

Realities Toolkit #10

Using Blog Analysis

Helene Snee, Sociology, University of Manchester

July 2010

1. Introduction

Blogs are a relatively new form of internet communication and as a result, little has been written so far on how they might be used in social research. For my PhD thesis, I used blogs alongside interviews to explore young people's representations of their gap year experiences. This toolkit outlines the methods I employed in my research and suggests the sorts of questions that can be answered using blog analysis alongside the issues that might arise.

What are blogs?

Blogs are a form of internet communication in which the author writes dated entries that appear in reverse order (i.e. earliest first) that can link to other webpages and that usually allow readers to comment. There are lots of different types of blogs. Originally 'web logs' were collections of links, but increasingly the term has come to be applied to online journals. They are often quite personal forms of writing. The blogs that I looked at for my research were a bit like travel diaries, as they were a record of a gap year, but they were also a way of communicating events to other people.

2. Why use blogs?

The aim of my research was to understand how young people told stories about their gap years and the resources they drew upon to describe their experiences. Blogs were a good way of examining these stories because they offered *narratives* of the gap year experience that could be analysed.

Blogs, along with other internet research methods, have a number of practical advantages. They are easy to access and the sample can be extended to those from a wide geographical area. For my research, I could include young people who were still overseas or located in different places all over the UK. This meant that I could examine more stories using blog analysis than might be possible with 'offline' methods.

As with any project, practical advantages can be useful, but it is important to make sure that the methods employed match the research aims and objectives. In the gap year study, I wanted to examine what the young people felt was important about their

experiences. The data was not driven by what I wanted to know but what the bloggers wanted to share.

Using blogs like this wouldn't be suitable for all types of research. For example, using unsolicited blogs would not have been useful if I wanted to know why the young people had taken time out in particular countries, because they might not have blogged about their reasons.

Another key factor to bear in mind is that it is not always possible to verify the identity of a blog author. For my research, it was not crucial to the research questions, because I was focussing on the stories. But if I had wanted to make claims about the link between the class background of the bloggers and their stories, for example, then blogs would not have been as appropriate. Even so, I did have to exclude one of the blogs I initially included in my study, because after contacting the author they did not meet the sample criteria (see Section 3). It might not always be possible to check who the author of a blog is, so it is important to think about the sorts of conclusions you can draw from your analysis.

It is also important to consider who you can access using blog research. Not everyone uses the internet and not everyone who does will write a blog. Even though younger people might be more likely to blog, I had to think about who might be excluded from my research. For example, are people who are less affluent less likely to blog? This meant that my research could not claim to represent all gap year experiences. However, what I could do was examine the types of stories that are told about the gap year.

3. Searching for blogs

The first step was to construct a sample of blogs that matched the aims and objectives of the research:

- I conducted an internet search for the phrase 'gap year', initially on two blog search engines: Google Blog Search and Technorati.
- These searches suggested some common websites that were in use for gap year blogs: Myspace, LiveJournal and Globenotes, so I also searched for the same phrase on these sites.
- This generated a large number of results, which included blogs that discussed gap years, but were not necessarily written by a gapper themselves. In addition, the results also included a number of 'spam' blogs (blogs that are set up as advertisements or to generate traffic to a particular website; see Li and Walejko 2008).
- Each search result was recorded in a table, which contained:
 - a reference number;
 - the gender, age and location of the author (if possible);
 - additional comments, for example the context of the reference to the term 'gap year';
 - whether the blog was suitable to include in the sample.

Sample of table used to record blogs found during searches

Ref	Link	Author Details	Comments	Inc. in sample?
1	[Link to website]	18 year old US male taking gap year between high school and college	Seems to be a 'high-flyer' – set up his own company at 12 and now has a book deal (info from link to another blog)	×
2	[Link to website]	50 year old female “taking time out”	Not travelling (recovering from hospital treatment)	×
3	[Link to website]	Company blog – advertising for xxx (gap year provider – charitable organisation).	N/A	×
4	[Link to website]	19 year old male – UK teaching / travelling around Thailand	Seems to have initially planned to go to more countries – need to read in depth.	✓

- Blogs were included in the sample if, on first reading, they met the following criteria:
 - the author was from the UK;
 - they took their year out overseas;
 - they took their gap year between secondary school and university;
 - they contained more than one or two posts (it seemed that some young people set up blogs with the intention of recording their gap year but then did not do it!)
- I then read these blogs more closely to double check if the blog met the above criteria.
- I also wanted to ensure that the final sample was balanced in terms of gender, and the type of gap year taken (i.e. there was an even spread of people who just travelled; who just undertook volunteering or work placements; and who did both of these activities).
- A total of 700 blogs were collected and recorded, which led to a final sample of 39 blogs.

Things to consider when searching for blogs:

- What criteria are you going to use? What is central to your research?
- Are there particular websites that seem to be used by your research population?
- Devise a way of keeping a record of your searches as you go.
- If you have a specific topic or type of blog that you are trying to find, be prepared to sift through lots of results.

4. Managing blog data

Once the sample had been finalised, the next step was to consider how to deal with the blogs when I was analysing them. Blogs are multimedia texts, and alongside the written elements, the gap year blogs I sampled also contained: photos; video clips; hyperlinks to other websites; advertisements; and audio files. I decided to focus on the written textual aspects of the blogs. Not all of them contained all of these multimedia objects, so it would have been difficult to compare them. Also, I was not interested in blogs as a

medium as such, but the ways in which young people told stories about their experiences. Furthermore, I had a lot of data, and needed to find an easy way to manage it. Taking all of this into consideration, I decided to convert the blogs into text files so that they could be easily imported into the version of the data analysis software that I was using (Atlas.ti).

Converting webpages to text (in Internet Explorer)

File -> Save As

Then select: Save as type -> Text File (*.txt).

Each journal was saved as a single text file in chronological order (earliest first, reversing the order of the posts as they appeared online) and then imported into Atlas.ti. This took a long time, but it was worth it in order to be able to code the data.

Different data analysis programs offer different ways of dealing with websites. Whole web pages can be saved, opened in Microsoft Word and then saved to be imported into NVivo, for example. Updates to data analysis software, like the latest version of NVivo, are increasingly able to deal with multimedia data. If you are interested in these elements of blogs, then it is worthwhile exploring the functions of such programs, and also the literature on visual methods. It very much depends on exactly what you are trying to explore with your analysis.

Another thing to consider is that blogs can be constantly changing. Some blogs may be 'finished' and never be updated, but often they are ongoing, and comments can be added to 'old' posts. I therefore had to take a practical decision to save the blogs on a specific date and time.

Even just focusing on the written text for my project was time-consuming. The total number of words I collected was over 1 million. Online methods provide instant access to lots of data but this has to be balanced with making sure that you have a manageable sample. I also had to cut some of the blogs down and exclude any posts written before / after the gap year itself that did not include any reference to the bloggers' time out to make sure I was focussing on the most relevant material.

It is advisable to read the blog online even if you decide to import it into a data analysis package. This helps you gain an understanding of the blog in context and you can see it how it is meant to be viewed, even if you then end up analysing particular elements of it in analysis software.

Points to consider:

- Are you going to focus on the written text only or is it important to your analysis to include multimedia objects like photographs and video?
- Explore the options in data analysis software for dealing with webpages.
- Try to be practical in terms of the amount of data you collect – it is easy to collect lots and lots of data using internet searches but this can be overwhelming

5. The ethics of blog research

Like any research project, blog analysis should be conducted according ethical guidelines. However, it can be difficult to apply these guidelines to online research. I had two overlapping issues that I needed to think about in my gap year project:

- *Are blogs public or private?*

Although the blogs I sampled were feely available on the internet, the difference between public and private is 'blurred' online. People might write something without thinking about how anyone with an internet connection can access it – in other words, they might have a false sense of privacy. It is also important to think about who is the intended audience of a blog – probably not a researcher! In addition, the gap year blogs I sampled also contained some personal information, like the young person's real name and email address.

- *Should I protect the participants or recognise their rights as authors?*

Given that there was some personal information in the blogs, it could be argued that the bloggers should be treated in the same way that you might treat interview participants, for example. This would mean:

- gaining informed consent to take part in the research
- anonymising their details and keeping them confidential. This is difficult with internet data because it can be easily found using a search engine.

The problem with this is that bloggers might actually want to be recognised for the work that has gone into their blogs, so they should instead be treated the same as someone who has written a newspaper article, for example. Published material like this does not require informed consent and so it would be important to recognise the author by providing a link to their blog when writing up the research.

Resolving ethical issues:

- Consult the literature on internet research ethics (see further reading). What sort of advice does this offer?
- The Association of Internet Researchers (AoIR) has published a set of ethical guidelines that can help with establishing privacy.
- Is there any research that has been conducted on similar issues using blogs or other online methods? How did these authors resolve any ethical dilemmas?
- What is the technical privacy of the blogs – are they open access?
- What is the perceived privacy of the blogs – do they seem very private, or do the authors directly speak to people they might not know?
- What is the content of the blogs – for example, are they on sensitive topics?
- What is the context of the blogs – for example, have they been written by vulnerable people?
- Are you quoting directly from the blogs? How easy would it be to identify individuals?

There was a strong case for considering the gap year blogs as public data and therefore that informed consent was not required. They were publicly accessible and there was a sense that they were perceived as being so. I established this as many of the blogs contained biographies of the author (so that unknown readers would know something about them) and when I conducted interviews with some of the bloggers, they were aware that the blogs were not totally private.

Despite this, I felt that the fact that there was personal information contained in the blogs meant that there should be some level of disguise and that protecting the young people's identity was more important than recognising them as authors. My final approach was to quote from the blogs but change any identifiable details. These were 'personal but not private' accounts (Hookway 2008: 105). These decisions seemed the most appropriate for this particular project, but it is very important to think about your specific research context and how these issues might apply to your own study.

References and further reading

Ess, C. and the AoIR ethics working committee (2002) *Ethical decision-making and Internet Research: Recommendations from the AoIR Ethics Working Committee*.

Available from: <http://aoir.org/reports/ethics.pdf>

A very useful starting point for identifying the key ethical issues in internet research

Exploring online research methods in a virtual training environment (2006) Available from <http://www.geog.le.ac.uk/orm/site/home.htm>

A valuable resource that provides useful bibliographies, training materials and case studies of online research (including blog analysis).

Hookway, N. (2008) "'Entering the blogosphere": some strategies for using blogs in social research' *Qualitative Research*, 8(1) 91-113.

A case study of some key issues associated with using blogs in qualitative research, including searching for blogs, building your sample and ethical dilemmas.

Li, D., and Walejko, G. (2008) 'Splogs and Abandoned Blogs: The Perils of Sampling Bloggers and Their Blogs', *Information, Communication and Society* 11(2): 279-296.

Discusses the problems that can arise when sampling blogs and reviews different sampling strategies.

More publications from this study

Snee, H. (forthcoming) 'Youth Research in Web 2.0: A Case Study in Blog Analysis' in S. Heath and C. Walker (eds) *Innovations in Researching Youth* Basingstoke: Palgrave.

Feedback welcome! If you have any comments on this toolkit or if you can tell us how you have used it in your research or teaching please do drop us a line at realities@manchester.ac.uk and let us know.



Join our **email newsletter** for updates on our toolkits, events and other activities:
www.manchester.ac.uk/realities/aboutus/newslist



Realities is part of the National Centre for Research Methods, which aims to improve research methods across the UK social science community. We are based in the Morgan Centre for the Study of Relationships and Personal Life at the University of Manchester.

Realities, Morgan Centre, Sociology, University of Manchester, Manchester M13 9PL
+44 (0) 161 275 0265 | realities@manchester.ac.uk | www.manchester.ac.uk/realities