

How 'Public' is Public on Social Media?

Ethical issues in accessing individually identifiable data from YouTube videos, Facebook pages and Twitter feeds

Researchers in many fields of study are becoming increasingly aware of the rich sources information publicly accessible from popular social media venues such as YouTube, Facebook, Twitter, Pinterest (and many others). But is this information truly 'public', in the sense that it is fair game for access, analysis and academic discussion?

There are, in fact, two quite separate issues involved here, from a Human Research Ethics point of view:

Is ethics approval needed at all?

Firstly, do these researchers require ethics approval at all? The answer to this question is quite straightforward – yes, they do! [Chapter 5.1 of the National Statement](#) clarifies the only circumstances in which ethical approval is NOT needed:

Research that can be exempted from review

5.1.22 *Institutions may choose to exempt from ethical review research that:*

- a. *is negligible risk research (as defined in paragraph 2.1.7); and*
- b. *involves the use of existing collections of data or records that contain only non-identifiable data about human beings.*

5.1.23 *Institutions must recognise that in deciding to exempt research from ethical review, they are determining that the research meets the requirements of this National Statement and is ethically acceptable.*

So pulling videos from YouTube, or personal data and opinions from Facebook or Twitter or personal blogs does not fall within the area of exempted research, as this information is unquestionably identifiable. So the answer to the first question is quite simple – researchers in this situation **do** need to submit an ethics application.

How ethical is such research?

Now for the more complex question: is obtaining data in this manner ethical? Or, to put it a different way, does such research involve any risk for the researcher's unsuspecting 'participants'? Here, the answer is anything BUT straightforward, but it's possible to tease out the issues and clarify what matters ...

An excellent article by Michael Zimmer, entitled "But the data is already public": on the ethics of research in Facebook" (Ethics Inf Technol (2010) 12:313-325) summarises the issues thus (my numbering):

"... the emerging challenges of engaging in research within online social network settings. These include (1) challenges to the traditional nature of consent, (2) properly identifying and respecting expectations of privacy on social network sites, (3) developing sufficient strategies for data anonymisation prior to the public release of personal data, and (4) the relative expertise of [HRECs] ...

... future researchers must gain a better understanding of the contextual nature of privacy in these spheres (Nissenbaum 1998, 2004, 2009), recognizing that just because personal information is made

available in some fashion on a social network, does not mean it is fair game for capture and release to all (see, generally, Stutzman 2006; Zimmer 2006; McGeeveran 2007; boyd 2008a).

Similarly, the notion of what constitutes “consent” within the context of divulging personal information in social networking spaces must be further explored, especially in light of this contextual understanding of norms of information flow within specific spheres” (p.323).

The issues, then, according to Zimmer, include both privacy and consent. The author also identifies (p. 322 – my added numbering) “... *four salient dimensions of privacy violations, as organized by Smith et al. (1996)¹ and based on thorough review of privacy literature: (1) the amount of personal information collected, (2) improper access to personal information, (3) unauthorized secondary use of personal information, and (4) errors in personal information”* .

While it is highly unlikely any reputable researcher would deliberately engage in any of these errors, not everyone is sufficiently familiar with the implications of social media to consider just how the people uploading YouTube videos/Facebook pages/Twitter feeds/blog posts would feel about discovering they were unknowingly the subject of a research project. It is, after all, only too easy to upload an opinion on social media which gives entirely the wrong impression to its readers – adding a layer of academic analysis to social content may well make things worse ...

Further useful advice can be found in “Ethical Issues in Qualitative Research on Internet Communities” by Gunther Eysenbach & James E. Till (BMJ. 2001 Nov 10; 323(7321): 1103–1105: available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC59687/>). This helpful article makes the important point that: “*Researchers and [HRECs] must primarily consider whether research is intrusive and has potential for harm, whether the venue is perceived as “private” or “public” space, how confidentiality can be protected, and whether and how informed consent should be obtained”* .

Note that many more recent articles from the US (e.g. Davidson, 2013²; Gelina, Pierce, et al., 2017³) state explicitly that it is not necessary to obtain consent when observing a public social network. This view may well be relevant in the US but the National Statement takes a very different approach to ‘limited disclosure’, stating in the Introduction to NS 2.2: “*Research involving limited disclosure covers a spectrum ... Examples along the spectrum include: observation in public spaces of everyday behaviour ... Depending upon the circumstances of an individual project it may be justifiable to employ an opt-out approach or a waiver of the requirement for consent, rather than seeking explicit consent”* .

Putting the ethics application together

So, if it isn’t possible to obtain ‘informed consent’ from those posting information on social media (and that is often the case), how do researchers go about putting together a sound ethics application to use such material? The authors of the material may well be ‘Internet celebrities’ and thus difficult to contact; or they may be hiding behind pseudonyms or anonymous links; or the researchers may need to access many, many sites to gain cross-sectional data; or ...

¹ Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals’ concerns about organizational practices. *MIS Quarterly*, 20(2), 167–196

² Davidson, B. (2013) ‘The Ethics of Online Observational Research’, Ogilvy CommonHealth WorldWide blog, Jan.8, Accessed 12 May 2017: <https://blog-archive.ogilvyhww.com/2013/01/08/ethics-online-observational-research-2/>

³ Gelinas, L., Pierce, R., Winkler, S.I. Cohen, G., Fernandez Lynch, H. & Bierer, B.E. (2017) Using Social Media as a Research Recruitment Tool: Ethical Issues and Recommendations, *The American Journal of Bioethics*, 17:3, 3-14, DOI: 10.1080/15265161.2016.1276644

The issues to be considered are risk and consent:

- Risk: will the use of the author's social media content create what [NS 2.1](#) defines as 'risk', i.e. the potential harm which may occur + the likely severity of that harm? NS 2.1 discusses risk in some detail, identifying many different types of possible risk (physical, psychological, social, economic, legal, etc.) and researchers hoping to access social media will generally spend some time considering just what risks (if any) are likely to occur as a consequence of their study
- Consent: since, by definition, the author/s of the material will not be consenting to its use, researchers will need to apply for a 'waiver of consent'. [NS 2.3.10](#) outlines the circumstances in which waiver of consent is ethically acceptable (it is also worth noting the provisions of NS 2.3.12 which requires all HRECs to report back to the NHMRC on all projects where waiver of consent was approved).