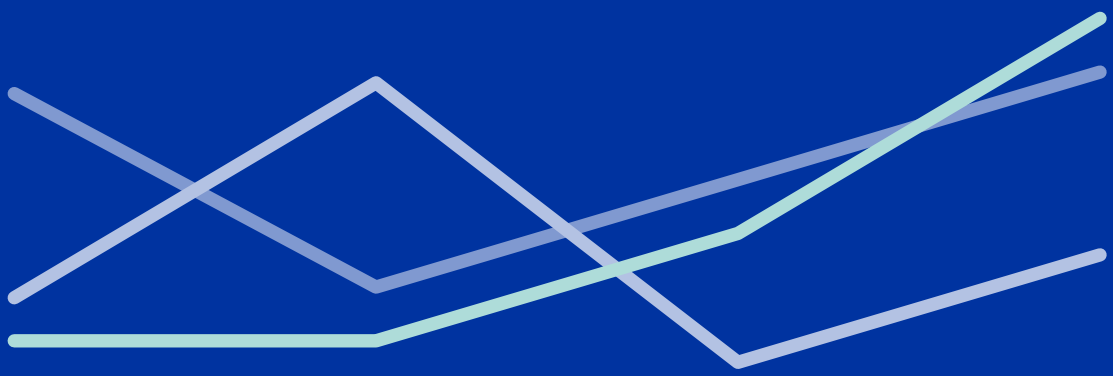
IOM • OIM DTM
THE UN MIGRATION AGENCY

Workshop Report on

Forecasting Human Mobility in Contexts of Crises

Berlin, 22–24 October 2019



A collaboration of the German Federal Foreign Office (FFO) and the International Organization for Migration (IOM)

Federal Foreign Office

International Organization for Migration (IOM)
The UN Migration Agency

EXECUTIVE SUMMARY

The co-hosted IOM-FFO Workshop on Forecasting Human Mobility in Contexts of Crises convened 67 key actors from academia, research, the humanitarian sector, policy and government involved in developing predictive analytics. The aim was to assess the current state-of-the-art in forecasting approaches regarding human mobility in the context of crises. In total 35 different organizations were present for the three day workshop from 22 to 24 October 2019. The workshops sought to address the lack of shared standards for employing automated and model based analytics to inform decisions on where, when and how to respond and assist crisis displaced populations.

One of the key concerns identified was the striking absence of agreed upon core principles to be adhered to for modelling. In the examples presented, the focus on modelling initiatives highlighted an observable trend of a geographical focus on Africa raising questions of scalability of such humanitarian action. [1] Accountability was widely discussed as well as challenges attributing it. The different stakeholders, such as data scientists, humanitarian programme staff, donors and decision makers across the different areas intersecting with humanitarian operations have different accountability and ethical frameworks. Ultimately, one of the biggest issues in modelling is the confusion between mathematical accountability and accountability to affected populations, or different definitions of risk. The difficulties of making a one-size-fits-all accountability framework were discussed. While the practicalities pose several challenges, there was agreement that do no harm principles should be followed. Another observation underlined that ethics discussions are often focused more on the legal aspects of data protection than "do no harm" principled approaches. Participants acknowledged the importance of involving local populations/communities in the development of models. This includes the participation of communities in terms of model inputs but also the algorithmic accountability of modelers. Nevertheless, practical solutions remain unclear. Likewise, it remains an open question how to balance actionability and ethical concerns – generally more granular data tend to be more actionable, leading to greater ethical concerns about misuse/weaponization. Ethical concerns further relate to the possible diffusion of responsibility when developing and applying models, since developers can argue the responsibility lies with end users and vice versa.

It was agreed that models are there to inform but not to replace human decision making for all decision making tasks where the concerns raised apply. To improve the development and use of models, greater cooperation and engagement among stakeholders including data scientists, responders, and people in need is required. However, even among experts (be it technical, humanitarian, or ethical) of the same domain a common definition of terms is often lacking. This can result in poor communication and ultimately misapplication of models with potentially harmful consequences. To ensure greater utilization of data and models with an integrated ethical approach, there is also a need for more involvement with policy makers as well as a need for research on how more integrated approaches can be implemented. This could serve as a basis to enhance mutual understanding and communication processes. Furthermore, it is necessary to conceptualize and define criteria to evaluate progress in better decision making.

This report offers a summary of the discussions held, challenges and best practices identified, and ideas for potential solutions. The introduction lays out the rationale for convening this interdisciplinary workshop. The following sections (II.-IV.) cover the workshop objectives, the rationale and the process of predictive analytics, including its opportunities and limitations. Section V on ethical concerns and possible solutions builds on group work and brainstorming sessions during the workshop. Sections VI. and VII. on best practices outline what has already been done (for example OCHA's Peer Review work stream[2]) and where possibilities exist to further build upon existing efforts. The final sections VIII. and IX. cover community needs, namely requirements to improve practices for forecasting human mobility in contexts of crises and to map a potential way forward. The content is not attributed to specific participants since discussions were held under Chatham House Rules.

[1] More data and analysis are needed in order to determine if this a selection effect due to preferences of the presenters or if this reflects a more general trend.

[2] For more information, please visit: https://centre.humdata.org/predictive-analytics/

I. INTRODUCTION

The International Organization for Migration (IOM) jointly hosted a workshop with the German Federal Foreign Office's (FFO) Early Warning Unit from 22 to 24 October 2019, gathering key actors involved in developing predictive analytics for forecasting human mobility in situations of crises, as well as ethics experts, in order to explore the juncture between ethical principles and data science practice. The workshop included representatives from the following institutions: 510 Red Cross, ACAPS, adelphi, Brunel University London, Bundeswehr University Munich, Danish Refugee Council (DRC), Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO), the European Commission's Joint Research Centre (JRC), European Asylum Support Office (EASO), Federal Ministry of Defence (Germany), Federal Ministry of the Interior, Building and Community (Germany), Federal Office for Migration and Refugees (Germany), Flowminder, Freie Universität Berlin, German Council on Foreign Relations (DGAP), Federal Foreign Office (Germany), German Institute for International and Security Affairs (SWP), German Red Cross, Hertie School, International Committee of the Red Cross (ICRC), Internal Displacement Monitoring Centre (IDMC), International Organization for Migration (IOM), Joint IDP Profiling Service (JIPS), John Hopkins University, Kiel Institute for the World Economy (IfW Kiel), Map Action, Ministry of Foreign Affairs (Netherlands), Peace Research Institute Oslo, Stiftung Neue Verantwortung (Berlin), U.S. Department of State, UN Organization for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data, United Nations High Commissioner for Refugees (UNHCR), University of Cambridge/Winton Center for Risk and Evidence Communication, Uppsala University, and the World Food Programme (WFP). The workshop aimed to compile the state-of-the-art in forecasting approaches for human mobility in crisis situations and to address the lack of agreed upon standards, when employing automated and model based analytics for informing decisions on where, when, and how to respond and assist to crisis displaced populations. A special emphasis was put on the communication of findings of predictive analytics to non-technical audiences such as senior decision makers and risks arising from a miscommunication of the scope and limitations associated with such findings.

One of the most common consequences of crises, including intra- or international conflict as well as natural hazards, is the displacement of populations from their habitual places of residence. Hence, a detailed understanding of both the mobility and the needs of populations displaced by crises is required to effectively assist these populations and transit or arrival countries or communities. In the best case, such understanding – which relates observed mobility patterns to particular triggers and circumstances – feeds into early warning mechanisms and the forecasting of movements.

Such early warning approaches rely on firstly, the quantity and quality of available data and secondly, the analytical capacity to extract and unravel the relevant information from that data. In principle, the extraction of relevant facts from a bulk of data, as well as the forecasting based on these facts, can be conducted in a qualitative and anecdotal manner. However, factors such as the ever increasing volume, interoperability, and complexity of available data as well as advances in predictive models and algorithms, have led to a transition from qualitative analysis and intuition based decision making towards quantitative, data driven, model based forecasts and decision making support. This transition generally bears the potential to significantly increase both impartiality and transparency of decisions made in assisting crisis affected populations. Yet, it also magnifies existing ethical dilemmas related to such decisions and further creates new ethical challenges. Great care and regulatory considerations are hence required in order to fulfill the aforementioned potential.

Currently, the full consequences of employing large scale, automated, and combined analysis of datasets of different sources in order to predict human mobility in crises situations, and the increasing reliance on artificial intelligence (AI) in that field cannot be foreseen. So far, systematic evidence and evaluations of employment of AI in this particular field is missing. Humanitarian and political actors, who base their decisions on such analytics, must therefore carefully reflect on the potential risks.

The latter include, *inter alia*,

- **Systematically discriminatory action**. This can arise due to biased algorithms. Biases may emerge from incompleteness, ill-representativeness or inaccuracy of input data, or be reinforced by the training of predictive models on data sets that were labeled and systematized in a biased manner.
- **Obscured accountability**, due to gradual reduction of human oversight and resulting dehumanization of decision making processes.
- **Inadequate use of predictive models**. Models may be missing context within them, leading to irrelevant or incorrect findings. Inadequate use can also be rooted in a **lack of or misleading communication** about the scope and limitations associated with findings of predictive analytics towards nontechnical decision makers. Furthermore, scope and limitations of model outputs can be purposefully disregarded by end users, for example in order to legitimize decisions already taken.

The current posture towards such concerns is reactive rather than proactive. There is a striking absence of agreed upon core principles to be adhered to, as well as standards and benchmarks to be met, when employing any kind of model based or automated analytics in forecasting crisis related human mobility to inform decisions on where, when and how to respond to crisis displaced populations.

Hence, there is the need for a detailed discussion about the scope and limitations associated with findings of predictive analytics as well as for an ethical discussion accompanying the changing nature of political and humanitarian decision making brought about by the advances in the field of data science. These discussions ought to involve technical experts, ethical experts as well as operational staff, beneficiaries, governments and the wider donor community.

II. WORKSHOP OBJECTIVES

The following questions guided the conceptualization and implementation of the workshop:

- What are the minimum professional standards for forecasting?
- How can we strengthen interdisciplinary exchange?
- How to discuss and incorporate a do no harm approach in our work?
- How to mitigate harm of unintended consequences?
- How do we place local communities within the modelling process?
- How do we communicate forecast findings?

The aims of the workshop included to identify frameworks and standards, to enhance literacy among attendees on the subject, and to offer a forum for discussions on ethical considerations in forecasting human mobility in contexts of crises. The workshop sought to identify potentially helpful technical solutions to mitigate and communicate ethical concerns in the application of data science methods for a principled approach to forecasting. It offered a forum to introduce the work of the IOM-steered inter-agency Data Science and Ethics Group (DSEG)) and the opportunity to convene with further actors working in related fields. These included colleagues from OCHA's Center for Humanitarian Data, the Big Data for Migration Alliance (BD4M), which includes IOM – GMDAC and the European Commission's Knowledge Centre on Migration and Demography (KCMD), the UK DFID Data Science Community of Practice, and the Computational Social Science for Policy (CSS4P) group of the EU's European Asylum Support Office (EASO).

III. CURRENT USE OF PREDICTIVE ANALYTICS

Predictive analytics are requested and used by a wide variety of stakeholders such as UNHCR, IFRC, WFP, FAO, including donors (e.g. CERF), governments (including internally – for example the German FFO), humanitarian agencies, political think tanks (e.g. KCMD) and universities. It supports decision making processes such as prioritization of governmental and intergovernmental assistance, including forecast based financing, quality control and enhancing strategic documents to inform humanitarian interventions in crisis.

Current forecasts and research presented in the workshop cover topics such as why some types of conflict lead to internal and cross-border displacement, destinations of IDP and refugee movements in conflict regions, global conflict risk, climate as conflict driver, machine learning and participatory food security analysis, cholera outbreaks, one-sided violence, call detail records analysis for understanding and predicting individual disaster driven displacements, refugee or IDP caseloads, monitoring of SPHERE standards by means of computer vision, and early signs of population displacement.

A rich set of data sources is used as part of this research. This includes data on conflict events, social media, population characteristics as well as nonconventional data sources. Examples of conflict events databases include the Armed Conflict Location & Event Data Project (ACLED) and the Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset (GED). Geospatial information is provided among others by FEWSnet, Open Street Map, and further includes Integrated Food Security Phase Classification (IPC) projections, the UN Office for Disaster Risk Reduction (UNDRR) model and satellite imagery by UNOSAT, NASA, ESA and others. Social Media sources, such as Twitter, Facebook and LinkedIn are also utilized to define target groups or develop disaster maps. Data sources on population characteristics range from official sources (e.g. census data) to humanitarian assessments (e.g. by DTM, UNHCR, REACH, JIPS, to name only a few), mobile phone surveys (including computer assisted telephone interviews) and more qualitative web surveys, as well as the Gallup World Poll. Some examples of nonconventional data sources include google search logs and call detail records.

IV. LIMITATIONS OF PREDICTIVE ANALYTICS

There are several limitations concerning the generalizability and applicability of predictive analytics as currently applied in the field of forecasting mobility in contexts of crises. Models are often extremely context specific and hence not transferable/scalable. Nevertheless, users may apply them to other contexts, with at times negative consequences. In one instance, a research group reported finding models largely transferable among African countries, but not to countries outside the continent. Similarly, an overfocus on African countries was also noted in the research and initiatives presented. This raised the question of whether most testing of humanitarian predictive analytics is being focused on African countries or if this was linked to the specific focus of the invited participants. If present, geographical biases may have concerning consequences. For example, if the use of modeling becomes a condition for funding, the lack of good models for other regions with affected populations would potentially prevent responses/interventions.

Furthermore, it was noted that predictions can (in most cases) only be made for phenomena for which there is data available. Additional data is often needed to make a robust, reliable, evidence based, and valid prediction. There is a risk of introducing biases, for example when mixing data of different quality for different regions/populations causing systematically worse performance for a certain region/population. It is partly for these reasons that there is a growing discussion about the need for peer review frameworks in the context of predictive analytics. More information on this can be found at: https://centre.humdata.org/from-analytics-to-action-the-need-for-peer-review.

One-off drastic political events have been shown to cause and influence population displacement much more significantly than structural changes. Sudden events, such as border closures, are often difficult to integrate into models. In order to get a holistic grasp of a situation, a mixed methods approach and the use of quantitative and qualitative data might be needed. This could be hard to convey to decision makers at the operational or policy level who show a preference for clear thresholds to trigger action. Further biases which constitute limitations are inherent to the method used to produce the models. There is no universally valid method to make predictions based on data. In brief, data science methods are optimization algorithms subject to rules or constraints. Therefore, all methods are inherently subject to limitations and biases associated to their 'decision strategy'. Additionally, biases can arise from underreported regions and missing values in the data for training the models. The use of 'Big Data analytics' confronts challenges of integration, uniformization and availability, where the above mentioned biases are quite likely present.

V. ETHICAL CONCERNS

Multiple ethical concerns regarding the application of data science methods for forecasting human mobility in contexts of crises were identified in consideration of the humanitarian principles. In contrast, it was noted that technical experts may prioritize other ethical concerns, such as the accuracy of the models they develop rather than their potential unintended and harmful use. These concerns relate to dimensions such as accountability, transparency, neutrality, risk exposure, efficient use of resources, quality concerns as well as end users and their motives.

Concerns regarding transparency and accountability arise especially where multiple layers of predictive analysis are potentially accumulated in a model, thus further obscuring how decisions were ultimately made. Questions of neutrality also arise, for example, where analysis of humanitarian stakeholders and conflict parties becomes conflated, or where humanitarian organizations rely on private sector partners to undertake analysis on their behalf, in absence of inhouse capacities for technical work aspects. When collaborating with private entities, concerns can arise about traceability and transparency. A significant number of private sector companies with important user databases, such as social media providers, benefit from algorithms that were developed with the advertising industry in mind, rather than a secondary humanitarian application. Likewise, if non-technical staff want to use sophisticated models to respond to displacement crises, they might struggle because they cannot understand how a certain classification was derived. This could raise concerns linked to distributed (legal) responsibility and shared accountability.

Additionally, ethical concerns are related to the potential of putting affected populations at risk. For example, when utilizing scenario/stimuli based predictions of how conflict will evolve, this might raise the question whether something like an evolution of conflict ought to be predicted at all. Another risk factor exists concerning re-identifiability across (potentially unknown) datasets and there are questions around how to measure re-identifiability.

To mitigate data protection concerns, sandboxing may at times be deployed – here, the entity desiring to run an algorithm does not obtain direct access to the data, but instead, provides the data holder with the algorithm and then receives an analysis output. The approach is used for example when working with call records data. Whilst sandboxing may seem to be an effective way to protect data, this technique itself can carry risks for data security. In a worst case scenario, the algorithm deployed could contain a copy-algorithm that creates a copy of the entire database without the data holder's knowledge.

Concerns identified also relate to the narrative created, for example, around specific population groups as part of a model. For example, one model uses IDPs as a variable for drivers of conflict, which stands in stark contrast to the humanitarian understanding of IDPs as victims/survivors. This may be grounded partially in a lack of translation of

the terms used by data scientist versus humanitarian actors and highlights the need for more knowledge exchange to familiarize each with the work of the other and exchange perspectives. Sometimes, models may rely overly on risk literacy of the recipients of a forecasting output and may make people feel overly safe (with potentially dramatic consequences) if they misinterpret the findings of a model.

Another concern relates to the amount of data collected versus what data is used and how. To draw an analogy from medicine, if a doctor were to collect data on a patient without being able to provide a solution to the problem they collect data on, they would have to inform the patient that this is an experimental study and the patient is, *de facto*, a research subject. In the humanitarian sector, however, there is a recognized trend to continuously collect data without always having an analysis plan and clear links between analysis and response in place. Data should be collected for a given purpose, with clear objectives, including an engagement mechanism with the affected populations.

Furthermore, if models are not being used due to a lack of trust or legal feasibility, questions arise as to whether it is ethical to invest oftentimes substantial resources for model development, which may mean less availability of resources for other interventions. Also related to funding, participants noted that whilst some funding channels require applicants to have a section on ethics, this section deals primarily with data protection and fails to require input on the more salient ethical concerns in the context of modelling.

Regarding end users, it was noted that data is often downloaded in locations where Big Data private companies are located, raising further concerns about unintended use and consequences, including potential commercialization or even weaponization. For example, it is conceivable that insurance companies could come across risk and preparedness assessments and as a result increase fees for the most vulnerable populations. The concerns thus reflected back on some of the key questions that triggered the conceptualization of the workshop in the first place, namely how to discuss and implement the do no harm approach; (a) how to mitigate harm of unintended consequences and (b) how to place local communities within the modelling process to ensure greater degrees of adaptation to local needs. There was an acknowledgment that data science, whilst carrying promises for improved action, can also harm people, since additional insights can be used not only for humanitarian purposes but also carries the potential for malicious use. Relatedly, data science in the worst case can harm humanitarianism, undermining access and acceptance of the latter.

VI. POTENTIAL SOLUTIONS TO ETHICAL CONCERNS

As part of the workshop, participants sought to identify a set of possible solutions to ethical concerns in the application of data science methods for forecasting human mobility in contexts of crises. These range from technical measures to improved communication strategies and quality assurance.

Regarding technical solutions, participants highlighted the importance to indicate the range of uncertainty on forecasting results, as well as the sensitivity of results to specific underlying assumptions. Slightly varying key assumptions in model setup such as exchanging or iteratively excluding input variables can show where predictions may be greatly affected by single factors and give a general account of model uncertainty/robustness of the model. This highlights a discussion from the third day about the critical importance on developing clear and transparent communication when it comes to model accuracy and limitations that can be understood by a diverse audience.

Agent-based models which are not trained on preexisting data and therefore, cannot be subject to training enforced bias, were proposed by some, with the benefit of being transparent in their assumptions and thus open to review by different stakeholders. Another key technical solution related to rigorous anonymization. It was also observed that the risk of modelling is different when scale comes into play, i.e. it matters whether research is done at a very

granular level, or on broader geographic areas, and at which time intervals. Less granularity translates into fewer ethical concerns for misapplication, for example, in the forecasting of potential conflict developments, but also less actionability. As mentioned in the previous section, sandboxing is sometimes used as a solution that enables running analysis without sharing datasets, though it may itself pose some concerns.

In general, the importance of ensuring good accessibility to data within the humanitarian community (and with vetted counterparts) was highlighted. To ensure that data is not abused, user management via humanitarian ID on HDX can be a solution.

From a communication side, ethical concerns can be partially addressed by developing an adequate framework for communicating uncertainty. It was observed that whilst not communicating uncertainty in the scientific community will lead to a distrust of results, in the humanitarian and political sector it can sometimes be the opposite. Notwithstanding, it was broadly agreed that there is a responsibility to educate the users of analysis output to generate a better understanding of scope, limitations and possible applications of the outputs. It was noted that some organizations have already developed a system for methodology and uncertainty communication to enhance transparency. It was also noted that transparent communication of analysis outputs, where the resulting messages are technically sound but politically difficult to defend, can be difficult but is still critical.

The role of the OCHA Data Centre was noted in providing modelling support, quality assurance particularly on predictive analytics, and fostering a community of practice and interest, including at field level, to jointly engage in further improvement of quality assurance processes. Furthermore, OCHA's leading role in promoting a new approach of peer review for predictive analytics that seeks to assess ethical and technical dimensions, as well as humanitarian relevance of newly developed models, was illustrated. Outcomes of the peer review process are recommendations that propose different ways to mitigate the risk acknowledging that reducing the risk to zero is impossible. Highlighting the risk, and identifying the accountable stakeholders, are key outputs of this process.

VII. BEST PRACTICES

Best practices to mitigate ethical concerns in the application of data science methods thus comprise joint risk assessments, ensuring transparency and traceability, establishing benchmarks, not only relying on models, encouraging mixed analysis, ensuring well defined processes and undertaking peer reviews. Some best practices and limitations within current applications on each of these dimensions are further detailed below.

Joint risk assessment can set out from categorizing ethical concerns into epistemic risks (inconclusive, inscrutable or misguided evidence) and normative risks (unfair outcomes, transformative effects). An additional aspect are concerns of traceability. Data scientists then develop accuracy measures, fairness measures, causality measures, statistical validation, and measures to determine quality of input data. Through this, possibly unfair outcomes can be traced and quantified, establishing the degree of representativeness, integrity and conclusiveness. Meanwhile policy makers, donors and practitioners, assessing for example the possibility for political impact, reputational damage, and the impact of the decisions on affected/vulnerable populations, scrutinize transformative effects. These actors are also best positioned to identify the political/humanitarian relevance of the outputs.[3]

Quantitative analysis in general, and predictive analytics in particular, are often challenged based on false positives and false negatives. However, as a best practice awareness should be raised to the fact that expert judgement is subject to false positives and false negatives as well. The only difference being that experts generally do not trace and quantify their respective tendencies towards certain error types, which leads to less transparency and limited possibilities to evaluate and improve analytical outputs. Best practice includes establishing realistic benchmarks

[3] Further reading in this regard includes Signal Code: Ethical Obligations for Humanitarian Information Activities; and Mittelstadt et al., "Ethics and Algorithms: Mapping the Debate".

based on an understanding of what has been tried and achieved before and comparing new developments to said benchmarks. One option is using the cost/benefit approach, or even the Ockham's razor principle. The latter is based on the premise that where there are several hypotheses for the same thing, all of which perform equally well, the simplest one ought to be used. As such, the challenge becomes to provide meaningful comparisons/performance metrics, to be more accurate and try to beat simple forecasting models. For example, if one were to forecast more migrants coming from a more populous country than a less populous one, a simple model looking at fixed proportion of total population may perform just as well as a more complex one and should be preferred.

Models always ought to be used in combination with expert judgement as a means of triangulation. If experts confirm model findings, then this is a good indicator for reliance. If not, then both the model and the expert should be further questioned, with the questioning process itself constituting part of a learning process. Expert input can also be integrated into modelling (for example via Bayesian analysis[4]) where qualitative data can inform modelling. However, one may want to keep the model and expert opinion separate to be able to compare at the end.

It was observed that in most countries more traditional data, such as census data and multisectoral needs assessment data, are often the primary data source to be complemented but not replaced by Big Data analysis. Furthermore, there always needs to be a stage in the research process where data outputs are cross-checked to ensure applicability to field operations.

Well defined processes are also critical, to ensure how analysis translates into action, to ensure regular incidence reporting, feedback loops and evaluation in the entire cycle of predictive analytics. To draw another medical analogy: evidence based decision making has century-long history in medicine, based on a joint effort of researcher, doctor and patient. In the humanitarian sector these participant groups can be seen as the data scientist, practitioner and beneficiary.

Another key aspect is to ensure usability of models. There is an abundance of top-down solution oriented approaches, but it seems sometimes unclear what would be needed from the field perspective. It is, thus, key to promote more bottom-up approaches and interrogate if models at the global level are really what your users need, or rather locally developed context tailored models. Here, user needs may vary between for example governments, who use models for strategic global prioritization purposes, and use of models in international organizations, that are more focused on a specific operational application.

Whilst peer review is widely agreed to be a critical best practice, it was noted that sometimes its application might not be possible because of the level of confidentiality of data inputs, especially where working in conflict environments. A possible solution can be to establish a scientific advisory board and establish other mechanisms to ensure institutions are able to have exchanges about data and methodological approaches, at least on the technical level.

VIII. COMMUNITY NEEDS

In order to implement the best practices listed above, developing a common language, enhanced partnerships, exchange platforms, professional standards, data access, and well defined roles and responsibilities, are all considered critical.

With regards to speaking a common language, even among experts (be it technical, humanitarian, or ethical) of the same domain a common definition of terms is often lacking. This frequently leads to misunderstandings with potentially dramatic consequences. To help resolve this challenge, it was agreed to create a glossary of terms that

[4] Bayesian inference techniques specify how one should update one's beliefs upon observing data. Accessed via: https://seeing-theory.brown.edu/bayesian-inference/index.html

have dual meanings / are open to misinterpretation by different audiences. The glossary will be published at a later stage to complement this report. There are also existing resources that help to clarify concepts when discussing with counterparts who are coming from a different field[5]. It was also considered critical to develop shared tools and procedures to evaluate risk. This cannot just be a process for data scientists to undertake in isolation, but needs to involve representatives from data science, ethics, legal domains, decision makers and, practitioners alike.

Enhanced partnerships with external counterparts are also critical for making progress. Data from traditional sources such as statistical offices (e.g. national census) may be available less regularly, not available in some cases or not recently updated. Tech companies provide the opportunity to use novel and innovative data sources to estimate population and displacement numbers or even flows. Thus, companies have a potentially significant role to play, but partnerships need to be transparent and address ethical concerns satisfactorily. The private sector's incentive for getting engaged needs to be critically questioned and assessed to avoid unintended consequences and weigh risks against benefits.

It was also noted that data scientists occasionally find themselves isolated in their organizations leading to duplication of efforts, a lack of effective peer review, and decreasing product quality. Such separation is counter-productive, and critical to overcome. Linked to this, breaking down barriers also in the accessibility to data sources is required to obtain more cross-country, comparable data, sub-national, high-frequency real-time data, and more historical data, to name a few. Making such outputs available on a sustainable basis is also critical so that model outputs can be utilized regularly, are reviewable and traceable. Whereas some data may simply not yet exist, some might be held in silos and could be accessed through greater cooperation.

Participants further agreed that professional standards for the application of data science methods in forecasting human mobility in contexts of crises is needed as a technical precondition just as in any other professional area or humanitarian sector. For example, there is no humanitarian systemwide analytical framework on social media analysis. Ethical red lines ought to be established, and an equivalent to the Common Humanitarian Standards / SPHERE standards, a kind of "Digital Sphere standards" should be developed. In the last decade, multiple organizations have developed their own data protection guidelines but there is a general lack of sector wide standards for self-regulation comparable to the SPHERE standards[6].

Roles and responsibilities also need to be clearly defined and outlined to deal with chains of responsibility. There are "perceived" different levels and stages of accountability: the data scientist is accountable for the accuracy of the model, the manager for the "product", and the overall coordinator for the implementation. Every actor is responsible for correctly estimating and communicating the risk in their domain and is accountable for harm arising from underestimating the risk they were to evaluate. However, responsibility where it is shared, can easily be shifted from one actor to another, leading ultimately to nobody feeling accountable. As a rule of thumb, the stakeholder hardest affected by risk, must have most weight in decision making.

IX. FUTURE DIRECTIONS

The application of data science methods such as predictive analytics in the humanitarian sector remains very much a nascent workstream. For 2020 and beyond, efforts might continue to move the sector more towards a predictive rather than reactive approach and to utilize forecasting methodologies to support humanitarian decision making and outcomes. This requires being able to assess how events are likely to unfold and evolve to move beyond

[5] See for example: Getting Ahead of Crises: A Thesaurus for Anticipatory Humanitarian Action (June 2019), available at https://cerf.un.org/sites/default/files/resources/Thesaurus_single%20column_WORKING_DRAFT.pdf

[6] SPHERE Standards are an identified set of principles and minimum humanitarian standards to be applied in humanitarian responses. For more information, please visit: https://spherestandards.org/humanitarian-standards/

traditionally reactive responses. Considering existing constraints of humanitarian access, the use of models also promises in potential of filling blackspots where information cannot be directly obtained on the ground.

To be able to move effectively in that direction, partnerships will be critical within the sector and beyond. For example, it would be helpful to invite representatives from other sectors to engage in knowledge exchange and learning, building expertise with the support of external actors. These could include experts working in weather forecasts, medicine, insurance, epidemiology, to name only a few. It will also be critical to involve ethics experts, decision makers, academia and humanitarian field staff more effectively. Continued investment is required to build additional capacity within the humanitarian sector, for example, engaging more data scientists and ensuring sufficient data handling capacity at both organizational and field level. The humanitarian sector will also need to continue to strengthen utilization of the data it uses, including ensuring increasing interoperability, and integration of nonhumanitarian data sources such as Big Data to support the sector's work. There are already efforts to create a stronger community of practice. This community needs to be multi-skilled (including data scientists, field operations, donors, senior management) as well as interdisciplinary (covering humanitarian, policy, and academic spheres). In addition, more work is needed to engage the private sector adequately whilst preempting unintended or potentially harmful consequences. The future work around private partnerships has been highlighted to have huge potential, but there are significant efforts to be placed to ensure transparency in the process.

To move ahead constructively, problems to be addressed by forecasting methodologies need to be identified by end users in the field and should not be technology/data driven. This makes it paramount to learn more about where humanitarian actors need support, and then use technology/data science as a tool in addressing identified challenges, rather than applying models without any prior insight as to how they will be applied and their outputs will be used. Forecasting displacement including potential attacks as triggers of displacement raises significant ethical challenges that are further exacerbated if there is no clear linkage to how model outputs will be used at field level. Therefore, it will remain critical to coordinate future work with operational humanitarian organizations, such as IOM, WFP, UNHCR, and ICRC to name a few. These can act as interlocutors to communicate what happens at the field level, and engage the wider community of practice on the topic and potential solutions.

Future workstreams could/should be closely linked and build on that of other entities, such as OCHA's Predictive Analytics workstream, and their peer review process, as well as the Data Science and Ethics Group (DSEG) and the IASC Results Group 1 Subgroup on Data Responsibility which OCHA, UNHCR and IOM co-lead.

ANNEX

SPEAKERS

Sarah Dryhurst is a postdoctoral research associate at the Winton Centre for Risk and Evidence Communication. Her research forms part of the EU RISE project and focuses on developing and evaluating tools to communicate seismic risk to a variety of public and policy making audiences. She also researches how people understand and respond to uncertainty about evidence, and the efficacy of common forms of risk communication such as risk matrices. She uses both quantitative and qualitative approaches in her research.

Maria Teresa Espinosa is a Monitoring Expert at the Internal Displacement Monitoring Centre (IDMC) since November 2016. She supervises the compilation of data and information on internal displacement in the Americas and Europe. Maria leads and supports different information management projects and provides technical support for the development of the Spatial Data Infrastructure, the development of data protection internal policies and practices.

Derek Groen is a Lecturer in Simulation and Modelling since 2015. He specializes in multiscale simulation, high performance computing and distributed computing, migration modelling, and has published over 25 peer reviewed journal papers. Derek is PI for Brunel and also lead developer of the FabSim toolkit, the Flee agent based migration modelling code and the MPWide communication library. Derek has experience with a wide range of simulation methods, including lattice-Boltzmann, stellar and cosmological N-body dynamics, molecular dynamics and agent based modelling.

Matina Halkia is Team Leader in Peace and Stability, Disaster Risk Management Unit, of the Space Security and Migration directorate in European Commission's Joint Research Centre. Since 2016 she has been responsible for quantitative early warning for conflict prevention, crisis monitoring and situation awareness, as well as post conflict urban recovery using remote sensing and integrated geospatial and statistical tools. She has been a JRC senior scientific officer contributing to evidence based policy support using remote sensing and media monitoring technologies since 2001.

Lisa Hultman is Associate Professor of Peace and Conflict Research at Uppsala University, and a research associate of ViEWS (Violence Early-Warning System). In her research, she uses quantitative methods to explain patterns of civil war violence, international responses to civilian atrocities, and the effect of international interventions. A recent article (published in the Journal of Politics) uses simulations to provide a comprehensive analysis of UN peacekeeping and the conflict-reducing effect of various policy scenarios.

Juliane Klatt is data science and quality control officer in the International Organization for Migration's (IOM) Displacement Tracking Matrix (DTM). With a background in theoretical physics, and data analysis in the medical sector, she leads global analyses on urban displacement, supports DTM country operations in employing predictive analytics for early warning and data quality control, and develops transition and recovery oriented analyses of internal displacement.

Véronique Lefebvre is a researcher in data science and signal processing. At Flowminder she has worked with mobile phone usage data to identify and predict disaster driven displacements, and to optimize the placement of services based on population mobility. She has also worked on predictive modelling of poverty, on analyzing the factors limiting financial inclusion, and on automating the process of designing national sampling frames. Her core research interest is in optimizing information extraction and data collection to support decision making for humanitarian and human development efforts.

Stefanie Lux - is the coordinator of the Forecast based Financing (FbF) programs at German Red Cross (GRC). Based on forecast information and risk analysis, FbF releases humanitarian funding for pre-agreed activities aiming to reduce the suffering and losses caused by extreme events. Stefanie Lux is responsible for advancing the work in this field in

GRC's twelve pilot countries, as well as for the shaping the methodology at global level in cooperation with other Red Cross partners, scientists, UN agencies and NGOs. Before joining GRC in Berlin, she has worked for different humanitarian organisations in Switzerland, Sierra Leone, Liberia and Haiti. She holds a Master's degree in international law from the Graduate Institute of International Studies in Geneva.

Thomas Mayer is a Data Scientist at the German Federal Foreign Office and supports PREVIEW since 2016. His current work focuses mainly on the conception, development and implementation of data driven visualization applications. Besides he works on several forecasting and prediction projects for the German Government with a focus on predicting political violence and building interpretable machine learning models.

Leonardo Milano leads the Predictive Analytics Team at the UN OCHA Centre for Humanitarian Data. He previously worked as Senior Data Scientist for the Norwegian Refugee Council (Internal Displacement Monitoring Centre). After obtaining his Ph.D. in physics he worked in two of the leading scientific research organizations worldwide: CERN, the European Organisation for Nuclear Research and LBL the Lawrence Berkeley National Laboratory in Berkeley, CA, U.S.

Rebeca Moreno is the founder of "Oportunidades para Internacionalistas", cofounder of @UrbanWatchers and currently working as Innovation Officer – Data Scientist of the UN Refugee Agency (UNHCR) Innovation Service. She supported field operations in Latin America, Europe, East Africa and MENA, focusing on the potential of data science and artificial intelligence for enhanced protection and decision making. Additionally, Rebeca collaborated with WFP Emergency Operations by conducting research on the funding shortfall on WFP Support to Syrian Refugees in Jordan.

Ryan Mukherjee is a senior researcher at the Johns Hopkins University Applied Physics Laboratory with over ten years of experience in machine learning, computer vision, and remote sensing. Most recently, Ryan led a large research effort developing satellite image classification algorithms and has been working on bringing similar state-of-the-art technology to the humanitarian domain.

Nuno Nunes is the International Organization for Migration's (IOM) Global Displacement Tracking Matrix (DTM) Coordinator who oversees the implementation of displacement tracking and field data collection in humanitarian emergencies. Working with partners at different levels of the private sector, civil society, academia and international organizations, Nuno is responsible for the organizations' efforts toward building collaborative approaches to management of displaced populations. He has been managing the team involved in the design and implementation of mobility tracking operations enabling a more data driven approach to humanitarian relief and policy making.

Elisa Omodei is a data scientist and the Predictive Analytics Lead at the World Food Programme's Research, Assessment and Monitoring division. Previously, she worked at UNICEF's Office of Innovation. She holds a BSc and a MSc in Physics and a PhD in Applied Mathematics for the Social Sciences, and before joining the UN she was an academic researcher studying Complex Systems science.

Carlos Santamaria is a scientific project officer at the Joint Research Centre (JRC), the European Commission's in-house science service. As a member of the Knowledge Centre on Migration and Demography (KCMD), his main responsibility is to address existing data gaps in the area of migration, as a way to strengthen evidence informed EU policymaking. Part of this work involves research into innovative data sources that can provide new insights into migration and human mobility, including during crises.

Tobias Stöhr is Senior Researcher at the Kiel Institute for the World Economy. He is experienced in quantitative research and data collection, both in the field (household surveys, experiments) and based on digital sources ("big data"). In his work, he combines these with traditional data such as administrative datasets to study migration decisions, attitudes towards migration and the effects of migration. He has both published academically on migration prediction and is involved in policy oriented projects that relate to migration decisions and migration policy.

Diana Suleimenova is a Research Fellow in Multiscale Migration Prediction. Her research concentrates on migration modelling and verification, validation and uncertainty quantification of multiscale applications deployed on emerging exascale platforms. Her dissertation focused on quantitative data analysis of forced displacement and the development of an automated agent based modelling technique to predict the distribution of incoming refugees across neighboring camps. Diana has previously codeveloped a simulation development approach for migration modelling.

Andrew Thow works on quantitative methods for understanding crisis risk and severity in the Information Services Branch of the UN Office for the Coordination of Humanitarian Affairs (OCHA). He is coordinator of INFORM, a multi stakeholder partnership for developing shared crisis analysis. He previously worked in humanitarian policy and was managing editor of OCHA flagship publications, "Saving Lives Today and Tomorrow", "Humanitarianism in the Network Age" and "World Humanitarian Data and Trends". Prior to this he worked in the private sector, developing methodologies to quantify non-financial risks and advising corporate and public sector clients.

Jasper Tjaden works as Data and Impact Analytics Coordinator at IOM's Global Migration Data Analysis Centre (GMDAC). Jasper joined GMDAC in 2016 and works on projects with a focus on data driven, evidence based migration policies. He is leading GMDAC's work on rigorous (experimental and quasi experimental) impact evaluations of IOM interventions and migration scenario and forecasting work. Jasper has authored more than 30 publications on migration and involved in many international migration projects. Prior to IOM, he worked for the World Bank and the Migration Policy Group.

Andreas Forø Tollefsen is a Senior Researcher at the Peace Research Institute Oslo (PRIO). His current work focuses on how local institutions moderate the relationship between grievances and conflict, as well as the impact of political violence on local migration processes. His research has been published in Journal of Peace Research, Journal of Conflict Resolution, Political Geography, World Development, Demography and Conflict Management and Peace Science.

Wilhelmina Welsch heads the Information Management and Innovation Unit of the Joint IDP Profiling Service (JIPS) based in Geneva. She leads the development and implementation of JIPS' information management strategy and promotes innovation in its work processes. A Master of Political Science graduate, Wilhelmina first started working on data and analysis within the UN system in 2009 and has since specialized in the strategic research, design and analysis of information management systems for development and humanitarian affairs as well as human rights.

PARTICIPATING ORGANIZATIONS

510 Red Cross, ACAPS, adelphi, Brunel University London, Bundeswehr University Munich, Danish Refugee Council (DRC), Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO), the European Commission's Joint Research Centre (JRC), European Asylum Support Office (EASO), Federal Ministry of Defence (Germany), Federal Ministry of the Interior, Building and Community(Germany), Federal Office for Migration and Refugees (Germany), Flowminder, Freie Universität Berlin, German Council on Foreign Relations (DGAP), Federal Foreign Office (Germany), German Institute for International and Security Affairs (SWP), German Red Cross, Hertie School, International Committee of the Red Cross (ICRC), Internal Displacement Monitoring Centre (IDMC), International Organization for Migration (IOM), Joint IDP Profiling Service (JIPS), John Hopkins University, Kiel Institute for the World Economy (IfW Kiel), Map Action, Ministry of Foreign Affairs (Netherlands), Peace Research Institute Oslo, Stiftung Neue Verantwortung (Berlin), U.S. Department of State, UN Organization for the Coordination of Humanitarian Affairs (OCHA) Centre for Humanitarian Data, United Nations High Commissioner for Refugees (UNHCR), University of Cambridge/Winton Center for Risk and Evidence Communication, Uppsala University, and the World Food Programme (WFP).