Interview: participant-8

**I: Can I start off by asking you what discipline or subject does your research fall under?**

M: Well, in British terms it's politics, the Americans would call it Political Science but it's on the border between political analysis and sociology really.

**I: So that kind of answers my second question which is: Do you consider your research within the field of Social Sciences?**

M: Yes, definitely, yes.

**I: Can you describe the objectives of your current research activities?**

M: Well we've just finished a piece of work which, as far as social media is concerned – which is what you are interested in, isn't it?

**I: So regarding the data capture, do you use any specific tools to achieve this?**

M: Well what we're thinking of doing in the future is to use Twitonomy. But up till now we've been working with people in the Computing Science School who have got a programme, let me find out – I'll tell you exactly what it's called…and they can capture Twitter feeds through this computer programme. I can't remember offhand what it's called, I'll find out for you.

**I: Was it something that they developed for you?**

M: Yes. No, not for us, it's something (*they)* developed independently, this had been developed and then, if you like, in a way we piggy-backed on this; we made contact with them and used the capacity that they have to collect these Twitter feeds that we wanted. That was basically how it worked.

**I: And did you decide to approach them from the beginning when you decided to do this research?**

M: Yes, we did. It was part of the bid that we put in

**I: Thanks very much, that's great. There's something very interesting you mentioned regarding your research methods because you've obviously got this tool for collecting Tweets but you also mentioned that you do some interviews and additional research. So my question is, would you always seek to cross-validate any findings you have from analysing social media data with, say, interviews or a survey?**

M: It would depend what we were trying to do. I think if we were trying to get an overall picture of political engagement, yes, we would. I have to be honest with you – we’ll probably get onto this later – our analysis of the Twitter information has not been very sophisticated so far and we would want to get a broader view of how people thought they were using Twitter and what they were getting out of using it from a political perspective. So we probably would, yes, we probably would do both.

**I: Can I ask, how do you store the Tweets once you've captured them?**

M: They are given to us on an Excel spreadsheet and we just store them on the computer.

**I: Ok. And moving on to the analysis of the Tweets.**

M: This is where we're not very sophisticated!

**I: Well, I suppose you are the tool, in this case, you are the one running the analysis so do you use any software at the moment or packages?**

M: We haven't done as yet, no, we haven't. What we did – to show you how crude it is – we got the information on the different number of Tweets that different people had made and what we did was we went through and we came up with words that we thought related to politics and then we saw how often those words were used within the Tweets and then if they weren't used then we assumed that they weren't talking about politics. So it's not very sophisticated; it's a manual effort to try and find out how often they've used politics in their Tweeting so we get an idea of how the political element of the overall Tweeting activity, how that relates to other sorts of activity. So it's very, very crude at the moment.

**I: So were you using the search tool within Excel to do this?**

M: Yes, we were, yes.

**I: So do you use any other form of software for analysing social media data?**

M: We haven't done, as yet.

**I: Are you aware of the software package, NVivo?**

M: Yes.

**I: So it's kind of like – promises all bells and whistles and delivers very little *(of it*)?**

M: I think we would use it if we were confident that it could do – within, this isn't a comment about NVivo generally, just the way they use it within. I know a number of people who have lost a lot of data that they've analysed on NVivo and they've had to go back and do it again.

**I: You have mentioned that you were using this word frequency count, did you have a pre-defined criteria or…?**

M: We came up with search words that we thought would be used for political Tweeting.

**I: And would you – if, say, someone could write you a macro in Excel to automate this process, would that be something that…?**

M: That would be good, that would make it easier.

**I: And if it kind of – if you could automate this process and that saved you some time, what would you do with the time you gained to have to work on these Tweets?**

M: Well we do want to do more, we do want to go in and look more at the actual Tweets. I mean, one of the things we're looking at is how often people re-Tweet things and we found that that happens an awful lot, previous Tweets are re-Tweeted a lot.

**I: So frequency of re-Tweets, that would be something useful?**

M: Well it's useful because it suggests that people are not putting their own ideas across, that they are just reproducing what other people have said and one of the things that we were looking at is Twitter an opportunity for people to put forward their own views, what they think and then engage in dialogue with other people and what we found, overwhelmingly, is that it isn't, that isn't how it's used.

**I: How do you perceive it being used?**

M: Well, it's much more monological, it's much more – it's based around an awful lot of re-Tweeting of existing Tweets saying 'isn't this great?' or 'this is what so-and-so said', and there wasn't an awful lot of people saying 'this is what I think' and other people responding to that.

**I: There's some very interesting questions in terms of a tool which we are trying to design. What I'd like to do now is discuss some of the ideas we have for a tool and what you would think about them.**

M: What I would say as well - the number of Tweets we were dealing with, it was thousands, it was just too much for us to go in manually in a lot of detail, we couldn’t do it, we didn't have the capacity to do a lot of in-depth analysis of actual Tweets so we did this rather crude – we applied these fairly crude criteria and saw whether they were used within the Tweets. So nobody is suggesting this is sophisticated and if we could find ways to make it more sophisticated to understand whether things really were political and know more about what people were saying, then that would be great.

**I: Ok, so what my next question is: Would you be willing be to have a look at some of the tools that we're trying to develop at some stage because I think from what you've told me, this aligns heavily with what we have in mind for what we're developing…**

M: Again, I would talk to Iain, Grahame and Lizzie about this but, yes, we would be very happy to do that.

**I: The first kind of finding I'm really discovering here is that, as you said, there's kind of like a trade off between how much manual work you have to do and then the size of your dataset because you don't have unlimited time. So in cases where you can automate some of the initial analysis, that perhaps means you have the potential to capture a larger dataset which would increase your sample size and everyone is a winner, kind of thing. So would you agree with that?**

M: I would, I would, absolutely, yes. We wanted to go back, it limits how much you could collect and I don't know if there's any easy way round that.

**I: What we have in mind is we find that the actual process of capturing social media data, in this case, Twitter, is actually a little bit daunting if you are not willing – if programming is not your thing or, say, you are not willing to invest. I think the latest version of NVivo I costed at five hundred pounds, for something which is, on the grapevine, mentioned as unreliable and not very stable, it's a significant investment. So what we find is there are ways to capture the data and some social scientists, they actually have the data delivered to them, as someone from a public account is actually posting things and they keep a record of it and they give them that record. But then you have the problem whereby the dataset is very small, so it's a very small sample. Going back to your research question about trying to look at the bigger picture, it's obviously very limited. So what we have an idea to develop is a tool whereby the analysis remains in your hands but what we do is we capture the data and then we provide you the means to annotate it and code it yourself so these types of words and this frequency of words you are interested in. And then we produce what they call a knowledge graph which just gives you some descriptive details and maybe some descriptive statistics about the type of data sample you've got. So this might include things like the number of re-Tweets, the number of times people have revealed their location when they are making a Tweet, what kind of hash-tags are used and then perhaps word frequency in terms of the hashtags used over a certain period.**

M: Yes, that would be very useful, very useful. It would make it far easier for us to analysis things and something like that would be – yes, we'd benefit from something like that.

**I: And with some of your other research, say if you were doing an interview like with me, you recorded it and had the transcripts transcribed, would you then follow a standard coding procedure where you would code the responses to your questions?**

M: Yes, yes.

**I: And then in terms of the answers, you then go into more open coding, is that right?**

M: Yes, yes.

**I: And then would you – if, say, you were doing this with Tweets and we provided the facility whereby, again, you've got a huge dataset so you can automate certain preliminary analysis, say the number of re-tweets, frequency of words but when you actually really want to look at what people are saying, there's still a manual process there and again if you have hundreds of thousands of Tweets…**

M: Then it becomes impossible.

**I: So what we're investigating or considering is a means of partially automating that process, so the system would give you the facility to annotate a portion of the dataset and then after, say you've done 10% of it, it would say 'I am confident that I can annotate the rest of the dataset with this percentage of accuracy based on..' Would that be something you are interested in?**

M: It would. Would that be something that NVivo should be able to do that now, should it?

**I: Not exactly, no. What you can do with NVivo is you can do this thing called auto-coding but that’s a little bit misleading because what it does is it annotates the data but in terms of the structure of the data you've got. And then what you can do is you can run queries so that will give you re-Tweets and word frequencies and things like that. But what we're proposing is the machine actually tries to learn how you are annotating the data, as in the open coding. And then what we try to do is let you – you have to teach the machine what you are interested in and then it will try to replicate it.**

M: I understand what you are saying, yes.

**I: And then the second part is – because depending on who you talk to, some Social Scientists are very sceptical about this idea, I've even heard it say 'you are turning us into a script' but the way we see it is the sort of trade-off is we look at it as, if we can automate this process then you can increase your sample size.**

M: Yes and it saves a lot of time, yes, you are right.

**I: And then we would always argue that the machine will never be as accurate as a person however will give you a degree of confidence and will also try to incorporate a future where it says 'these Tweets, we don’t understand them, you are going to have to manually do that.' Does that seem something..?**

M: It sounds very good, it would give us access to much more material and you know, if we were looking at things statistically it would be much more robust because we are dealing with bigger numbers and able to make more conclusions based on the bigger numbers. No, that would be really useful.

**I: Ok, so there's some – I've got a few other questions but this has been really helpful because you've literally ran through the requirements that we need.**

**So I'd like to ask you: Are you aware of any kind of theoretical frameworks or methodological frameworks for analysing social media data?**

M: No. I'm not, but that doesn’t mean other people in the team might – Lizzie might be, for example, but I'm not.

**I: Ok, so I've come across this in my research, it's called the honeycomb framework so I'll just show that to you and if you could let me know what you think about that. I came across that in terms of – you can find that sort of thing on Wikipedia but it's just giving you a criteria for how social media is defined. So I was wondering if you can see in that something which could help you…would that potentially help you..?**

M: I think it probably would.

**I: Don't let me persuade you but do you think it would guide your objectives and maybe help you form research questions, perhaps?**

M: Yes, it could do, yes. I think that is quite interesting. Yes, it could. I'd need to think about it a bit more but it could. It could be useful. It depends entirely what we were trying to find out but I think on the sort of research that we've just finished, it would have been useful.

**I: When you mentioned the objective of your research in terms of first-time voters engaged before and after, it would maybe want to ask you how do you know the people on Twitter are eligible to vote?**

M: Well we don’t entirely...

**I: Yes.**

M: And we also looked at some youth related organisations and the Tweets that they'd received and so on. So we can't be 100% certain but we think we got as close as we could do.

M: Yes, we'd use that to identify people who we could then follow afterwards, yes.

**I: Ok, that's very interesting.**

M: I'm not saying that's necessarily the best way to do it but it is one way.

**I: I think the – outside of this approach, I think this is fairly new to all of us because the social media hasn't been around that long. So we do kind of have these problems where there's the reliability and trust of the data but you have to do the best you can. Would you ever consider using social media to recruit people or to distribute a survey, say you had people who agreed for you to monitor them, perhaps?**

M: Yes, we didn't do any of that. On the Twitter analysis we didn't get people's agreement for this because it was public knowledge. But if we did get people we could ask them, yes, 'would you be willing to be interviewed or would you complete a survey?' Yes.

**I: But again, it's another problem, isn't it, in terms of directly recruiting people versus public available knowledge?**

M: Yes, it is, it is. Because one thing we would have really liked to do is do some work on Facebook. Because we think that probably more – it's more expressive on Facebook rather than people just saying what other people think. From what people have said, there's more chance that people will be putting forward their own views and perhaps there being dialogue with other people. But there are all sorts of problems in getting access to Facebook accounts and some privacy issues and various other things and so we've not been able to do that so far but we would like to if it was possible.

**I: So at the moment, well I'll ask you this question first and let you know what we've developed so far. So if you had a dream research tool for helping with social media research, what would it do?**

M: Well it would let us get the sort of things you just raised; it would ensure that we would have access to the right people, it would ensure that we could get the information that we wanted and it would help us to analysis the information and that would, probably for me, go beyond simply Twitter because I think Twitter is quite limited from a political point of view and if it could include Facebook and some of the other new social media that's in (…) that I don't know anything about, that would be brilliant. I may be asking for the impossible!

**I: Oh no, no, not at all, I think this is the way it should work; users should ask for everything in the world and a computing scientist should deliver as best they can. Because if you leave it to us we'll give you something that, for us, is very interesting but it's hardly usable so…**

M: I am right when I say that getting hold of Facebook accounts is difficult, isn't it?

**I: Yes, yes. The good news is I've developed a very prototype tool at the moment which can actually capture Facebook data but you are absolutely right in terms of the limitations because again, you have to find a public profile.**

M: We would find it useful, yes, to see what we got out of it would be interesting.

**I: I think there is a limit to how much you can pull…**

M: Yes, but if you could find a way of doing that, that would be brilliant. We have thought about perhaps going back and trying to do it through Twitonomy.

**I: There's a technique called 'argument mining' which I've been looking at recently where people are able to extract the type of arguments people are making for a particular topic. Would that be something..?**

M: Yes, that sounds really interesting. That would certainly take it further than we've done it so far but it's something I'd be very interested in.

**I: And there's one other thing which I think someone mentioned to me yesterday and I'd like to hear your opinion on it: if, say, you've identified, I think the only way I can describe this is with an example, so let's say you have a load of Tweets about the referendum and you have arguments in favour for independence and these are centred upon one particular argument regarding the economy and you have like a polarity of people, positive comments about it, almost like a Likert scale so you know these people that are making the same comment about the economy but some people are like 'Yes, yes, this is brilliant, we should go for independence', other people are in the middle and then…would that be something that you'd be interested in?**

M: Yes, yes, that would be very useful.

**I: I just have to come up with a way of defining it like a polarity Likert open coding, something like that!**

M: Yes, that would be useful if you could do it.

**I: Well you've not only answered all my questions, I think you've given us a very precise and very useful requirement spec so I'll stop this now.**

END