Article

The Role of Emerging Predictive IT Tools in Effective Migration Governance

Cristina Blasi Casagran [1,*,†], Colleen Boland [1,†], Elena Sánchez-Montijano [1,2,†] and Eva Vilà Sánchez [1,†]

[1] Department of Public Law and Legal History Studies, Autonomous University of Barcelona, Spain;
E-Mails: cristina.blasi@uab.es (C.B.C.), colleen.boland@uab.cat (C.B.), eva.vilas@e-campus.uab.cat (E.V.S.)
[2] Department of International Studies, Center for Research and Teaching in Economics (CIDE), Mexico;
E-Mail: elena.sanchez@cide.edu (E.S.-M.)

* Corresponding author

[†] These authors contributed equally to this work

Abstract
Predicting mass migration is one of the main challenges for policymakers and NGOs working with migrants worldwide. Recently there has been a considerable increase in the use of computational techniques to predict migration flows, and advances have allowed for application of improved algorithms in the field. However, given the rapid pace of technological development facilitating these new predictive tools and methods for migration, it is important to address the extent to which such instruments and techniques engage with and impact migration governance. This study provides an in-depth examination of selected existing predictive tools in the migration field and their impact on the governance of migratory flows. It focuses on a comparative qualitative examination of these tools' scope, as well as how these characteristics link to their respective underlying migration theory, research question, or objective. It overviews how several organisations have developed tools to predict short- or longer-term migration patterns, or to assess and estimate migration uncertainties. At the same time, it demonstrates how and why these instruments continue to face limitations that in turn affect migration management, especially as it relates to increasing EU institutional and stakeholder efforts to forecast or predict mixed migration. The main predictive migration tools in use today cover different scopes and uses, and as such are equally valid in shaping the requirements for a future, fully comprehensive predictive migration tool. This article provides clarity on the requirements and features for such a tool and draws conclusions as to the risks and opportunities any such tool could present for the future of EU migration governance.

1. Introduction and Theoretical Framework

Migration is unpredictable. Since 2015, with the advent of the so-called "migration crisis," different institutions within the European context have allocated extensive resources and funding in seeking migration forecasting or predictive tools (European Commission, 2021). The main objective is to respond to the continual demand for strategic and effective global migration governance (Robinson, 2018; Triandafyllidou, 2020).

Indeed, international migration flows have become more diversified and globalised, and are notably marked by increasingly restrictive admission policies, relative change in labour market dynamics, issues of legal irregularity and human trafficking, and new transnational networks and spaces (Arango, 2018). Both institutions

and academia have identified different drivers that could impact these migration flows (Castles et al., 2014; Massey, 1999). Whereas some contend that factors such as conflict, the economy, and climate are the main reasons behind a decision to emigrate, others hold that sudden disruptive elements such as political events or social unrest in a territory could lead to a decision to leave a country (Carling & Collins, 2018; Raleigh, 2011). As international migration flows are complex and uncertain, the need for equally complex tools that help understand and manage migration becomes imperative.

Meanwhile, incorporating algorithmic governance has become common practice internationally, and today new technologies are increasingly embedded in everyday decision-making. The potential of computational power can aid in addressing various gaps in understanding migration flows, and thus benefit policy. For example, new technologies enable, inter alia, conducting subnational forecasts within international migration, considering dimensions like labour force status and ethnic groups, examining micro and macro factors involved in international migration (by harnessing the potential of micro models alongside macrosimulations), and simulating sets of scenarios to understand migration drivers for policy (Wilson & Rees, 2005, pp. 340–341). It should be noted that from a temporal point of view, forecasting is a process of predicting or estimating future events based on past and present data, whereas predictions consist of estimating the outcomes for unseen data (i.e., new or test data); for the purposes of this article, the two terms (forecasting and prediction) are used interchangeably.

Most importantly, however, the way international migration governance is understood and practiced can be mutually configured and shaped by technologies for migration forecasting. In this sense, international migration governance is a contested field with competing interests and stakeholders, and predictive tools exercise the potential to introduce or reinforce unequal power relations. In utilizing these tools, those states with more technological capabilities can further solidify their position in setting the international migration agenda (Beduschi, 2020). Such tools can also be appropriated to further securitize or bolster non-entrée policies and human surveillance, at the expense of those rights protected by international human rights frameworks (Broeders & Dijstelbloem, 2016). This is important to bear in mind in seeking a truly effective global migration governance oriented towards adequately benefiting all stakeholders, especially migrants themselves.

In the case of the EU, efforts are underway to achieve early warning of migratory movements, to forecast them or to predict mixed migration flows in particular (Sohst et al., 2020). Mixed migration flows are characterized by their irregular nature (often in breach of regulatory norms of countries of origin, transit, and destination), and consist of distinct types of migratory populations, including groups of refugees, asylum-seekers, forced and voluntary migrants, and others. Most recently, an EU-commissioned assessment concluded that a forecasting and early warning tool based on artificial intelligence (AI) technology evaluating the intensity and direction of these mixed migratory flows is feasible (European Commission, 2021). As the scope of this article portends to considering mixed migration flows in light of addressing migration to the EU, it does not extend to all predictive analytics or tools for humanitarian action, although there are extensive, global efforts underway in this regard (Hernandez & Roberts, 2020).

Within this context, this study analyses how existing supranational predictive IT tools address these issues to achieve effective migration governance in the EU. This two-fold inquiry first asks: What are the main predictive migration tools and what is their scope? Here we focus on identifying the variables and data sources used to create the tool models, as well as their underlying objectives and rationale, exploring how this relates to the governance of migratory flows to date in their respective target countries. Additionally, this study further examines how existing prediction or forecasting tools might affect migration governance within the EU. Hence, the second research question is: To what extent can these existing tools facilitate effective migration governance? In essence, this study seeks to provide a comprehensive understanding of relevant past and current IT prediction or forecasting tools for migration, as well as to determine the extent to which such tools can offer improved migration governance and policy solutions in the EU.

This study includes three main parts. Firstly, it describes the qualitative methodology used for this investigation. Secondly, it conducts a comparative examination of three IT prediction tools or projects in the field of migration. Lastly, it suggests what a valid prediction tool for migration should include in terms of variables of analysis, user interface, data accuracy, modelling, and mechanisms for explaining and communicating predictions, in order to translate outputs into effective governance policies.

2. Cases, Methodology, Data, and Categories of Analysis

In order to answer the two main research questions, we reviewed an extensive number of models, projects, and tools with the potential to predict migration flows arriving to the EU. The IT predictive tools were selected from a 10-year time range from 2010–2020. Originally, in the analysis, we identified 18 tools and projects incorporating AI, which were relevant for EU migration governance purposes. However, data were incomplete for many of these, as some were no longer in use, further information was not accessible, others were not predictive tools as such, or those responsible for tool operation were unavailable for interview. As a result, we ultimately selected three tools and projects that could be useful for forecasting or predicting EU mixed migration flows as a sample for this comparative examination:

(1) the Jetson tool, funded and operated by UN High Commissioner for Refugees;
(2) the Early Warning and Preparedness System tool (hereafter EPS-Forecasting), funded and operated by the European Asylum Support Office;
(3) Foresight, currently funded and operated by the Danish Refugee Council (DRC). It was initially funded by the Danish Ministry of Foreign Affairs, with the model and user interface developed in collaboration with IBM.

It should be noted that the Internal Displacement Event Tagging and Clustering Tool (IDETECT), funded and operated by the Internal Displacement Monitoring Centre (IDMC), is also examined in-depth in Section 4.1 below, although this tool is not considered predictive as such. Moreover, in listing the final selection of tools above, it is equally important to signal that the Jetson project does not describe itself as a "tool" as such, but rather a proof of concept, as over the course of a year it conducted scoping, research, and piloting, but never produced a user-facing, interactive tool. At the same time, as it uses machine learning for predictive purposes, and pertains to all of the categories of analysis, it is used here as one of the selected "tools," with these qualifications.

To answer the research questions, a qualitative analysis has been carried out. The analysis used two types of research techniques: document analysis and in-depth interviews. The document analysis provides access to empirically objective, common, and reliable evidence (Franzosi, 1998, p. 547). We focused on analysing all the publicly accessible information published about these tools, including user-manuals, websites, reports, and academic publications, totalling approximately 50 sources.

Regarding the interview analysis, two kinds of in-depth interviews have been conducted. We would first highlight those carried out from February to March 2021, where we interviewed five experts on forecasting or AI tools. The interviewees were selected according to their expertise, using the snowball sampling method, and included: the three developers of the forecasting tools reviewed in this article, an IDMC representative regarding the AI tool IDETECT, and the founding developer of the Global Database of Events, Language, and Tone (GDELT) project (which monitors the world's broadcast, print, and web news in over 100 languages). These interviews first explored the scope of the different tools, and secondly inquired as to what extent the existing pre-

dictive tool or project was effective in predicting migration for effective migration governance. Furthermore, they informed the selection of the three tools for this article's in-depth analysis.

The other type of interview conducted included a group interview with 13 representative European NGOs specialising in migration and based in Bulgaria, Denmark, Greece, Italy, and Spain, on 20 January 2021. The NGOs were selected according to their expertise on the ground. The main objectives of this interview were twofold: firstly, to understand to what extent this kind of predictive tool can be useful for them as end-users; secondly, to learn what they would like and expect from this type of tool.

These interviews enabled us to validate and complete the information examined via the document analysis technique (Corbetta, 2003), lasting between 40 minutes and two hours. The interlocutors were all informed about the purposes of our research and were given the opportunity to review this work. The general sample has been sufficiently representative to enable saturation and triangulation of the information obtained.

We have predefined various categories of analysis to improve the consistency of the research. As no previous literature has identified these, we have extracted different categories from the data collected via an inductive process. All the data collected via the document analysis as well as the interviews were systematically entered into an Excel spreadsheet and organised by the different categories according to the study's two key objectives, as Table 1 illustrates.

3. Comparative Analysis of the Existing Prediction Tools

This section focuses on the study's first research question as to the main predictive migration tools and their corresponding scopes. As explained in the methodology, we analyse: (1) the Jetson tool; (2) the EPS-Forecasting tool; and (3) the Foresight tool. To carry out this comparative analysis, we identified four categories of analysis in Figures 1, 2, 3, and 4: target migration flow, timeframe, variables of analysis, and data sources.

Much of recent literature acknowledges that there is no "one model as the 'best model' for all situations" (Bijak et al., 2019, p. 12; Sohst & Tjaden, 2020). In fact, using partial or different models with the same data would produce different forecasts. Providing a comprehensive forecasting tool entails assessing the type of

Table 1. Categories of analyses.

Comparative descriptive analysis (Section 3)	Component of a valid predictive tool (Section 4)
Target flow	Variables of analysis
Timescale	User-friendly interface
Variables of analysis	Accuracy of migration data
Data sources	Best suited model(s)

Source: Own elaboration.

migration flow analysing the corresponding hard data available, and cross referencing with varied contextual data (qualitative and quantitative). The modelling should then be tailored to each of these different flows and accompanying data to improve accuracy and minimise potential bias.

As illustrated in Figure 1, none of these tools forecast regular migration, presumably because this information can be more easily obtained, as it can take months to years in applying for admission to another country. By contrast, irregular and forced migration can either occur quickly via a sudden event, which could be detected and signalled with an early warning system in place, or over longer periods of time with varying uncertain events that affect migration flows. The Jetson project formerly predicted forced, internally displaced people as pertained to the case study of Somalia. Meanwhile, the Foresight tool focuses its forecasts on only forcibly displaced asylum seekers and refugees from a given country. Finally, the EPS-Forecasting tool seeks to anticipate flows of refugees and asylum-seekers arriving to the EU, which can be unpredictable if migrants arrive via irregular routes. It should be noted that forced migration is referring to migration that is not voluntary,

while a refugee or an asylum-seeker is a legal term signifying those individuals that lack protection from their own country and are entitled to international rights or protections. As Figure 1 demonstrates, while they are distinct, sometimes these terms and concepts may overlap or are combined in the migration flow being targeted for prediction.

Modelling corresponded to the type of migration flow under examination. The Jetson project used machine learning gravity modelling and multivariate time series analysis (a total of 11 modelling techniques), the Foresight tool uses machine learning and Bayesian network models, and EPS-Forecasting uses adaptive machine learning algorithms, further compared in Section 4.4.

The types of movement to be predicted or forecasted relate to how each of the tools have different prediction timescales, portrayed here in Figure 2. The Jetson tool worked with both monthly predictions and additional short time frames, including three-month predictions, as they sought to test the assumption that sudden conflict events or external factors like drought and floods would cause population movement towards areas of humanitarian assistance. The EPS-Forecasting tool
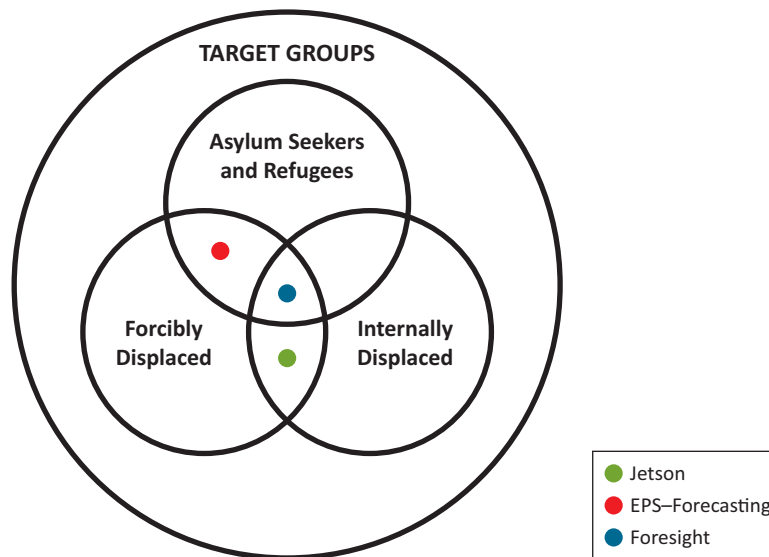

Figure 1. Target migration flow by predictive tool. Source: Adapted from Carammia et al. (2020); R. Jimenez, Jetson project, interview, March 11, 2021; A. Kjærum, DRC interview, February 18, 2021; and C. Melachrinos, interview, April 23, 2021.

Figure 2. Timescale of the predictions according to each tool. Source: Adapted from Carammia et al. (2020); R. Jimenez, Jetson project, interview, March 11, 2021; A. Kjærum, DRC interview, February 18, 2021; and C. Melachrinos, interview, April 23, 2021.
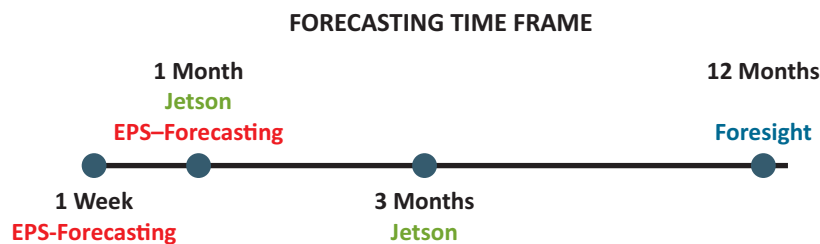
provides weekly predictions, as well as offers short-term predictions of up to one month that can be expanded via a user-selected variable. Finally, the Foresight tool initially predicted mixed migration flows one to three years in advance but was then redesigned (among other reasons) for better accuracy to focus on forced displacement via one to three-year predictions.

As displayed in Figure 3, these tools include different variables of analysis; this relates to how selecting tool variables depends on the underlying migration theory or assumption, and the identified migration flows or tool objective, further detailed in Section 4. The most common variable, used by all three tools, is information on violence or conflict in a particular territory, due to a consensus that this represents one of the most influential factors in migration flows. In addition, both the JETSON and EPS-Forecasting tools use information on political events to feed their predictions. In particular, the EPS-Forecasting tool identifies political events and social unrest, which could be caused by riots or rebellion within a specific country or territory.

It is of note that both the EPS-Forecasting and the FORESIGHT tools take into account the area and/or country of origin's economy or governance to refine their predictions. In addition, FORESIGHT uses data on geological, hydrological, and meteorological events to inform predictions.

Figure 4 shows that several data sources are used for the predictive tools' models. Here it is noteworthy that Internet usage can assume an active role in irregular migrants' choices and routes (Lif, 2016). The EPS-Forecasting tool incorporates real-time online data sources in its predictions. However, none of the tools incorporate data from social media platforms like Twitter, Facebook, or Instagram; this is perhaps due to an ongoing debate as to the ethical concerns behind use of migrants' personal data.

Figure 4 also demonstrates how all of the predictive tools rely mainly on open-access data. When it was in operation, the Jetson project used data from the Armed Conflict Location & Event Data Project (ACLED), the Food Security and Nutrition Analysis Unit, the Food
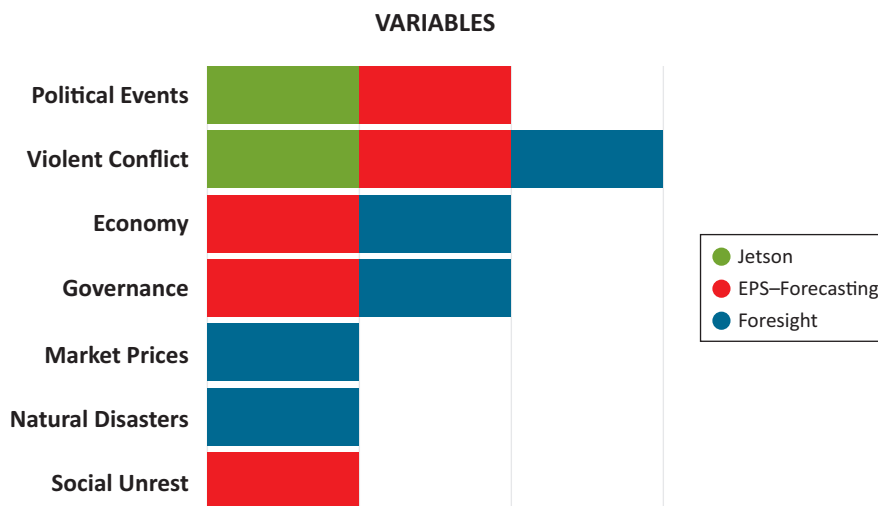

Figure 3. Variables of analysis included in each predictive tool. Source: Adapted from Carammia et al. (2020); R. Jimenez, Jetson project, interview, March 11, 2021; A. Kjærum, DRC interview, February 18, 2021; and C. Melachrinos, interview, April 23, 2021.

Figure 4. Data sources for the predictive tools. Source: Adapted from Carammia et al. (2020); R. Jimenez, JETSON project, interview, March 11, 2021; A. Kjærum, DRC interview, February 18, 2021; and C. Melachrinos, interview, April 23, 2021.
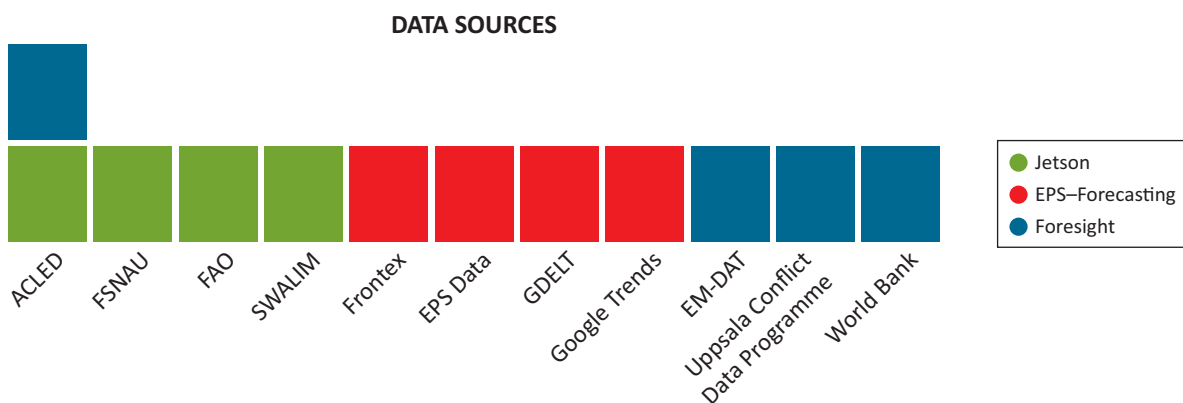
and Agriculture Organization of the UN, and the Somalia Water and Land Information Management Project. In addition to making use of GDELT, Google Trends and Frontex data, the EPS-Forecasting tool focuses on predicting asylum applications using their own weekly EPS-Data acquired from exchange with EU+ countries (EU member states and associated countries), the latter of which provide data on 19 standardised, disaggregated indicators, or the EPS-Data. A subset of those data is then used for the EPS-Forecasting tool (Albertinelli et al., 2020). The Foresight tool utilizes 18 different sources of data, with main sources including EM-DAT, World Bank, the Uppsala Conflict Data Programme, and ACLED.

4. Towards a Valid Predictive Tool for Migration Governance in the EU

Following this brief comparison of three main predictive migration tools, this study fleshes them out further and assesses the most relevant criteria in determining what a valid predictive tool for effective EU migration governance should look like. As such, this section partially responds to the second research question: To what extent can those existing tools facilitate effective migration governance? In particular, the following aspects are analysed: necessary variables for the models, interface of the tool, data and prediction accuracy, and adequately suited models.

4.1. Necessary Variables to Be Included in the Tool

What follows is further exploration of how and why to select variables and establish parameters for these tools or their models, providing context for why the tools analysed above did so. First of all, a valid predictive tool should include variables of analysis related to the demography of the selected countries of origin. For instance, variations in population size within countries of origin, transit, and destination can point to the existence of migration flows in those territories, and even provide data regarding migration routes. There are many public sources of information that provide quantitative data related to migration, including, among others, population statistics (e.g., City Population, 2021; UN, 2021), number of displaced individuals (e.g., International Organization for Migration, 2020; OECD, 2020), number of asylum applications (e.g., Eurostat, 2021; UN High Commissioner for Refugees, 2020), and main migratory routes (e.g., Frontex, 2020). Other relevant data sources could provide information on the given theoretical drivers of migration, including armed conflict and violence (ACLED, 2021), climate disasters (EM-DAT, 2021; European Centre for Medium-Range Weather Forecasts, 2021), low levels of development (The World Bank, 2021), food insecurity (Integrated Food Security Phase Classification, 2020), irregular governance (Rulers, Elections, and Irregular Governance, 2021), unaffordable food (Food and Agriculture Organization of the

UN, 2021), policy changes (European Country of Origin Information Network, 2020), or other specific events (GDELT, 2021). As illustrated in Figure 4, two of the three tools studied here use the public dataset ACLED, which offers the "dates, actors, locations, fatalities, and types of all reported political violence and protest events" in real-time throughout the world (ACLED, 2021).

Additionally, a valid predictive tool working with quantitative data could benefit from a monitoring team that can simultaneously examine or provide qualitative data to validate or correct the tool's quantitative results. An example includes the work of the IDMC, which operates an AI tool and a real-time monitoring platform as part of its work in investigating displaced individuals worldwide. IDMC enlists a team of about eight to 10 experts to monitor the raw quantitative data, ensuring it corresponds to real-time displacements.

Ideally, when assessing successful integration strategies, predictions of migratory flows should not only include parameters related solely to countries of origin, but also variables pertaining to the country of destination, such as macroeconomic indicators (unemployment, job vacancies) or migration and integration policies (migration caps, visa regulations, etc.). This consideration is adequately incorporated into the EPS-Forecasting model, and Foresight tool developers also tested this origin–destination relationship when initially looking at mixed migration in its preliminary modelling. In using such variables, both macro and micro synergies must be considered to sufficiently address populations with multiple characteristics.

Big data has demonstrated advantages in that it could allow for identifying a particular phenomenon affecting migration flows, can permit measurement of variables at a regional or even local level, and provide input in updating a corresponding algorithm (open-access data sources usually provide information only at the national level). For instance, if a conflict occurs in a specific region, this could be accurately measured by identifying users' reactions in that particular location via social media. In recent years, several universities and research centres have been working with big data in forecasting displacement globally, examining social media for sentiment analysis, and in evaluating economic and social variables (European Commission, 2017; Singh et al., 2019). In this sense, an ideal model could include behavioural and sentiment analysis collected online. In utilizing data on sentiment, for example, a tool could identify posts, "likes" or interactions shared by migrants on their social networks, which could then be properly anonymised (according to the terms of the EU General Data Protection Regulation, as further described in Section 4.2) and codified as positive or negative reactions towards potential destination countries, routes, and relevant migration topics. Although several models that incorporate data on individual preferences and opinions have been proposed in recent years (Alam et al., 2020; Papakyriakopoulos et al., 2018, p. 9), there is no evidence that this type of

data has been used to feed an actual predictive migration mechanism to date. EPS-Forecasting uses Google Trends in generating their predictions, and a valid predictive tool could similarly incorporate insight from previous studies exploring how variables from "Google queries" can predict migration flows (Böhme et al., 2020; Lif, 2016).

Finally, a valid predictive tool could also benefit from GDELT data, which provides real-time, open source, spatio-temporal data sets on crises, drawn from the world's news media. The project codifies news and information about events via the (political science discipline) Cameo system and provides a range of users with the various methods with which to analyse this georeferenced, globally contextualised data. As mentioned above, the EPS-Forecasting tool already uses GDELT data, selecting a set of 240 events potentially driving migration and displacement, divided into five macro-categories (political events, social unrest, conflicts, economic events, and governance-related events).

4.2. Tool Interface and End-User Involvement

A valid predictive migration tool must be designed so that its selected end-users can easily operate it. In this regard, we arrive at three main conclusions based on our analysis of the interviews conducted with 13 NGOs based in Bulgaria, Denmark, Greece, Italy, and Spain.

Firstly, end-users should be able to select and deselect the main demographic criteria. In this sense, at the very least, an origin–destination–age–gender migration matrix should be made available. In addition, potential end-users in our study find it relevant to have criteria such as nationality, language spoken, ethnic group, and skills of migrants. Although it would be ideal to include all these criteria, concerns could arise in terms of data protection and ethical requirements. Particularly, article 5(1)(c) of the EU General Data Protection Regulation establishes that the processing of personal data must be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed" (European Union Regulation of 27 April 2016, 2016, p. 35). In essence, migrant's personal data should be processed at a minimum, only as needed for the purpose. In particular, as migrants meet the definition of "vulnerable and minority groups," any tool processing their personal information should ensure that access to this information is limited to specific selected end-users, and that this information access' purposes comply with the EU Charter of Fundamental Rights.

Secondly, according to the surveyed potential end-users, the interface of the tool could include interactive user features like the potential to compare previous periods, or scenario-building capacities. For instance, users find it particularly attractive if a tool can allow them to break down migration to the EU by routes, and to identify and visualise population groups according to specific attributes.

Lastly, potential end-users find it important that such a tool could provide predictions automatically, without having to rely on inputting their own data, which is often insufficient or even non-existent. In this regard, the tool could have a mechanism in place via machine-learning techniques, or a human monitoring unit responsible for updating information. Such a mechanism would guarantee, for example, that one year after the tool is operational, predictions and functionalities of the tool remain accurate, considering all recent phenomena impacting migration flows.

To ensure the effectiveness of the tool, it is essential to engage with experts or conduct pilot tests among end-users in real-world environments, as part of the validation process. Two of the three tools examined and still in operation receive formal validation from in-house versus external users.

4.3. Accuracy of Migration Data and Predictions

While there are many sources of information, at the same time the inherent uncertainty and complexity of the migration phenomenon, and its study, present several issues with data. The first problem encountered includes access: Given the nature of migration, as well as the multiple actors involved, there may simply be a lack of data, it could be incomplete, or it could be impossible to obtain (Felkai Janssen, 2020; Kjærum, 2020). For example, as some of the events being anticipated are sudden or unpredictable, a frequent possibility with mixed migration flows, the short life of the process makes data collection itself difficult. Another example includes how, initially, Foresight sought to use household level information, but ended up using national level data, as obtaining the former dataset was not feasible.

Moreover, raw data are not always accurate. Although accuracy of migration data has improved over time (Wilson & Rees, 2005, p. 339), some of this available information can often be inconsistent or incomplete, especially due to existing bias in the way data is collected (for example, there is currently a lack of gender disaggregated information in migration research). Assuming that data sources are sufficiently accurate, the next challenge lies in achieving accurate predictions. Many scholars have sought to minimise uncertainty through their models and studies (Bijak et al., 2019; Wilson & Rees, 2005). However, there is still a large margin of error and various biases in migration predictions, due to events that are very difficult to predict (Bijak et al., 2019). An example includes the economic cycle of a country. Although many studies have aimed to identify vulnerable economies facing risk in recent years (Brei et al., 2020; Stamer, 2019), in general, economists are not always capable of accurately predicting country recessions (Thomson-DeVeaux, 2020). Consequently, emigration driven by an economic downturn is even harder to forecast.

In the last 10 years, many methods have studied how to explicitly identify errors (Wilson & Rees, 2005, p. 340)

and improve predictions. One strategy for reducing high levels of error includes assessing data sources according to the concept of "true flow." True flow is understood as the number of migrants a flow would amount to according to a given definition of migrant, if immigration could be monitored perfectly. Another strategy could focus on assessing the margin of success by pulling everything backwards (e.g., before 2015 in the case of migration predictions), and looking at the ratio of error to the success of "past" forecasts. Through this method, the model could include benchmarking against the actual change in migration flows from one year to the next, or benchmarking against the previous prediction technique in place.

4.4. Best Suited Model(s)

Finally, it is important to recognise how lack of or errors in data, as well as differences in conceptions and theories of migration, relates to how a tool is designed or models should be chosen. The fact that a tool relies on only one forecasting model could also lead to uncertainty (Bijak, 2016; Disney et al., 2015), as any single model will have only a limited number of variables. Even if a particular model works well for a certain period, one sole event might change everything, and from that point on the predictive tool might have to consider a different degree of uncertainty. Advocates for the advantages of Bayesian models, a model used in Foresight, argue that they theoretically can provide for the three elements of uncertainty inherently characteristic to migration prediction: the uncertainty of future events, of migration data, and of different forecasting models producing different results. Moreover, the Bayesian approach can allow for incorporating historical trends, expert judgements, and various model specifications (Bijak & Wiśniowski, 2010).

Meanwhile, a gravity model like that used in the Jetson experiments employs demographic variables, using population size to map people over spatial distance; this type of approach may offer insight into future flows' structures, rather than magnitude (Bijak et al., 2019). In fact, it has been increasingly used in forecasting numbers of people who may want to move along with their potential destination, as has been conducted in a World Bank report exploring potential future climate migration (Rigaud et al., 2018). Finally, the literature has identified how agent-based modelling approaches in combination with machine learning are promising to provide predictions based on indicators or decisions to migrate at the individual or household level (corresponding to micro-level theories of migration; Searle & van Vuuren, 2021). These agent-based models are often used to explain migration, rather than predict (Klabunde & Willekens, 2016).

As a way to mitigate the uncertainty of some of the variables, model-based estimations could be complemented by expert judgment. Experts are especially necessary for assessing past forecast errors, identifying random variations in the models, enriching policy debates, and offering long-term perspectives. Both at the EU and international level, scenario-building constitutes a form of anticipating and preparing for future migration patterns, or understanding alternative future flows (Kjærum, 2020). This involves using models in an explanatory capacity, with experts predicting how different scenarios affect the composition of migration flows in a systematic process. However, disagreement among such experts and the difficulty of providing actionable conclusions complicates this method. A strategy to mitigate lack of consensus and best incorporate expert opinion includes the Delphi survey, which surveys experts in multiple rounds to reduce individual biases and promote consensus (Sohst & Tjaden, 2020).

5. From Predictions to Policy

In addition to the considerations outlined in Section 4, the feasibility and effectiveness of predictive tools and projects for migration governance require extensive assessment of who manages them and how. These new technologies present a set of tools to understand and anticipate migration, as the information they provide can be analysed to inform long-term, good governance efforts. However, there are still several potential challenges and gaps in first providing these tools for policymaking, and then in translating predictions into policy decisions.

For one, economic and political considerations remain inextricably linked. Policy makers may view economic costs as a barrier to even utilizing such tools in the first place. This could be addressed by incorporating open-source data and engaging in resource and knowledge sharing, as the Jetson project did by using open-source code and volunteer developers (R. Jimenez, Jetson project, interview, March 11, 2021). Furthermore, sufficient political will could assist in obtaining funds. Still, the earlier described competing interests of migration governance do present continued concerns with political risk, and multi-stakeholder collaboration is a complicated negotiation. As a result, these tools may remain for internal use by the developer and funder to mitigate political sensitivities.

For example, while the EPS-Forecasting algorithm is both replicable and transparent, it remains internal in compliance with the dissemination policy of some of its data, which is provided in a two-way exchange with governments under the condition it is not used publicly (Carammia et al., 2020). However, in addition to being used by its initial, internal stakeholder, the Foresight tool is also used by external stakeholders who have provided positive feedback and is made available to other stakeholders or governments (A. Kjærum, DRC interview, February 18, 2021).

In all cases, translating the available and significant wealth of analysis, information, and technological potential to actual decisions for migration governance requires continuous efforts towards enhanced communication, participation, and transparency, as Figure 5 illustrates.
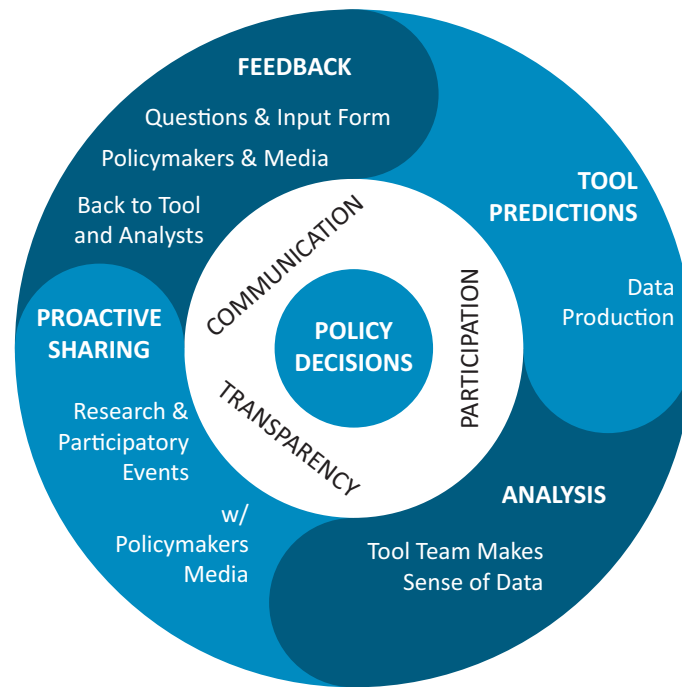
Figure 5. From tool predictions to policy decisions. Source: Own elaboration.

To utilise the explanatory and empirical value of migration predictions for governance, the nature of translating research into policy necessitates both human analysis in the form of a research team, as well as a policy unit or team corresponding to a given tool, in order to ensure the tool is useful for governance. In effect, while the tool analysts and even invited outside experts make sense of tool results, the engagement unit could serve as the liaison with policymakers. The latter team can ensure information is provided in a palatable, non-technical format, as overly scientific explanations can ultimately not prove useful (Albertinelli et al., 2020). Such an engagement or policy team would also maintain an established link with policymakers and stakeholders, so that the tool's analysis unit could communicate results in a two-way dialogue. For example, the European Asylum Support Office has maintained such an ongoing, two-way conversation and engagement with EU+ countries that seek early warning and forecasting of asylum seekers, and provides them with up to date, limited release, comprehensive reports regarding the state of early warning and forecasting to date or the impacts of recent trends like the Covid-19 pandemic; they are working towards a joint predictive exercise among such stakeholders in 2021 (European Asylum Support Office, 2017; C. Melachrinos, interview, April 23, 2021).

Evidently, the dialogue between scientists and policymakers is not simply a matter of submitting palatable desk research resulting from the tool, and the proposed engagement team's role would mean organizing participatory, collective, ideally public, events like workshops, focus groups, and webinars, where policymakers could present questions and input. This transparency would demonstrate the impact that this technology and data offer, and it would act as a check on decision-makers: It would either monitor that they do not assert their agenda or self-interest in the direction of one tool or monitor that they do not compare and select different tools to simply match tool results to their desired policies. Moreover, recent literature suggests that elites across countries—including those drawing from the business, media, and civil society in addition to those from government and research—are moderately open to and confident in select global institutions (Scholte et al., 2021); increased transparency and engagement with such elites in this prediction to policy dialogue could in turn further legitimise global migration governance.

Finally, the entire process of creating policies for populations would be incomplete without visibility into the lived experiences of migrants themselves, either obtained by a tool's analysis unit in tool design or interpretation of results, or by the policy team in engaging with decision-makers. The Jetson project, for example, incorporated this at the analysis stage, conducting qualitative interviews among the displaced individuals being monitored, to validate variables utilized in running their simulations.

Rather than reactive, short-term decision-making, long-term policy provision that incorporates interrelated policy areas beyond those strictly related to migration and integration measures are key (Szczepanikova & Van Criekinge, 2018). For example, rather than solely anticipating and preparing for migration in the short-term, translating migration predictions could mean reforming or bolstering public health or welfare policy for inclusive growth, as was indicated in this study's

interviews with relevant NGOs; in short, it is key to account for all stakeholders in human mobility. Given that migration governance should resemble a horizontal process, the tool and its proposed units should engage all relevant multilevel governmental and non-governmental actors, including international organisations, EU institutions, states, regional or local governments, civil society representatives, academic experts, and even the private sector. Finally, such a long-term view could mean also engaging origin and transit countries, enacting a specific policy or developing capacity-building projects, rather than focusing exclusively on host countries.

6. Conclusion

This study explores existing developments in predictive tools for migration. After reviewing the literature and available documents, following up with tool developers and interviewing potential end-users, three predictive migration tools and projects were studied and compared, namely, the Jetson tool, the EPS-Forecasting tool, and the Foresight tool. The analysis demonstrates the challenges in providing for effective interaction and feedback among tool developers and end-users, and how each of these tools has a different scope, data sources, models, and validation mechanisms, according to their goals.

The discussion fleshes out the difficulties to date encountered by those managing predictive tools in offering predictions that could serve the totality of relevant stakeholders in their intent to develop strategic migration governance (Robinson, 2018; Triandafyllidou, 2020). In illustrating this, the study emphasises four main aspects that should be taken into consideration to create, or transform a predictive tool into, a valid predictive tool for effective migration governance. These main elements provided here include guidance on variables to incorporate into the models, ways to involve end-users in the process, adequate levels of accuracy, and tailoring modelling to the prediction or governance objective. Furthermore, mechanisms to convert predictions into policy decisions were ultimately emphasized.

Again, there are unresolved challenges related to predictive migration tools at all stages of the development process: While some existing tools have managed to validate their predictions and achieve an acceptable level of accuracy, others still struggle to obtain accurate predictions even over a few weeks' time. Even if this first issue is overcome, the later stage of converting such predictions into decisions for governance remains an ongoing process for all pre-existing tools. The ultimate goal of predicting migration flows for governance should be to enable policymakers and appropriate stakeholders to make prudent and robust decisions, by illustrating a clear causal relationship between migrant arrivals and necessary policies for managing future migration. Section 5 explores this crucial step of engaging with policymakers, noting how it is essential to offer clear explanations of comprehensive policy solutions based on the predic-

tions, properly communicating the uncertainty of migration forecasting.

Most importantly, the way international migration governance is understood and practiced can be mutually configured and shaped by technologies for migration forecasting. The literature to date has pointed to the risk of new technologies and IT predictive tools intensifying global or regional asymmetries, and curtailing human rights, which is at odds with effective migration governance (Beduschi, 2020; Broeders & Dijstelbloem, 2016; Crépeau & Atak, 2016).

At the same time, this analysis of current predictive tools and related projects has demonstrated that if developed and operated in a transparent and accountable process, such tools can alternatively be leveraged as an equalizer in the field of migration management. Firstly, knowledge and resource sharing via open source and access can reduce costs or barriers that otherwise exacerbate power differentials in global migration management. Secondly, these tools boast the capacity to incorporate a broad and diverse range of actors—individual states, intergovernmental organisations, civil society, interdisciplinary researchers, and technical experts, as well as migrants themselves—that can offer the most comprehensive data and insight for effective governance.

In conclusion, this study indicates that there are currently several initiatives underway in pursuit of an IT predictive tool that could assist in migration management. The EU and international organisations are working with more data and insight than ever before understanding migration flows for governance purposes. Nonetheless, to render them both useful and relevant, predictive tools require continual monitoring and re-assessment, consistent and expanded multi-stakeholder collaboration, and further efforts toward sufficient communication and translation of outputs, in order to better assist concrete policy decision-making and outcomes.

Acknowledgments

Conflict of Interests

The authors declare no conflict of interests.

References

Alam, M., Gesese, G. A., Rezaie, Z., & Sack, H. (2020). MigrAnalytics: Entity-based analytics of migration

tweets. In K. Taylor, R. Gonçalves, F. Lecue, & J. Yan (Eds.), *Proceedings of the ISWC 2020 posters and demos track, co-located with 19th International Semantic Web Conference* (pp. 74–78). ISWC.

Albertinelli, A., Alexandrova, P., Melachrinos, C., & Wilkin, T. (2020). Forecasting asylum-related migration to the European Union, and bridging the gap between evidence and policy. *Migration Policy Practice*, *10*(4), 35–41.

Arango, J. (2018). Explaining migration: A critical view. *International Social Science Journal*, *68*, 105–117. https://doi.org/10.1111/issj.12183

Armed Conflict Location & Event Data Project. (2021). *Disaggregated data collection, analysis & crisis mapping platform*. https://acleddata.com/#/dashboard

Beduschi, A. (2020). International migration management in the age of artificial intelligence. *Migration Studies*. Advance online publication. https://doi.org/10.1093/migration/mnaa003

Bijak, J. (Ed.). (2016). *Migration forecasting: Beyond the limits of uncertainty*. Global Migration Data Analysis Centre. https://gmdac.iom.int/sites/gmdac/files/gmdac_data_briefing_series_issue_6.pdf

Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W. F., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*, 470–487. https://doi.org/10.1002/for.2576

Bijak, J., & Wiśniowski, A. (2010). Bayesian forecasting of immigration to selected European countries by using expert knowledge. *Journal of the Royal Statistical Society*, *173*(4), 775–796. https://doi.org/10.1111/j.1467-985X.2009.00635.x

Böhme, M. H., Gröger, A., & Stöhr, T. (2020). Searching for a better life: Predicting international migration with online search keywords. *Journal of Development Economics*, *142*, Article 102347. https://doi.org/10.1016/j.jdeveco.2019.04.002

Brei, M., Borio, C., & Gambacorta, L. (2020). Bank intermediation activity in a low-interest-rate environment. *Economic Notes*, *49*, Article e12164. https://doi.org/10.1111/ecno.12164

Broeders, D., & Dijstelbloem, H. (2016). The datafication of mobility and migration management: The mediating state and its consequences. In I. Van der Ploeg & J. Pridmore (Eds.), *Digitizing identities: Doing identity in a networked world* (pp. 242–260). Routledge. https://doi.org/10.4324/9781315756400-13

Carammia, M., Iacus, S., & Wilkins, T. (2020). *Forecasting asylum applications in the European Union with machine learning and data at scale*. ArXiv.

Carling, J., & Collins, F. (2018). Aspiration, desire and drivers of migration. *Journal of Ethnic and Migration Studies*, *44*(6), 909–926. https://doi.org/10.1080/1369183X.2017.1384134

Castles, S., de Haas, H., & Miller, M. J. (2014). *The age of migration: International population move-ments in the modern world*. Ethnic and Racial Studies, *38*(13), 2355. https://doi.org/10.1080/01419870.2015.1050048

City Population. (2021). *Population statistics for countries, administrative divisions, cities, urban areas and agglomerations—interactive maps and charts*. https://www.citypopulation.de

Corbetta, P. (2003). *Social research theory methods and techniques*. SAGE.

Crépeau, F., & Atak, I. (2016). Global migration governance: Avoiding commitments on human rights, yet tracing a course for cooperation. *Netherlands Quarterly of Human Rights*, *34*(2), 113–146. https://doi.org/10.1177/016934411603400203

Disney, G., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2015). *Evaluation of existing migration forecasting methods and models: Report for the Migration Advisory Committee*. ESRC Centre for Population Change.

EM-DAT. (2021). *The international disaster database*. https://www.emdat.be

European Asylum Support Office. (2017). *Quantitative assessment of asylum-related migration: A survey of methodology*. https://easo.europa.eu/sites/default/files/publications/Quantitative_assessment_of_asylum_related_migration_1.pdf

European Centre for Medium-Range Weather Forecasts. (2021). *Advancing global NWP through international collaboration*. https://www.ecmwf.int

European Commission. (2017). *Big data and alternative data sources on migration: From case studies to policy support.* European Comission Knowledge Centre on Migration and Demography. https://knowledge4policy.ec.europa.eu/migration-demography/big-data-alternative-data-sources-migration_en

European Commission. (2021). *Feasibility study on a forecasting and early warning tool for migration based on artificial intelligence technology*. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/5afa29f0-700a-11eb-9ac9-01aa75ed71a1

European Country of Origin Information Network. (2020). *About ecoi.net*. https://www.ecoi.net/en/about/about-ecoi.net

European Union Regulation (EU) No 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). (2016). *Official Journal of the European Union*, L 119.

Eurostat. (2021). *Asylum and first-time asylum applicants by citizenship, age and sex—monthly data (rounded)* [Data set]. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_asyappctzm&lang=en

Felkai Janssen, Z. (2020). Migration monitoring and preparedness at the European Union level: The need

for multi-stakeholder cooperation. *Migration Policy Practice*, *10*(4), 45–47. https://publications.iom.int/es/books/migration-policy-practice-vol-x-number-4-september-december-2020

Food and Agriculture Organization of the UN. (2021). *Domestic price warnings*. http://www.fao.org/giews/food-prices/en

Franzosi, R. (1998). Narrative analysis-or why (and how) sociologists should be interested in narrative. *Annual Review of Sociology*, *24*, 517–554. https://doi.org/10.1146/annurev.soc.24.1.517

Frontex. (2020). *Migratory map*. https://frontex.europa.eu/along-eu-borders/migratory-map

Global Database of Events, Language, and Tone. (2021). *The GDELT story*. https://www.gdeltproject.org/about.html

Hernandez, K., & Roberts, T. (2020). *Predictive analytics in humanitarian action: A preliminary mapping and analysis* (K4D Emerging Issues Report 33). Institute of Development Studies.

Integrated Food Security Phase Classification. (2020). *The IPC population tracking tool*. http://www.ipcinfo.org/ipc-country-analysis/population-tracking-tool/en

International Organization for Migration. (2020). *Displacement tracking matrix*. https://displacement.iom.int

Kjærum, A. (2020). Foresight: Using machine learning to forecast and understand forced displacement. *Migration Policy Practice*, *10*(4), 26–30. https://publications.iom.int/es/books/migration-policy-practice-vol-x-number-4-september-december-2020

Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: State of the art and challenges. *European Journal of Population*, *32*, 73–97.

Lif, O. (2016). *Migration in the internet age. The 21st century: When refugees went online* (Lund Student Papers). Lund University. https://lup.lub.lu.se/student-papers/search/publication/8873467

Massey, D. (1999). International migration at the dawn of the twenty-first century: The role of the state. *Population and Development Review*, *25*(2), 303–322. https://doi.org/10.1111/j.1728-4457.1999.00303.x

OECD. (2020). *International migration database*. https://stats.oecd.org/Index.aspx?DataSetCode=MIG

Papakyriakopoulos, O., Hegelich, S., Shahrezaye, M., & Medina Serrano, J. C. (2018). Social media and micro-targeting: Political data processing and the consequences for Germany. *Big Data & Society*, *5*(2). https://doi.org/10.1177/2053951718811844

Raleigh, C. (2011). The search for safety: The effects of conflict, poverty and ecological influences on migration in the developing world. *Global Environmental Change*, *21*(Suppl. 1), S82–S93. https://doi.org/10.1016/j.gloenvcha.2011.08.008

Rigaud, K., de Sherbinin, A., Jones, B., Bergmann, J., Clement, V., Ober, K., Schewe, J., Adamo, S.,

McCusker, B., Heuser, S., & Midgley, A. (2018). *Groundswell: Preparing for internal climate migration.* World Bank. https://openknowledge.worldbank.org/handle/10986/29461

Robinson, C. (2018). Making migration knowable and governable: Benchmarking practices as technologies of global migration governance. *International Political Sociology*, *12*(4), 418–437. https://doi.org/10.1093/ips/oly020

Rulers, Elections, and Irregular Governance. (2021). *Rulers, elections, and irregular governance dataset* [Data set]. https://oefdatascience.github.io/REIGN.github.io/menu/reign_current.html

Scholte, J. A., Verhaegen, S., & Tallberg, J. (2021). Elite attitudes and the future of global governance. *International Affairs*, *97*(3), 861–886.

Searle, C., & van Vuuren, J. H. (2021). Modelling forced migration: A framework for conflict-induced forced migration modelling according to an agent-based approach. *Computers, Environment and Urban Systems*, *85*, Article 101568. https://doi.org/10.1016/j.compenvurbsys.2020.101568

Singh, L., Wahedi, L., Wang, Y., Kirov, C., Wei, Y., Martin, S., Donato, K., Liu, Y., & Kawintiranon, K. (2019). Blending noisy social media signals with traditional movement variables to predict forced migration. In A. Teredesai & V. Kumar (Eds.), *KDD '19: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1975–1983). Association for Computing Machinery.

Sohst, R., & Tjaden, J. (2020). Forecasting migration: A policy guide to common approaches and models. *Migration Policy Practice*, *10*(4), 8–13. https://publications.iom.int/es/books/migration-policy-practice-vol-x-number-4-september-december-2020

Sohst, R., Tjaden, J., de Valk, H., & Melde, S. (2020). *The future of migration to Europe: A systematic review of the literature on migration scenarios and forecasts*. International Organization for Migration.

Stamer, M. (2019). *Country Risk—Which countries are the most vulnerable to the global slowdown?* Euler Hermes. https://www.eulerhermes.com/en_global/news-insights/economic-insights/Country-Risk-Which-countries-are-the-most-vulnerable-to-a-global-slowdown.html

Szczepanikova, A., & Van Criekinge, T. (2018). *The future of migration in the European Union: Future scenarios and tools to stimulate forward-looking discussions*. Publications Office of the European Union. https://doi.org/10.2760/274425

The World Bank. (2021). *World development indicators*. https://databank.worldbank.org/source/world-development-indicators

Thomson-DeVeaux, A. (2020). *Even without a pandemic, it's hard to forecast a recession*. FiveThirtyEight. https://fivethirtyeight.com/features/even-without-a-pandemic-its-hard-to-forecast-a-recession

Triandafyllidou, A. (2020). Decentering the study of

migration governance: A radical view. *Geopolitics*. Advance online publication. https://doi.org/10.1080/14650045.2020.1839052

UN High Commissioner for Refugees. (2020). *Refugee data finder*. https://www.unhcr.org/refugee-statistics/download/?url=pwN2

UN. (2021). *United Nations population division*. https://www.un.org/development/desa/pd

Wilson, T., & Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place*, *11*, 337–360. https://doi.org/10.1002/psp.389

About the Authors

Cristina Blasi Casagran is assistant professor in EU law at the Autonomous University of Barcelona, and she holds a PhD in Law at the European University Institute (Florence, 2015), in which she specialised on EU privacy and data protection law. She is currently coordinating the H2020 project IT Tools and Methods to Manage Migration Flows (ITFLOWS) and the Erasmus+ Jean Monnet Module EUCONAS.

Colleen Boland is a postdoctoral researcher at the Autonomous University of Barcelona, where she works on the EU Horizon2020 funded ITFLOWS project investigating the management of migration flows to the EU and co-coordinates the project's Gender Committee. She holds a PhD in Sociology and Anthropology from the Complutense University of Madrid. She is currently a 2021–2022 re:constitution fellow, a programme funded by the Forum Transregionale Studien and Democracy Reporting International.

Elena Sánchez-Montijano is research professor at the Center for Research and Teaching in Economics (CIDE, Mexico) and a researcher at the Autonomous University of Barcelona, where she collaborates on the EU Horizon2020 funded ITFLOWS project. She holds a PhD in Political and Social Science from the Pompeu Fabra University (Barcelona). Currently, she is co-coordinating the project Towards an Integration Policy for Migrants in Mexico, funded by the Open Society Foundation and is a partner coordinator at the Atlantic Network 2.0 project, funded by the European Jean Monnet programme.

Eva Vilà Sánchez is a graduate student at the Autonomous University of Barcelona. She is currently conducting the MA on European Integration and has completed an internship at the EU Horizon2020 funded ITFLOWS project. She holds an undergraduate degree in European Studies and French from Royal Holloway, University of London.