

NTDS_027

Key:

I: Interviewer

R: Respondent

I: Actually it is really useful. Okay. Work. Yeah. Okay. Great. So, really I would like to start from your involvement. You're in the MEDMI in advice capacities.

R: Okay. Well, I don't remember the full details of how this all got assembled, but what I do remember is very clearly that I've had ideas for quite a long time about the notion of putting together data on climate.

[Interviewer pours glass of water]

R: Thanks very much. Data on climate, the environment and human health and wellbeing, so I don't know if it was my idea or if we all came to this separately, but anyway, it's something that I've felt strongly about for a long time. However, I am not a modeller, nor am I a data scientist. That's not my view. But I was very keen to see this kind of project go forward. And then we heard, Laura and I heard, from the Medical Research Council that there may be some possibility of getting funding for this kind of area to create a platform, so if we could create the infrastructure for putting together climate, environment and health data as a platform, then other people in the future would be able to put more data into it and do analyses on it and it would be a resource.

I: Yeah.

R: So that was also a very appealing idea. I won't go through all my roles in the university, but suffice it to say that I don't see myself, at least, as being a research professor in the typical sense anymore because when I joined the medical school, I did so from being the former chief scientist of the Environment Agency of the UK, so I was very useful to... was very used to having a high level view on what was needed and on the diversity of information that's out there and the need to bring it together, and then how this information could fuel policy, guide policy. So my role in this was not to be an active modeller or anything like that, but I did agree to chair the science advisory board for the project and to recommend a bunch of, what I think are, high level individuals, and Laura had some ideas and probably other people chipped in, and we assembled the board of advisors, which I have chaired, and I can't remember how many meetings we've had, at least three, maybe more, and at which point, people have brought information with the progress of the project forward and I have chaired the discussions and added in my own comments on how we felt it was progressing, what could be done differently, what were the challenges to overcome, so that's kind of where it came from in the first place. I'm very aware that the spatial and temporal scales over which climate data, environment data and health data are taken are very different and so there's a technical problem of how you knit those things together, and I would suggest that none of us who were involved the initial phase of the project quite realised just how difficult that was going to be to do.

I: Right.

R: I kind of assumed that this was a common problem and that there are algorithms and ways of doing this all the time. It turned out to be actually considerably harder than we realised, and we did get a lot of good, high quality advice from the various, different members of the advisory board, I felt. And Laura was pleased with the meetings that we've had. So, anyway, yeah, that's the initial thing.

I: Yeah. Okay. So I guess that you are in a position to... if you want to maybe open up more what were the topics of discussion, the challenges that came up and emerged as most prominent.

R: Yeah, well, the most prominent was the one I just mentioned about how you put data from different spatial and temporal scales together. Probably another very big problem was access to data. Some of the data was available free and some of it, you had to pay huge amounts of money for, which we didn't have, so that was another challenge, and I don't think, again, we'd realised quite how difficult that was. And then, from a scientific point of view, there are some serious questions to ask. For example, if you saw that the climate had changed over a period of time and that that had resulted in changes of weather, you might see some change in the environment as well that followed on from that, and as the environment changed, you might also see that there are some change in human health.

I: Yes.

R: The problem is that those things might not be related at all.

I: Yes.

R: They may be... it may be thirty or forty years later before you saw an effect, and we were just coincidentally seeing these things. So this is a shotgun approach. You see all these relationships among things, but you don't know that they are causally related.

I: Right.

R: So that's a very, very big challenge. But you can also do another thing, which is test hypotheses. So, for example, if you have in your mind that as the climate changes, there will be more rain in south west Britain and so there will be more runoff into the sea as the rain washes down, it washes soil and nutrients into the sea, and the hypothesis might be that the algae in the sea, the plankton in the sea, will grow more rapidly, some of them will become toxic, toxins will be taken up in shellfish which filter these things, people eat the shellfish and they get sick.

I: Right.

R: So that might be the cause and chain hypothesis. Then, you can go and look at the data and see if you can that in it to confirm or... and that, in a way, is a much more robust and better way to do it, and so in the study we had, I think, three different examples of hypothesis testing, and I think that was quite useful. The problem, though, is that also restricts you because you already know what you're looking for whereas what you really want this thing to do is

to illuminate new problems that you hadn't realised, that a change in the weather was related to a change in the environment to a particular change in health. So that scientifically is a very challenging area.

I: Because the expectation was more like also to have a more inductive... you see correlations pop up and then you...

R: Yeah, and then you go away and investigate.

I: Come back and investigate.

R: Yeah. Yeah. So those were... the project emerged like that more and more as we went on.

I: Yes.

R: But we... I mean, the problem that I would say is that compared with what we heard at the first meeting of the advisory board when the project leaders for each of the different subsections described what was going on... what we really achieved was far less than that because of all these problems getting in the way.

I: Okay. So from the expectations that were said in the first meeting.

R: Yeah. Yeah.

I: So these kinds of challenges emerged gradually or...

R: Yeah. They emerged in bursts really, so there would be rapid progress and then hit a wall for a bit with each one of them, and then something else would happen that would slow you down, so probably the first one was the data. Could we get it. A lot of the data we thought was available and that perhaps it had been indicated that we could get hold of maybe it really wasn't available and it wasn't easy to get hold of.

I: Yeah.

R: And then we got hold of some of the data and then we found the spatial scales, the temporal scales problem that... for the Met Office, it might be 5 square kilometres and for the Environment Agency, it might be a couple of kilometres or something, square kilometres, and knitting those things together was technically difficult. And then, there were practical issues like personnel. Could we hire the right people to do this work for the money we had available? And just to focus on that aspect for a while, firstly, it was very difficult to find people who had the different skills, the several different skill sets embodied in one person, and we only had money for one person, if you see what I mean. And secondly, there were, then, issues with the particular individual involved who went off to another job and did things, and then it was hiring somebody else to replace him, so that's a disruptive effect on the whole process. Anyway, then, we got somebody back in again. And then, part way through the project, I came across a big data company called Black Swan and described some of the problems we were having, and they said that they'd be willing to try and help out with those problems, which they tried to do, but given the timescales available and the costs that could have accumulated, we didn't get too far on that. It wasn't a lack of willingness on either part. It was

just practical time and coming up towards the end of the project. But one of the things that they pointed out was that they would have three individuals with their particular skills sets working together rather than trying and embodying it in one person.

I: In one.

R: So I suppose perhaps we were a little bit naïve at the beginning thinking that this one person would be able to do all these clever things, whereas, in fact, you probably needed three individuals with different skill sets.

I: Right. So this collaboration with Black Swan was initiated or it never took off really.

R: It never... I mean, it tried to take off, but it never quite made it because the time... this was towards the end of the project, and there simply was not time for Black Swan to get everything in place. It was too late. But we did learn from it.

I: Yes.

R: And possibly that would be my conclusion about the whole project because although we perhaps didn't achieve the goals that we set out to achieve, what we did do was learn a huge amount on the journey about how to do this, how to do it well and what the kind of issues were, so I think that was very valuable. So I don't think it was a waste of money. I think it absolutely wasn't. It was how you do this and what are the issues that come up, and I think the reports that were produced from MEDMI will be very, very valuable to other people coming through.

I: Mmmhmm. So if you summarise it, what would be the lessons learned mainly for you? Does it come back for...

R: Yeah, I think... well, it's a bit of a subtle lesson in a way because I think on the one hand, it's an expectation from the funders that when you describe a project, that every piece in the jigsaw will work out just as you describe it when you make the proposal for the project and it will all go along. If we had described it saying that we envisage these stages in the project, but we'll have to see how it goes as we learn and develop because nobody has tried to do this in quite this way before, they wouldn't have funded us.

I: Yeah.

R: So I think there's some sort of contradiction or conflict between those two kinds of approaches to it. So, I think that's important. Probably we should have spent more time trying to find out in advance about these problems, but then again, that's difficult because who has the time to go and have all these meetings unless you're funded to do it, so there's another tension there, I think, that comes in. The... I suppose the other thing is that... I guess the funders in some ways are fairly relaxed about where we got to with it. We did find out a lot of things. We didn't perhaps achieve the original goals, but we did create some kind of platform and it's been cited, the MEDMI project, as an innovative pilot of how to do this. And we have learnt a lot from the pilot. The key thing now from my perspective, and I don't know if this will happen or not, is... I think there ought to be MEDMI 2, which goes now forward with another

set of funding over the next few years and says, 'Okay, we've learnt all of these things. We now know how to do it much better. We know what we would have done. Let's have another go and let's try and wrap it up to the next level -

I: Yeah.

R: - and let's get a few more experienced partners with solving some of these data issues involved.' For example, I suggested that we could get Black Swan involved in the next version or not Black Swan, but another big data company. It could be anybody, but let's get people who are used to handling big data. Let's get them involved a bit more and... yeah, I mean the other thing is that we could probably also learn how to generate questions in a better way, so from the data that we've got, we've already seen some sort of potential relationships between things. Maybe we could now move those things forward as well. So I think when you fund projects like MEDMI, you have to really be prepared to fund a second and a third go to really get the value for the investment you've made in them because we're just expecting perfect results over too short a time period when we've first set it off.

I: Yeah.

R: So a little bit more sophisticated approach to... this might take quite a while and it is research. We don't know what the outcomes are. So let's move forward in a more cautious but committed way, I'd say.

I: Yeah. Yeah. So if you were to envision how the MEDMI 2 would be structured and what you would need, what would you... so the big data company...

R: Yeah, I think that, and interestingly some of the members of the advisory board had enormous expertise, who, I think, we'd get some of those as actual collaborative partners in the next round, because once they'd really realised what we were trying to do, and this isn't true of all members of the advisory board, all were helpful, but some were more helpful than others, and the ones who were the most helpful, once they began to engage with it and heard the feedback from the researchers, they had some really good ideas and I think that they would help in going forward to make a more sophisticated application, to overcome some of the problems that we faced.

I: Right.

R: And probably the other thing would be to say to the holders of the data... because we often had to go to people who weren't strongly involved in the project, maybe try and get holders of the data more involved, and we've actually been doing that.

I: In order to overcome what kind of question?

R: Well, with the holders of the data, to simply make sure what data they've got, in what form and are they willing to put it into the project or are we going to spend six months negotiating whether there's a fee to go with it and what kind of data they have and when it is available and those kind of things. We were told early on with one of the... I'm not going to name the data source, but one of the data sources had multiple sets of data that we could use, but as it

turned out, those sets of data hadn't really been gathered together in a useful way that we could access and include in the project.

I: Yeah. So, what happened?

R: It was unusable. It might be for MEDMI 2. Once they got their databases all sorted out, and in a form that we can use and if they say that they're available to use and there's no cost with it and they want to share in the results from the analysis that come out from that data, if we can get that kind of agreement better in place on the clear understanding that they would actually do this -

I: Yes.

R: - then it would make the whole project more robust.

I: So in a way it's like thinking about getting things more integrated.

R: Yes, it is. Absolutely.

I: More interdependent.

R: Yeah. I think... there's this great grey area between proposing a project and getting everybody to get involved in it and think about it and do all of that, and then when you get funding, it changes everything. People then start to think, 'Hang on. How much funding are we going to get out of this?' And they get more committed to it because their name is on it and they're going to be involved in it. So I guess it's partly the psychology of the fact that when you're putting in these applications, the chances are... you know, if it's a one in ten chance that you'll get funded, people will put something into it, but they're not going to put their whole body and soul into it because it's not being funded. When it gets funded, that's when they turn around and say, 'Okay.'

I: Good start.

R: People say a lot of things in that early phase, I think, with good intention, good will, but when it comes to actually delivering it, maybe they're not quite at the point they thought they were to be able to deliver it.

I: Right. Yes.

R: [coughs] Excuse me.

I: And in terms of instead the advisory board expertise, is that... why do they have valuable expertise? Is that more because they were involved in other kinds of similar projects or...

R: Yeah, so we took a.... I took, because I think I chose most of the advisory board with Laura, together we chose them, but we had people with different skill sets to try and cover a broad range of areas, so we had some who were specific... had specific expertise in the analysis of data and modelling and putting it all together. Others we had who were much more influential at the science policy interface, so what we would do with MEDMI as a platform, how would you get it widely publicised and known. For example, we had Dame Julia Slingo, the chief scientist at the Met Office, who was very clear about we should... if we could get this kind of system working, how we would inform

policy with it and how we would make it more broadly available. We also had another chap who had worked a lot with the Natural Environment and Research Council who knows an awful lot about data analysis and he was very helpful in the early stages. You do get a little bit of attrition, should we say, from the advisory board. It's very difficult to get these top class eminent people altogether on the same date.

I: Yeah.

R: So whilst some attended all of the meetings, a few just attended one or two.

I: Right. And how did these various people with different kinds of expertise chime in in the process... also, the process itself was reckoning with these sorts of challenges at different levels at different times.

R: No, they were good. They all offered suggestions and they offered additional contacts and they did suggest ways round it and where to put the emphasis. A couple of the projects, the marine one and the one that was antibiotic resistance... actually, no, it was disease outbreaks, bacterial disease outbreaks and so on, we got some very good advice on those and good contacts, and the advice was, 'Play up on these ones, but if the other thing isn't working so much, leave that one for now.'

I: Yeah.

R: So there was some of that going on. That was useful.

I: Right. Okay.

R: Yeah.

I: I was wondering again about the main three kinds of problem that you mentioned. Would it be fair to say that there was also consequentiality between these? Like you mentioned the access of the data was the first roadblock and once you got the data, then you would start to realise that there was also that challenge of the spatial... or is this more sort of post-doc... is this something that I'm projecting on because that's just the way I...

R: No, it's not. No. I think we were learning in each of these different categories. I don't want to give you the impression that we weren't aware that there would be challenges, just simply that we weren't aware of the scale of the challenges. A lot of it was caused by... the data in some of the areas, say climate change or perhaps we could talk about climate and the environment data, was probably easier to deal with, that's my impression, than the health data which was a lot less accessible and gathered in different ways. And I don't think we quite appreciated how bad that was. We knew that would be difficult, but we didn't know how difficult, and particularly I think the health data, from my recollection anyway, was the most difficult to get robust data that we could use.

I: Yeah.

R: There was one other thing that I haven't talked about, and that's because I'm not sure if I entirely understood it, but I know it was a bit of issue was... and perhaps, in a way, it was kind of a core issue at some level and that is that there was a way of putting all the data together so that it could be accessed from people outside the project. Because when you're creating a platform of this data, you need to get people and there seemed to be a great deal of difficulty in finding a system whereby everybody in the outside world could get through to access this data and yet at the same time, keep the data intact, not messed up by anybody, and if it were anything to do with health data then it was confidential and what have you. That too posed quite a challenge.

I: **Right. And why? Because...**

R: Well, particularly in relation to confidentiality and things, I think, because one of the things you can do with confidential data, and perhaps it was part of the discussion with Black Swan, you can write pieces of software that can go into the data and can ask a question without actually revealing the data. So just as an example, it's perhaps not entirely relevant to this, but it's a way of illustrating it, supposing you could write a piece of software that went into a health database and asked the question, 'If you have had these two diseases, just randomly two diseases, before the age of 15, are you more likely when you're over 65 to get dementia?' And so it can dive in, it can go through all the data and it comes out and it gives you an equation that says, 'Yes, the relation is this.' So that's not made any of the health data available, there's no breach of confidentiality, but it has answered the questions.

I: **Yes.**

R: So I think that kind of thing we had to learn a lot about.

I: **So this was functionality that you were expecting to...**

R: Yeah, we thought...

I: **You had planned to deliver.**

R: Yeah. But then the databases were not in the form to allow us to do that. Some of the databases that we wanted to get access to. And also the technicalities of doing that is quite difficult, which is why had we had longer with Black Swan, they may have been able to help us more with that kind of approach as well.

I: **Right. Okay. So because you realised that...**

R: It was just too late.

I: **And they would have provided...**

R: Yeah. They would have... we could have worked with them to learn how to do that.

I: **Yeah. Okay, and how did you leave this?**

R: Well, I think to be... this is just my impression because I haven't heard anything to contrary, we may be having one more Black Swan, sorry, one

more MEDMI meeting as a kind of summary but I'm not entirely sure whether that's going to happen or not -

I: Right.

R: - and I think that Laura has been having some sort of general discussions about MEDMI 2, but again, I'm not sure that that's actually going to happen, so as I say, I'm chairman of the advisory board, so my contact points have been informally with Laura, just chatting, and at the advisory board meetings. So I've not been involved in all the routine meetings of the team, from London School, Public Health England and other places, London School, did I mention London... yeah, and so on. So I haven't been involved in all those meetings, so I don't know. I'm kind of awaiting whether I'll be called again to chair anymore of them.

I: Yeah. Yeah. I would like to understand more again about this difficulty in preparing a system that would access... provide access and confidentiality with data intact. What kind of problem with data intact?

R: Well, it's a very technical issue, is this, so I'm not sure if I can tell you more. What I'm trying to do is conceptualise it for you, but if you were talking to Laura and, in particular, the guy who is tasked with doing this, then he'd be able to explain to you about what the difficulties are with doing that, but a lot of this is about who owns the data, how publically available is it, could anybody mess with the data, if they can get access to the data, can they mess with it, so you almost need a copy that cannot be fiddled with. You know, nobody can change any of the data in it, but a copy that people can take away with them and do what they want with the data. They can discard some of it if there's too much or they can amalgamate it, they can fiddle with it, but you don't want that to happen to the original set, and doing that over an interface with the whole world, as it were, is quite a difficult thing, so they were coming up with discussions on how to do that better. But I think you have to ask Laura to tell you exactly where we've got to do with.

I: Yeah, I'm trying to survey most of the people who were involved anyway.

R: Sure.

I: No, I was asking this because it seemed that it could also be related to the other kind of issue that you mentioned which was the idea that... the constraints that you found that are involved in the researching through this kind of infrastructure which is, for example, parsing out from the correlations, possible causations...

R: I mean that's a fundamental problem of it that we just have to kind of live with in a way.

I: Yes.

R: And I think you just have to be clear what you're doing, so hypothesis testing is very valuable, but you have to know what the hypothesis is, whereas the other approach where you see these correlations, you have to realise that this is a shotgun approach really. It's a fairly random approach but some of these

things that stand out might actually be important and then you'd want to design studies that would test them specifically.

I: Yeah.

R: So part of the project is about illuminating new ideas, new correlations that you hadn't thought of and another part is testing hypotheses, and if you keep that clear, that's useful.

I: Yeah. Yeah. Yeah. I understood that the way that you explained to me the change in the weather, the changes in the environment and then the algae, so it's like you hypothesise a causal chain and then you try to cut it at the middle or at the end points and see if you find...

R: Yeah, that was just an illustrative example, but there are things that you can with, for example, air pollution and those kinds of things to see... and we did have some air pollution case studies, I think, with the London School and they tell you all about what happens with respiratory issues when air pollution is bad, and do climate events and particularly weather events, climate related weather events, do they exacerbate the air pollution and so on. So I think that that works quite well.

I: Because why I find this interesting is because I think it's quite challenging obviously to prepare all this in terms of the complexity because I imagine you have very complex datasets and then also you need to agree on a set of methods, progression, and then there's infinite varieties of it and stuff, you have to then code them and prepare them and their functionality. Was that also coming up as a problem to decide...

R: Yeah, a lot of that, I suppose, was kind of core work of what people were doing and... but, yeah, I mean it's challenging but it's kind of doable. Yeah, I suppose the other thing is that this illuminates is that you can only go so far with this data approach before eventually you probably have to do experimental approaches that look at mechanism, so whilst you might see that if there's a climate and weather event that exacerbates air pollution and you see that respiratory problems get worse, I mean, that is quite a strong causal chain and if you see it happening repeatedly every time there's one of these weather events, that air pollution gets worse and their respiratory cycle gets worse, then that's fairly compelling. But ultimately, you want to know, 'So what is it?' Is it that the particles in the air cause some sort of cellular damage in the lungs that produce disease? Do you see what I mean? The really detailed mechanism. So looking at data, this kind of data, will never tell you that mechanism, but it will take you part of the way to finding what you should be looking at.

I: Yes. Yes. So this kind of realisation was that something that came up more towards the end as you were trying to just what we are doing now. You know, find out what are the main takeaway points.

R: Yeah, I think that there is a process where maybe some of the individuals in the programme who appreciated some of these things at the beginning, but by the time you assemble it, think, 'So what the hell did we do here?' and 'How did it all work out?' Then you create a much more... it is a bit of a retrospective view but you get a more global view about exactly what's

happened, what the weaknesses were, what the strengths were, what we achieved and what you'd want to do next, so I do think that comes towards the end of the process.

I: Yeah, so in terms of... one thing I think I haven't asked you, in terms of MEDMI 2 or any future imaginary... what kind of skills and how would you compose a team in the way... less more in reference to specific, like we said Black Swan, the computer company, but how do you close the gaps? What are the gaps that...

R: Yeah. Well, again I don't know if I'm the right person to ask that because I haven't been so intimately involved in the day to day doing of the research, but certainly, I think that you'd want to get advice from people who handle big data all the time. The discussions that have taken place between the big data people in the project and some of the expert advisors and so on, there's been a lot of good dialogue between, them so I think that they will be clearer about what's needed.

I: Yeah.

R: And that's not a very good answer, but I'm afraid that I'm not sure that I can give you a more detailed, technical answer, but.... yeah, there are some people who are really clever at writing ways of putting these data events on different spatial and temporal skills together and you'd want to find out who the best people are for doing that and then, there would be other people who would be writing these bits of software to interrogate relationships and what have you. So there are different people. They are not one person.

I: Yes. Yes. So in that place, it's connected also, I guess, to size of the funding or how sizeable the funding needs to be.

R: Yeah, absolutely. So the funding that we asked for was quite a lot and... but it probably wasn't enough to take on this challenge, so that's another very important point, I think, that if people really want to make these things work, then you have to put a lot of resource behind them, but I understand... well, in fact, I can tell you that NERC, Natural Environment and Research Council, have now got a new call out and in that call they actually cite MEDMI as being an example of the kind of thing that they want to do more of. So they seem pretty happy that we've got so far down the line because as I say, they actually say that MEDMI ... maybe this is NERC and MRC, Medical Research Council. Medical Research Council were the primary funders of this, and NERC topped it up as far as I recall, but NERC are now very keen to do more of this, and Laura would be able to supply you with the email that actually... had a snippet from the... had a piece from the NERC call. I know she had that. Somebody had sent it to her -

I: Okay.

R: - because it particularly mentioned MEDMI, so she's be able to let you have that. That might be useful to you.

I: Yeah. Great. I think we have asked all the issues that I had -

R: Good.

I: - into the programme and I had... without needing to ask you, I had foreseen that we could talk about these things that I wrote: why only hypothesis generating or is it easy to overinterpret and stuff.

R: Well, I realise that my comments have not been as detailed as perhaps would be more useful, but as the chairman of the advisory board, I haven't been involved in the day to day thing. I've been involved more at the conceptual level.

I: Yeah, but birds eye views are also very helpful because of course, I'm going to visit other people involved.

R: Yeah, you will get the details.

I: Appropriate people, I will try to get into more details.

R: Excellent.

I: Great. Thank you.

R: Good.

I: Thank you very much.

R: You're very welcome.

(End of recording)