MRS Market Research Standards Board

Online Data Collection and Privacy

Response to submissions

The Consultation

In August 2011 MRSB published a discussion paper. The primary aim of the paper, as set out in the opening paragraph was to explore "the extent to which personal data could be legally and ethically collected online without taking steps to obtain the consent of the individual concerned."

The paper appeared the same week as ESOMAR's Guideline on Social Media Research[1] and CASRO's draft Social Media Research Guidelines[2] which were drafted in consultation but are in a number of respects subtly different. MRS made detailed submissions to ESOMAR during their drafting process, resulting in a number of significant changes to that document.

MRSB for its part did not wish to publish a guidelines or a new set of rules for this area. Instead, recognising that this was a complex area of law and a rapidly changing technological landscape, MRSB sought to lead a discussion about the underlying principles that inform the work of researchers and how these apply to research both online and offline.

The resulting debate has been vigorous and wide ranging. We have received some detailed responses, incited some lengthy blog and comment pieces, as well as an online debate including practitioners and professional bodies from both sides of the Atlantic.

MRSB believes that this debate is necessary and important to the future development of research in this area. The fabric of the future is being constructed online and the issues raised are not restricted to market and social research. As our lives increasingly take

[1] http://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Guideline-on-Social-Media.pdf
[2] http://www.casro.org/pdfs/1011/Social_Media_Research_Guidelines.pdf

place in the online environment, the decisions we take now on privacy and on responsible professional behaviour will have a lasting impact in the way we live and work for decades to come.

Researchers in all their work need to consider the risks they are taking in their own business, but also the risks to which they are exposing their clients. The eighth principle of the MRS Code of Conduct is that Researchers shall exercise independent professional judgement in the design, conduct and reporting of their professional activities. MRSB seeks to maintain a self-regulatory framework that is neutral both in regard to methodology and technology. This however places a significant burden on professional researchers to take responsibility for the technical aspects of their work. To this end MRSB's discussion paper and this response document is part of an ongoing dialogue to help researchers work through these issues in regard to online research, to make decisions that are in the best interests of respondents, clients and research.

Research does not exist in a vacuum. It exists in a landscape of law and regulation. Many of the related disciplines in the marketing communications sector which use data, such as direct marketing or online behavioural advertising, are subject to specific legislation or political attention in ways which research is not.

Many submissions we received expressed a desire to start from a blank sheet of paper, to tear up the Code of Conduct and to ignore principles which are held to be out-moded, old fashioned and out of touch with the modern world.

In almost every respect the MRS Code and the principles reflect the law which applies to all individuals and organisations engaged in data collection and processing. Deleting the Code would not change the underlying obligations of researchers, but would lead to the direct enforcement of these obligations by other regulators, who do not have any experience of, or appreciation for, research.

The nature of reality and the online world

This is a complex area and the law works and applies in ways that are not ordinarily expected. Some of what we say feels counter intuitive, so perhaps it is best to start from first principles.

In the real world, within reason, we can observe individuals in public spaces without their permission. It is possible to stand in a public place and watch people come and go, count them and observe how long they spend there. Under current legislation, this is acceptable as there is no **data** in that interaction.

Under the Data Protection Act 1998, ***data*** *means information which—*

(a) is being processed by means of equipment operating automatically in response to instructions given for that purpose,

(b) is recorded with the intention that it should be processed by means of such equipment.

As soon as a recording device is introduced, the information captured will be data and the Data Protection Act 1998 engages.

In the discussion paper we said virtual life is real life. This is incorrect in one major respect and that is with regard to the nature of reality itself – what is the virtual world made of?

The virtual world is made of one thing – data. Everything we see, hear or experience online is dependent on the underlying coding of the medium and the way data is copied, transmitted to and rendered on the devices we use to experience it. In the midst of all that data, there are strings of data that relate to identifiable living natural people – **personal data**, and that is where the problems begin.

Personal data *means data which relate to a living individual who can be identified—*

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual;

The Data Protection Act 1998 may not have been written with the internet in mind but it is clear that where personal data is being **processed** the Act will be engaged.

Processing *in relation to information or data, means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data, including*

(a) organisation, adaptation or alteration of the information or data,

(b) retrieval, consultation or use of the information or data,

(c) disclosure of the information or data by transmission, dissemination or otherwise making available, or

(d) alignment, combination, blocking, erasure or destruction of the information or data;

Taking the observation example, in the online world the researcher is not observing the individual, but rather they are observing that individual's **personal data** - observation which requires **processing** of the data. This means that under the strictest reading of the Act the law could apply immediately to that observation.

Comment: **"This is public information"**

The Data Protection Act 1998 does not contain a distinction between personal data in the private and public spheres. It refers to the purposes for which data is processed. The availability of personal data in a public space may help researchers understand the purposes for which it may be used but mere availability is not a sufficient criterion to justify collection and further processing.

Further, the reality of being "in public" has changed[3]. Thirty years ago, anyone who could see or hear you, you could likewise see and hear. Now "in public" means visibility to countless strangers, and we are watched much beyond our own capability to watch back. Also, domestic behaviour - the music you're listening to, the TV you're watching - is broadcast through social media.

This rapid change in meaning of "in public" should give researchers pause before they decide that it is fair or reasonable to collect and process information without consent.

Comment: **"…a man stands on a rock in the middle of a square and starts talking to people who pass by"**

A number of parallels were drawn between online activities and real world activities. This may well be the case, or they may be entirely new activities with no real parallels, given the durability of the medium (all public Twitter posts are archived by the Library of Congress, for example). What is determinative however from the legal point of view is not the fact that this data is publically available but rather the purposes for which it was originally processed. A tweet is in all likelihood purpose free, given that Twitter is in

[3] http://hautepop.tumblr.com/post/17118800100/surveillance-drone-industry-plans-pr-effort-to-counter

effect a broadcast medium. A comment on a friend's photo on a Facebook page could however be far more limited in purpose.

Comment: **"…the comments about masking or cloaking are unhelpful"**

Clients have stronger views on the use of masking than researchers. In the various discussions of the paper, some clients have said that they want either insights based on actual user data, or verbatim comments with the consent of the commenter. There were concerns from clients of using altered comments to illustrate findings.

The discussion of public and private spheres makes sense when the issues are wholly ones of avoiding harm to individuals. The same applies to the issue of masking, which may provide a solution to that problem. Masking however does not address the underlying and prior issue of whether personal data has been collected and processed lawfully.

Every researcher has a duty to ask data suppliers – where did this data come from? How has it been collected fairly and lawfully? Working with service providers can address a lot of these problems; accessing data under licence via a social graph or an Application Programming Interface (API) would be in line with the terms of use of the service that users have agreed to. Harvesting or scraping of data from sites will be less clear cut and due diligence is required before obtaining research sample from a vendor.

Comment: **"The suggestion that explicit research consent should be collected in order to track online buzz is out of touch with reality"**

The concluding point of the paper was that researchers must [obtain] the informed consent of all persons from or about whom data is collected. In the UK at least, informed consent can be a fairly low standard in practice, with signs in public areas being used to inform individuals that they are being recorded in a public area and their presence in that area being taken as consent. We would also note that the Act engages where personal data is being collected and the relevant provisions of the Code relate to information about an identifiable Respondents or attributable comments.

The key issue in buzz monitoring is the nature of the data required to keep track of consumer responses to commercial services and products. In many cases, the data collected will be items like name mentions of a client or their product or service, or measures of positive or negative sentiment through analysis of the surrounding text. In isolation, these items would not be considered personal data and their collection and processing would not engage the relevant terms of the Act or the MRS Code of Conduct.

The underlying principle then is one of data minimisation. Researchers should consider what information is strictly necessary to achieve the aims of their project. If personal data can be avoided then requirements of fair and lawful processing will not be triggered, and buzz monitoring could be conducted without the need for consent.

Comment: **"…we oppose the MRS's standpoint as we understand it but would support guidelines in line with ESOMAR's"**

Perhaps the most frequent concern raised regarding the paper was that it was out of line with the guidelines published by CASRO and ESOMAR. There was a common belief that CASRO and ESOMAR permitted the collection of personal data from public social media without the consent of the individuals concerned and that MRSB has taken an unreasonable stance in querying this.

The CASRO Social Media Guidelines set out a general principle of complying with the law, which mirrors Rule A1 of the MRS Code of Conduct:

> 4.4 Compliance with Law and Regulation
>
> Researchers must comply with existing state, federal and international statutes and regulations governing privacy, data security and the disclosure receipt and use of personally identifiable information.

Later, the guideline deals directly with public social media and allows for the collection information subject to certain qualifications:

> 6.1 Research in the Public Space
>
> In public spaces, normal conventions should apply. In the absence of facts that indicate otherwise (e.g., statements made by the participant or other participants), researchers may make use of this information and collect/copy/use content **subject to ToU policies, applicable data privacy laws, and the CASRO Code**.[emphasis added]

One should note that the US does not have a federal data protection act. It does have some data privacy legislation which target specific sectors such as healthcare. For a researcher working wholly within the US, engaging an applicable data privacy law may be no more than a theoretical possibility. In the UK and European context however, applicable data privacy laws are unavoidable, as we have discussed previously.

The ESOMAR guidelines do not contain an equivalent section to 6.1 of the CASRO guidelines. There is no explicit default assumption in the ESOMAR Guidelines that personal data can be harvested without consent even in public social media areas.

The ESOMAR guidelines do however contain the following passage:

> 2.3 Consent and notification
>
> The ICC/ESOMAR Code states that users' co-operation must be based on adequate information about the purpose and nature of the project and their agreement to participation obtained. **In addition, in some countries, existing data protection laws may also require users to be informed when personally identifiable data is collected**.
>
> Although it is potentially easy to obtain consent from members of market, social and opinion research communities, it poses more issues for other social media where users will generally not have been informed in advance or have consented to its use for research unless this is covered in the ToU.
>
> As noted in the ESOMAR Online Research guideline, researchers must remain mindful of concerns about privacy and intrusion if sending an email requesting such consent. They must reduce any inconvenience such an email might cause to the recipient by clearly stating its purpose in the subject heading and keeping the total message as brief as possible.
>
> If consent has not been obtained (directly or under the ToU) researchers must ensure that they report only depersonalised data from social media sources. **If researchers are using automated data collection services, they are recommended to use filters and controls to remove personal identifiers such as user names, photos, links to the user's profile, etc. Where this is not possible or they are manually collecting data from websites, their analysis must only be with depersonalised data and no attempt should be made to identify people** – see section 2.4 for a discussion on when identifiable quotes can be potentially used. [emphasis added]

To summarise, ESOMAR sets out a general principle of requiring agreement for participation in research and further notes that informed consent is required for the collection of personal data in many jurisdictions. (This would be the case in the UK and throughout the European Union for example).

The last paragraph quoted does not specifically mention collection but rather reporting. However, to be consistent with the opening section, a reasonable interpretation would be:

a) In the case of automated systems, these should be set to filter out the collection of personal data.

b) In the case of manual collection, the researcher should not take copies of personal data they have access to.

MRSB believes that the guidance that has been produced by CASRO and ESOMAR are compatible with MRSB's position, setting out the same standards albeit with different language and emphasis.

Life is complicated

People have secrets. They have spheres in which they act, say, write, and confide in particular and different ways. People are not rational – behaviour is not calibrated on a scale. It is unreasonable to expect human behaviour to conform exactly to Facebook privacy settings. Ethical behaviour should not be defined by the current state of the art or by budgets available.

A researcher should not assume that an attributable comment has no meaning or no significance to the user, or that its use could have no negative consequences. "I like Big Macs" may seem perfectly innocuous but what if it was attributable to an outwardly devout Hindu, an orthodox rabbi or a vegetarian?

Privacy is a real concern - in the time since the original MRSB paper was written, Facebook have been sanctioned by the US Federal Trade Commission for misleading users and will be subject to FTC privacy audits for the next 20 years.

The concerns expressed by MRSB in its original discussion paper also reflect wider societal views. In September 2011, two researchers, Danah Boyd from Microsoft Research and Kate Crawford from the University of New South Wales, published a paper entitled Six Provocations for Big Data[4]. The fifth provocation is the statement "Just Because it is Accessible Doesn't Make it Ethical". They go on to say:

> With Big Data emerging as a research field, little is understood about the ethical implications of the research being done. Should someone be included as a part of

[4] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

a large aggregate of data? What if someone's 'public' blog post is taken out of context and analyzed in a way that the author never imagined? What does it mean for someone to be spotlighted or to be analyzed without knowing it? Who is responsible for making certain that individuals and communities are not hurt by the research process? What does consent look like?

It may be unreasonable to ask researchers to obtain consent from every person who posts a tweet, but it is unethical for researchers to justify their actions as ethical simply because the data is accessible. Just because content is publicly accessible doesn't mean that it was meant to be consumed by just anyone. There are serious issues involved in the ethics of online data collection and analysis. The process of evaluating the research ethics cannot be ignored simply because the data is seemingly accessible. Researchers must keep asking themselves – and their colleagues – about the ethics of their data collection, analysis, and publication.

Where do we go from here?

MRSB supports the work of researchers and encourages them to develop new methodologies and to explore new areas of research. It is important that our regulatory framework is pragmatic and enabling while conforming to basic legal requirements and high ethical standards.

MRSB recognises that this is a fast moving area, and we are keen to engage on an on-going basis with researchers to understand the methodologies and technologies in use, and to provide timely and relevant support and advice.

As a first step MRSB has issued revised MRS Guidelines on Online Research and Research with Children and Young People incorporating some of the issues discussed here. These guidelines will be kept under review and will be subject to more frequent revisions than our other standards documents.

In parallel, MRS will continue work to influence regulators and legislators to ensure that new legislation is as positive for research as possible. We have developed a very good relationship with key regulators in the UK, such as the ICO, and for the last two years we have worked via EFAMRO, the European Research Federation, to monitor and lobby on revisions to the Data Protection Directive. EFAMRO has also been involved in work by the Council of Europe on guidelines for search engines and social media, which we hope will lead to better (and lawful) access to social media information by researchers.