NTDS_030_1

Key:

I: Interviewer
R: Respondent

I: When I was doing my PhD I tried to transcribe some of the interviews myself and I understood how painful it is.

R: That's exactly what my wife did. She first asked me to transcribe it and then we just got a professional typist to do it. That's how we learnt to use Dropbox, because the typist was really up to speed. She just said, "Oh, just drop me the recording. Put it into my Dropbox here." "What's Dropbox?" Then we went and, fine, it was good. A long time ago.

I: It's been a while since Dropbox has become default. The way I would like to start is maybe just by you recounting your involvement, how you started getting involved, at what point, for doing what, and then a little bit the story of MEDMI from your experience.

R: The story really starts about three or four years ago when I changed roles inside the Met Office and I became an IT fellow. In a sense that's like being an emeritus professor. Office stopped trying to get me to do things I was no good at, like paperwork, and ruin budgets, and doing the things I was good at, which is international collaboration. I don't have any team. There are two or three of us as IT fellows. There are science fellows as well. We've had a proven track record with a global reputation in a number of areas. "Just go and do your own thing," and what's what I'm doing. That allowed me to focus on the geospatial aspects.

I: Your background was then...?

R: My background is mathematics, two degrees in maths, then after leaving university and working in developing countries I joined the Met Office and have been here for 40 years. I've had three careers, so one was in research and maths. The second was in computer graphics and data visualisation, a lot of technology, and supercomputing was part of both of those. The third one was international collaboration and telecoms, mainly in the meteorological arena where the WMO is part of the United Nations, 193 countries all agree to do things, so interoperability and standards. So I've had this big theme for decades going through my work. So when I look at what the British government does, saying, "We have a standards institute with national standards." There's no point having a national standard in anything. There's no point having a European standard like the INSPIRE directive. It's got to be global. Global or nothing. So I've always had a global perspective.

Act locally, think globally. That kind of thing. So got involved in the geospatial things in 2009. I established that MetOcean Domain Working Group, got involved, started to convince the geographers there's a dialogue to be had. That's fine. Then it was about three years ago when MEDMI started. I heard about it. There's a big geospatial aspect here. Of course we just have the internal reports of what's going on, but of course I've known Brian a long time. Internal Met Office report, just what's going on, on the grapevine as it were. I

said, “Ooh, this is interesting. I wonder if they’ve thought of all the geospatial aspects,” and by that time I’d had two years of being involved with the international geospatial community, so I thought I was going to contribute. So I talked to Brian and others. I was in two minds as to whether to do any work about it, when I found out that the board overseeing the project had a representative of Oracle.

This brings out my prejudice because we’ve got Oracle databases. We pay a lot of money to Oracle. I’ve been in IT long enough to know the real name of the game is not lock-in. It’s what’s your exit strategy from that big player? You may trust them for this contract, but you don't trust them for the next contract. For example, for quite a few years Met Office used Google Maps. Then Google Maps, last year, changed the terms and conditions, and we decided to leave. We no longer use Google Maps on our public website, because Google’s terms and conditions, they’re not free, there’s a cost, there’s lock-in. So that’s always been the name of the game in IT. So I was a little bit concerned. So I got involved. I got less involved once the project started taking shape, and also Kerry was appointed as the IT person who was recruited, and once I realised he had a strong inclination towards open source software to make it more sustainable and portable.

It was sustainable, to be honest, and cheaper. That’s when I stopped getting less involved and just took on much more of a watching brief. Up until then I was actively involved in the meetings. “The IT infrastructure, just don't take what’s given. Just don't take what university has because that may be inappropriate for the project and for going forward.” So once I realised that Kerry was kind of committed and then he sort of convinced Laura and the others that open software strategy was the strategic direction. It’s fine. I kind of stopped having a day-to-day interest in it.

I: Right, because you were satisfied with that idea?

R: Yes, that’s right, and it’s gone in that direction. That’s fine. One of the things I found interesting compared to other projects I’ve been involved in, and other research in the rest of my career, is how unnecessarily diverse the health community is. In some sense diversity gives you strength and resilience and a rich ecosystem of doing things, but at some point the diversity gets in the way. I definitely feel the health and medical community are much, much too diverse. In a sense, the things I found interesting about the project were nothing to do with the technology, whether medical technology or epidemiology, or geospatial or meteorological, but the fact that where’s everybody coming from? You get somebody who represents what ostensibly is a public sector, good organisation, but they’ve got loads of private contracts on the side. So that is a really funny mixture that we get in the National Health Service between private and public.

Having been in the public sector for quite a long time, private contracts are very well clearly delineated to one side. So we have people to run the canteens. We have people to run some of the IT hardware, but the critical things in the Met Office, the key business things that are really important to us we own. They’ve not been contracted out. I get very worried when I can’t tell whether somebody’s who’s meant to be looking after my medical data is interested in selling it. Public Health England fell into that trap a couple of years ago. That’s another reason for me keeping a watching brief. That’s got nothing to do with my professional interests, but as a public citizen I think I’m

concerned about what happens to my data and I don't want my data sold to marketing people. That's kept my interest because I found the health community and the medical community is quite different from what I was used to, it's quite surprised me, and individuals just pursuing projects.

Whereas, here, you'd see somebody involved in an international project and the paper gets published in a prestigious document, like Nature, and it's got 20 authors on it. That's normal. You feel medical research is not quite the same. It's much more about personal careers. Does that make sense?

I: Yes. I would like to know then in what way these things came to shape MEDMI.

R: I don't know. I think it only shapes it in terms of the individual studies and projects within the framework. So I think the framework is enabling those, but whether they've got the right attributes or not I don't really know, because I don't have time to keep track of all of that. I'm more interested in the IT framework, the data framework, and make sure that's technically right and sustainable into the future. I think one of the problems I have with the technology is there's a very fundamental thing, which is most epidemiological data is a big hypercube.

I: Hypercube, the four-dimensional cube?

R: Or in-dimensional, right. In-dimensional hypercube. Just a big spreadsheet, right, with loads of axis. Most of the data manipulation is just taking out a slice or extracting a slice or slices you can do things with, right? That's fine. That's all well-understood, and there's lots of technology that supports that, and there's also international standards that support that because there's even ISO standards that support that that have come out of the statisticians who are accumulating information about countries, the developmental level, and exchanging data within the United Nations to see how countries compare to each other, through the OACD and organisations like that. Fundamentally it's the big spreadsheet model. The problem with geospatial data is you need an extra layer, an extra component of that model. Because you if you take your hypercube and you say, 'One dimension is location, x and y,' or maybe two dimensions, maybe three dimensions, XYZ, it doesn't matter, or four dimensions.

Actually when you start then taking out slices, a lot of the geospatial data, when you take it out, is still too big to handle. You extract all the data and you get all the data for the globe, so therefore you need to chop it up. So not just slicing, but you have to dice the cube so it's kind of chopped up into, 'Just give me the data near where I am,' which is not the same as a slice. A lot of that structure is not available yet. Nobody's really recognised that. A fundamental aspect of a lot of geospatial data, there's so much of it you can't handle the whole globe at once. You can't even just handle one aspect of it all at once. You have to just divide it up into a grid and say, 'Give me the tile where I am, or the tiles surrounding me.' So I'm assuming you've all agreed to share the data globally, which again is an interesting contrast with medical data and hydrological data. There's no long tradition of sharing data globally, so you have lots of isolated datasets and then you can't join them all up.

I: Is that problem that you were talking about, the slicing and dicing, is that referring to a problem of linking that to other datasets, for example?

R: It's related to linking.

I: So obviously it matters to me to know here and now, because I'm here and now and I'm not in Honiton. So is that what you were talking about?

R: Yes, in a sense. So you're only interested in data around here, not the data around Honiton. When you talk to linking, you're then immediately getting into linked data and the web 2.0, which is instead of having humans following links, you're having software following links, and that's the kind of the name of the game for the future. Then you say, 'How do you link all of this data and how do I go from one piece of data and then say, 'Actually, there's other data related to this,' and the software just follows it automatically and finds it. So there's a lot of assumptions there about you've got the data in the right structure. Basically things are either cubes or trees. Most data structures can be condensed down to trees or cubes. Actually you can make a tree into a cube, and vice versa, but it's something you don't want to do really, because it's expensive. The linked data community, i.e. the World Wide Web, the W3C in Geneva, they've recognised the existence of the data cube's a fundamental object, and slices of it are a fundamental object.

So there are lots of mechanisms. If you have got a cube or a spreadsheet, you can take a sheet out of it and you can create a URL for the sheet, and that's the persistent or whatever, and that all works, but nobody's got any kind of infrastructure that says, 'Well, within that cube I want localisation.' There's no scheme for that that's available to all browsers in the world. Basically we're talking about getting browsers to do things. So one of the things that interests me about MEDMI is in a sense that we want to get away from proprietary software, client server software, software that knows about a domain, like, 'Oh, this is medical software. This is a medical browser,' or, 'this is a meteorological browser.' We have lots of meteorological browsers. The forecasters use them for forecasting weather, but in five or ten years I don't think there should be one. It will just be a browser, whatever the browser is in ten years' time. So browsers have become the mechanism of access of data.

So the future of MEDMI is how do you make browsers available and do all the security, know all about the structure of the data, knowing about the terminology, whether it's meteorological or medical or about digging holes in roads. So there's got to be publicly available control of the (unclear 0.16.20), which search engines have, but that's all still in the private domain. Google obviously has big tables, as it were, of controlled vocabulary in hundreds of languages, which it uses to respond to search queries, and knows where to look. Meteorology has had controlled vocabularies for a long time. Well, traditionally since 1853. In medicine in the UK they can't even agree a vocabulary to use. 'Oh, we're not using the European one.' All this kind of stuff, the American one is different, and until there's this slightly more global infrastructure in healthcare you can see it's going to be problematic. So there are pressures there, and there should be pressures there to push us in that direction. It's all very strategic and long-term. It's not answering your question at all, is it?

I: No, but I find it's extremely interesting, also because I was finding very interesting your way of talking about it as a hypercube and stuff. So is that your terminology?

R: No. The hyper bit is standard mathematical terminology. In terms of a data cube, that's very... World Wide Web Consortium have done recommendations about cubes. That's a fundamental thing for linked data and for the Semantic Web. It's all there.

I: I know something about that.

R: I'm also on standards groups for World Wide Web as well.

I: For the W3C?

R: Yes.

I: When you were saying that the geospatial data is a cube, just for clarification, by geospatial data we mean...?

R: Well, the major aspects are X, Y, Z and T. When you start putting lots of things from things you'd have on maps or whatever in that cube, it'll get very big, and especially when you put in meteorological data or graphic data, environmental data, there's a vast amount of data every day, and it changes all the time. So it's too much to cope with the whole cube. We can't cope with the whole cube. We're producing a fraction of a petabyte every day. That's a lot of data. When you get to that stage, you can't move it around easily, even if it's in a cloud. Once it's in a cloud you can't get it out because it's too much. If you look at the space community, they're already doing things like having lasers satellite to satellite, so the traditional way of doing things with a satellite is you've got somebody observing earth, getting back the temperatures and the colours, or whatever, of the oceans and the areas, and then just beaming them down to a ground station.

But then there'll be only so many ground stations that could cope with that amount of data, so therefore the satellites try and store it until it got a third of the way round and the ground station squirted it all down. So, well, okay you address that by having ground stations that a satellite can always see. It can be quite a lot. Now we're getting to the stage where satellites are starting to transmit to other satellites using laser, which is very high bandwidth, before they're going down to the ground station because there's so much data because they can't keep it. So the amount of data's vast. Actually medical data is vast, because most of it's still in silos and it's still local. It's still local to the GP, or to the patient, or the hospital. In a sense, some of that ought to be more widely-shared, provided it's been anonymised, whatever that means.

I: That's got also very many dimensions. That's the problem, isn't it?

R: Yes.

I: Going back to MEDMI, so you were involved more at the initial stage?

R: Yes.

I: So after Kerry's appointment you didn't have basically any involvement?

R: When he was doing his plans for the framework and how it was going to be built, that's fine, I kind of watched that, and then from then onwards, once the

projects have kicked off using data, I've not paid much attention. I've also been too busy and I only work part-time now.

I: You said you were concerned and looking at the data security. The data security is set-up in a sensitive fashion and the structure of the data.

R: Yes. I've got strong views in that area as well. Again that's come from my meteorological background. We made a decision many years ago, about 20 years ago, to say, 'You don't want to embed security in the data.' That's a mistake, from our point of view. You could see somebody constructing patient records say, 'Here's the public bit, and here's the private bit.' Well, actually, that's a slightly flawed approach because once you've got a lot of public data, you can deduce other things from it, and the security no longer holds in the same way. So in some sense the security's got to be kind of an (unclear 0.22.36) direction. It's got to be application level. Is it an authenticated user of a system? Is it an authorised user? That's a layer completely separate from the data. It's a difficult issue to know how much you can anonymise data without actually destroying its value.

I: Are you thinking about a discussion that you had about MEDMI?

R: No, it's just the way some of the existing databases are. The fact that they address their security by letting in a very small number of people rather than thinking very clearly about what... because some of those databases were constructed a long time ago, basically some of them were on paper originally, and then digitised, so we've inherited those kind of structures. It's not at all clear what the best approach is for... it's not clear that anonymising works anyway, and then you've got the issue of authorisation and authentication. People confuse the two. They also think obfuscation is actually good enough security and the answer is, 'No, it isn't.' We've been through all those struggles and other communities have as well. So I think it's still another question.

At present, the medical approach... well, it's got a mixture of anonymisation, which I think doesn't work, because you have this issue of you kind of remove some of the patients' personal details and say, 'Well, actually, that's the only person within that postcode, so we know who it is.' So it's very easy for an outsider, if they really wanted to, to de-anonymise it. If you anonymise it, to a certain extent you may actually be removing the value because you anonymise it so much you've gone past the scale at which you see the correlations. The fact that for some reason let's say a remote mountainous district of the UK, 'There's only three farms, it must be him.' So you have a bigger area, but then the area encompasses some lowland, and then you've removed the correlation with 'It's high, it's cold, it's wet, that's why he's got bronchial diseases,' or whatever. So there's an issue of does the anonymisation of the medical data occur on the same scale as which the meteorological data is (ph 0.25.44) *significant*.

I: Was this an issue that came up?

R: I don't know whether it's been discussed at all. Anonymisation has just been put on the table as an issue, and I don't think the project as a whole has really tackled it. I think it's a big issue for the whole of medicine, actually.

I: So why is it not really tackling it?

R: It's very difficult. I think it's difficult because you've got the very old datasets and a lot of infrastructure around them. Certainly in the early part of the project it took Laura and the others quite a long time to get permission to say, 'We're a valid user and we're not going to sell it,' and this, that and the other. Owners of data tend to be a bit like this; not very open. So I think there's quite a long way to go, because even in the Met Office we've had a history of open data and some data is freely available to other Met services yet we sell it to them. It's freely available. It's meant to be free. We've been selling it for convenience. Rather than have it in an inconvenient form, we'll give it you in a more convenient form and charge you. That's part of our business model. Trying to get away from that. This data's freely available.

So if we're charging them, they just go to somebody else and get it for free, which is what they're entitled to because everybody's internationally agreed it's freely available data. It's essential data for safety-critical things. Selling data that's meant to be free is not a good business model, but then there's also this... I don't think it's a belief. I think what's happening now is freely available data engenders more business. Not only have you got the open source software model, you've got the open data model, which engenders new business and innovation. I see those medical databases still being, 'This is my data, get off, because we've been doing research on it and I've got PhDs coming out of it and it's mine.' I still see that.

I: Going into the history of MEDMI, do you recall any particular discussion or challenge that arose that or discussion related to a challenge that you have been involved with, in particular in respect to security or data structures?

R: I was aware of one or two people offering up their data, but on very stringent terms, which actually kind of stops a project working. To be honest, I think that data wasn't taken up, because they were in the old model of data. So because the security issues got in the way of them saying, 'Well, let's be slightly more relaxed about security, even though it's there,' or maybe they just needed harder work, but then the project's got a finite life and you can't change everything. So Laura and the other people more directly involved in the individual projects have just got to make a decision, and I'm happy for them to do that. So I can't give a concrete example. I can remember a person quite clearly, we were meeting in London but I can't remember the names.

I: Do you want to recall the episode?

R: We had a meeting of a presentation for potential databases and potential candidate problems and one or two people... the barriers were too great to be involved in the project. It was a couple of years ago.

I: What about the task of linking relating meteorological data and health data on a spatial scale? I know that there's been a lot of thinking and learning around this kind of issue and a couple of different approaches have been tried. So what's your experience of that?

R: Well, the problem we still have is our systems are still geared to saying, 'Here's the cube. Here's the half a terabyte of data. Pick out the bits you want,' as opposed to saying, 'Here's the bit for you near Honiton.' We haven't got an infrastructure for doing that on a sustainable basis. We can do ad-hoc

one-off solutions. 'Okay, here's the bit and you can pick out all the bits for Honiton and UK but if you change to another parameter, or you change to some input from a forecast model from the European Weather Centre, or from Météo-France, it'd have to be re-built.' There's no generic solution. It's just a tailored solution.

I: You think you should have been able to find a generic solution?

R: No, this was part of the process. That's what informed my thinking. So I didn't think like that when we were thinking about it. Having thought about it, with hindsight, that's one of the lessons learnt, that we need that kind of structure.

I: That we need...?

R: Across met services, right, to make met data more readily available to projects like MEDMI, cross-domain. That's why I was going back to the fundamental dicing of the data cube.

I: Can you be more concrete in the example?

R: There are only 20 institutions in the world that can do global forecasts. Forecasts for the whole globe, and they are the big met services, civil defense in NASA. There are perhaps about 70 countries in the world that will do numerical forecasting for their country, and they can do it at a high resolution. That still leaves 100 countries that are dependent on that first 70 or 90 countries for the forecast. Of those 70 countries, most of them are aiming to do resolution of... well; the global ones are doing a resolution of about 15kms. That's state-of-the-art. Got a big supercomputer, costs millions, uses megawatts of electricity, can forecast the globe every 15kms around the globe and over 100 levels. That's fine, but that's not really giving you some of the detailed things you may want to look at in medical aspects. A 15-by-15 kilometre square could have some very high ground and some very low ground, some very damp ground and drier ground on the other side of the hill.

So you may not pick out the correlations. You may want something finer because especially if you're looking at things like pollution plumes from factories or whatever. Therefore these 70 or other odd countries are aiming to get their forecast models, their regional models, down to a kilometre. Lots of them are 15, 10, 5, down to 1 kilometre. That's the kind of resolution which you can forecast for and therefore then accumulate a climatology for, and have quite detailed 1 kilometre resolution. That enables you to start doing I think different kind of studies than you can in MEDMI at present because it's coarser. Is that going to change in the future? Well, the resolution is... more people will get to 1km resolution, but getting below a kilometre is hard, and it's going to take 10, 20 years, if at all, below that. So that's the kind of resolution the data's going to be held. I've forgotten what the whole point of that was.

That means if you want to look at some say across Europe, you're not going to get it from one country. You're going to get a 1km grid from one institution, Météo-France, one from Germany from Deutscher Wetterdienst, and one from Italy, from Italian Air Force. That's the only way you're going to get that kind of resolution across a number of countries. If they all do it the same way, fine, and if they all agree to join up, fine, but in practice you will see that they'll end up a bit like the medical databases that they're sort of the same but they

don't quite join up. So it's getting that infrastructure where you can blend all of that kind of data at that resolution. So that's where it becomes important, when you get down to that scale. If you're happy to just do things at 20-25km resolution, or coarser, fine, you just take the global model data from those 20 institutions.

I: Were you also implying that the current solution in MEDMI for the linkage is more sustainable in the long-run because it can better join to future resolutions?

R: I don't know the answer to that without looking at the detail, because certainly yes and no. Some aspects of it in terms of the IT should be sustainable, but in practice a lot of the devil is in the details. So when the metadata is constructed, colleagues may have constructed the metadata for MEDMI and not thought about, 'Well, how would we plug this into the German initiatives into health?' because we know they all do it, and they have projects going on at German universities. Have they chosen the same set of standards or not? Not quite clear, without doing some research. So the potential is there. Whether we've actually succeeded in doing that, I don't know. That's an aspect. How much can you take the MEDMI solution and lift and shift it to France or Italy or Holland or whatever and apply it there?

I think there's a reasonably good chance of doing that with meteorology because meteorologists have these global standards and maybe a bit of tweaking and twitching because your grid doesn't quite join our grid and you've chosen a slightly different parameter. We could sort of fudge it and make it work. Medical I don't think will happen at all. I think that's a useful test with which to assess the whole thing, whether it's going to be possible to lift and shift to another country.

I: Now that the project is approaching its end, if you look back, what were your expectations for the project at the beginning, and how you see these expectations now?

R: Well, I don't think I had expectations for it to be quite as global as I've just described. I think that's what my expectations are now. I'm just saying, 'Let's be bold and take the global perspective,' and that's what I'd like to see. At the time I thought it would be much simpler and I thought it would be much more about the technology rather than attitudes and politics and entrenched positions, and people having different perspectives.

I: If you were to do a MEDMI 2 or simply to do it again, what would you do differently? What would you wish had been done differently?

R: I think I would have done a much tighter and deeper technical review. So, for example, metadata, try and use existing standards rather than just make something up. Try and use controlled vocabularies that are maintained by other people rather than make up MEDMI's own vocabularies. A third thing would be I'd quite like to have seen a road map for UK epidemiological databases. I don't know if there's any chance of that, because of politics and all this. I'm so used, from my community, to have an understanding of where we want to go. So when I was in Washington colleagues were talking about the National Weather Service. They routinely have been involved in a collaborative project with us, France, Korea and Australia on building infrastructure, to which the Finns and the Americans said, 'Ooh, we want to

join,' and put serious money into it, along with the money we put in. We're talking about seven figures each. 'If you want to join, a million Euros, please.'

So we're used to doing that kind of thing. So I think I've learnt a lot about the healthcare community and how it hasn't really got a handle on IT in the way that we have. I hope that doesn't sound too arrogant. So that kind of road map and push towards global standards for vocabularies, structures, policies, data policies, security policies as well, I don't see that at all. I don't see WHO getting involved in that at all. We've got an international framework to do things. Not quite sure what WHO could do, or whether anybody in the UK would pay any attention anyway.

I: So you found that the solutions that technical review of the project.

R: I think they've moved the communities away from where they were, which I think is good, but I thought the steps would be a bit bigger. I thought it would move more. I guess I was going to be over there.

I: Sorry?

R: My eyes are always bigger than my reach, you know?

I: Longer than your arm?

R: Yes. I don't know how that translates in Italian.

I: Okay. In terms of organising collaborations, how were the two ends collaborating? Do you think anything could have been done better in the way...?

R: Not sure really. No, because I think it was... when I first got involved I had a much simpler view of what was going on. It turned out to be much more complicated. I think Laura and her colleagues have actually done quite a lot to move the community forward and make the community aware of the wider issues. Any change is long and hard. I think the real issue is the next stage. How we can build on it and improve things.

I: If we can go back a little bit, we hinted to this in the discussion a few minutes ago. I'm talking with other members of the project and I understood there was this discussion between building this interpolated read to make things 'equal' versus moving to this situation where the system where you can do that on-the-fly in a more configurable way, and stuff. So did you participate in this discussion, and what do you think about the two possibilities?

R: Not really, but I'm aware of it, yes, because the simple answer is if it's complicated and difficult you probably need both ends of the spectrum. People say, 'Oh, it must be this. No, no, it must be that,' and in the real world it's actually a bit of both. The problem with interpolation is it depends on the parameter, and interpolation sometimes is sensible and sometimes it's not. The classic example is... let's take the weather analysis, the wind speed and direction. Over here the wind is that, over here the wind is that. What's the wind in-between? The answer is... people say, 'Obviously it's this.' Well, no, it could be... if there's a front here, the wind is like this all the way over here until it gets to there. If the front is here, the wind's like this all the way here

until it gets to there. So interpolation in the simple sense doesn't work. So you're better off not interpolating. That applies particularly to rainfall. Rainfall is very noisy.

You can't interpolate. You can't say, 'If it's raining here and it's not raining here, it must be raining half the amount there.' It doesn't make sense. That's a better example. It's either dry or it's wet. It's either raining or it's not raining is a better outcome rather than an average. So there's got to be a range of interpolation techniques where the simplest interpolation is you take the nearest value unchanged, which is not really interpolation, but we think of it as one end of the spectrum of interpolation, and at the other end it's extremely sophisticated and you just run a model. You'd only want to interpolate if you can't run a model, because it's cheaper. That only applies to things like sometimes temperature, usually applies to pressure. Pressure's usually safe to interpolate. So the answer is you need both.

I: Another question that I wanted to ask you is...

R: Just to follow-on from that, and therefore in a sense we ought to have in the metadata some sense when we're exposing this data at one level we say, 'Here's a set of parameters. Here's a load of values. If you want an intermediate value, we allow you to interpolate,' or, 'here's the interface to interpolate.' Another parameter will say, 'We do not allow you to interpolate. You must use the values given.'

I: These are two approaches?

R: Yes, two approaches, but of course we may actually do it by saying, 'Well, actually, if you want to interpolate to another value we'll give you another interface and we give you back the right answer, from our point of view, if you want the interpolation.' Of course we may not.

I: Is this...?

R: I don't know what we've done.

I: On another aspect do you think that the project factored enough resources, especially on the technical end of things?

R: Probably did, but the problem was the changing staff. The fact that Kerry was given a temporary job and then he looked for a permanent job and then he came back. That probably held things up. If the person who'd been appointed stayed, it would have been better. The insecurity of the position kind of undermined work I think.

I: I would like to ask you one conclusive question. I hope I'm not taking too much of your time.

R: No, that's fine. We have to leave by 7:15.

I: Okay. My question is I heard from other members of the project that there's been quite a bit of learning also in terms of setting the expectation about what you can achieve in terms of results from this linkage. If you go to the website, for example, it presents the project and it says, 'We can tell that this dataset can be helpful for hypothesis

generation, but things need to be tested further." Then, "You need to be aware of time-lag between the datasets," and there were a couple of other.

R: Probably relative resolutions of the medical data versus the meteorological data.

I: Yes, and not over-interpret correlations with causation. So do you have any reflections on these comments?

R: No, I think that's fine.

I: Were these considerations surprising?

R: No, not at all. I don't think so. Because, in a sense, I think it's learning from both sides, in the sense that we're not aware of the issue of all sorts of complications, so you end up with some medical data and it's all been assigned the location of the laboratory where the blood test was done, as opposed to where the patient actually lived. Now, we weren't aware of things like that. Actually I'm not sure some of the medical people were aware of that either. So all those little details that have come out, and I think that's where a lot of the value is.

I: You mean in the learning?

R: In that learning, yes. So, for example, I have a series of tests against a series of patients distributed across a region and you either get the patient's home address, the doctor's address, the laboratory's address, or the administrative address of the NHS trust for that data. Obviously anything other than the patient's location is going to undermine any kind of deduction or hypothesis you want to make.

I: Thanks very much for your time. I think I've asked you all the questions that I wanted to.

R: I don't think my answers were particularly concrete, a bit of arm-waving.

I: It's nice also to link the specific project to broader questions like you were doing, especially about standards and the global scale and collaboration. Thanks for these considerations.

(End of recording)

NTDS_030_2

R: No, the previous meeting I was at, which was in Holland, and geographers always talk about two kinds of things. They talk about features and coverages. Features are things on a map, anything of interest on a map. Usually they're points or lines, places on a map, roads, telephone wires, or whatever, and coverages, which are an area, a blob on a map which may be a varying value. It may be the height of the ground or it may be the concentration of obstructive pulmonary disease, or whatever. That's coverage an area. So geographers always talk about the coordinate reference system, which you do these and is it the National Grid? Is it various map projections,

this, that and the other? Somebody pointed out actually a very valid use case for discussing features and coverages and areas on a map on the macro scale. They're inspecting macroscopic slides of samples and they want to identify blobs of things and features on the macroscopic slide.

So the coordinate reference system is not latitude and longitude, but microns in X and Y on the microscopic slide. The geographers say, 'Ooh, I never thought of that.' So they're using geographic software to analyse microscopic slides.

I: Wow. Okay.

R: It's rather obvious when you think about it, but they never thought of doing that.

I: Amazing. I didn't know that. This came up at the MEDMI meeting?

R: No, this came up at the previous meeting I was at in Holland. We were just discussing this as a valid use case, because geographers think, 'Nobody can measure anything on the earth's surface to better precision than about 1cm or 1mm, why do you want to do less than 1mm?' They said exactly this.

I: It's comparable also precisely because the geographic software thinks that the earth is flat, isn't it?

R: Yes, and of course what happens on the web is lots of people do something called WGS84. World Geodetic System 84, and when you look at maps on the web, that's what you normally see, and people just say, 'Latitude, longitude, and that's where I am.' So if you're in the north of Finland or in Antarctica it's not very helpful, because of the way they project it, and then people just say, 'If I put loads of decimal places, seven decimal places, it gives me my location to the centimetre.' Well, no, that's the precision. We want to talk about accuracy. You may have specified it to a centimetre, but actually it's over there, not here. It's over there. The accuracy is metres, because, for example, America and Europe are moving apart by about 2cms a year. WGS is sort of fixed to America. So your GPS coordinates in Europe are getting wrong by 2cms a year.

So the last ten years they are this far out. Australia is moving several centimetres a year northwards, because of tectonic plates and things. So over ten years it's a couple of metres out. So if you're trying to do things precisely on one scale, it doesn't work. There's confusion between precision and accuracy. So there are all kinds of interesting geographical things that we have to learn about, that the professionals know but they have difficulty explaining outside of their... they don't see why they have to explain it. 'We're professional. We know this. We're experts, so why should I explain it in simple terms to you?'

I: Right. That's interesting.

(End of recording)