Tracking and trawling: Theorising ‘participants’ and ‘data’ in qualitative e-research

Katrina Pritchard* and Rebecca Whiting

Department of Organizational Psychology, Birkbeck, University of London, Malet Street, London, WC1E 7HX

*corresponding author: k.pritchard@bbk.ac.uk

Developmental Paper submitted to Research Methodology Track

Word Count (excluding tables and references): 1446

Summary:
We offer a reflexive examination of the challenges of reconfiguring two core concepts in (here, qualitative) research: data and participants; within the context of an e-research project. We use two related but different techniques deployed during our data collection: trawling and tracking; to highlight practical issues but also to draw attention to broader methodological concerns.

Note to organisers: This developmental paper relates to work in progress as part of our research project “The discursive construction of age at work” (funded by the Richard Benjamin Trust, Early Career Award Number 1103, also see our project blog http://ageatwork.wordpress.com/). As at submission, we have completed the data collection phase of the project but by the conference we will be able to extend the discussion presented below based on the data analysis conducted between February and August 2012. At conference we will be able to illustrate this discussion with ‘live’ examples from our data set.

Developmental Paper:

We offer a reflexive examination of the challenges of reconfiguring two core research concepts: data and participants within a qualitative e-research project. Using two related but different techniques deployed during our data collection, tracking and trawling, we highlight practical issues and draw attention to broader methodological concerns.

We use ‘e-research’ as an overarching term that incorporates varied online and internet-based approaches (Fielding et al., 2008) covering a “wide range of activities” (BPS, 2007, p. 1) from a broad range of research orientations. Developing alongside the proliferation of the internet itself (Evans et al., 2008), researchers are exploring internet research tools and techniques which have “often been research not just about the Internet but also on it and through it and constituted within it” (Hine, 2005, p. 205).

Our study examines the language of age at work. Our e-research approach is qualitative and discursive. We consider the overall construction of age discourses, unpacking particular conceptions (such as generations) and examining the relationships between them.
interesting analogy here is that e-research could be described as the ‘young upstart’ at the research methodology table, given the broader media association between youth and technology (Bennett et al., 2008). Data collection has involved technical tools to gather stories, accounts and discussions about age at work published on the internet, for example, on-line news media and blogs.

Despite the growing interest in e-research, it is important to bear in mind that it has not been universally welcomed. Travers (2009, p. 172) comments “it is hard to see how new technologies add much that is really new to qualitative research....more worryingly they flatter us into thinking that, because the methods are new or innovative, no further thought about methodological issues or how one analyses the data is required”. Our approach is to address these methodological issues head on and use our e-research as the basis for examining both pre-existing and emergent assumptions we have as researchers (here in respect to participants and data). In doing so, we seek to go beyond simply making the rhetorical choice of ‘internet research’ “defined to evoke the qualities that the Internet held as a cultural artefact as the cutting edge technology of the moment” (Hine 2005, p. 245).

Key texts (Fielding et al., 2008) and chapters (Evans et al., 2008) already examine various aspects of e-research with ethics emerging as a particular focus of discussion (Ess, 2009). Whilst critical in informing our own research journey to date, we specifically seek to examine challenges that have prompted us to review our action plan and caused us broader methodological angst (Bryman, 2008). Our aim is to use our project to offer a reflexive examination of the challenges of reconfiguring expectations and understandings of two core notions in research: data and participants. Existing research appears to have avoided a direct confrontation with these issues. Researchers may discuss ‘material’ or ‘texts’ and avoid the notion of data altogether (e.g. Spicer, 2005) while participants are often depicted in relation to the broader technological context under investigation rather than in respect to the research process (e.g. Pablo and Hardy, 2009).

**Tracking and Trawling**

Through development in our pilot, our e-research project has incorporated two types of data collection activities: ‘tracking’ and ‘trawling’ (see Table 1 below).

<table>
<thead>
<tr>
<th><strong>Approach</strong></th>
<th><strong>Description</strong></th>
<th><strong>Tools deployed</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tracking</td>
<td>Uses a variety of means to follow particular people or groups of interest due to their engagement with a specific topic of relevance to the research.</td>
<td>Twilert Newsletter, press release and website sign-up Robot (website change detection)</td>
</tr>
<tr>
<td>Trawling</td>
<td>Uses specific key word search tools to provide daily alerts of potentially relevant material across a variety of source types (e.g. blogs, news sites, twitter)</td>
<td>Google Alerts Twilert Nexis</td>
</tr>
</tbody>
</table>

As an example, within ‘trawling’ we used Google alerts (a free Google tool) to run daily searches on key topics. The example below shows the results of one such search (for the terms ‘generation’ and ‘work’). These terms were selected after piloting and reviewing ‘hit’ rates for various combinations but, as shown, not all returns are relevant. Some ‘hits’ may be
inaccessible (e.g. if a subscription is required), false or malevolent (increasing security settings was an early lesson learned) or deemed irrelevant on closer investigation.

Examples from Google Alert ‘generation’ and ‘work’ 12th January 2011

Workers warned not to share Facebook secrets with their colleagues (active link to content)
Daily Mail (source)
Four-fifths of young people have made Facebook friends at work, and the average Generation Y-er has 16 colleagues as friends ...
See all stories on this topic »

Toyota Reveals Next-Generation Hybrid Concept (Slideshow) (active link to content)
Wall Street Journal (blog) (source)
The system includes numerous touch-screens and is meant to work intuitively in the same way a smartphone does...
See all stories on this topic »

Retirees ‘to challenge work-shy youngsters’ (active link to content)
Money News (source)
Furthermore, 67 per cent believe the younger generation moan too much about their work-life balance....
See all stories on this topic »

‘Tracking’ is potentially more focused and was introduced to follow specific individuals, or more usually groups or organizations, actively discussing age at work using the Web as a forum through which to engage others in debate. They are therefore ‘participants’ in the arena of our research. Whereas when trawling the ‘key word’ is critical, when tracking the name or identity is the basis for the search protocol. These names emerged from both prior knowledge and trawling activity; they included HR professional groups, consultancies, campaign groups and government agencies.

Participants and Data

In line with the social constructionist thinking underpinning our research, we began fully open to the notion of the ongoing construction of research objects and relationships. This reflects the belief that “the research enterprise itself is a cluster of negotiated relationships” (Hoskins and Sholtz, 2005, p. 100) and that, as Alvesson and Deetz suggest, “the researcher constructs the data to interpret” (2000, p. 21).

However, as data collection got underway, we were challenged in our previous understandings of these notions, developed primarily through engagement with more traditional qualitative research approaches.

Tables 2 and 3 below map our methodological concerns as experienced to date through our engagement with tracking and trawling respectively. They are in the early stages of development and we anticipate they will be refined as data analysis commences.
Table 2: Participants

<table>
<thead>
<tr>
<th>Tracking</th>
<th>Some level of prior identification of a ‘name’ is necessary to be effective.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>An active decision to ‘track’ has to be made – usually done via periodic discussion between the two authors.</td>
</tr>
<tr>
<td></td>
<td>The decision to track implies the significance of the participant to the research, we have a vested interest.</td>
</tr>
<tr>
<td></td>
<td>The participant is not anonymous but individuals remain largely unnamed (though specific communications may involve identification of ‘key players’).</td>
</tr>
<tr>
<td></td>
<td>Productive tracking generates a sense of achievement, and may impact our view of the participant – and vice versa.</td>
</tr>
<tr>
<td></td>
<td>The participant is unaware they are being tracked. However, they actively seek an audience for their message and invite us to sign up; we receive regular newsletters and press releases. They have our email address. Who is tracking who?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trawling</th>
<th>Multiple potential participants can be raised by a key word search; participants may have an active or passive role in the data gathered.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>How active are the characters in the news stories reported? They often seem passive compared to the comment-posters (some of whom are so prolific their comments can far exceed the word count of the original article). Who are the comment-posters talking to?</td>
</tr>
<tr>
<td></td>
<td>A range of peripheral potential participants is apparent each time a web-search is followed. Who are the advertisers surrounding discussions of age at work? Who are the models posing in the ‘library photographs’? Some are identified by name but the degree to which this is an active choice is unclear. A staff reporter on the other hand may not have the organizational right to a by-line.</td>
</tr>
<tr>
<td></td>
<td>Comment-posters and other may use a range of ‘identities’ so tracing an individual in ‘real life’ is often virtually impossible. Should we concern ourselves with trolls, trolling and other such aspects of internet-based communication?</td>
</tr>
</tbody>
</table>

As with other forms of qualitative research, the decision to identity and label ‘participants’, ‘respondents’ ‘subjects’ or even ‘actants’ (Austrin and Farnsworth, 2005) is complex and worthy of debate. In this research approach, our ‘relationships’ are technologically mediated such that this becomes both practically and methodologically difficult territory to negotiate.

Issues associated with data are similarly complex. In one sense, e-research opens up access allowing the researcher to disregard physical boundaries but requires negotiation of technological ones instead. (We flinch in memory of malicious software downloads...
A particular struggle however has been re-negotiating the concepts of ‘primary’ and ‘secondary’ data. Primary data in e-research is usually defined as that which is obtained directly and actively from participants through direct questioning (e.g. via an on-line interview or questions posed by a researcher to prompt responses from participants within a discussion forum). But does the broad delineation of primary and secondary make sense in terms of our approach to tracking and trawling?

### Table 3: Data

| Tracking | Named organizations or groups often have an established web presence that is the basis for our tracking. This may be a blog or more formal website, depending on the size and financial status of the group.

Different web presences offer different types of data, some of which may be interactive either asynchronously (e.g. discussion forums which may be hosted by a member of the organization) or synchronously (e.g. via webinars).

Most offer a form of ‘sign up’ to receive information generated by the group, often in the format of regular newsletters. These are constructed by the group for a generic (and probably largely unknown) audience. The group is talking to us; the emails are personally addressed to us, in a fashion (‘Dear ageatwork@gmail.com,...’). They are often written in an informal and conversational style, inviting us to get more involved in their activities. While not constructed for our research, these communications are therefore attempting to engage in conversation, a feature more like the typical definition of primary data.

| Trawling | Much of the ‘data’ collected via trawling activities is posted on the web rather than sent to us directly. This may include press releases and news stories positioned as sharing information or views with an unspecified audience (which just happens to include us with our research interests).

Other sources are more personal: blogs or commercial material such as advertisements for training courses. In each the notion of our interaction with the author is subtly different, depending on their approach (including the degree of interaction encouraged) and our reading. The data is there to be collected but we could also engage in a conversation.

While for ethical reasons we are not gathering data from subscription discussion boards or chat rooms, many sites offer a ‘comment’ function. Usually commenters must be registered to post their views but the comments are displayed alongside the original article for all to read.

Commenting is a complex area for our research. Above we considered this with respect to participants but it also raises issues of understanding data. Comments often develop into threads and conversations, albeit virtually and (usually) asynchronously, in which aspects of areas of the original article – and beyond - are developed and examined. Individuals share experiences and ‘voice’ opinions. This is not static data, but develops and grows often in quite complex ways. There may be many tangents and also arguments (the latter about opinions expressed but also spelling and grammar).
Secondary data is usually assumed to exist and await collection by the researcher. However in our experience the issue of researcher involvement is challenged in both tracking and trawling. While we may not be actively asking questions (as in an interview), the notion of a static text awaiting collection does not fit with our experiences to date.

**Conclusion**

We use our qualitative e-research to offer a reflexive examination of two core research concepts: *data* and *participants*. Our data collection techniques, *trawling* and *tracking* highlight, practical and methodological issues that challenge definitions of primary and secondary data and the notion of who is a participant in research. Notions of data and participants have yet to be fully explored within e-research yet are central to our understandings of research issues (including ethics) more broadly. Our experience to date suggests that e-researchers be wary of writing out *data* and *participants* from their research vocabulary.

**References**