator is also included as main indicator in the Social Scoreboard for the European Pillar of Social Rights.

Increasing female labour market participation and the equal economic independence of women and men is one of the priorities of the Strategic Engagement for Gender Equality. This strategy sets the framework for the EU's future work towards gender equality.

Category:	Gender equality	**Data code:**	SDG_05_40
Indicator:	Inactive population from 20 to 64 due to caring responsibilities by sex		
Time coverage:	2000-2029	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 5 - Gender equality		

Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_05_40/default/table?lang=en		
Indicator description and rationale: The indicator measures the reasons why individuals are not actively seeking work, so they are neither employed nor unemployed and considered to be outside the labour force. This definition used in the European Union Labour Force Survey (EU-LFS) is based on the guidelines of the International Labour Organization (ILO). While several reasons may exist why somebody is not seeking employment, only the main one is considered. "Inactivity due to caring responsibilities" refers to the reasons 'looking after children or incapacitated adults' and 'other family or personal responsibilities. The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 5 on gender equality and SDG 8 on decent work and economic growth. In order to promote women's social and economic empowerment, SDG 5 calls for, among other things, recognition and value of unpaid care and domestic work, equal rights and access to economic and natural resources, technology, basic and financial services and property. One of the principles addresses those women and men shall have equal access to special leaves of absence in order to fulfil their caring responsibilities and be encouraged to use them in a balanced way.		
Category:	Gender equality	**Data code:**	SDG_04_50
Indicator:	Employment rates of recent graduates from 20 to 34 by sex		
Time coverage:	2006-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 5 - Gender equality		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_04_50/default/table?lang=en		
Indicator description and rationale: The indicator measures the employment rates of persons aged 20 to 34 fulfilling the following conditions: first, being employed according to the ILO definition, second, having attained at least upper secondary education (ISCED 3) as the highest level of education, third, not having received any education or training in the four weeks preceding the survey and four, having successfully completed their highest educational attainment 1, 2 or 3 years before the survey. The indicator is calculated based on data from the EU Labour Force Survey (EU-LFS). The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 4 on ensuring inclusive and quality education for all and SDG 5 on gender equality.			

SDG 4 seeks to ensure people have access to equitable and quality education through all stages of life, from early childhood education and care, through primary and secondary schooling, to technical, vocational training and tertiary education. In addition to promoting formal qualifications, SDG 4 also aims to increase the number of youths and adults with relevant skills for employment, decent jobs and entrepreneurship. SDG 5 aims at achieving gender equality by, among other things, ending all forms of discrimination, violence, and any harmful practices against women and girls in the public and private spheres.

The indicator provides information on the transition from education to work and helps analyse access to the labour market among recent graduates.

Category:	Migration	**Data code:**	MIGR_EMI2
Indicator:	Emigration by sex		
Time coverage:	1990-2018	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Demography and migration; Emigration		
Link	https://ec.europa.eu/eurostat/databrowser/view/MIGR_EMI2__custom_501207/default/table?lang=en		

Indicator description and rationale: The action by which a person, having previously been usually resident in the territory of a Member State, ceases to have his or her usual residence in that Member State for a period that is, or is expected to be, of at least 12 months.

Migration statistics and demographic statistics in general are widely used to plan actions, evaluate programmes and monitor trends in a number of social and economic policy fields.

Category:	Poverty and social exclusion	**Data code:**	ILC_PEPS01
Indicator:	People at risk of poverty or social exclusion by sex and age group		
Time coverage:	2004-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.

Source in Eurostat database	Database by themes; Population and social conditions; Living conditions and welfare; Income and living conditions; People at risk of poverty or social exclusion; Main indicator - Europe 2020 target on poverty and social exclusion		
Link	https://ec.europa.eu/eurostat/databrowser/view/ILC_PEPS01__custom_508483/default/table?lang=en		
Indicator description and rationale: The collection "People at risk of poverty or social exclusion" houses main indicator on risk of poverty or social inclusion included in the Europe 2020 strategy as well as the intersections between sub-populations of all Europe 2020 indicators on poverty and social exclusion.			
Category:	Poverty and social exclusion	**Data code:**	ILC_PNP1
Indicator:	At-risk-of-poverty rate of older people		
Time coverage:	1995-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Living conditions and welfare; Income and living conditions; Income distribution and monetary poverty; Monetary poverty for elderly people		
Link	https://ec.europa.eu/eurostat/databrowser/view/ILC_PNP1__custom_509188/default/table?lang=en		
Indicator description and rationale: The collection "People at risk of poverty or social exclusion" houses main indicator on risk of poverty or social inclusion included in the Europe 2020 strategy as well as the intersections between sub-populations of all Europe 2020 indicators on poverty and social exclusion.			
Category:	Poverty and social exclusion	**Data code:**	ILC_MDSD01
Indicator:	Material and social deprivation by activity and employment status, and sex		
Time coverage:	2014-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Living conditions and welfare; Income and living conditions; Material deprivation; Material and social deprivation		
Link	https://ec.europa.eu/eurostat/databrowser/view/ILC_MDSD01__custom_509268/default/table?lang=en		

Indicator description and rationale: The collection "Material deprivation" covers indicators relating to economic strain, durables, housing deprivation and environment of the dwelling.			
Category:	Poverty and social exclusion	**Data code:**	ILC_MDDD11
Indicator:	Severe material deprivation by sex		
Time coverage:	2003-2020	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Living conditions and welfare; Income and living conditions; Material deprivation; Material deprivation by dimension		
Link	https://ec.europa.eu/eurostat/databrowser/view/ILC_MDDD11__custom_509358/default/table?lang=en		
Indicator description and rationale: The collection "Material deprivation" covers indicators relating to economic strain, durables, housing deprivation and environment of the dwelling.			
Category:	Poverty and social exclusion	**Data code:**	YTH_INCL_030
Indicator:	Young people not living with parents at-risk-of-poverty or exclusion by age		
Time coverage:	2004-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Youth; Youth social inclusion		
Link	https://ec.europa.eu/eurostat/databrowser/view/YTH_INCL_030__custom_510951/default/table?lang=en		
Indicator description and rationale: People between 16 and 29 years of age that do not live with their parents and that find themselves in a situation of risk of poverty or social exclusion. The collection "People at risk of poverty or social exclusion" houses main indicator on risk of poverty or social inclusion included in the Europe 2020 strategy as well as the intersections between sub-populations of all Europe 2020 indicators on poverty and social exclusion.			

Category:	Poverty and social exclusion	**Data code:**	SDG_01_60
Indicator:	Population living in a dwelling with a leaking roof, damp walls, floors or foundation or rot in window frames of floor by poverty status		
Time coverage:	2003-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 1 - No poverty		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_01_60/default/table?lang=en		

Indicator description and rationale: The indicator measures the share of the population experiencing at least one of the following basic deficits in their housing condition: a leaking roof, damp walls, floors or foundation, or rot in window frames or floor.

The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 1 on ending poverty in all its forms everywhere and SDG 11 on making cities and human settlements inclusive, safe, resilient and sustainable.

SDG 1 calls for an eradication of extreme poverty and for a reduction of relative poverty over the next 15 years. It envisions shared prosperity, a basic standard of living and social protection benefits for people everywhere, including the poorest and most vulnerable. SDG 11 aims to renew and plan cities and other human settlements so that they offer opportunities for all, with access to basic services, energy, housing, transportation, green public spaces and others, while improving resource use and reducing environmental impacts.

Category:	Poverty and social exclusion	**Data code:**	SDG_06_10
Indicator:	Population having neither a bath, nor a shower, nor indoor flushing toilet in their household by poverty status		
Time coverage:	2003-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 1 - No poverty		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_06_10/default/table?lang=en		

Indicator description and rationale: The indicator measures the share of total population having neither a bath, nor a shower, nor an indoor flushing toilet in their household.

The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 6 on clean water and sanitation and SDG 1 on ending poverty in all its forms everywhere.

Among other things, SDG 6 calls for ensuring universal access to safe and affordable drinking water, sanitation and hygiene, and ending open defecation. It also aims at improving water quality and water-use efficiency and encouraging sustainable abstractions and supply of freshwater. SDG 1 calls for an eradication of extreme poverty and for a reduction of relative poverty over the next 15 years. Considering the indicator, a bath, shower and indoor flushing toilet are basic facilities of a household. Their availability plays an important role in healthy livelihoods and people's well-being. A lack of these facilities is considered as housing deprivation. Furthermore, their accessibility, specifically within one's household, contributes to the end of open defecation and therefore promotes a healthy environment and reduces human health risks.

Category:	Poverty and social exclusion	**Data code:**	SDG_07_60
Indicator:	Population unable to keep home adequately warm by poverty status		
Time coverage:	2003-2020	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	Data for 2020 not yet available
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 1 - No poverty		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_07_60/default/table?lang=en		

Indicator description and rationale: The indicator measures the share of population who are unable to keep home adequately warm. Data for this indicator are being collected as part of the European Union Statistics on Income and Living Conditions (EU-SILC) to monitor the development of poverty and social inclusion in the EU. The data collection is based on a survey, which means that indicator values are self-reported.

Indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 7 on affordable and clean energy and SDG 1 on ending poverty in all its forms everywhere.

SDG 7 among other things calls for ensuring universal access to modern energy services, improving energy efficiency and increasing the share of renewable energy and to accelerate the transition to an affordable, reliable, and sustainable energy system. SDG 1 calls for an eradication of extreme poverty and for a reduction of relative poverty over the next 15 years and seeks among other things to ensure equal rights and access to

economic and natural resources as well as technology.			
Category:	Poverty and social exclusion	**Data code:**	SDG_11_10
Indicator:	Overcrowding rate by poverty status		
Time coverage:	2003-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Tables on EU policy; Sustainable development indicators; Goal 1 - No poverty		
Link	https://ec.europa.eu/eurostat/databrowser/view/sdg_11_10/default/table?lang=en		
Indicator description and rationale: The indicator measures the share of people living in overcrowded conditions in the EU. A person is considered to be living in an overcrowded household if the house does not have at least one room for the entire household as well as a room for a couple, for each single person above 18, for a pair of teenagers (12 to 17 years of age) of the same sex, for each teenager of different sex and for a pair of children (under 12 years of age). The indicator is part of the EU Sustainable Development Goals (SDG) indicator set. It is used to monitor progress towards SDG 11 on making cities and human settlements inclusive, safe, resilient and sustainable. Among other things, SDG 11 aims to renew and plan cities and other human settlements so that they offer opportunities for all, with access to basic services, energy, housing, transportation, green public spaces and others, while improving resource use and reducing environmental impacts.			
Category:	Health	**Data code:**	HLTH_SILC_14
Indicator:	Self-reported unmet needs for medical examination by sex, age and main reason declared		
Time coverage:	2008-2020	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Cross cutting topics; Quality of Life; Health; Access to healthcare		

Link	https://ec.europa.eu/eurostat/databrowser/view/hlth_silc_14/default/table?lang=en		
Indicator description and rationale:			
Category:	Health	**Data code:**	HLTH_SILC_02$DV_381
Indicator:	Self-perceived health by sex and age		
Time coverage:	2008-2020	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Cross cutting topics; Quality of Life; Health; Health Status		
Link	https://ec.europa.eu/eurostat/databrowser/view/hlth_silc_02$DV_381/default/table?lang=en		
Indicator description and rationale:			
Category:	Work life	**Data code:**	LFSA_EWHUN2%24DV_603
Indicator:	Average number of usual weekly hours of work in main job by sex, professional status, full-time/part-time and economic activity		
Time coverage:	2007-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Cross cutting topics; Quality of employment; Work Life Balance		
Link	https://ec.europa.eu/eurostat/databrowser/view/lfsa_ewhun2%24DV_603/default/table?lang=en		
Indicator description and rationale:			
Category:	Natural and living environment	**Data code:**	ILC_PW01$DV_527
Indicator:	Average satisfaction by sex and age		
Time coverage:	2013; 2018	**Frequency:**	Annual
Unit of	Rating	**Coverage:**	Europe

measure:			
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Cross cutting topics; Quality of life; Natural and Living Environment; Landscape and Built Environment		
Link	https://ec.europa.eu/eurostat/databrowser/view/ilc_pw01$DV_527/default/table?lang=en		
Indicator description and rationale:			
Category:	Asylum and managed migration	**Data code:**	MIGR_ASYAPPCTZA
Indicator:	Number of first-time asylum applicant by sex and age group		
Time coverage:	2008-2020	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Asylum and managed migration; Asylum and Dublin statistics; Applications		
Link	https://ec.europa.eu/eurostat/databrowser/view/MIGR_ASYAPPCTZA__custom_503564/default/table?lang=en		
Indicator description and rationale:			
Category:	Asylum and managed migration	**Data code:**	MIGR_EIPRE
Indicator:	Number of third country nationals found to be illegally present by sex and age group		
Time coverage:	2008-2020	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Asylum and managed migration; Enforcement of Immigration Legislation		
Link	https://ec.europa.eu/eurostat/databrowser/view/MIGR_EIPRE__custom_504205/default/table?lang=en		
Indicator description and rationale:			

Category:	Asylum and managed migration	**Data code:**	MIGR_EIRT_ASS
Indicator:	Number of third country nationals who have left the territory by type of assistance received		
Time coverage:	2014-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database	Database by themes; Population and social conditions; Asylum and managed migration; Enforcement of Immigration Legislation		
Link	https://ec.europa.eu/eurostat/databrowser/view/migr_eirt_ass/default/table?lang=en		
Indicator description and rationale:			

Category:	Asylum and managed migration	**Data code:**	MIGR_RESFIRST
Indicator:		First permits by length of validity	
Time coverage:	2008-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Database by themes; Population and social conditions; Asylum and managed migration; Residence permits; Residence permits by reason, length of validity and citizenship	
Link		https://ec.europa.eu/eurostat/databrowser/view/MIGR_RESFIRST__custom_506015/default/table?lang=en	
Indicator description and rationale:			
Category:	Asylum and managed migration	**Data code:**	MIGR_RESVALID
Indicator:		All valid permits by length of validity	
Time coverage:	2008-2020	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe

Source	Eurostat	**Comments:**	Data for 2020 not yet available
Source in Eurostat database		Database by themes; Population and social conditions; Asylum and managed migration; Residence permits; Residence permits by reason, length of validity and citizenship	
Link		https://ec.europa.eu/eurostat/databrowser/view/MIGR_RESVALID__custom_506128/default/table?lang=en	
Indicator description and rationale:			
Category:	Asylum and managed migration	**Data code:**	MIGR_RESFRPS1
Indicator:		First permits issued for family reunification with a beneficiary of protection status by reason	
Time coverage:	2016-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	Data partially available
Source in Eurostat database		Database by themes; Population and social conditions; Asylum and managed migration; Residence permits; Residence permits by reason, length of validity and citizenship	
Link		https://ec.europa.eu/eurostat/databrowser/view/MIGR_RESFRPS1__custom_506176/default/table?lang=en	
Indicator description and rationale:			
Category:	Asylum and managed migration	**Data code:**	MIGR_RESFRPS2
Indicator:		Permits valid at the end of the year for family reunification with a beneficiary of protection status	
Time coverage:	2016-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	Data partially available
Source in Eurostat database		Database by themes; Population and social conditions; Asylum and managed migration; Residence permits; Residence permits by reason, length of validity and citizenship	
Link		https://ec.europa.eu/eurostat/databrowser/view/MIGR_RESFRPS2__custom_506230/default/table?lang=en	
Indicator description and rationale:			
Category:	Education and training	**Data code:**	EDAT_LFSE_03
Indicator:		Population by educational attainment level	
Time coverage:	1992-2019	**Frequency:**	Annual

Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Database by themes; Population and social conditions; Education and training; Education and training outcomes; Education attainment level; Population by educational attainment level	
Link		https://ec.europa.eu/eurostat/databrowser/view/EDAT_LFSE_03__custom_506663/default/table?lang=en	
Indicator description and rationale:			
Category:	Quality education	**Data code:**	SDG_04_50
Indicator:		Employment of recent graduates from 20 to 34 by sex	
Time coverage:	2006-2019	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Tables on EU policy; Sustainable development indicators; Goal 4 - Quality education	
Link		https://ec.europa.eu/eurostat/databrowser/view/sdg_04_50/default/table?lang=en	
Indicator description and rationale:			
Category:	Quality education	**Data code:**	SDG_04_60
Indicator:		Adult participation in learning from 25 to 64 by sex	
Time coverage:	2000-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Tables on EU policy; Sustainable development indicators; Goal 4 - Quality education	
Link		https://ec.europa.eu/eurostat/databrowser/view/sdg_04_60/default/table?lang=en	
Indicator description and rationale:			
Category:	Peace, justice and strong institutions	**Data code:**	SDG_16_40
Indicator:		Perceived independence of the justice system	
Time coverage:	2016-2019	**Frequency:**	Annual

Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Tables on EU policy; Sustainable development indicators; Goal 16 - Peace, justice and strong institutions	
Link		https://ec.europa.eu/eurostat/databrowser/view/sdg_16_40/default/table?lang=en	
Indicator description and rationale:			
Category:	Peace, justice and strong institutions	**Data code:**	SDG_16_50
Indicator:		Corruption Perceptions (source Transparency International)	
Time coverage:	2012-2019	**Frequency:**	Annual
Unit of measure:	Index	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Tables on EU policy; Sustainable development indicators; Goal 16 - Peace, justice and strong institutions	
Link		https://ec.europa.eu/eurostat/databrowser/view/sdg_16_50/default/table?lang=en	
Indicator description and rationale:			
Category:	Peace, justice and strong institutions	**Data code:**	SDG_16_60
Indicator:		Population with confidence in EU institutions by institution (source DG COMM)	
Time coverage:	2000-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Tables on EU policy; Sustainable development indicators; Goal 16 - Peace, justice and strong institutions	
Link		https://ec.europa.eu/eurostat/databrowser/view/sdg_16_60/default/table?lang=en	
Indicator description and rationale:			
Category:	City statistics	**Data code:**	lfst_rimgpnga
Indicator:		City population by citizenship and country of birth	

Time coverage:	1990-2019	**Frequency:**	Annual
Unit of measure:	Individuals	**Coverage:**	Europe (cities)
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; City statistics	
Link		https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do	
Indicator description and rationale:			
Category:	Social Inclusion	**Data code:**	URB_CPOPCB
Indicator:		People at risk of poverty or social exclusion by broad group of citizenship	
Time coverage:	2003-2020	**Frequency:**	Annual
Unit of measure:	Rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Social inclusion; People at risk of poverty and social exclusion	
Link		https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do (first link works for Daniel (UAB)), for us in Kiel only the following works https://data.europa.eu/euodp/en/data/dataset/mSBr8zbYzXySNiV8mItuw	
Indicator description and rationale:			
Category:	Social inclusion	**Data code:**	ilc_lvho25
Indicator:		Housing cost overburden by age and sex	
Time coverage:	2003-2020	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Social inclusion; Living conditions	
Link		https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_lvho25&lang=en see above, does this work for you?	
Indicator description and rationale:			
Category:	Active citizenship	**Data code:**	migr_resshare

Indicator:		Long-term residents among all non-EU citizens holding residence permits by citizenship	
Time coverage:	2008-2019	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Active citizenship	
Link		https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_resshare&lang=en see above	
Indicator description and rationale:			
Category:	Active citizenship	**Data code:**	migr_acqs
Indicator:		Residents who acquired citizenship as a share of resident non-citizens by former citizenship and sex	
Time coverage:	2008-2018	**Frequency:**	Annual
Unit of measure:	Percentage	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Active citizenship	
Link		https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_acqs&lang=en	
Indicator description and rationale:			
Category:	Recent immigrants	**Data code:**	lfst_rimgpnga
Indicator:		Recent immigrants by sex, age and citizenship	
Time coverage:	2008-2019	**Frequency:**	Annual
Unit of measure:	Percentage of total population	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Recent immigrants	
Link		https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_rimgpnga&lang=en	
Indicator description and rationale:			
Category:	Recent immigrants	**Data code:**	lfst_rimgenga

Indicator:		Employed recent immigrants by sex, age and citizenship	
Time coverage:	2008-2019	**Frequency:**	Annual
Unit of measure:	Percentage of total recent immigrants	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database		Cross cutting topics; Migrant integration and children in migration; Migrant integration; Recent immigrants	
Link		https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_rimgenga&lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	UNE_RT_M
Indicator:		Unemployment by sex and age	
Time coverage:	1983-2020	**Frequency:**	monthly
Unit of measure:	Percentage of active population	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/une_rt_m/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	ILC_DI03
Indicator:		Mean and median income by age and sex	
Time coverage:	1995-2019	**Frequency:**	yearly
Unit of measure:	Euro	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/ilc_di03/default/table?lang=en	
Indicator description and rationale:			

Category:	Demography	**Data code:**	DEMO_PJAN
Indicator:		Population on 1 January by age and sex	
Time coverage:	2008-2019	**Frequency:**	yearly
Unit of measure:	number	**Coverage:**	Europe
Source	Eurostat	**Comments:**	here we care about the age, so we need to know what age inhabitants have
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/demo_pjan/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	EARN_MW_CUR
Indicator:		Monthly minimum wages	
Time coverage:	1999-2021	**Frequency:**	half-yearly
Unit of measure:	Euro	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/earn_mw_cur/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	TESSI190
Indicator:		Gini coefficient of equivalised disposable income	
Time coverage:	2009-2020	**Frequency:**	yearly
Unit of measure:	n.a.	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/tessi190/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	UNE_LTU_Q
Indicator:		Long-term unemployment by sex	

Time coverage:	1992-2020	**Frequency:**	quarterly
Unit of measure:	Percentage of active population	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/une_ltu_q/default/table?lang=en	
Indicator description and rationale:			
Category:	Work life	**Data code:**	HSW_MI01
Indicator:		Accidents at work by sex, age and severity	
Time coverage:	2008-2019	**Frequency:**	annually
Unit of measure:	standardised incidence rate	**Coverage:**	Europe
Source	Eurostat	**Comments:**	separately for "4 days and over" and "fatal"
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/hsw_mi01/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	NAMA_10_GDP
Indicator:		GDP and main components	
Time coverage:	1975-2020	**Frequency:**	annually
Unit of measure:	Euro	**Coverage:**	Europe
Source	Eurostat	**Comments:**	please generate an annual growth rate in percentage points as well: (GDP_{2020}-GDP_{2019})/GDP_{2019}
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/nama_10_gdp/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	NAMA_10_LP_ULC

Indicator:		Real labour productivity per person	
Time coverage:	1975-2020	**Frequency:**	annually
Unit of measure:	Index, 2010=100	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/nama_10_lp_ulc/default/table?lang=en	
Indicator description and rationale:			
Category:	Social security	**Data code:**	SPR_EXP_SUM
Indicator:		Social protection: Expenditure: main results	
Time coverage:	1990-2018	**Frequency:**	annually
Unit of measure:	Euro per inhabitant	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/spr_exp_sum/default/table?lang=en	
Indicator description and rationale:			
Category:	Social security	**Data code:**	SPR_EXP_PENS
Indicator:		Social protection: Pensions	
Time coverage:	1990-2019	**Frequency:**	annually
Unit of measure:	Euro per inhabitant	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/spr_exp_pens/default/table?lang=en	
Indicator description and rationale:			
Category:	Economic conditions	**Data code:**	PRC_PPP_IND

Indicator:		Purchasing power parities (PPPs)	
Time coverage:	1995-2019	**Frequency:**	annually
Unit of measure:	n.a.	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/prc_ppp_ind/default/table?lang=en	
Indicator description and rationale:			
Category:	Work life	**Data code:**	TESEM110
Indicator:		Temporary employees as percentage of the total number of employees	
Time coverage:	2005-2019	**Frequency:**	annually
Unit of measure:	percentage of the total number of employees	**Coverage:**	Europe
Source	Eurostat	**Comments:**	n.a.
Source in Eurostat database			
Link		https://ec.europa.eu/eurostat/databrowser/view/tesem110/default/table?lang=en	
Indicator description and rationale:			