

NTDS_39Key:

I: Interviewer
R: Respondent

I: **The way we could start is maybe if you want to quickly introduce yourself and the background that led you to be in a position to be involved in MEDMI, and then we can move on to basically the story of MEDMI.**

R: Okay, fine. So I've had a long-standing interest in global environmental change and health, starting more than 20 years ago actually. When I first became aware of the problem of climate change and began to think how it might affect human health. Then subsequently I was involved in many other issues and also had a major administrative role here at the school as the director of the London School. I kept an interest in this topic. When I stepped down as Director, that was over five years ago, I decided to focus on this area for my research. At that time I also had a very part-time appointment with what was then the Health Protection Agency, now Public Health England, and so I was the chief of climate change for a time. Then I transferred a very part-time honorary position really with Public Health England to keep a fairly small advisory role around environmental change and health.

So I originally became aware of MEDMI, I was approached because I had the link with what is now Public Health England. Laura Fleming approached what was then the Health Protection Agency and so I was involved through my links with HPA, now PHE. Then I became a co-investigator on the MEDMI grant, because of my position.

I: **So basically you were involved since the start of it?**

R: Yes.

I: **Writing on the grant and all that prefaces?**

R: Correct.

I: **So with what expectations did you start?**

R: Well, I've long had an interest in the research on global environmental change in health, and a key part of that is how to link data on health and environmental issues. It seemed logical to start in a country like the UK because we have quite good data in this country in general. So this country could be an example for others as to how one might go about these matters. There's now growing interest, I would say, in linking health and environmental data, which I could talk about later, if you wish, but a growing international interest, and major international initiatives to do this. So in this respect MEDMI is a very important opportunity for experiential learning about what the challenges and opportunities are of doing this kind of work.

I: I've read already the grant application and several reports that have been then written also. So as part of the expectations or the goals for this platform there was the idea of developing methods, right?

R: Yes, developing methods, using the data obviously for research purposes, both on the direct effects of climate change, or climate variability and health. So things like changes in temperature, both cold and heat, in relation to direct effects like death, mortality rates, but also because of the links with Public Health England, the link between a range of infectious diseases and fluctuations in weather/climate. Obviously with an interest to the longer-term trends, but they're difficult to pick up over short periods, so one has to look at more time series, more fluctuations in temperature and other meteorological variables in relation to health, or to the transmission of diseases, depending on what you're interested in. So obviously in other countries, and to some extent in the UK, one's also interested in extreme events, not just heat waves but also floods and so on.

So there's growing interest on the impact of these more extreme events, including not just on infectious diseases but also things like mental health. For example we know that flooding populations are more likely to experience sometimes quite severe mental health problems that can last for quite long periods. So this is a very important aspect of planning adaptation to environmental change, and assessing the resilience of communities. So we also hope that in the longer run there might be an opportunity, a kind of tool, if you like, for assessing the effectiveness of adaptation strategies to promote resilience of communities and specific strategies that would help them to adapt to a changing climate.

I: How did the project then develop over time, if you recount the trajectory of the process and some of the events that shaped it?

R: Well, I think we were probably, as always happens in these complex projects, maybe a little bit over-optimistic about the speed at which this could move, because these are complex issues of obviously confidentiality and ownership of data. So it takes time to shift and to change major organisations [REDACTED], in which there are a range of different interests, I would say, in different groups that have different interests in data. Everything from the conventional environmental epidemiology through to infectious diseases, and indeed more specifically climate change, heat waves and those kind of very specific issues. So it's been quite a long learning curve, I would say, knowing who exactly to approach [REDACTED] and how to get buy-in or agreement to share data. Then, once one's got that, of course, then there's the actual mechanics of making those links.

The objective that we had of course was to essentially free up this data to make it available for the research community through the website, which again took a long time, perhaps predictably. Just because not just the technical aspects of writing the code and so on to get access to this data, but also the natural concerns about confidentiality and so on that have to be addressed. Then the question of, 'What is the purpose of the data?' Is it really going to be sufficient for new research or is it for training young researchers, or training people in the use of these kind of data? So one needs to clarify precisely what this data is going to be used for in the future I think in order to exploit it to its maximum capacity.

I: You mean this from an access negotiation perspective, or is it also design?

R: It's also design as well I think, because obviously if it's for training purposes then the data might be simplified, for example. Some of the data on the website, as you probably know, is quite simplified data. So it allows people, who may not be experienced statisticians, to analyse data using simplified statistical approaches. So again it depends very much on what the purpose of the website... well, making it open access to this data is. I think initially we probably were a bit naive in what we expected could be achieved in a relatively short time, and with relatively limited resources. Nevertheless I think it's been a very useful exercise overall; and we've been able to actually publish... through the theme as a whole there's been quite a lot of publications come out of it. The purpose was not largely to do new research but to make available data to the community.

So I would say in some respects it's perhaps been more productive than one might have predicted. In other respects it's taken a long time to get the data into a stage where it could be exploited. Then there's always the question of once the grant's finished who will maintain that. Ultimately it has to be, I think, government organisation or government-funded organisations like Public Health England, like the Met Office, and others that can really act as long-term custodians of that data.

I: If you don't mind to come back to this point a little bit, which is in some respects it's also been more productive than expected. So why and what are the conditions that make MEDMI possible for it to be..?

R: Well, I think particularly, and I wasn't involved directly so I can't take credit for it, but I think in particular the people involved in the infectious disease data were able to undertake more analytical work than we'd initially anticipated. So people like Gordon Nichols and so on at Public Health England with Majeed, did quite a lot of statistical analyses of data, which turned out to be quite productive, and has resulted in a number of publications. I think when we embarked on this, when we applied for the money, we didn't anticipate perhaps so much primary scientific data coming out of it. So I think it's partly the motivation and interest of the individuals concerned, and also the fact that there were some... once the data had been fully... because people like Gordon Nichols know the data very well, they were in a good position to ensure that it was fully exploited as much as it could be. So I think that's been very positive.

Some of the other data, for example we always had this idea of linking with primary care data, which is particularly interesting, firstly to me because I was formerly in primary care, but also because in the UK we have unusual opportunities to do that because virtually 100% of family doctor records are computerised. Of course not all of them are... the systems are not provided by one company, so you have to negotiate with different companies. We've been quite fortunate in finding one company, TPP, which is particularly open to research, and they have started... it's taken a long time to actually get the data, but they've now started releasing that data. So, with my colleague, Shakoor Hajat, who I guess you'll be interviewing probably?

I: On Wednesday.

- R: He's starting to look at that in terms of looking at the relationship between some common diseases, like for example diabetes, our consultations for diabetes, are they related in some way to environmental variables? Are people with a common disease like diabetes more vulnerable to environmental change? These kinds of things we'll be able to look at. At least the short-term relationships, maybe not the long-term.
- I: **These works then will carry on for..?**
- R: I don't know to what extent they'll continue after the end of MEDMI. Obviously we won't know unless we get additional funding. It's possible but you'll have to ask Shak, because I'm not sure he'll be able to continue working on this beyond the end. It may be that he would be able to finish some things off. He and I are going to be talking about that fairly soon.
- I: **Okay. So how important is the knowledge of the data? For example Gordon Nichols said (ph 0.13.15) *excessive knowledge in the data* and that sort of felt..?**
- R: Well, I think it was in the case of the infectious diseases, because there are so many infectious diseases one can look at. So first of all it's important to understand which ones are likely to be related to meteorological variables. He's done quite a lot of work in that area already. So he was in a very good position, I think, to exploit that data. I think some of the other data, I'm not sure quite... it may not be so important. Shakoor Hajat, who you're talking to, has had quite a lot of experience in the past of analysing primary care data, and using times series analysis on primary care data. We did some work together back in the 1990s, in fact, on that. So he has experience of doing those kind of analyses, and he's a very experienced statistician at running times series analyses.
- I: **Going back instead on the main learnings, so you've mentioned basically, if I'm correct, three main areas, so maybe if you want to open these up just a little bit, the three are knowing who to talk to and agreeing on access to datasets. The second one was the mechanisms of the linkage and the third was how to make available these data to the browser. Would you like to review again this, opening that up with also what you can recollect, and with possible examples?**
- R: In terms of knowing who to approach, I guess when we first did this work it was with Anthony Kessel who was the lead person at what was then Health Protection Agency. Anthony of course has got a major role now in global health. He's got a long-standing interest in environment. Of course there are many other stakeholders within what's now Public Health England. I suppose it isn't until quite recently that we've realised that Giovanni Leonardi's group... I don't know if you're meeting with him?
- I: **I haven't fixed a date yet. We've been in touch.**
- R: They're the environmental epidemiology group. It wasn't until quite recently that we suddenly realised that they were probably the best equipped to actually act as long-term custodians for the data. It turns out they're very interested in doing so. So that's been I think beneficial to make those links, even perhaps a little bit more slowly than we should have done. I think

because there are these different obviously groupings within PHE, as I mentioned already, I mentioned three of them, and they all have different interests. So knowing which are the best equipped and most likely to be able to continue this work was, for an outsider, even someone like myself who knows a bit about Public Health England, I still don't know the details. It's a big organisation. So it just took a long time to understand who were the appropriate people to work with. I think in terms of the long-term looking after, being a custodian of the data, then I think that Giovanni Leonardi's group will be very well-placed to do that.

I: So what are now the kinds of questions that you are thinking about in terms of the long-term custody of the data, how do you hand-off the work?

R: I suppose ensuring that the data is fully exploited. Let's take the example of the primary care data from TPP. So TPP is an independent company. So that's an interesting ex because that's not Public Health England data. It's external data. TPP want to keep control over that, understandably. They have set-up a small foundation, as I understand it, to enable the analysis of the data that they hold. So I think that they're doing this work kind of pro-bono through their small foundation, so it'll be very important for the future to ensure that Giovanni Leonardi, for example, has those contacts that those kind of links can be built on and further exploited. So I guess there's a difference between the data that's owned in a sense by PHE, like disease outbreaks, disease incidence. Changes in mortalities comes from the Office for National Statistics, but PHE I think have got some close link with ONS, so they can get that data.

They've got all the syndromic surveillance data, which allows them to look at different syndromes which might... if there's a disease outbreak and they don't know what it is, but they suddenly get an increase in people consulting family doctors with a certain group of symptoms, then they can look to see, 'Well, what's causing this?' It's often a viral disease that's causing this kind of outbreak. So all that data can be linked to the environmental data to see if whether it's some other environmental trigger that might cause that. So there's the syndromic surveillance data, there's the laboratory data, which is where they have a laboratory confirmation of a particular disease, so an organism has been grown, and we're doing quite a lot of work to see whether using the laboratory address is as good as using the locality of where the patient lives.

If you can use the laboratory then it gets round a lot of problems of confidentiality, because obviously for any one laboratory there'll be thousands of people in their locality being dealt with by one laboratory. If you've got to use a detailed measure of where the patient lives, then that's much more difficult to keep confidentiality. So we're hoping to get some results from that work in the next couple of months to see whether we can use the laboratory data as a kind of surrogate, if you like, for the patient, where they live. If that's so then that will make research much easier for the future, because then you can just use the data from the proximity of the laboratory, or any other environmental data. So that will be useful for the future. Then the other area that we haven't really explored but will be important for the future is the health impact of strategies to reduce environmental change, by which I mean what we call 'co-benefits of low carbon policies'.

So in the future what would be very useful to do would be to evaluate the benefits of policies to reduce environmental damage, like for example to reduce greenhouse gas emissions, which may themselves have health benefits. So for example if you reduce burning of coal, that reduces air pollution, that will also have health benefits, over and above the benefit for climate. If you use private cars less, that could have benefits for physical activity, as well as reducing greenhouse gas emissions. So this is something we haven't got into in MEDMI, but in the future this will be quite important areas for development, and something that could be exploited for the future, I would say.

I: What I wanted to also ask you was if you think about these kind of learning about the learning curve about accessing the data, where to access it, in the project you were following on the temperatures, what did you have to do in order to adapt to this... if you had to do anything at all, did it require you to shift a bit the focus or the process or the planning of your project?

R: Well, really on the practical side Shak was more the person that was involved in that. I think initially he has done quite a lot of work on deaths in relation to temperature, both heat and particularly heat but to some extent cold I think. The death data was available through the Office for National Statistics, and obviously Public Health England also monitors the death data. As far as I know, I don't think it took... I think there's a problem with accessing the data in the very near-term because obviously death data, it takes time to confirm death and confirm the cause of death and so on. So depending on what you want to do, if you want to look at it in real time, that's very difficult. Well, it's impossible really. I think you can get some approximate numbers but of course until you get the final confirmation, which might take quite a few months, then you're always quite a few months late in terms of analysing the data. These kind of problems are fairly well-known I think.

So as far as I know there were no new difficulties that were found, but it does mean that you're basically always having to look at the historical data for deaths. In terms of the primary care and hospital data, again of course you would be a little bit behind, but there I think the delay was really first of all we initially wondered whether we might be able to get national data through something called the GP Extraction Service, GPES, which is a national service that was going to extract GP data from all the different computer systems. Then we found out that that was extremely slow, because they seemed to themselves have experienced major difficulties, and I don't know the details of why. They've had real problems in providing that data; and I think that's been a common experience of many researchers.

That is why we approached this company, TPP, because we'd heard that they were less bureaucratic, and also because it was a single company, and you could move at the speed of that company, rather than waiting for the slowest company, waiting for the national register to work. I don't know whether GPES is working yet. It may be, but I don't think it is. I think it's still having troubles. So that's why we met with the TPP people, who were very helpful, and we're now focusing really on them. In the future, if the GPES system gets off the ground, then it would be possible to use national data for all the different companies.

I: How did you... just what comes to my mind in this moment, how did you discuss and think about what it means to talk to only one company, because they are approachable? I know that in your discipline you are concerned with issues of sampling.

R: Yes. You could argue it's not a random sample of the UK population, but it has a very wide geographic dispersion. To some extent, the choice that the family doctors made as to which computer company to go with were not very strongly related to... I'm not sure how much bias they really introduced into the selection, because their main reason for selecting that company wasn't because... for example poorer people were linked to practices which used this particular system. I don't think there was any bias in the selection. I think it must have been partly random chance, depending on which company was marketing most effectively in those areas, partly history, so I don't know the answer. They cover very large populations. TPP covers millions and millions of people. So it's not like it's just a few thousand. It's a large population.

I: In terms of learning about the mechanisms of the data linkage itself instead, was there also some new learning, or also not new learning, and what it meant for the project?

R: As I understand it, and I wasn't involved directly, but I think that in terms of for example the website, I think we hadn't quite realised that you need quite specialist coders, so people who can use Python and so on. Those skills are not very widely available in the academic sector. So they had to use someone who was, as you'll have probably heard, this individual was very good but he was also in demand commercially, so went off and got another position, and that caused a lot of delays in the work, because he then had to do that work almost as a freelancer in addition to his paid work. So I think we hadn't fully realised that these kind of schools are not that widely available and they're commercially very attractive, so people can get good salaries in the commercial sector, and that causes problems and delays that we hadn't fully, I think, anticipated. So it's a different kind of set of skills than you might get, say, in the average university academic or even statistician. These people are quite specialised.

I: Talking more generally about... you've mentioned in the beginning the opportunities of data linkage, research, do you want to say something more about that, what's the potential there? What are the challenges? So we're moving onto a more general level but if you've got examples from MEDMI that's great as well.

R: I think that at the moment we are still kind of exploring the potential of these changes, because we've tended to look at fairly relatively simple things like fluctuations in temperature or air pollution. What is a much bigger challenge is when you're dealing with multiple environmental changes, and also the scale of those changes, and particularly in a country like the UK, of course, we can be fairly well protected against some of the effects of environmental change, because for example we can afford to buy our food on the world market. So if there's a sudden collapse of crop production in country X, we can buy our food from somewhere else, we have enough money and we trade a lot of our food. Whereas if you're a poor subsistence farmer in Africa, then what you're growing is what you're going to eat, or what you watch your neighbours grow.

So in this country we're kind of insulated from some of the direct effects of environmental change, and also because we're not living at the edge of our environmental capacity, which some countries are. Then we are relatively buffered or protected. The countries that we really want to study, where we anticipate that the most serious effects will occur most quickly, are the ones where there's often the poorest data, so that's why we have to use a country like the UK because the data's pretty good here, even if the challenges are not as great. I do think that the lessons that we learn in the UK could be extrapolated or used for other countries in due course, including those at low middle income levels, because as the price falls of electronic information systems, and they're taken up more frequently, more widely, then we can see those opportunities might develop, even in quite low income settings. So the lessons that we learned here could be of much broader value, I would say.

Certainly looking at these complex interacting environmental changes is quite challenging scientifically, because we're dealing often with long causal pathways, and pathways that could be geographically quite, you know, outside the country. So at the moment we've started with relatively simple things, where there are direct linkages, but obviously, as these systems become more sophisticated and more widely used then one can start to look at more complex issues, like extreme events for example, or big disasters and so on, or changes in the distribution of vector borne and other diseases, which are a particular problem in low income countries. So the lessons that we learn here could be more widely used in future.

I: One thing that I forgot to ask you earlier when you mentioned these three points, so the first question is, is it correct to say that those are the main lessons that you suggest you take away from the experience?

R: I would say the lessons are partly about concerns about confidentiality, which we knew about, but they are quite dominant at the moment, and so we have to take those into account, and to some extent the possible trade-off between loss of confidentiality and loss of scientific precision, because we can't study individuals, we have to study populations. Or at least we can't through this particular work. You can study individuals in some contexts, where you've got, for example, Biobank, which is a big... this is something we did actually have a workshop on the use of Biobank. You've probably heard of Biobank.

I: Yes. I actually was sitting in there last November.

R: The workshop. Yes, that's right. So there are ways of using individual data where you've got volunteer populations like Biobank with half a million people. We are very interested in linking environmental data to Biobank, and some people are starting to do that, in fact. Hopefully our workshop will help to stimulate that discussion. If you're dealing with routine data then you can't do that, and you have to use anonymised population data. Obviously there will be some loss of precision as a result of that, and some things will be very difficult to study. So, for example, flooding is a very difficult thing to study because it's very localised, and how do you define it? Is it the fact that water enters your house, or is it where water damages your house, or is it where water forces you out of your house? Or is it where water enters your neighbour's house?

So defining what exposure to flooding is, is difficult, and then getting the data on a very localised level is extremely difficult, particularly when people may

have to move. So, for example, after a flood they may be re-housed somewhere, either temporarily or permanently. So how do you study those people? So we've certainly... this is not just MEDMI, but other studies too, have shown that it's quite difficult to study some of these very localised phenomena, partly because it's difficult to trace the people because they're being moved, secondly the issue of confidentiality, thirdly the issue of linking their personal data to their use of health services, particularly if they're moving around, they're moving from one temporary accommodation back to another one, so how do you get a full assessment of their use of the health service?

Well only really through their National Health Service number, which is unique, but then the problem is these are patients, they're not necessarily people in research studies. So how can you use... you know, it would be difficult to get permission to use their individual data without their consent, but it would be very difficult to get their consent. So these kind of issues.

I: I've got the impression that there are interactions between these issues. So like you were just deliberating.

R: Yes, there can be. Some things are not so much affected and others would be affected, depending on the problem and so on.

I: So norms and customs around confidentiality can make the mechanics of the data linkage more sort of difficult or complex, or just maybe more it needs to find an agreement to be talked about.

R: Yes, and in general I think it's going to be really difficult to get that kind of data, in the UK anyway, that very personalised data. Unless you could somehow recruit a cohort of people with their permission, but then you'd have to go in almost as they were being flooded, or perhaps study people that live in an area that's frequently flooded, but then the problem is they may not be typical either, because when you've been flooded several times you may adapt in some way. So it's a complex area, but it certainly raises all sorts of very interesting difficult questions.

I: If I can ask you the conclusive question, it's again recollecting the MEDMI, if you mentioned either preparing MEDMI or if you prefer if you were to go back to the start and change the way you organised and provided for resources, what kind of changes would you apply?

R: Probably we would have perhaps put more effort into really building links across Public Health England, or Health Protection Agency as it then was, from the beginning. Maybe we didn't fully appreciate the range of people that would be interested in this work. So, yes, I think we could have been more effective at mapping out all the people who could be interested and who would be in a position to support this kind of work, and perhaps to understand better the stakeholders within Public Health England that might have a particular interest in seeing it continue. Now I think we have those links, but it's taken some time to get them. So that's something could have been done maybe more. We could have been more effective I think at doing. Also in practical terms it was quite difficult because Public Health England was going through lots of changes, staff changed their roles and people were away for personal reasons, so it's easy to be wise now, but at the time...

I don't think we made any major areas, I think it was just that it took longer than one anticipated, as often happens. I think in the end we've finished quite strongly, although earlier on it seemed to be moving very slowly, but I think now we're finishing in quite a good place.

I: Great. Thanks a lot for your time.

(End of recording)