NTDS 034

Key:

I: Interviewer
R: Respondent

[Irrelevant conversation]

I: Okay. I think obviously we've talked before and so it's not necessary to ask you to present yourself and the kinds of introductions that I ask normally to people that I meet for the first time [overspeaking] for the first time about MEDMI. I thought more like… and also because we keep talking about this anyway. I thought that I wanted to sort of go at the heart of some of the things that I've been looking so far -

R: Okay.

I: - and ask your view on these things and then you will also help me to bring these also to obviously examples or to the storyline and stuff like that.

So it seems like, obviously, as you were saying, the project started with a high ambition also because I think you had multiple goals, established the infrastructure that can become a legacy for the future and then also do the demo projects and the pilot projects of research, so buff the sort of infrastructure and then the research as a demonstration and things.

R: There is also a political aspect too, of the university was in the midst of developing their… or expanding their relationship with the Met Office, so that was really important to them, and the medical school was coming into the university, so it was bringing kind of another strand of research collaboration between the Met Office and the university around environment and health.

I: Yes.

R: I mean, I was literally called into Janice Kay's office and told I will do this particular type of grant application with the Met Office. It grew from there. But it was an interesting experience.

I: Okay. So it was sort of a top-down…

R: A-huh. And I was relatively new. I think I'd only been there five months or something.

I: Okay. But I think it probably made a lot of sense to you anyway because you had been doing these kinds of…

R: Well not… I mean, I had to run large projects which were very interdisciplinary, but I'd never worked with the Met Office before. I'd work with NOAH in the UK but not on this level, and I had never thought of myself in that context of big data necessarily. I mean, I've also run a cancer registry. But again, that was sort of a different way of looking at the world. I didn't frame the world by big data, and now I would say that one of the things I do is

around big data. I mean, part of it is just the whole language has happened since that time, right. Data mashup, the first time I ever heard of it was… the expression was from Brian and it was when we wee trying to come up with an acronym name for this project [laughs] and he said, "It's a data mashup," and I was like, "Okay," and I looked it up on Wikipedia.

I: Yes, that's a nice choice of the [overspeaking] definitely. Okay, yes.

R: And then the other part that happened was we had the environmental stuff because we wanted the environment and health and then the question came, "Well, where are we going to get some health data?" without having to go to particular health groups like SAIL or Biobank or something like that but a more global group, and Public Health England had access to a lot of health data and used it and the Met Office was already interacting with Public Health England, so they brought Public Health England to the table and Public Health England brought London School, Andy Haines, to the table because there's a very close relationship between them and the individuals. And at that point it was like we've got enough big players in place and it's a partnership grant, there's not that much money on the table or resources and it's a really short period of time, so that's enough, let's move forward.

I: Okay. I see. Yes. So it was sort of in this sense much more complex than previous projects that you had…

R: It was on a different level because there was this feeling for me at least… and remember, I was new to the UK, I was a new professor in the medical school. In my old medical school I was not at that level. So I was at a hierarchy level, starting a whole new centre and I was basically told by… I guess she was like one of the Vice Deans of the (unclear 0:05:34.9) "Now, you will put this grant in and this is what it's going to be." So that was an interesting experience. And it was also dealing with the MRC, a funding agency. I mean, lots of experience with NIH, but no experience with MRC. A whole new culture. The Dean of the medical school was very helpful, he was involved a lot with MRC at that time and Stephen Holgate who is a professor at Southampton has been incredibly… has been a real mentor too, he also was involved in the MRC and he was like… There's an interest at the MRC now around environment health and you should push that. So it was learning all new culture. And again, Andy Haines had just stepped down from being Head of London School. Anthony was global… it was actually HPA at that time, Health Protection Agency, he was head of that and Brian was third in line at the Met Office. It was a high-level grant. And then the question was what were we going to do and how are we going to afford it given the really small resources on the table?

I: Right. So this is interesting because, yes, I haven't… I mean, I've talked about sort of the origins of the project with Brian, I remember. He was also telling me basically the same kind of story from the other side, he was telling me… "Well, from the Met Office, basically I was called in to [overspeaking]."

R: [Laughs] Yes. Actually it probably came from Julia Slingo. We approached Julia first, that's right, because she and Mike know each other and also she's the main contact for Exeter, yes. You are right, she must have gone to Brian – "You are going to do this." [Laughs] He was very good about it. I mean, they all were basically. So there was this whole political thing, diplomatic political

aspect to it, which I have to say I was new to. And the other thing is we put in a proposal and got rejected and got feedback and then we had to go back in, and that was also really useful. I have to say that those guys were very patient with me and I got a lot of support both from the other three… the co-PIs and Mike, but also in the medical school I got a lot of support, where they would like read endless… you end up having this proposal that you kind of whittle down to make it… it's like a… I think of it as polishing or putting facets on a diamond and it becomes a standalone house of cards. That was a process of probably a year.

I: Yes. So basically… and this actually sort of I think suggests a question which is then how did you… you had all these people in the room, right, as you said, enough big players in the room, let's start thinking about what we are going to do. So how did the ideas take shape?

R: Yes. That's an interesting part. I also think that what we originally thought, we realised after it was funded we probably couldn't do. There was this push around creating a linked database where we would have it available, linked. It was going to be this thing where we were going to do the linking, so you would have human health data and environmental data and it would be linked and then it would be available to the researchers, but the fallacy of that is if it's linked, you can't necessarily look at lags or manipulate the data if it's all linked. So, for example, if you have a flood, people can be killed right away or they can be harmed from that exposure. For example, their mental health, it doesn't actually give evidence of that for months. So there's lags from exposure to health onsets. If you link the data ahead of time you've almost set the conditions for what that lag is, you've defined that lag. If you want people to really use the data, explore the data, you can't do that. You see what I mean?

I: Yes.

R: Or - and this is something actually that Christophe is doing a lot of - you actually have to give them sort of pre little bits of different types of lags because you are ultimately coming down the hypothesis that some environmental exposure is causing subsequently a health outcome and that infers that there's a time period between the hit and the outcome. So there was that aspect. The other thing is… and we came up with a list of, to the original proposal, of sort of the other thing we had to imagine, was how would the health community - because this was MRC at that point - how would the health community envision using this resource? And I think we came up with like six different things. You could explore hypotheses, you could use it for a valuation, you could use it for planning, and so that took a lot of sort of talking and writing among each other. And as a side note, NERC came on board because Dan Bloomfield was helping us put it together, he does research knowledge transfer stuff, and he knew the person at NERC and sent her the proposal and said 'this might interest you', so they came back and said 'if you get that funded we would be interested in contributing money to it'. That's a total side note.

I: Yes.

R: So anyway, a lot of it was not necessarily… and I'm trying to think. We had a lot of phone calls. I don't think we had very many face-to-face meetings before it was funded. In fact, if any. We had a lot of group phone calls and we

had a lot of written product and this is what… I've done a lot of grants this way, a written product that you circulate and you set deadlines and you change it and you hone it and you diamond facet it, and in the process you kind of work out the ideas. There was one big part where I remember Clive Sabel, who really hasn't participated that much but was key in helping us initially, he and Brian kind of sat down and said, "Okay, we have to get concrete, what do we mean by how we are going to link it?" and started to talk about things like the postcode on the human side and how that could be translated to a lap-long point in time and space for the environmental data, you know, how are we concretely…? So we actually had… because you only have like ten pages to write this whole thing up. How are we concretely going to do that?

And the other part that came up was we decided that we needed some demonstration projects, that even then we realised that in 33 months [laughs], with the kind of resources we were talking, we were not going to conquer all of environment and health big data, so we decided to set up three projects, and the projects evolved a lot from work that was going on among the three groups, especially the first two. So Shakoor Hajat was already working in the area of project one, which was temperature and mortality, and project two, which was infectious diseases and climate, at that point HPA, now Public Health England, Anthony and his group were working on that. And the last one on harmful algal blooms, I was doing work on that in the US, although we definitely called that a blue sky project because the oceanographic modelling, and the Met Office helped with that one.

So we set those up as working off of leveraging already existing research. The whole idea of the browser and stuff really wasn't on the table until Ceri came on board much later after we got funded. I think I was thinking of the projects at that point as kind of more traditional epidemiologic projects, research projects, where you'd end up with a paper, where you could show that the… and you'd say I used the MEDMI data and here's what I was able to do and I found this new or interesting aspect where I confirmed something or I extended something. It was much later once we got funded and we got Ceri's unique skill set on board and he was like, "Well, you know, we could try doing this," and he started to work with Shakoor and they came up with the idea of can we make this kind of a web based almost like analysis function? And when that happened then we thought, well for project two can we do something different? So Shakoor and Ceri worked on it, and Shakoor does this particular type of time-series analysis and they tried to make it accessible and menu-driven and there was some visualisation so you could actually look at how the analyses turn out and explain what it is. So it's the kind of thing you could actually use with students, for example. So that became project one.

And then project two evolved into… the only health data that we've been able to really get was this SGSS data, which is all the infectious disease cases reported like 2,000 different types of infectious diseases that were reported to Public Health England and so Gordon, who came on later - thank God he came on - he was like, "Well why don't we do a screening where we link some environmental atmospheric whatever you want to call them, weather data, with these infectious diseases and use it as a quick screen to see whether or not there are diseases that clearly have some sort of climate thing going on just using regression analyses?" And then we thought, well how can we do something different from what Ceri is doing in project one? Why don't we

have a visualisation thing where people can interact with the data not just in terms of analysis but they can interact with it directly? So that's what Neil has put together based on the work that Gordon and Mark Cherrie did. And actually Mark did his own little thing also as well, another part of it.

And the project three became we really need money to do this. Brian brought Rose of RCL on board and she had been doing the oceanographic modelling and she's very open to new collaborations. The problem is she has no time. And so I went off to try and get… it took me four tries to get the NERC funding. I tried PHE, I tried NERC, I tried PHE and finally we got it at NERC, so that PhD actually officially started this month and it's obviously going to go beyond the life of the project but at least we can talk about the fact that oceanographic modelling is going to explore the issue of where there are harmful algal blooms on human health bringing in other data, the HES, the Hospital Episode Statistics data. In the long run we hope to look at whether or not we can see a connection in the UK between harmful algal blooms and human health. Which hasn't been done here. Everyone keeps telling me it can't be done. We did it in Florida. I think we can do it here. So that project has sort of been kicked down the road.

I: Yes, but now it's getting really started.

R: Yes. And it will go beyond the life of the project. So I'll report what's been happening, but it clearly won't be done yet.

And then we realised… I mean, I had another person who was Harriet before, Nicky Cocksedge, she was very good, but she left to take another job and we got Harriet. It's been just fabulous. I mean, she's amazing. She really embraced this project and she really got the website going and we realised – and I'm still not clear why – but we realised that we had extra money, and I think part of it was because Ceri left and then we were hiring a consultant and blah-blah-blah, so we decided, well, in some of my other centres that I've run, we've had pilot monies, and especially for early career researchers or people who are already involved in a group, we've given them a little extra money to do something interesting on the side, can be incredibly productive. So we had a very short call and we've reviewed them officially and that kind of thing, and I think the pilot projects have been incredibly productive because they show that… well for one thing it's broadened some different types of researchers and they've given us feedback on what it's like to use this data and this facility but they've also illustrated very different types of ways of using it, so it ranges from Jess's connecting it in with her Ménière's disease to the SAIL guys who gave us a lot of feedback like, you know, this is not an easy access thing. They've actually driven us to say that the last piece of work that Ceri or supposedly someone else will do is to make the interface between Christophe who kind of holds most of the data on the server and stuff, make that easier for researchers to use. So these are people who have been approved to use the data but it will give them like a… presumably, some sort of dropdown menu so you can define 'I want these dates, these variables and possibly – I don't know if this is possible – linked or not'.

I: Yes.

R: And that's useful because Giovanni and those guys at PHE, that helps them with the possibility that we might be able to intercalate MEDMI into their environmental tracking in the long run. We are talking to PHE right now

about… they have a whole environmental tracking programme that they've been doing for a while and they've actually spent a lot of time just talking to their stakeholders. This is tracking… identifying conditions where you have environmental and health data that you have on a regular basis and you can actually watch things happen, the way people watch influenza rates, there's the potential to get on top of them early, to give early warnings, that kind of thing. So the same thing for environmental and health conditions. So they are setting up a system at Public Health England about that. They seem very interested in the MEDMI. You'll talk to Giovanni, and there is a possibility that MEDMI maybe absorbed by them, if the Met Office is okay with that, which would be great. It's a form of sustainability for me.

I: Yes, yes, absolutely. Also because there's a big incognito what to do with MEDMI after.

R: Correct. Correct. So that's why I'm really interested in… we are trying to do an MRC programme grant, but I feel like I've had to push that along. There's now a new call that just came out from MRC NERC and it actually mentions MEDMI in the call and it mentions the work we did with Biobank and it's around environment and health data. It's small, it's like 16,000 per project, but the implication is there will be more work in this in the future. So I'm going to try and apply for some funding for that. Possibly one of the ones I want to apply for is to say… because it supposed to be for feasibility projects and not pilot projects, use that possibly to carry forward this idea that Public Health England might take on the environmental tracking and maybe use a disease that they are interested in doing as sort of the pilot project of that.

I: Right, yes. Actually I was wondering about the pilot projects. I know that they have provided reports or something [overspeaking]

R: And the understanding is that most of them will also publish. They finished in December. So one of the things we are doing is pushing on them to actually do peer review publications. Some of them it's not going to work. There's one that Christophe did with Rachel McInnes and I don't think it will be… It was on pollen data, and I think the one that Sujit is doing, it's on air pollution, may or may not be (unclear 0:22:29.5). We are really kind of helping them to make the data better, more accessible, richer, whatever. But Jess's has already been submitted as a publication. The one that Mark Cherrie did with Gordon around the infectious disease regression modelling, they've written it up and they are finishing that off. The one that… Mark finally got hold of the Biobank data to be able to do the solar irradiance in osteoporosis with Nick Osborne.

I: Yes, I've seen that.

R: And there's another one. Oh, the SAIL guys. I don't think that will turn into a project, but they are planning to apply for the NERC MRC funding and they are interested now. I mean, the nice thing is that's sort of made them interested in using the Met Office data, so they are looking at childhood obesity and green space and I was like, "Okay, can you layer in the Met Office data? Put in temperature, do something," and they are now interested in pursuing that, which is great.

I: Yes, absolutely.

R: And I think eventually they'll get a paper out of that.

I: Okay.

R: So it will end up probably three to four extra papers that we weren't going to have before. And you guys. You are one of our pilots. [Laughs].

I: I was wondering, do you think it could be useful for us to have a look at the reports if it is not… I don't know the content of the reports, so if it's not confidential or…

R: No, no, no, they are very brief, they are like one page. No, you'd be welcome to. I'll ask Ceri. Okay?

I: Yes.

R: And actually they might be up on the website.

I: Harriet, she provided me with some pages from the SAIL because she was saying that was very useful feedback.

R: They gave us feedback.

I: She said that was very useful feedback, so I think you could have a look at that.

R: Good. And they are the ones, like I said, who were driving us to say the last piece of work that Ceri or someone else is going to do is to make that interface. Because right now you either have to know Python, which most epidemiology types don't know or public health types, or you have to have somebody who does it. If we had this thing… And also the other thing is it's a lot of work for Christophe. He has to interact individually with every single group.

I: Yes.

R: And if they can get the kind of data they want out of it… And the other thing, it illustrates it. I mean, other places do this. The Cancer Registry I used to work on had something like this, the CDC, Centre for Disease Control in the US has something like this where you can say I want to look at mortality. It's a very high level, but you can download a dataset to look at and play with.

I: Yes. And this would be a piece of work that would be done at the end?

R: Well probably we'd like to start it now. Remember, we end in November. Ceri can't do it till June. So I'm talking also to people at PHE to see if they could do it before then. I don't know. I'm not sure. We've just started it.

I: Yes. Because I had (ph: 0:25:47.7) this sensation but I'm not entirely sure we are talking about the same thing, this interface between the work of Ceri and the work of Christophe.

R: So one…

I: It is kind of challenging.

R: Yes. Well there's two others at a whole nother level and it has to do I think with the hugeness of the data, that we have a server that's dedicated to this, but ultimately - and this is where you get into an area where I'm not that clear - it's about processing the data and the speed of it. So right now the two browsers, the visualisation one from Neil on project two and the statistical one that Ceri and Shakoor put together actually do not do their analysis in real time off of the database and that's been a real… some people are upset by that, some people aren't. At one level it does provide a certain security, right, because it's off of preformed data, so people can't get back to confidential data, and particularly on the browser, Ceri's browser that's going to be on the website outfacing, I think that's important, we don't want people to be able to get back to the original data. So this means that they just can't because it's preformed. But in terms of a concept, in other words demonstrating feasibility, it would be important… it would be lovely if we had more time and money. And this was one of the things on the table for the last couple of months – what are we going to do in the last couple of months is to be able to actually make, in real time, for example, the browser working off of the real data on the server. And according to Ceri – I can't say this for the visualisation, although Neil also said the same thing at the last meeting we had about this – we could spend a lot of time doing this and we might end up with an incredibly slow browser, either for the statistical analyses or for the visualisation because there's a lot of data that gets cranked out by these things. Visualisation requires a lot of the data fast, and ditto for the statistics. So the alternative was to focus on, okay, we've had feedback from SAIL and also talking to Giovanni with the tracking group, that there is a group of researchers out there who want to interact directly with the data. They could go to Christophe, but that's very labour intensive, and they are not stupid people, they just don't know Python. Could we create an interface with some kind of system in it, I'm assuming some sort of dropdown menus or something where you could define the kind of data you want to use for MEDMI and get it for yourself, for you to use? And that's what we are thinking of putting our last little effort in is to do that.

I: So it wouldn't be browser-based?

R: It could be browser-based. It depends on who does it. If Ceri does it it will be browser-based. All the work that Christophe has been doing is allegedly - and again you are getting into an area that I don't understand [laughs] - he's been setting up the database with Python so that it is much more easily callable and for this particular interface apparently, because it won't be asking the data to do much, it will really be about calling chunks of the data and it would be, on some level, predetermined what you could call, right? It's not analysing things in real time, it's just calling the data, so that then it could download and you could use that data to do whatever you want. And I'm imagining you would be able to do a download of unlinked data and possibly linked according to specification. The second one, I'm not so sure. So, do you see what I mean?

I: You download the unlinked…

R: …data, so that I could link it myself. So it does assume that the person knows what they are doing. Remember, this is a group that had to get permission to do it. These are people who are in research institutions and do this kind of work. But it means you don't have to know Python, it means that you have some idea of what you are doing but you are not at the level that the SAIL

guys were complaining about. So that seemed to appeal to both possibly the SAIL guys but also definitely the Public Health England guys.

I: Right, yes, yes, of course.

R: So these things come up. And by the way, Ceri says - and I can't remember at this point - he says this was something he thought was important from the very beginning. He actually wanted to do that when we started. I can't even remember anymore how we then moved over into doing the browsers. And it means different things…

I: Okay. So it somehow drifted…

R: Yes. Or drifted or… I don't know what happened.

I: Right.

R: I can't remember how these things evolved at this point. And the other thing that's been clear to me is, or maybe not so clear, no matter how much we talk about it, we are so interdisciplinary relatively that often we would be having these meetings where different people would be talking completely differently about the same thing or their understanding of it was completely different and it would take months to sort of figure that out. I mean, the whole browser concept was completely new to me, but for other people I think they had very set ideas of what it was going to be and it may have turned out to be something not what they thought, and there was this whole aspect of trying to do it on the web and this whole aspect is, as I've just described it, trying to do it in real time off of the database. Those were all things that we were trying to do. And I think we've found some challenges with it.

I: Yes. It's interesting because it sounds like this is like a twist to… I think it's well known when you have interdisciplinary collaborations, like remember also you told us last time also of your work back in the US, you got the environmental people, the health people, they talk different languages and so there's lots of coordination work that needs to happen, but it seems like here you are also talking about…

R: It's theoretical not just language.

I: Right. Yes. But I think also in respect of that difference of bringing these two sciences to find a point of convergence - here you were also mentioning that these two sciences need to talk about a third entire aspect which is the computing or [overspeaking] web browser -

R: Yes.

I: - which is not at least thought of or considered core in the health or core in the environment. Nobody really owns that space.

R: Yes, yes. And that was another promise that also identified a whole lack of expertise.

Going back to the thing which I think we've talked about before, that somebody like Ceri is very unusual. We ended up interacting, for example, thanks to Mike Depleasure with Black Swan, and they confirmed for us that

there are very few people that can kind of go across the outward facing web, the software development and the analysis. In fact, probably almost no one, because even Ceri will say he doesn't do the analysis. He had to collaborate with Shakoor to do that. But he can do the other two bits and he can talk across that. You look at Black Swan. They have three separate groups at least doing that and then a few people that can kind of talk the language above it. So no, you are right, that whole piece… I mean, I had no idea we were getting in there when we went that way. I mean, I was very much approaching this from, I guess, both as a user, an epidemiologist that would try and use this data, but imagining that person, so maybe like the SAIL group is how I was kind of approaching it. But also… Did you want to meet the head of the medical school?

I: No, I wasn't planning…

R: Can you turn it off?

I: Yes.

Part 2

I: Okay. So we were saying, yes, obviously the lack of expertise kind of problem.

R: Yes. And also we had a lot of high-level people, but not very many people who were actually going to do the work.

I: Yes. So that made the project depend also on some…

R: So there was Shakoor, who really is pretty high level. Majeed came on eventually. It took a long time to hire him because HPA became PHE and there was a hiring freeze and it was a nightmare. And then Ceri. Really those were the only people that had any substantial portion of their time, and then later Harriet, dedicated to the project and everybody else was sort of there to make sure that things happened and there was ideas and that it was moving forward, but those were very, very busy people.

I: Yes, yes, of course. Yes. This has been something that comes across often in the interviews, if you want, the project was too dependent on a few individuals and then, for example, what happened with Ceri's own choices and that obviously had an impact on the progress of some aspects.

R: There was like a six-month period where almost nothing was moving forward, it was really frustrating, but I have learned a lot of patience [laughs]. But the other thing that happened, which was very good, was (unclear 0:01:44.7) Advisory Board, and I think Julia Slingo came to it and Laura Burke, who is the head of EPA for Ireland. She's great. And David was there. And they basically said you have bitten off way more than you can chew. Focus on delivering the demonstration projects. I'm not sure if we discussed pilot projects at that point, and I'm happy to share those notes too, if you are interested or not. And that helped a lot because it made… personally, it made it seem like less overwhelming and being able to kind of focus on, okay, what are we going to try and do with each of the projects. It just made it a lot more deliverable.

I: Right.

R: It focused it down and said we are not going to deliver everything, we are going to have some feasibility aspects, something to demonstrate, to capture the imagination and, I don't know, it just helped a lot for us at that point in the project.

I: Yes. Absolutely. And then also I guess it focused you on sort of making sure that the linkage infrastructure sort of…

R: Yes, correct.

I: Which I imagine is the most important piece of tool…

R: And Christophe, by the way, is the other person. Christophe was the other big piece of that. And I think when we started we weren't realising… that was a whole nother part. And again, Ceri, of course, raised this, and I was being naïve about this, because when I was running a cancer registry, again, I was at this level and not really hands-on. A lot of this whole issue of data processing was done by a huge group of people and I was kind of taking that for granted and I was thinking the Met Office data is primarily raw data in that oh we can easily take that on, but it actually turns out that a lot of the Met Office data that they actually use is model data and they are actually kind of loathed to release the raw because it actually doesn't make sense to most people, it's not really useable. So there was that aspect. And then it turns out, thank God for Christophe, a huge amount of making that data available so that it could be easily linked with other data, be it health or other environmental data.

I: A huge amount of work was involved.

R: Processing, yes. So part of it was this Python piece, which makes it more organised and callable, but it was also figuring out well at what level… one was turning it into gridded date, so actually making it so that it can be a space and time point that could be linked to other data but then there was also the whole callable issue.

The other thing we did was, and this was in January a year ago, we had a meeting where we brought in some other groups to interact with us and that was where we brought Biobank in to talk to us because we were trying to identify sort of major health resources that might want to play with us. A GP one called TPP that Shakoor is actually… his last piece of work is actually… he's finally got access to TPP data linked with, I think, temperature.

I: TPP GP practice?

R: Yes, it's GP practice. I can't remember what TPP stands for. They are an interesting group. They are a private company but they've actually set up a sort of research branch or even a separate company, I'm not sure, and they actually make available an anonymised dataset for researchers to work with, and they were interested in interacting with us. So we are still exploring that. Whether they might be in the long run willing to make accessible some portion of that anonymised database, I don't know. What's happened now is… currently what they've done is they took temperature data from the Met

Office and linked it just like Biobank does and then they've given it back to Shakoor to maintain confidentiality, and presumably control.

We also had Jan Semenza who works at the European Centre for Disease Control and they have a thing called the E3 portal. Gordon had spent a year there in Stockholm working with him.

I: E3?

R: E3 portal. It's at ECDC. And they have acquired European country-level data, both environmental and a lot of it is infectious disease. So it's like they are trying to do the MEDMI thing but on a much larger scale, because we were only UK-focused.

I: What's the relationship with them?

R: Well, Jan Semenza who works with them and Gordon Nichols collaborate a lot and Gordon had spent a year with them. So they came over and told us about some of the stuff they've been doing and some of the issues they've had. They've avoided a lot of the issues. Well for one thing it was kind of EU mandated, that these countries give them their data, and also it's very non-threatening to give people your data relatively [laughs] on the country level. It can be threatening too, for example, I think one of the papers that Jan and maybe Gordon were involved in was around the first reported cases of locally acquired malaria. That can be threatening to the Greeks for tourism. So that was just to talk with them. Also, Wayne Elliott, who is on our advisory board and is in the Met Office, was spending time at WMO, the World Meteorological… and he's also tried to feed stuff into us. But so far we haven't gone international because we've been having enough to do just with the national stuff. So it was more like talking to groups and seeing how they… and out of that arose a workshop on the Biobank stuff that Sabina came to.

I: Yes, I was also there.

R: That's right.

I: Okay. So it looks like obviously since you've now got the core of the work on its way, the sort of legacies or the opportunities are now sort of coming up.

R: I think so. And we did try to also interact with groups that we thought might be interested. And also there was this sort of quest for human data, because when… so Health Protection Agency, when we put this in, they were the player, not Public Health England, and at that point they thought that they had access to a lot of data, human health data. For a variety of reasons. They didn't once they became Public Health England, either because Public Health England had this huge reorganisation and they've been trying to figure out what data they have and don't but also it's become clear that a lot of the data that they feel that they have are really only for use within their organisation, that they are not really at liberty to give it to MEDMI. So that led to, for example, our endless quest for ONS data. When we put this proposal in it was under the assumption that they had access to the vital statistics data and that we could just stick it up on MEDMI. It became clear… and maybe all this stuff was evolving at the same time, do you know what I mean, because vital statistics keeps doing its thing. It became clear that we couldn't just give away

that data. They have access as does the Met Office to HES data, the Hospital Episode Statistics data but again it's given to them, my understanding is, under the understanding that it will only be used within their walls for specific public health purposes and they can't just make it available, that there are processes that you have to go through to get those data as an individual researcher. [redacted] the Met Office has really given a lot of data to MEDMI, it's almost, I would say, 98% their data right now, and SGSS, which we finally acquired, is the only real health data we have at this point. And I've sort of resolved this by saying, you know what, at least we can show in demonstration project two how health and environment data can interact and also demonstration project one, too, as well, that's vital statistics data but at a very high level. It's also led to other studies. So, for example, with SGSS, the highly confidential part is that they have people's residence. We are using laboratory of diagnosis as the variable that will be linked with the environmental data and they are now doing a separate… Gordon and Majeed and Gianni are doing a separate analysis of sensitivity to see how big a difference it is, what a bigger hit it is in terms of lab versus home for the environmental link, and my bet is it won't be a very big difference because the environmental data in many cases has a large radius as well, so it will incorporate both places. It will be the same environmental data no matter where you are.

I: [Overspeaking] evaluation of the cost, epistemological cost of protecting the residents or something?

R: Well, what they are going to look at is what is the concordance? If I have the housing data, the residents' data, how different is my environmental data from if I use the lab data, and the hope is that it will be very similar. Other groups are trying to look at it for things like socioeconomic consensus type data, and I have a feeling it will be much more, obviously, it will be a less concordance for that, but luckily we don't have to look at that right now [laughs].

I: Right, yes. Absolutely. And this is an analysis that is going on now. I think it's interesting then to come to the question that I wanted to ask you. I think it's come through a different way, which is, okay, so in the interviews so far it's come up that the major stories of things that sort of slowed down the progress in some respects were… one was maybe the access, the accessing [overspeaking].

R: Oh yes, yes. Well, particularly the human, totally.

I: The other was to figure out the link, right, the levels of the link, what should be the standard solution.

R: And who gets what.

I: Finding an agreement.

R: Right. And the other was the personnel.

I: Right, yes.

R: Expertise, skill set and resources available.

I: Yes. So, for example, if we just now focus on the accessing the datasets things, what I'm trying to understand also is if this implied some kind of sort of thinking about how to find solutions differently, so the same… like what you were just saying, for example, what if we do the same research using the lab location instead of the postcode location -

R: Oh I see what you mean.

I: - since we won't get the data we would like ideally to get.

R: Okay. So has that become a regular thing, is that what you are saying?

I: No, no, I'm just asking. If you think in general that the story of not finding it easy to access datasets has implied a lot of this kind of, you know, work.

R: Workarounds.

I: Workarounds, if you want.

R: So again, their original idea, which was wrong, was to have this accessible linked anonymous database that was going to have all different types of data linked up. It doesn't actually make sense, okay, but I'll just say that's where at least some of us… well I can't speak for all my colleagues but for some reason that was what was in my head. So the first issue, again, is that you can't have it all linked up or different groups are going to want different linkages.

I: Right, yes.

R: And really the ability with the first browser to go in there and look at it differently, change the lag periods, is one of the things you want to be able to do.

I: Yes.

R: So then we realised… so that was just one big one. So then you are going to have to have raw data, model data and linked data, variables, depending on what the person can do and what they want.

I: Right. So that's the solution now?

R: Well, that's one issue. So that's what's in MEDMI right now is model data, raw data and Christophe is making these variables that have these different lags so it would be easy to link with them. And if we do that interface maybe we can appoint…

I: So kind of pre-selections that are figured or recommended.

R: Correct. Exactly.

I: [Overspeaking] default options.

R: Like a week, a day lag or whatever. But the other thing that became clear is that… and I think we did sort of understand this in the original proposal, but I

don't think we really grasped this was going to have to be… that there's going to be a group of data owners who will only interact with us if we give them our data, they make the linkage and maybe they'll give us something back or maybe they are going to tell us we have to go through their systems, we and others will have to go through their systems to then access those data, those MEDMI linked data. And that's Biobank, TPP. It's the health databases really. Oh, by the way, we also approached Centre for Ecology and Hydrology, CEH, and they were also at that investigator meeting in January and they were talking about their databases that they have. They have a lot of NERC databases, because we were trying to expand the environmental data. Anyway, that's just a side note.

At one point we had talked about having the MEDMI on the cloud and at one point we were actually going to be joined up with the IBM project that was funded by TSB. I can barely remember this. Brian might remember it better than I. And TSB, which is now called Innovation UK, I think, had gotten money with the Met Office to set up a cloud and that you could use the cloud allegedly even for confidential data, and so our idea would be to put MEDMI out there and there be a group of researchers who could just download the data they wanted but then there might be other researchers who would need it linked for them. So rather than having a server and that kind of thing, it was actually going to be on the cloud. And Alberto Arribas sort of has done that, or tried to do that with the Met Office. I met him after he started that, but he's tried to do that. And that went away because I think the funding may have been taken away from that project, it wasn't progressing. So then we went back to the idea also of having it on a server in a place with limited… and then we also evolved into this issue of limited access. That was another thing. At one point I was hoping we could have an anonymised database that anyone could look at, and two parts came from that: one was that even though SGSS is a fairly anonymised database, there's still this feeling that it needs to be behind a wall, and then there was this other feeling of, even with the Met Office data, if people use it improperly they could come up with… and this happens in the cancer registry too, people will suddenly publish, for example, that Florida has the highest rate of cancer in the world, and what it really is it's unadjusted for age and they have an old population. So just for age they just have a lot of cancer, but it's not because they are the cancer capital of the world. So it's the kind of incorrect interpretation of the data, that also became an issue for us and that's why we sort of went to that idea of a firewall. Not a firewall, but of people having to get permission to use the data. And we also thought that that part of governance might make it easier to approach groups like Biobank and TPP to convince them to let us have data, but I don't think it has made a big difference, or ONS.

I: Yes. On this one it's come across really before, this idea of like obviously there's also a need to, yes, control against misuse or misinterpretation of the data.

R: I'm not as worried about that.

I: Yes. I'd imagine a journalist could say, 'Oh Florida is the cancel capital of the world,' but then I think because it's so accessible it would be also easier to take it down or just to… for other journalists, you wouldn't probably even need to… or like somebody else could come up, sort of at least that's the open access data movement kind of argument, they

say, 'Oh yeah, people can make lies, but others will actually be policing…'

R: Well it's kind of like a Wikipedia thing, that it will correct. Again, I'm not as worried about it, although I do remember having to deal with the fallout from that, and not just from media but from researchers. I mean, there was a guy who published this garbage while I was involved in the cancer registry in Florida using Florida cancer registry data and it was just a very bad analysis. Eventually, presumably, there's some sort of peer review pushback, but it is out there, people find it and periodically they'll decide that's an issue. I mean, I think that you really can't control those things and I think data (ph: 0:21:56.8) will out. But in order to… I guess the other thought we had was we are trying again… If you think as MEDMI as a sort of giant demonstration project at lots of different levels, they are not giant but large, then having a website where you have a front, an outward facing thing with things that people can see how their data can be manipulated with the sort of modified demonstration project one browser and then a part which is only accessible to people that allegedly are in institutions and have skill sets that can really use the data, that we felt was an important thing to model because it was clear that for certain types of data that was the only way they would ever end up in MEDMI and I think we always felt that our audience was much more kind of researchers, people that use public health data and environmental data rather than the general public.

I: You would have to model in order to get them on the website (unclear 0:23:00.7) they need to be prepared.

R: Right. And also they need to know that that is a route. Going back to what you I think originally asked, that that is one way to access and use MEDMI data. The potential to do it with some kind of confidentiality or firewall, we were trying to model that as well. You see?

I: Yes.

R: Because I think there will probably always be datasets that are under the impression that they can keep it confidential [laughs] and this is the kind of thing they expect to see, the kind of structure and process they expect to see in place if they are going to interact with you.

I: So you were saying we were thinking about how we could model this.

R: And that's what we are setting up with the website. The website is set up such that there is a forward facing thing with a limited browser and then you get permission to go to the back and then you can use the other browsers and you can get access to the data.

I: Right, yes, yes, yes, yes. So you were referring to when and what point in time of the project?

R: I think it really happened when we finally started and that was thanks to Harriet to really imagine the website as the platform. Originally I was not planning with this. I felt like 33 months we were not going to spend a lot of time on the website because we didn't have much resource for it and time and we kept talking about a platform but it was sort of either a cloud or the server where the data were. I think we started to focus more on the website because it was something tangible and it was like a structure plus a process

and we could use it as a platform to demonstrate different aspects of the things that we've already talked about, so the website became more important once Harriet was in place to actually make it happen and it kind of dovetailed.

The other thing I remember that drove us to having this kind of other… this (unclear 0:25:26.4) behind the forward facing thing was the ONS data, because they basically made us list which MEDMI groups, which MEDMI individuals, researchers, would have access to their data. So if we were going to put that data on MEDMI we had to have a process where you got screened and you are told if you want to use the mortality data, you have to fill out this paperwork and we have to let ONS know that you are using it and then you can access MEDMI data that will include the mortality data. We are still in the process of making that happen. I think the ONS thing has been going now like a year and three months or something.

I: I think that is interesting, also in other kinds of cases, for example, SAIL, that I've been looking into.

R: I mean, they took a very different tack and they also had a lot more resources [laughs].

I: Yes. So in what way then sort of security requirements sort of have then shaped what you have done?

R: I don't think we spend as much… I know we haven't spent as much time as the SAIL guys have. I remember when David came down to talk at our big data conference he was just like, "That's the issue." If you are going to deal with the health data it's about confidentiality. We've been in some ways lucky in that we only had this environmental data and really then the issue becomes how much the MET office are willing to let be available to outside people, and they've been extraordinarily open and generous about that data, albeit you have to know Python. But still, they have not made things… and we don't have much, we just have like a one or two-page thing on governance and our criteria is basically that you are in one of the partners or in an established research institute. And remember, we think of this sort of MEDMI one, if you go on with this and you do have confidential data, and clearly the ONS data is the example, our initial example of that, you are going to have to put a lot more in place.

The other part where it came up was I really wanted to put… I felt like the browsers might be a way to deal with the whole confidentiality issue. And this is back when we were thinking the browsers were going to literally do data analysis and real time off of the database and I thought – and this is probably my own ignorance – that if you use the browsers, be the visualisation one or the analysis one, you are not getting to the original data. You cannot drill down to an individual and therefore you are on some level looking at anonymised data endlessly, right, and for me that substantiates confidentiality. Ceri and others have said you would have to actually test that. There's issues about whether you really could drill down. I don't think you can with what we have. But they said you have to test that, and also you'd have to sort of try and hack it. And again, we don't have the time or resources to do that.

One other part of the whole browser thing is we also had originally an idea, and one of the reasons we used open source software to create the browsers, and this was from Ceri, was we also had a kind of, I don't know, hippy nirvana idea that we might even get to a point where people who are knowledgeable could add code to the browsers and expand, for example, the statistical analyses, and that was the sort of things we talked about at the beginning, but it also became clear that to really… not police it but sort of curate that kind of thing would be a lot of work. So if you made it open, the code that he's created to do this, to make the browsers, my idea would be, well, couldn't people come along and go, 'Well, I want to do this part of it, another part of analysis and here's a piece of code that fits in'? But you have to have very knowledgeable people to do that kind of coding and probably you need someone to just oversee that it's not going… just because of ignorance or because of malintent. But around the browser that was my idea way back.

I: Yes. And I was also thinking about maybe if you considered about building the infrastructure over the cloud as opposed to having the server.

R: Yes, we did initially, particularly when we were talking with IBM, but then that seemed to go off the table. I didn't really know Alberto, and when I did meet him his thing was on a much higher level than what we were doing and so we just ended up going with the server because we kept… again, it was when you ratchet down your ambitions you start getting very concrete [laughs]. And remember, we started this project three years ago. So the cloud just wasn't on the table anymore, it became more about having a server that we could, in essence, monitor and deal with. But clearly that's also a major limitation. Like for me to have Public Health England potentially or possibly take this on, am I going to give them the server? [Laughs] I mean, ideally, it would be better to put everything on a cloud; it would make it much easier to transfer.

I: But it wasn't also related to security where the data's stored and things like that?

R: I'm sure that was part of it, but the cloud just sort of went off the table when we left the IBM group.

I: Right. Because it was just inaccessible.

R: Yes. It became a non-issue, for whatever reason. I can't remember at this point.

I: Right, yes. And the other question I'm asking myself is, as you learnt along the way, about how difficult it would become to access the health datasets, and I have understood that it was already there, as you also mentioned. It's not that you were completely naïve because I've read the grant application and there is discussion obviously that the health data needs to be negotiated and so on, so it doesn't just come with a click -

R: Right.

I: - but it was just slower than expected.

R: Yes.

I: Did this sort of shape then the way you conducted the projects or timing of projects, timing of pilot projects or something like that?

R: So demonstration project one, he already had the mortality data, he'd already done that. Demonstration project two, we got SGSS and that work has actually been done in part as a pilot, it's been done over the last six to nine… well actually it's nine months at this point. We put in the ONS application over a year ago. We interacted with Biobank over a year and ended up with that workshop. We decided to make it a bigger issue not just about us but about Biobank environment and health, in other words we brought people in from a bunch of different areas, if you remember. We talked for a while about Census data, we also talked with Black Swan about possibly bringing them in to do some of this. I'm trying to think of other databases that we've looked at. We also spent about nine… I mean, the first year - and it's not anybody's fault or anything - but there was about a year there where we were kind of waiting for Public Health England to get itself… because remember literally the timing was terrible, they became Public Health England just as this thing got started and then there was a whole… we kept being told, "Well, we have to see what our databases is, we are doing (unclear 0:34:11.2). We kept waiting for that to happen because we thought, well, then we'll get all the health data we need. Also I, it was a separate project, applied for the HES data. That's still taking time. TPP has taken a year. So, you see. And with all them, the other thing that started evolving was… and this may or may not be completely upfront, one of my ideas was, well, rather than get them to say no right off, 'There's no way we are going to do that,' why don't we ask them for data for a specific pilot project, do something, and then they see what it is we have, then we have something tangible to offer them and then see if we can come to some kind of agreement about something anonymised, some piece of data that they might be willing to have as, I don't know, a taster on MEDMI, that was my other idea. But until we have something tangible to show them it's a really theoretical conversation and their default is going to be, 'Yeah, give us your data, we'll be happy to link it and then you have to ask us for it back.' So I thought if we could show them tangible products and then say, 'Look, we can do this and it can be useful for you…'

I: There's the potential that we can do more.

R: Exactly. And even if it was that we had a taster on our website.

I: And you applied this strategy?

R: Well, that's basically what ONS, TPP and… Biobank is completely off the table, but that's still out there for ONS and TPP.

I: Right, okay. As we were mentioning the collaborations that have sort of started to develop from the MEDMI, one thing that I wanted to ask you was also what about the collaborations with the private sector or other kinds of… because there was [overspeaking].

R: Yes.

I: But actually you haven't talked about it [overspeaking] other interviewees.

R: The only group that I would say that… well TPP is private sector. Black Swan was another group that we sort of went down this… although in many ways that collaboration was going to be probably more us buying a service.

I: Right, yes. It was you need a…

R: Yes. And then there was something else I thought about. Now I can't remember what the other one was. And that probably has been about it. Because one of the things has been… I mean, most of the human health data is not owned privately, it's owned by sort of pseudo groups like SAIL or Biobank, right. [redacted] There's another group that we interacted with, maybe it was TPP. We had a lot of conversations… It must be TPP that I was thinking about and that they work with Birmingham… Yes, it must be TPP. But beyond that, I would say we have not really gone that way. We haven't really investigated as much as we said we would.

I: And so this was the result of sort of also these kind of complexities…?

R: Probably I was on this thing for 8% and I only have so much time and energy and I clearly gravitated towards groups that I felt comfortable with, I know their language. When I've interacted with businesses it's been more about like well what money is on the table for us to do that? And even the Met Office has a certain aspect of that because they are sort of a pseudo business. So we've actually given them more money. And so I think I sort of shied away from more of those conversations [laughs] and it sort of stayed in my own comfort zone, probably.

I: Yes. But were you expecting to engage more of, you know, sort of businesses there or organisations?

R: I mean, Oracle was involved in our… like I said, Jeremy Nettle, but he never really indicated that there was collaborations there. I have to say, I probably just haven't done due dilig… IBM originally was really interested but their whole thing kind of went down the tubes. So I think… And I just haven't really investigated as much as I probably should have. I just ran out of time and energy.

I: Yes. I would like to ask you some… not conclusive but sort of, yes, as we move towards the conclusion of the interview, things like say, yes, having a bird's eye view. So one question is thinking about the pilot projects, which has been the most successful or the one that has had the smoothest process?

R: The demonstration projects or the pilot projects?

I: I would do one and one.

R: One and one. I mean, there's really only… So there's three demonstration projects and they are very, very distinct. It would be hard for me to say… I mean, both of them have… really demonstration project one and two are the only ones that are relevant because three has been about, you know, we've finally got funding for it, it's going to go on, I'm happy about that, but it's impact on MEDMI right now is going to be pretty slight. And projects one and

two, they've been resource intensive, both of them, and I think part of it we're sort of feeling our way with them. And the one regret I have is that particularly demonstration project one has not resulted in peer-reviewed papers. Shakoor really tried, but what we got back, the feedback we got back is, 'It sounds like a really interesting idea, this browser and all the rest of it, but have you evaluated it and do you have… have you shown that it works? Have you worked with users around it?' And we don't have the time and he didn't have the interest and resources for following up on that. So that was a little disappointing, but I still think we have a tangible product that we can show and it's interesting when people try it out, they enjoy using it and like it and it might be a teaching tool. So I don't necessarily have a strong feeling with either of them. The second demonstration project will result in a paper but it's going to be a summary paper of, you know, here's 2,000 different infectious diseases and we identified the following ones as ones that should be explored further for climate change.

I: Yes.

R: But that's a publishable peer-review paper, it's a tangible thing and it's the kind of thing the MRC wants to see. Okay?

I: Okay.

R: In terms of the pilot projects, I mean, I still think they are evolving because they are actually less than a year and they finished in December but they each tell a different story. The nice thing about Jess's one is it seems to sort of capture the imagination but also it's nice because it uses more clinical data and it shows a very strong case for environment and health interconnections and why it's important to have these data available to explore these connections. She had already been talking to the Met Office and Christophe but this allowed her to look at an environment and health hypothesis, which the patient community had been talking about for years and prove, in essence, I think that it's a real thing, like doctors are going to take them more seriously because she's got this objective evidence.

I: Yes. And instead of the pilot projects, which is the one that has had more… a certain path or [overspeaking]?

R: Well I think that every pilot project is again different. I mean, the SAIL I think there was a lot of frustration, but we got a lot out of it because we learned that we need to take a different direction at the very end in terms of making the data more useable and I think it will lead to them doing more environmental stuff, but I don't think that's going to turn into a paper. The other two, like I said, the pollen and the air pollution, I never expected them to do papers, their stuff will contribute… there will be a paper some day that has an attribution. And actually I think the two that Mark's involved with, they will end up as papers. In fact I know they will.

I: Yes. And why are these projects not turning into a paper, is that simply just, I mean, it's part of the research?

R: The pollen and the air pollution ones?

I: Yes.

R: Because what they really did is made data more available. They may end up contributing to papers. But Jess's idea was a hypothesis driven thing, very well defined. In fairness to everyone, she had done work on it before, right, and what it allowed her to do was get the person power to actually make it happen. The other pilot projects, like I said, the pollen and the air pollution ones are leveraging off of ongoing projects and what it will do is for those ongoing projects give them better data. And maybe if I went and did the pilot projects again, maybe I would change the criteria and make them more… I might make it more focused so it only brings in a Jess type project where you have to have a strong hypothesis that will result in a publishable paper. I don't know, we had three months to do it and we had enough money to fund eight projects and I think if we can get four, or six, I think six projects, but anyway, if we can get four or five papers out of that, I think that's great, and especially since they demonstrate different ways of using that idea of interconnection in environment and health.

I: Yes, yes, absolutely. And the same way… yes, the same way I understand also about the demo project one, that it contributed to a tool that can be used for teaching and…

R: And that's my hope. But if we had more time and money we would try and do a formal evaluation of it with users, which was the feedback we got, and I agree with that.

I: So it seems like… I mean, all of these projects can then lead to something that then matures on and can be sort of used.

R: Hopefully.

I: So why were you saying that it would change the criteria to have stuff that is… is that because you need impact, or why would that be more preferable if it is [overspeaking]?

R: Well, I mean, I've done pilot project – what do you call it – programmes, pilot project programmes before with other centres and mechanisms with funding and you can make them very specific so that they are purely hypothesis-driven research. The two that make the data richer, better, are part of hypothesis driven research but they are not in themselves going to end up in a tangible product, at least not for a while. So you can make it to say we will only accept hypothesis-driven or we could have funded only… I mean, we basically funded all but one project that we got and part of it was just I wanted to get some projects out there and people were using the data so we could learn lessons but also so we could get some tangible deliverables and part of it was we were running out of time. If you are going to give people money you've got to give them time too. [Laughs]

I: Yes. So, more generally, looking back at the whole project, at the whole adventure, what would you do differently?

R: [Laughs] Oh my God, I don't even know if I'd do it. No, I've gotten incredible benefits out of this on lots of different levels, and I've learned a lot. I mean, for one thing the budget, although we knew it, I mean, I had an inkling going into it that it was going to be a big problem, but the fact there was only one full-time employee was just ridiculous, and there's this tension when you have four strong partners, do you give each partner their own resource so they'll

come to the table happy or do you fund some sort of central highly controlled resource where you might actually be more productive? But on the other hand if the partners aren't happy they are not going to be productive. So there is that tension, right? When we could have funded two or three centrally held groups or individuals who might have done more. We might have been able to have a software person, a database person and a postdoc doing data analysis. We could have maybe funded this really differently, we could have set it up differently and it might have been more productive in that way. The alternative would be we could have brought in more health partners right from the beginning [redacted] but then the resources would have been even more thinly spread and we weren't trying to fund a particular health group, we were trying to fund a concept of bringing health and environment data together. And I go back to also what these things… If you read or, I don't know if it's still even up. If you read what partnership grants are supposed to be, they are either supposed to take an existing group and make it sort of a quantum leap better or it's supposed to bring a group of collaborators together to start an idea and show feasibility. And I think we've done that. It depends on what your threshold is. But I think we have done that. And I still think if you look out there there's not… there's very few groups trying to do this. I mean, there's a reason they are not trying to do it, because it's really hard, but there are groups doing big data on environment and climate and all the rest of it and there are groups doing big data around health but that linking part just isn't there.

I: Yes.

R: And by linking I mean the ability to access both types of data and be actually able to link them and look at those data.

I: Yes.

R: I still think that's a lot… it's not there.

I: No, yes. Yes. Absolutely. Just last two things. You said in the interview that two people, I mean, of others I think have been very useful when they got on the project, who were Gordon and Harriet.

R: Yes.

I: So can you just sort of explain how they helped?

R: Gordon, he has great enthusiasm and he has produced a lot of papers involving MEDMI data. Also his connections with ECDC I think have been really interesting and might contribute to a project, one of the projects going forward with the MRC. He also took ownership of project two, demonstration project two and really pushed that forward, and he also just kind of… he has been working at that interface of environment and health, although he's primarily a health researcher. But he originally, I found out, was trained as a laboratory microbiologist, so he really gets the other side, if there is another side, and he's somebody who I think sees the potential. I mean, I have to say everybody in the project, I do believe, sees the potential, sees the need and the potential, but he actually hands on is doing it.

I: Yes.

R: And then Harriet, what she provided was somebody who was… she works part-time but, well, she's very smart, very organised, great people skills and was able, even though this wasn't her background, to grasp what we were trying to do, and where she didn't know she would ask questions and learn about it and often would reveal that different groups were having these parallel but not connected, you know, we were all thinking that the other person knows what we are doing but were really not, you know, around the whole browser stuff. She would ask the question that would show us that we were not on the same page. She also committed to putting the website together and the Biobank meeting, the investigators' meeting, meetings that we've had, the advisory board, all of that has been because of her. So she's sort of underpinned the process. She's been also about the documentation and also just making sure that things happen, that subcontracts happen. It's been huge for me too because at one point there was all this day-to-day stuff that I was just not keeping up with and when she came that…

I: Got done.

R: Yes. So I couldn't see the forest for the trees. But she is also an unusual person. I mean, she's very… she can see the big picture and focus on what has to be done and the little details. And again, her people skills are amazing. I think everybody in the project really respects her as a person and her abilities. So we've been really lucky.

I: Yes, I've heard that from several people, that she was very helpful for this kind of input and…

R: The stuff we have, between her and… I mean, everybody has contributed. Ceri, definitely, huge impact, and she. And Christophe, too.

I: Yes. Great. Okay.

R: And I should mention Majeed as well but…

I: Yes. I've got that as well. Okay, great.

R: Thank you.

I: Is there anything you would like to add that came to your mind during the interview that you thought I'm going to say a bit later? Otherwise we've talked [overspeaking].

R: I don't think so. I feel like a lemon that's been squeezed.

[Laughter]

I: Okay.

(End of recording)