



FINAL REPORT

Project reference:

MYGEOSS Third Call For Innovative Apps in the environmental and social domains
(http://digitalearthlab.jrc.ec.europa.eu/mygeoss/info_thirdcall.cfm)

Name of the App:

MIGRATE – MIGRation pATterns in Europe



Responsible for the App:

Dr. Marco Minghini

via Boccioni 11, 22100 Como (Italy)

E-mail address: marco.minghini86@gmail.com

Expert contract number: CCR.IES.C393200.X0

1. Introduction: Description and purpose of the App

MIGRATE – MIGRation pATterns in Europe (<http://geomobile.como.polimi.it/migrate>) is a Web mapping application aimed at educating and raising awareness about the phenomenon of migration in Europe. This goal is achieved using a gamification approach, i.e. users interact with the application by playing a map-based game, where questions are asked, and answers are provided and explained, about a number of topics related to migration. Answers provided by the users will help to understand the current knowledge and perception of migration-related issues. During the MYGEOSS Final Event in Brussels on December 6, 2016, the first MIGRATE challenge was officially announced and launched. This challenge will be open until January 5, 2016 and will give awards (in the form of Amazon vouchers) to the top 10 players of the gaming app (see <http://geomobile.como.polimi.it/migrate/challenge>).

As the final report for MYGEOSS project, this document summarizes the work done, outlines its current and future development, and offers a number of considerations and reflections on different aspects of the development of the app, in particular data access and discovery and software requirements. Lessons learnt and recommendations for the European Commission for future programmes with similar aims as MYGEOSS are also provided, together with a list of all the materials produced for the MIGRATE project.

2. Data sources and licenses

As migration fluxes began to increase and, as a consequence, they strongly affected population growing rates, as well as economy of the hosting and departure countries, many national and international agencies started to collect data about human population and its movements all over the world. These databases provide a huge amount of complete and updated data, sometimes also going back in time, in order to show evolutions and trends. Thus, alongside data collection, also statistics, indices and estimates are retrieved from them. This information is often represented through static or interactive maps and graphs, which allow for a more intuitive and immediate understanding of the phenomena.

2.1 Overview of available data on migration

Available data on migration can be roughly divided into two categories, according to the specific mission of the agency or organization collecting them. The first category is mainly related to population statistics (i.e. growing rates, demography, social protection and living conditions) and how they change as function of migration fluxes, primarily the legal ones. Main data sources of this type are national and international census agencies such as Eurostat (<http://ec.europa.eu/eurostat>), US Census (<http://www.census.gov>) and Istat (<http://www.istat.it/it>). The second category includes data strictly connected to migration, regarding the number of migrants, refugees and asylum seekers arriving to host countries (sometimes also through illegal routes), the number of dead and missing people and main origins and destinations. Data collected from organizations as UNHCR (United Nations High Commissioner for Refugees, <http://www.unhcr.org>) and IOM (International Organization for Migration, <http://www.iom.int>) belong to this category.

The Internet has also plenty of web pages and blogs where not only important organizations but also single individuals as journalists or researchers collect data, information, statistics and write about migration and migrants' stories. An example of a very famous blog is Fortress Europe

(<http://fortresseurope.blogspot.it>). In some cases also interactive web applications have been developed to show collected data in a graphical way, see e.g. <http://www.global-migration.info> and <http://www.therefugeeproject.org/#/2015>.

2.2 Data access and discovery: achievements and limitations

The main criteria used for selecting the data to be used in MIGRATE were license, easy accessibility, and significance to the aims of the application, which means they provide up-to-date information useful to give a complete insight into migration phenomenon to let users really understand its dynamics and proportions (i.e. migration flows around the world, migration routes, types of migrants, deaths during their travels and population estimates in relation with migration). First, in order to fully comply with the guidelines of the MYGEOSS Third Call, all the selected data are available under the conditions of full and open access with no restrictions for reuse except for the obligation to mention the source of the input data. The data sources used, which are briefly mentioned below, are listed in detail in the Data Management Plan which was delivered in August 2016. Licenses of these datasets are ODbL, CC BY, CC BY-IGO, CC BY-SA, ODbL, ODC-BY, PDDL, CC0 and other customized open licenses.

The main open data source exploited is the United Nations High Commissioner for Refugees (UNHCR, <http://www.unhcr.org>), the UN agency which is mandated to lead and co-ordinate international action to protect refugees, safeguard their rights and well-being ensuring that everyone can exercise the right to seek asylum and find safe refuge in another State, return home voluntarily or to resettle in a third country. UNHCR provides data about population statistics on migration, as detailed in the Data Management Plan. What made UNHCR datasets an important source of information for MIGRATE is that these data provide details about the composition of populations of concern (refugees, asylum seekers, internally displaced persons, returned refugees and internally displaced persons, stateless persons and other people who do not exactly fall into one of the previous categories) and allow to understand structure, composition and evolution over space and time. Another organization providing up-to-date, open datasets about migration flows throughout the world is the International Organization for Migration (IOM, <https://www.iom.int>), the leading inter-governmental organization in the field of migration. IOM provides updated statistics on migration flows to Europe and on the number of dead and missing migrants. Sources and licenses are detailed as well in the Data Management Plan. MIGRATE makes also use of datasets derived from the Migrants' Files project (<http://www.themigrantsfiles.com>), a detailed database created by a consortium of journalists about migrants' deaths in their attempt to reach or stay in Europe from 2000 to 2016. The database includes data about the money spent by migrants to reach Europe, the money spent by the European governments to stop/limit migration and to deport migrants, and the count of migrants' deaths. Datasets exploited in MIGRATE include also open datasets on European populations provided by Eurostat (<http://ec.europa.eu/eurostat>), the EU statistical office which provides reliable and objective statistics at the European level enabling comparisons between countries and regions. Finally, all the geospatial data used in MIGRATE (i.e. the land borders of the world countries) are extracted from OpenStreetMap (OSM, <http://www.openstreetmap.org>), the global open geospatial database created by volunteers.

The main reason preventing the use of the other existing data on migration is their license. In some cases there is no clear statement of the license; as an example Frontex, the European Union Agency which promotes, coordinates and develops European border management, provides migration route maps which are only available after a written request

(<http://frontex.europa.eu/trends-and-routes/migratory-routes-map>). Similarly, two web applications provide interesting data about migration. The diagram developed within the “peplemovin” initiative (http://peplemov.in/#f_VN) shows flows of migrants based on the World Bank latest available data (<http://data.worldbank.org/indicator/SM.POP.TOTL>)¹, while The Refugee Project (<http://www.therefugeeproject.org/#/2015>) is a temporal map of refugee migrations since 1975 based on UNHCR and UN data. Unfortunately, in both these cases the data license is not stated at all. In other cases data are available for personal and non commercial use only, as population prospects from UN (see <http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates15.shtml> and <https://esa.un.org/unpd/wpp>), migration statistics from the OECD International Migration Database (<https://stats.oecd.org/Index.aspx?DataSetCode=MIG>), migration flows data from the International Centre for Migration Policy Development (<http://www.imap-migration.org/index.php?id=39>) and estimates of migration flows between and within regions from all over the world developed by three researchers (<http://www.global-migration.info>). Regarding geospatial datasets, an alternative to OSM is the GADM database of Global Administrative Areas (<http://gadm.org>), which was not used due to its license (it is freely available for academic use and other non-commercial uses, while redistribution or commercial use is not allowed without prior permission) and to its coarser detail compared to OSM.

3. Software requirements

As stated in the MIGRATE Technical documentation delivered in September 2016, all the software used for the development is FOSS (Free and Open Source Software). In particular, the application is based on the Django framework, which is built in Python and was selected as it manages in a standard manner all network communication between client and server and provides a good level of security for the game, while allowing an agile development. Open source software proved to be fully suitable for the requirements of the application and allowed a high degree of customization according to the needs.

MIGRATE was conceived as a gaming application since the very beginning, however the team did not initially consider to give real/monetary awards to the top players. Based on the team’s experience in citizen science and participatory projects, the latter strategy was then adopted for two main reasons: a) to motivate users in playing MIGRATE multiple times and as seriously as possible; and b) as a consequence, to collect more data, and more reliable data. This sharp change of strategy, which happened towards the end of the development phase, required to implement a number of non trivial changes in the code to ensure the highest level of security of the application. This allowed the team to cover security issues that may arise from malicious users that would damage the integrity of the surveyed data.

The collected data is available through the game platform for third parties. This is possible due to careful considerations about users’ privacy from the beginning of the project, since during the registration phase the players are not required to provide specific information about themselves that would make them easily identifiable (e.g. name, surname, e-mail address and postal address). In order to deliver the awards (in the form of Amazon vouchers) at the end of the challenge to the top 10 players, the contact with them will happen by e-mail by simply publishing a message on the MIGRATE website asking them

¹ MIGRATE does not rely on data from World Bank as they are already retrieved from UNHCR, see <http://data.worldbank.org/indicator/SM.POP.REFG?end=2014&start=1990&view=chart>

to contact the organizing team. The personal information submitted during the registration phase will be checked in order to verify the players' identities.

An important lesson was learnt about how to design an effective game. A complex algorithm was defined for computing the user's score so that: a) the top players will be neither those playing most times nor those playing in average the best (i.e. those scoring the highest average scores), but those who both play many times and with a high average score; and b) any user, even if starting to play later than the beginning of the challenge, can still become the top player. A non secondary aspect to take into account in the game design was the need to change part of the game questions at predefined times during the challenge. This is done to make the game more challenging and entertaining and to avoid that, after playing many times, users "have learnt" the questions and just answer without even reasoning about migration. The experience gained on this kind of design will be highly beneficial for the team in future works.

4. Lessons learnt and recommendations for the European Commission

After the successful experience of MIGRATE, we are happy to share a number of lessons learnt and offer some recommendations to the European Commission for future programmes similar to MYGEOSS.

First, as it is clear from Section 2, we experienced the difficulties and problems arising during the analysis of the policies on data access and use, especially considering the requirements on open access stated by the MYGEOSS call. In particular, we found that it is not always easy to find the license specifications of the datasets (which sometimes are hidden or not clearly shown) as well as understand the real "openness" of the datasets. Some datasets which at first appeared to be openly-licensed were then discovered to be not. In this regard, we suggest that the specifications of license conditions of the data which can be used within the application is explained in a more detailed way in the call, perhaps adding a list of data licenses which are compliant with conditions required. As a matter of fact, the call only stated that the data had to be "available under the conditions of full and open access with no restrictions for reuse except for the obligation to acknowledge the source of the input data", and that the data had to be compliant with the GEOSS Data Core license terms. However, when looking at the data licenses, it was sometimes hard to understand whether their licenses were compliant with these conditions. In addition, the call shall stress more the importance of this and ask proponents to carefully check the data policy of their data sources before submitting the proposal. In the case of MIGRATE, some of the data sources indicated in the project proposal were only later discovered to be not usable due to license constraints.

A second important aspect is about the data collected/produced by the application itself – in the case of MIGRATE, the answers submitted by the game players. The call shall inform proponents that all the data and information collected within the project will be released as open data, and clearly state the license, or the possible licenses, under which these data can be released. In the case of MIGRATE, we initially planned to release data under a CC BY-SA license, but were then suggested to switch to a CC BY license (as the SA is a restriction, not a condition). We feel this aspect is crucial and should be better stressed in the call, as most of today's apps are specifically aimed at collecting data.

A third aspect is about the privacy statement related to the user's personal information submitted through the application – in the case of MIGRATE, all the personal data submitted by users during the registration phase: gender, age range, country of origin, educational level. The call shall state that, in case the app aims to collect personal data, a privacy statement must be produced; the call may also make reference to an example of privacy statement that proponents can use as a reference.

Finally, some minor suggestions about the contents of the call are briefly listed in the following:

- proponents shall be invited not to rely on a single data source, in other words the evaluation of proposals should strongly consider the variety and significance of the data sources used;
- in the proposal, proponents shall clearly define who are/would be the potential users of the app, to quantify them, and to draft a plan explaining which strategies will be put in place to reach/engage these users. This aspect is crucial to avoid that the apps, although technically perfect, would be unused or underused after their final release. In the case of MIGRATE, we decided to open the challenge with monetary awards also to avoid this risk.
- proponents shall carefully check the quality of the datasets they plan to use in their project; they shall be required to define all the available quality parameters (up-to-dateness, scale/positional accuracy, temporal accuracy, semantic accuracy, etc.) in the proposal.
- regarding open data which are released from the European Commission (e.g. those belonging to the GEOSS Data Core), a specific functionality shall be implemented which allows users of those data to provide feedback about their quality or fitness-for-use. This can also help in keeping track of how much and for which purposes data are used.

5. Current and future work

As already mentioned, at the time of writing (December 2016) the first MIGRATE challenge is in progress and it has already achieved successful results: more than 220 players have registered, more than 150 have played at least once, and the 10 most active have played at least 100 times. As stated in the official page (<http://geomobile.como.polimi.it/migrate/challenge>), the challenge will end on January 5, 2017 and the top 10 players will be awarded with Amazon vouchers (for a total amount of 200 €) by February 15, 2017. All the answers submitted by users are available for download in JSON format, and under a CC BY license, from <http://geomobile.como.polimi.it/migrate/download>.

The work on MIGRATE will not stop after the end of the first challenge. The team will carefully analyse all the data collected (game answers, scores and user profiles) and use statistical techniques to identify trends in the perception of migration according to the personal information submitted by the users (gender, age range, area of origin and educational level). Before receiving the awards, the top 10 players will be also briefly interviewed to understand how their perception of migration has changed thanks to the game and what lessons they have learnt. Results of this analysis will be published in at least one academic paper and presented in at least one academic conference. The paper(s) and the presentation(s) will carry appropriate attribution to the MYGEOSS project and the EU's H2020 research and innovation program which has funded it. Finally, the team will also evaluate the opportunity of updating the MIGRATE game with new questions (perhaps based on new open data sources, if available) and launch a second challenge.

Finally, the following list provides links to all the materials produced for MIGRATE:

- MIGRATE official website: <http://geomobile.como.polimi.it/migrate>
- repository of MIGRATE source code: <https://github.com/kilsedar/migrate>
- MIGRATE Data Management Plan:
<https://github.com/kilsedar/migrate/blob/master/deliverables/DataManagementPlan.pdf>
- MIGRATE Technical Documentation:
<https://github.com/kilsedar/migrate/blob/master/deliverables/TechnicalDocumentation.pdf>
- MIGRATE Final Report:
<https://github.com/kilsedar/migrate/blob/master/deliverables/FinalReport.pdf>
- data produced by MIGRATE (CC BY): <http://geomobile.como.polimi.it/migrate/download>
- video explaining how to play MIGRATE: https://www.youtube.com/watch?v=CGWVm_XIXKo
- MIGRATE presentation given at the GEO-XIII Plenary, St. Petersburg (Russia) on November 8, 2016: <http://www.slideshare.net/mingo23/migrate-migration-patterns-in-europe>
- MIGRATE presentation given at the MYGEOSS Final Event, Brussels (Belgium) on December 6, 2016: <https://github.com/kilsedar/migrate/blob/master/deliverables/Presentation.pdf>
- video of MIGRATE presentation given at the MYGEOSS Final Event, Brussels (Belgium) on December 6, 2016: <https://www.youtube.com/watch?v=oxb26HTHdOk&t=10s>
- MIGRATE poster presented at the MYGEOSS Final Event, Brussels (Belgium) on December 6, 2016: <https://github.com/kilsedar/migrate/blob/master/deliverables/Poster.pdf>
- interview to Dr. Marco Minghini about MIGRATE at the MYGEOSS Final Event, Brussels (Belgium) on December 6, 2016:
https://twitter.com/EU_ScienceHub/status/806153538493628416
- MIGRATE presentation given by Prof. Maria Antonia Brovelli at Politecnico di Milano channel (in English): <https://youtu.be/Z5789ijmyxU>
- MIGRATE presentation given by Prof. Maria Antonia Brovelli at Politecnico di Milano channel (in Italian): <https://youtu.be/f3HDTTSqbhQ>

Como, December 15, 2016

Marco Minghini

