

# Data science episode - Final edit

## SPEAKERS

Rob McNeil, Christina Pao, Jacqui Broadhead, Emre Korkmaz

**Jacqui Broadhead** 00:00

Hello, and welcome to The Migration Oxford Podcast. I'm Jackie Broadhead.

**Rob McNeil** 00:03

And I'm Rob McNeil.

**Jacqui Broadhead** 00:05

Today we're talking about data science and all of the questions surrounding this central question to migration studies, I think about who counts and who does the counting. There's so much interest about numbers in migration, it can sometimes feel like it dominates the discussion. And yet, we don't always talk so much about some of the questions that sit underneath how that data is gathered and, and what we do with it. So, what was your kind of instincts around choosing this topic for the podcast,

**Rob McNeil** 00:37

I just think this is fundamental. Everything that we do really is about data in one way or another. And of course, we have to recognise what the limitations of that are. But some of the things that we've been dealing with over the last little while in particular have been really interesting. Like, for example, we've seen suggestions that during COVID, there was this massive decline in the number of migrants in the UK, which actually didn't necessarily look like it was true. And one of the reasons that the data was starting to show this is that the methodologies that people were using just to collect that information in the first place have shifted. So that instead of there being face to face interviews with people, for things like the Labour Force Survey, suddenly everything was done online, and through phone calls, and through various different things that involved not actually having that direct contact with people. And of course, some people were less likely to answer questions than others, and particularly people whose English wasn't great, were much more reticent about actually participating in these things. And that meant that you naturally got something skewed happening, which is that migrants were less visible in the data. And so, recognising just how small basic things about data collection can have an impact on the way that people understand the world is really, really important. And I think one of the things that's really key to both of the people that we're going to be speaking to in today's podcast, is the kind of, like, the overlap between the importance of quantitative data and qualitative analysis, which is really, really key.

**Jacqui Broadhead** 02:04

I think that's so true; I think sometimes we have this idea that numbers are automatically objective. And both have emphasised the importance of mixed methods and remembering the human within the data and vice versa. But it's also I think, really true that we have seen some of the traditional ways of

capturing data in particularly during COVID have suffered. And that's one of the things that we're discussing today. And actually the potential of new datasets that are available to migration researchers through technology, so things like data from mobile phones, even from the European Space Agency, all of these kinds of new forms of data, which on some level are quite exciting, because they give us the capacity to be able to understand mobility better but seem to come with this huge additional raft of ethical, legal and privacy related challenges.

**Rob McNeil** 02:57

That's right. And I think this is the thing that we all need to get our heads around the world is changing, the way that we collect data is changing, the questions about how and why we might use things like social media data, or information from mobile phone tracking, and stuff like that starts to become really, really important. And you know, as a sort of community of scholars, we can get very excited about the prospect of being able to understand vastly more. But of course, this comes with really complex ethical questions, particularly questions about consent, and how you actually understand whether or not the information that you're acquiring has been provided to you willingly and if it hasn't, what you do about that?

**Jacqui Broadhead** 03:36

Absolutely, I think it's such a fascinating discussion. And I'm so happy that we have two such expert researchers that are really able to dive into these issues with us.

**Rob McNeil** 03:46

Excellent. So I'm very pleased in that case to introduce Christina Pao, who is a PhD student at Princeton studying Sociology and Social Policy, and who organised an extremely brilliant measuring migration conference with Maksim Zubok and Domiziana Turcatti at Oxford while she was here, a year or so ago, and Dr. Emre Korkmaz, who's a lecturer in migration and development, and he was just published a brilliant new book on data science and migration studies, Emre, you've edited this new book data science for migration and mobility. Now, this is an absolutely enormous topic. So, can you just give me a quick overview of what sort of areas the book covers? And why you felt that this was particularly needed now?

**Emre Korkmaz** 04:30

Yeah, in the last two years, we have been working on this book, it has 19 articles and 48 academics contributed to this book, from migration studies, computer science and data science. So, we aim to bridge these two disciplines, migration studies and data science particularly the book introduces new data sources and have to analyse them for migrations because we have three main sections. In the first section we discuss why these new data sources are significant for migration and mobility studies and explain the main ethical and legal concerns. Then in the second section, we introduce these data sources, how you can access these data sets, what kind of expertise you need to analyse them. So, this is important for migration scholars to find co-investigators. Then in the third section, our contributors explain various case studies on using data sources to analyse migration and mobility. I think we need to bridge these disciplines and interdisciplinary team should be formed to analyse migration and mobility because in some cases, as a migration scholar, when I read some data analysis of migration, I find these datasets fantastic, the techniques used are fascinating. However, when you

share these data analysis with many governments, for instance, when you predict migration, they will use this to stop them, and they may cause death of many migrants. So, because data analysis help decision makers to predict and track migratory movements, this may allow them to stop before they reach to the borders. And in today's world, many governments will not build shelters or bring food when they have certain info. So therefore, if you do not want to harm people in need, technical expertise will not be sufficient. So, migration schools and data scientists should collaborate. So, this book aims to contribute to this process.

**Rob McNeil 06:32**

Absolutely, I can. So, Christina, similarly, you've also organised this conference recently here in Oxford, which brought together scholars presenting analysis dealing with a huge range of different topics, and covering the complexities of measuring and understanding migration. So, one of the key contextual factors that you've been dealing with here is that of COVID. So, can you tell us a bit about how COVID affected scholarship in analysing migration? Generally?

**Christina Pao 06:56**

Yeah, there are, I think, three different things that have come up in the process of developing this conference, one of which was how COVID has affected migration patterns themselves. So how have the motivations for moving changed? Are people moving in different ways? Can they move with the same family members as before, right, so all the risks that come with COVID also come with migrating with COVID in a COVID period. The second element is how COVID has changed the methods that we have to use, one of which is thinking for qualitative scholars of not being able to do field work. For quantitative scholars, there might be gaps in the datasets that they use. So ongoing panel data, for instance, might have just a hole in the time of COVID. So, we might not know the same things that we knew before from repeat datasets. And the third thing that I didn't expect was how on a very meta level COVID has changed the way researchers of migration can come together in a COVID period. So, for instance, we had struggles getting visas to migration scholars coming internationally, for instance, and it felt depressingly ironic to have a migration conference where people couldn't actually come who wanted to. And so, thinking about the ways that COVID has also impacted, the ways that we come together as a scholarly community has been very difficult, and also, I think, has provided lots of opportunities for us to be very innovative. So, we did a hybrid conference, we think that this is more feasible than we thought in the beginning. And so, there are ways to work around it. But nonetheless, it's been a very reflective process to think about, like the hegemonic nature of knowledge production and the ways that COVID actually exacerbate some of these things. Certain scholars that are already based in the UK continue to have in person opportunities to accessing certain conferences and scholars coming from all across the world might not have the same ability to come. So, we have to be very mindful of that.

**Rob McNeil 08:57**

Absolutely. Now, Emre, obviously, a new book is hot off the press. But one of the things that seems clear is that data science publications, and as a data science publication, this does naturally tend to skew towards quantitative analysis. Now, do you think that modern quantitative analysis techniques give us a fuller picture of migration issues than previous approaches? And how well can they capture the complexities and the human realities that underpin any social analysis like this?

**Emre Korkmaz** 09:27

So, I see three main challenges related to datasets. The first is the digital inequality. For instance, if you would like to analyse mobile phone CDR data or social media data, many migrants don't have phones or smartphones. So particularly based on gender, age and class, access to internet or mobile phone could be difficult. So, data collected could only be produced by young male, relatively wealthy and more educated migrants or in general, it might not represent all the population. The second issue is that data we produce does not represent all of our life. So, we evaluate many social, economic cultural issues when we decide on migration. And we may not decide migration individually, this could be a family decision as well. So, therefore, we need qualitative study to fill these gaps to understand the reasons to migrate. And thirdly, most of these datasets are collected by private corporations, such as digital platforms, social media companies, or most importantly, mobile phone operators. They analyse these datasets to send us commercial messages. And these companies are not willing to share these datasets. And you know, they are not audited due to privacy or commercial interests. So, accessing these datasets is a very difficult issue to in other social sciences, disciplines like economics and politics. You know, there are many academics working on large datasets using advanced statistical methodologies. In migration research, we are more qualitative people, you know, paying more importance to individual stories, exploring cultural and social networks. And in my personal opinion, quantitative analysis would assist us to and complement these qualitative methods. And it will also demonstrate us the power of like corporations, or governments in observing, influencing, and managing migratory movements.

**Rob McNeil** 11:34

That's a fascinating answer. And thank you very much, Christina. Similarly, I mean, your conference dug into the relative merits of quantitative and qualitative methods for understanding migration. So, from your point of view, I mean, what do you think the relative benefits of each approach are? And as Emre was saying, do you think that scholars who are embedded in one approach should be trying to broaden their horizons and add depth and nuance to their work by embracing different ways of doing things? Or do you think that really, there's also a value in really honing highly specialised skills, rather than trying to be, I suppose, a jack of all trades here?

**Christina Pao** 12:10

Sure, I loved Emre's answer. I also am a huge proponent of mixed methods. And I think most people who do quant or qual research often do mixed methods research. It's just that we're finding new ways to label things. And obviously, these categories are constructed, etcetera. But I'm thinking of several of the presentations that we had in the conference, some of which dealt with, for instance, social media data, as Emory had mentioned, and even the large scale computational social scientists that we had used some kind of hand code at some point, to use some qualitative methodologies in order to develop their models. And I think this speaks to the fact that the data are changing. With the changing data and with the changing questions that we have, our methods need to continue adapting to them. And I think a lot of the typical merits of quantitative and qualitative research for migration still hold. Quantitative research does really well, the kind of macro level, now casting of international migration, or thinking about population level trends, or doing international comparisons and seeing if they're statistically different and significant, can measure a substantive treatment effect, or a causal effect of one

intervention over another for a policy. But qualitative research does a really good job of filling those gaps, questioning what is missing in those data, because I really like what you said Emre, about the data that we produce is not us in our entirety. Qualitative research does a really good job of digging into the holes, figuring out what we don't know. And I think mixed methods research is going to be the future. I think it is already happening, most people are doing it, maybe not calling it that. But I think migration scholars are at the forefront of really ensuring that we're bridging these divides.

**Rob McNeil** 14:03

I like the fact that you've said that because I think that this is one of these spaces in which migration really does, if not lead the way, it's certainly part of a kind of vanguard of disciplines that are really, really focusing on the value of these kinds of things, methodologies. And I don't think you can really do the research without trying to cover both sides to some extent. So, look, I've got a question for both of you now. So, both the book and the conference addressed the issue of the ethics of migration data collection use. And obviously, we've covered a little bit of this, but it's an absolutely fundamental issue that we need to think about. It's also very well-rehearsed issue. So, I'd be quite interested in knowing what came out of the work that both of you did, that might have made you reassess or reimagine some of the ethical considerations of measuring migration. So, I mean, Christina, if I could start with you on that.

**Christina Pao** 14:51

This is such a tricky thing for the reasons that we mentioned before data are proliferating even when we don't necessarily know what those data are as scholars. And I think one of the questions that Maksim, my co organiser, and I wanted to highlight in developing our conference was, when should we not collect data? What are our ethical responsibilities of holding possessing individual data points, which are individual lives and stories? And when can we ethically use these for our analyses? When do we know enough to draw a conclusion? And so yeah, the exact question that we had written out was, what are the ethics and implications of what it means to track migratory flows? And when are these data appropriate, helpful, or harmful? And I think the driving point that I want to emphasise is that measurement for measurement's sake, can be harmful. And as long as migration is stigmatised, I think these data can be weaponized. I come from North America, there have been many times in which for instance, census data had been used to target migrant populations, my family was impacted by Japanese internment. So, thinking about what it means to actually ask questions that might seem harmless on the outset, but hold those data can be difficult and ensuring privacy, ensuring consent, all those things kind of continue to proliferate, the larger scale our data become.

**Rob McNeil** 16:20

That's a huge answer. Christina, very, very interesting. Emre, I'm um, I'm intrigued to know what the book has done that's kind of given you maybe a new way of thinking about the ethics.

**Emre Korkmaz** 16:32

Yeah, this is a very important question. So, when you convene qualitative research. Of course, we think about our interviewee's security and all their consent privacy, these are really important questions, so we have to design it. But in the quantitative study, you can receive these datasets from a company or from an open source and you are on your desk, and you don't know who these people are. So, it might

be also very harmful for them. So, the ethics is a very important topic. And in our book, we discussed ethics in a separate chapter. But we also asked from all our contributors to have a section on ethics in their chapters, because many readers may just download one paper, so they should be aware of the ethical concerns of that specific introduced dataset. Also, case studies section provides useful information to learn how other research groups for ethical considerations in their own research projects. There are many concerns on this topic, you know, these are privacy concerns and legal related issues, you know, are we respecting minors' privacy? Is there any possibility to receive consent? So, we are asking these questions all the time. And, you know, how could we ask consent from migrants if we analyse satellite data provided by the European Space Agency? So, this is a very difficult question, and I still don't believe that we are living in a post consent society. So, we should find some creative ways to overcome these challenges.

**Rob McNeil** 18:04

Excellent. Again, both really, really fascinating answers. I, again, have a question, which is for both of you. This is this question about policymaking media and public attitudes and migration, which are all potentially heavily influenced by data and analysis of migration statistics, and very much by reports from academia and from civil society, as well as from sort of government, statistical bodies, and things like that on various different elements of migration. So, what is your work on the book, and on the conference made you think about the responsibilities of the academy, in particular, in sharing and explaining data, which then go on to shape public debates? Emre, I mean if I could start with you,

**Emre Korkmaz** 18:45

You know, I think data-based analysis is very important in informing the public. However, we should not only rely on these data sets, you know, these are political and social issues and migration is a very dynamic field. Data sets might give a better picture of the current moment but should not be elaborated as the sole representation of the truth, you know, they are not very neutral to. So we should include various issues such as digital inequality and corporate interests and government policies. For instance, last year, Poland could be considered as hostile towards refugees. But after the war in the Ukraine, they are a welcoming country. So, these data sets will show opposite results when you compare these two years, but this is still very valuable if we try to explore dynamism of these different political agendas.

**Rob McNeil** 19:37

Thanks, Christina, what about yourself? I mean, what do you think the sort of responsibilities are of the academy in this sort of space?

**Christina Pao** 19:44

I think one of the best things for academics to do is say what we don't know. And so oftentimes, we're very good at writing up what we know what we recommend what the future steps for research are, but it is difficult to fill the holes and I like what you were saying Emre about, you know, having chapter editors, for instance, write about the ethics of the data sets. This is part of the ethical conversation we were having before; I think one of the best things that we can do is try to not keep people guessing not make people go too into the weeds of things be very upfront about the limitations that we have. Because every data set has its limitations is why we have so many different methods. It's why mixed

methods, I think, are the way to go. Nonetheless, even the best data sets don't have everything. The second part of that is what do we do when we know something? So, there are times in which I think researchers, myself included, want to continue gaining more data, we want a new data set, we want to launch an original survey, we want to do more field work. But maybe when we know something, spending a little bit more time ensuring that we make the findings accessible to a policy audience to a broader public. And that is, I think, really represented by, we had some papers in our conference that discuss specifically should we get more data? Or should we have better policy implementations with the data that we have? And so, these are kind of the questions that should be driving us, I think when we're trying to take on a new project. The last thing I'll mention is open-source code and data. So, one of the ways to make knowledge public is to ensure that when consent is there, when there are no privacy concerns, or ethical log jams, that we are able as scholars to not just hold on to our data, but also make it available for other migration researchers. And I think this is going to be part of the process of ensuring more equity also within academia, which also means that there's going to be more equity hopefully, in the ways that academia can affect policy.

**Rob McNeil** 21:44

Excellent. Okay. Well, the last question I've got, which is, again, for both of you, which is both the book, and the conference dealt a lot with the importance of data visualisation. So why does it matter so much? And what do you think constitutes data visualisation done right, in inverted commas? And, Christina, if I can ask you to start off with that?

**Christina Pao** 22:04

This was one of the most exciting parts of our conference, we were able to have our keynote speaker be Nikola Sander, invited by Oxford's own Will Allen, who's also a data vis expert. This was a huge element that we wanted to highlight. Because data vis like I said, I think in migration studies, and of all kinds of fields and substantive areas is one of the most progressive in this space. Thinking of how we visualise people and stories is a huge portion of making the academy accessible to the public. And I think some of the most thoughtful visualisations can really change the way that we think about solutions moving forward, they are representative and descriptive and potentially causal in their making. But I think they're very substantive and potentially impactful on a wide scale when they're done right. And that's just potentially what can be shown on an eight by eleven piece of paper. It's amazing to see what data visualisation done well can do in influencing minds, policy and what we know moving forward.

**Rob McNeil** 23:08

Excellent. Emre go on, what about you? What's data visualisation done right?

**Emre Korkmaz** 23:12

Yeah, so in our book we have some chapters dedicated to visualisation in migration management. And of course, this allows us to understand large data sets and complicated issues very easily. So, we have these large data sets, we should analyse them, and then at the end, we should demonstrate it in an understandable way. So, the visualisation is very important and many United Nations agencies and NGOs use them very actively. But maybe I could say we should be very careful not to feed some certain political agendas and not to oversimplify the process. We tend to receive these kinds of information like okay, 90 million people is doing that, 120 million people is doing another thing, kind of,

that might create some more problems instead of creating sympathy and empathy towards migrants. So, we should also discuss how we should visualise this data and what would be the political outcomes, or does it fit with our political targets? So, we cannot separate these from each other?

**Rob McNeil** 24:14

Absolutely. Okay. Well, look, guys, I'm so grateful to you for talking to us today.

**Emre Korkmaz** 24:19

Thank you so much.

**Christina Pao** 24:20

Thank you.

**Jacqui Broadhead** 24:21

You've been listening to The Migration Oxford Podcast. I'm Jacqui Broadhead.

**Rob McNeil** 24:25

And I'm Rob McNeil.