#  Interview: Participant 4

# I: So the first question is: Could you please tell me your discipline, your research field and what your current role is?

F: What is my discipline? My home discipline is probably Geography,

# I: Research field.

F: My research field for this would form into media and communications research, it would fall into the area of research about researchers but it's about the perception of risk and ethics and data use and re-use. So the whole thing about big and open data, the perception of risk of data**.**

# I: I think in that you kind of answer the next question which was: Can you give a brief overview of your research? If there's anything you want to add, you kind of (*gave it*) there.

F: So the other thing I've been doing as part of this which would fall into is about the field of best practice and getting them to be the researchers. Not overly analysing it but kind of doing what are the research terms, doing research, what is a really simple analysis of it, so I've been doing that as well which is kind of not as robust but it will probably get written up into research about research but it also could be a bunch of littleprojects and then also there's another project that I'll be involved. So kind of other bits of research that I'm doing or things that look like research even if it's not the full origination, writing up, dissemination thing, but I am presenting on the ethics risk stuff in July**.**

# I: How do you currently use social media in your research/

F: So there's a couple of things I use it for; I use it to source data, so searching for it. I use it as provocation when I'm speaking to people, I use it as an illustration of big data, either when I'm talking to people or engagement but also I use it as a recruiting tool,

# I: So what social media do you use?

F: Personally? So, me as a researcher, individually I have Twitter and Facebook, Instagram, Pinterest, which I started using early on and I tend to use – tend to be more personal, my Twitter is professional and personal, Pinterest, I have some professional stuff on there that is open, not that I promote it, so I'll kind of like save good images or save good papers with an image in and use it as my pin-board and (Ednote) in the same way and share that. But then I'm responsible for the kind of organisational Twitter account, so we have Twitter, we have a static Facebook page and we have online media. We did talk about doing other stuff like blogs but…

# I: What about for when you are doing your actual research?

F: Twitter. Twitter for the simple fact that you can do a really low-tech (dumper search, turn in to the Twitter's feed), copy and paste the information and dump it into *(Worvel*). So I have done some stuff where I've used a small *(API)* or got someone else to write and I've played with it but I've never actually had the time to kind of sort it out when I'm like, 'I've done everything properly' so there's some stuff I would really like to do, the longitudinal stuff over the summer where I would like to do something more sophisticated like being able to do multiple search terms or have a live thing running or look back at the section in time that my low-tech, not having to think too much, not having to programme – write any code so I'll do that and I'll also use the hashtag aggregators and things so use very, very low-tech dump stuff on the search bar, play around with the settings so it gives me what I want it to give and…

# I: You've kind of mentioned a couple of them but was there any specific tools you used to access social media data? Or with your social media data?

F: Tools, I've used some of the – I've used the Twitter API and cutted and pasted the code. But seriously low-tech: use the Twitter interface, dump the search and get it out. And I guess what I'd ideally like is something that is a cross between writing the API kind of code, where it says 'you can do this, this and this', but 'here's how to do this really simple thing', and I can get it to do that using API but I'd like something…that was a cross between the two where I didn't have to go and teach myself.

# I: How do you store your data once you've collected it/

F: In an Excel spreadsheet or a word document. I've got some nice Excel spreadsheets, I've got some word documents and I've got some nice pretty pictures.

# I: So basically, how do you analyse it?

F: So how I analyse it is because I haven't done this really in-depth stuff recently, I've just kind of used – done some textual analysis and done the frequency analysis and got Wordit to do it, so literally dumped it in this great big huge text file, got it to get rid of stuff and then, just telling the story – that really simple, 'here's really basic sentiment *(artist*), bigger words, here are the people's ideas that are coming up. ' So that's all I need. And it's really valid – you get a computer programme to count it and make it into a pretty picture for you. But it's provocation and using social media as a tool for provocation and then doing the research on the results. It's enough. What I would love to be able to do is do some more followers and some of the stuff that the useful blog does for the networks and stuff that Gephi can do. So that's my plan.

# I: Is there anything else you can't do?

F: Map things when people have turned their geo-location off, which I'm pot calling kettle black. Grouping by some kind of things that the data that’s in there – when there's things that would be really useful for me, kind of thinking about provenance and trust stuff, would be really good if it grouped the users for me. So you can really easily see is this someone who posts every day, is this someone who posted multiple times a day, is this someone who posts once a year. Is it currently an active account, is it not an active account… so it does some of that; what kind of users have you got the information coming from, it kind of dumps them into buckets for you. Because at the moment that's something that unless you go through with a pen and paper, it's just doing that *(stuff)* on analysis, you've got to work out what's their frequency of posting. But if you've just grabbed the Tweets and the Tweet feeds that mention the key names you don't have the rest of their Twitter profile and it would be really useful if it was not like classical but kind of – 'these people post hardly ever, these people post twice a week and these people post at least once a day' and then it says, 'ok, here's the groupings already done for you.' Without you having to go and look at their profile. So I guess – I'm not sure if it's possible or not. Just to give you an idea, because if you are taking opinions and looking at people's perception of risk and how much the Tweet about it, it gives – it does one of the groupings earlier and one…

# I: Would you want to know from that, because obviously if you are collecting data on a specific subject, that person may Tweet lots but only on that subject and they meet Tweet lots about lots of things. Would that be…?

F: Yes, that would be the – I guess it's the profile (*thing*) of it, which I'm not sure – it would be quite hard to do but kind of say 'this is someone who generally Tweets about x,y and z' or some way of flagging that if you are using a search term that's also in their profile description that you then say they might be – a really easy way that you know that they are interested in the subject because they mention it in their profile but also it's the search term you've looked for.

# I: Do you currently add any other data when you are doing this?

F: I add the time, I add the location. I add the location to where the people I asked for it was. I add, actually, thinking about it – with the (*00:12:12*) perception stuff, I've added any media attention to social media that was in the previous two weeks, that kind of – are there things that may have influenced it? So, it's more qualitative data that – it's not a number that gives you 'ok, so the prevalence of social media in the news was .2%' or something but it was just 'here's the information', where it was collected, here's the people, that kind of contextual stuff that is not so much – I'm adding in, there were fourteen traffic jams in that city, *that(s)* quantitative stuff but that qualitative contextual stuff that's really useful.

# I: Where do you get that qualitative data from?

F: there's a search database that you just dump the terms in and I have that to do it for me. I geo-locate it just because I have to for the open-data stuff. That's about what I add in. And my reflective stuff, so my notes from there which are sometimes a bit data-(*ry)* and sometimes not.

# I: What would you like to be able to do with social media data and other data so, with your notes and all this?

F: What would I like to do? You see, in some ways my low-tech way works at the moment. To be honest, what I would really like to do, which is a bit complicated, is this kind of like five or six word analysis processes. What I'd love to do is be able to come and say 'here's my social media dataset and here's the five different word (*clouds*)' without having to do it each time. And then be able to just go – 'I like that one', and I think there's some really powerful stuff I'd like to do with mapping but because of the geo-location issue and the fact that a lot of people don't geo-locate their Tweets, if we were in the States it would be a different matter. What I would really like to do is when you've got a Tweet that mentions a link, if it's an article or something it just automatically puts them in a database and does that link-data thing for you so you don't have to go through every one, click on it, save it and put it into the data-set. But just kind of automatically links to it and then there's twenty other things so if the social media comments or other things where it's mentioned in. So you know on web-science where you've got that thing where you can see stuff, this paper referred to and all that – so you can go backwards and forwards, in some ways it would be really good if you could do that with links and stuff and particularly with news articles where increasingly they are mentioning – Twitter mentioning something and doing that timeline rather than you having to physically do it each time.

# I: If you had a dream research tool for helping with social media research, what would it do?

F: It would be – it would look like a really simple *(00:15:25)* and it would say: Here are my search terms. You can dump them in. I would like Tweets between this date and this date, I would like them in this format. And I'd like you to also indicate frequency of Tweeting – are they an expert. The first pass of – pick stuff up. So some way that you could say – (you did say dream), so a lot of the stuff I do, I would love to be able to say, 'here's my search And download those as well because I think where we are missing – and where I'm conscious I'm missing out on the (*00:16:20*) is, if it's not in English it's really hard to pick up on. If we're using Twitter as a tool to do a barometer of a social group or *(00:17:09)* even though it's not, you've got to remember that that's not always English. Or something that does – ok, here are the people from something they say, we can tell you their age bracket. And you can just say, 'ok, I want age or nationality or location', that would take that first sort out. And take the spam out. So something that would get rid of that first couple of days of (…) sorting, 'that's crap, that's not', but equally get you more data or nudge you to ' have you thought about that?'

# I: So you quite like the idea of automatic context *(00:17:47)…*

F: Yeah and also, I'm not sure if it would be possible, is that I think in five years time where it would be useful is it searches the open data things and resources and says 'by the way, this seems to chime in with these open data sets' so in the same way that geosciences (…), you can just plug- in and goBut because of the open – it would be really nice if it said, 'we'll also search these open data archives for you' because obviously they can't hold the Tweets but it's really hard to find the *literature (in)* the space and it also it might be that *(00:18:47)* these search times and you can only speak to the researcher. You did say dream. But I think it's link-data stuff, like linking to where there's open-data sets or linking in to – 'would you also like to search these websites', 'would you also like to search for Twitter data search term'? And then it does it all in one place for you.

# I: So you like the idea of being able to find other data sources that are similar to what you've got and add the context?

F: Yes, or just because it's such a nightmare. And we've talked about people going 'we use social media data'. Yes, that's very nice of you. How did you get it? Did you search for it, did you actually check it or did you just graph it? Did you actually check that people were Tweeting about this who are actually in the area. Stuff that allows you to make that decision without it being called 'airy' provenance and getting into *(tautologies)* and (*W3 seat*) stuff and all that stuff that I know exists and I nod and can generally talk about for five minutes.

# I: You wanted to do that but it's kind of like 'under the (*car bonnet 00:19:55)'*?

F: Yeah, but it would take a button that says 'why don't we ask you this?' But yeah, because I think some people freak out and try and make it more complicated because the search tools in Twitter are actually pretty powerful and sometimes just doing that quick and dirty…search in a search bar is fine. Because, to be honest, the issue of this stuff and big data, is if you get too much data back then it's just helping you manage and classify it and do that first. First cut of data, is the important thing I think.

# I: Ok, so anything else you would like to add?

F: I would really like a way where it's easier to search Facebook which doesn't involve paying loads of money. Because more people in the use Facebook than Twitter and it's got a broader demographic. And something that would let me – when I've been thinking about this I've always been thinking about Twitter just because it's a lot easier to use but something that kind of lets you run something on multiple platforms at the same time, that if it says – say, for instance, you can run a hashtag search on Instagram or Facebook, so it could say, 'we'll convert these into hashtags and run them on Facebook and see what happens', but…the data is as such that means we can't do it. And that searching across multiple platforms I think is quite powerful. Yes, something that makes it easier to search the pictorial things from the descriptions because I think particularly with – Instagram has seen a huge growth at the moment and there's – so many people I know who are vaguely on Facebook, not on Twitter, but they are on Instagram and I think it's because Instagram has the – you can protect your account on Instagram and (so can you) protect your account on Twitter but in some ways it's a lot…I don't know, I think because of mobile phone coverage you can just take a picture, send it and it's got more archival value to you as a picture than a piece of text.

# I: But it's much harder to analyse.

F: It's really hard to analyse. But, increasingly people are putting hashtag and stuff in the descriptions but it's how easy it is to describe that stuff. Because – and also the fact is that unless you go into a random account while you search the hashtag, you need an Instagram account in order to search that stuff and I think Instagram I think is getting more important than blogs, blogs are fab because they are really easy to search but they are declining. And the number of people I know who have been quite high profile bloggers, lifestyle ones and and made lots of money from it, they are checking out, they are not updating as much because they can update on Twitter or Facebook or Instagram, update more regularly but not as much time and investment into writing a nice five hundred, eight hundred word piece of text. And I think we are missing out, it's kind of like we've got to assume that social media doesn't take the barometer of the nation. Yeah, we can use it to take the barometer of the nation on elections, for far more people than if you were surveying but you've got to take into account the population more, so I think if you can do it across more multiple platforms you get a more representative…because of the varying thing. One of the problems we have is that all the serious money into who uses what in what country, data and the protection means that they probably couldn’t do it in the same way. And it's changing. So I think that is…that would be the kind of…I think it's one of the biggest challenge in social media is the growth of the picture sites and I don't think there's going to be anything new, I think Google+ showed that it was a disaster, blogs that I think – as much as I love blogs but my reader is not as full as it used to be. So commercial blogs, and things like that are growing but individuals aren't. And a lot of the time we're looking at individual *(theft)* so I'm not sure, I could be wrong, but I don't think there's going to be another Twitter or Instagram or Facebook or Blogger. I mean, Vine has been more little videos and those havebeen more popular than I thought they might be, I don't think they are…

# I: Everybody has a bit of everything now.

F: I think it's – it would be kind of good, I could imagine something that was tabbed and you can pick your social media site that you want to – you put the search terms in and then it says, 'ok, we've run your search here, here's the caveat but here's the information on these sites, if you want it. But if you want more information, here's another tool somewhere else.' Or 'here's the link of a company you have to pay twenty grand to, to do the proper search' but that kind of one-stop shop so you are aware. But also something that kind of just combines the – has behind it the 'ok, so if you are a friend of this person, in this case you need to record this information now. Here's a nice text-file for you', which would be cool. Actually that would be quite cool, 'here's the search you ran for open-data for coding purposes. Here's your search terms, here's what you ran, here's how many results you got. Drop it in an open-data archive. Tick.' That would make my life a lot easier. And some (*safe-ing),* you could go back and run it again six months later. 'Or would you like me to email you the data of these results? We'll run it again and we'll do it in a month and we'll send you the new stuff.'

# I: That's really helpful.

F: Kind of automate the stuff so you don't have to run it yourself, it gives you the tools and gives you more than it being black boxes. A black box where you can open up the tabs or open up different lids of the process, if you need them or to guide you. Because I think a lot of people might assume that people want a total black box, but no, you need to be able to prove the data and use the same data-set for multiple different things.

# I: You want to be able to explore it to understand it.

F: Yeah, interrogate it**.**

**END**