

# Social media provides a data goldmine for third parties

June 16 2015, by Ansgar Koene

---

The history of human experiments often focuses on biomedical research and the gradual changes in acceptable practice and ethical considerations. But another class of human experiments that has had its own share of controversies is the study of human behaviour.

Internet Mediate Human Behaviour Research (IMHBR) is primarily defined by its use of the internet to obtain data about participants. While some of the research involves active participation with research subjects directly engaging with the research, for example through [online surveys](#) or [experimental tasks](#), many studies take advantage of "found text" in blogs, discussion forums or other online spaces, analyses of hits on websites, or observation of other types of online activity such as search engine histories or logs of actions in online games.

It's big business and the pervasive use of these methodologies is not only by academics but also corporations [and governments](#) seeking to support evidence-based policy decisions or [to nudge](#) societal behaviour.

Even though the basic principles of "respect for the autonomy and dignity of persons", "scientific value", "social responsibility" and "maximising benefits and minimising harm" are the same for this type of research method as for any other, the following issues often pose particular challenges for internet-mediated research: the distinction between public and private information, confidentiality, and [informed consent](#). There is an urgently need to establish clear codes of ethical conduct for IMHBR.

## Whose information is it?

The distinction between public and private domains is vitally important since this greatly affects the level of responsibility and obligation of the researcher. For human behaviour research online, however, it is often difficult to determine if participants perceive an online forum as "private" or "public". While almost all internet communication is recorded and accessible to the mediating platform, such as Facebook and Twitter, and much of it even publicly accessible, users of these platforms may nevertheless consider those communications to be private, despite click-signing the terms and conditions of the service provider.

To quote professor John Preston's testimony to the House of Commons science and technology committee [on responsible use of data](#):

*People treat social media a bit like they treat the pub. They feel that if they go into a pub and have a private conversation, it does not belong to the pub; it is their conversation. They interpret Twitter or Facebook in the same way – as a place to have a conversation.*

This was also one of the contributing factors in [the Samaritans' radar debacle](#) where they proposed an alert system to flag when people were tweeting potential distress and suicidal messages. In its post-investigation communication by the [Information Commissioner's Office](#) to the Samaritans, the ICO stated:

*On your website you [Samaritans] say that 'all the data is public, so user privacy is not an issue. Samaritans Radar analyses the tweets of people you follow, which are public tweets. It does not look at private tweets.' It is our view that if organisations collect information from the internet and use it in a way that's unfair, they could still breach the data protection principles even though the information was obtained from a publicly available source.*

## Confidentiality

Anonymisation is one of the most basic steps for maintaining confidentiality and showing respect for the dignity of research participants. It is also a requirement imposed by the Data Protection Act 1998 [when dealing with personal data](#). The need to protect the anonymity of participants is even more pressing when the research uses data from online sources where access to the raw data cannot be controlled by the researcher.

At the same time, the wealth of secondary information sources that can be mined in connection to any hint at the identity of a participant is making it increasingly easy to de-anonymise data. This [was publicly shown](#) by journalists for the New York Times who followed the web trail of user No. 4417749 in the AOL Search Log in 2006 and were able to identify her – and also by [the lawsuit against Netflix](#) for insufficient anonymisation of information disclosed in a prize competition database.

## Terms and conditions that no one reads

In order for informed consent to take place, it is vital that the participant is fully aware of what is being consented to. Unfortunately, current online business practice has heavily eroded the concept of informed consent by habituating people to click-sign terms and conditions forms that are too long and unintelligible to understand.

Sometimes driven by social pressure to join the network their peers are using, people readily skip over the details and give their consent for allowing corporations to access their data for a wide range of purposes. A hint at the dangers of normalising such attitudes towards the concept of informed consent was given by the statement in the controversial 2014 "[Facebook news feed manipulation experiment](#)" – a secret [study](#) on

"emotional contagion" that involved changing what 689,000 users saw from their friends' feeds to see if it influenced mood.

One of the researchers attempted to defend the study, saying that participants [had provided consent](#) because "it was consistent with Facebook's data use policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research". The data use policy, however, does not provide any information about the nature of that specific study, instead speaking only of "research" in general terms.

Various organisations and learned societies, such as the [British Psychological Society](#), the [Association of Internet Researchers](#), the [British Association for Applied Linguistics](#), the [Information Commissioner's Office](#), as well as our own [research group at Nottingham University](#) and many others are currently actively engaged in formulating and improving the guidelines for internet-mediated research.

As part of this work [we are currently running a survey](#) to ask citizens which conditions they would like to impose on researchers for making their social media [data](#) available to research studies. Ultimately, without clear guidelines and transparency, we're handing out decisions about us and our information to companies, governments and researchers, without us knowing what it will be used for.

*This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).*

Source: The Conversation

Citation: Social media provides a data goldmine for third parties (2015, June 16) retrieved 27 April 2024 from <https://phys.org/news/2015-06-social-media-goldmine-parties.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.