

Non-disclosure in Internet-based research: the risks explored through a case study

K Masters

Citation

K Masters. *Non-disclosure in Internet-based research: the risks explored through a case study*. The Internet Journal of Medical Informatics. 2009 Volume 5 Number 2.

Abstract

Introduction Ethics guiding the disclosure of information regarding research into online discussion forums stems from two models: the human subject model, which views the postings as extensions of humans, and the text as object model, which views the postings purely as textual objects separated from humans, and requiring no special consideration. Concerns focus on disclosure of too much information; this paper explores the risks associated with disclosure of too little information. **Methods:** An analysis of online responses (from emails, blogs, and elsewhere) to the author's previous publication in which the author withheld identifying information about the research site. **Results:** The non-disclosed information was requested, and reactions to non-disclosure on ethical grounds were mixed. While some accepted the ethical argument, others rejected it, calling into question the credibility and validity of the research, pressurising the author to reveal the details; others identified themselves as part of the research site. The desire for the information culminated in a detective game which ended only when a suitable site was identified as the research site. **Discussion and Conclusion:** As found by other researchers, the non-disclosure of information will not be unanimously acceptable, and the site will be sought. A model for further disguise, using a "Maryut site" is described.

INTRODUCTION

Since formalised ethics' rules regarding research into human subjects emerged with the Nuremberg Code [1], they have undergone several reviews [2], and are a complex mixture of social norms, values and legal issues. Of certainty, however, is that the role of ethics in human research is crucial to that research.

Different countries approach ethics in research from different perspectives. For example, ethics in the USA tend to focus on a risk / benefit model, where the aim is to maximize the benefits and minimize the risks [3]. One of the risks is the over-exposure of individuals and groups [3], and this is one of the reasons that the anonymity of the subjects is important and must be maintained [4]. While other countries have different bases for their ethics' rules, the protection of the subjects and their identities is a common theme [5], and is crucial to the World Medical Associations' (WMA) Declaration of Helsinki Declaration of Helsinki (See Clause 23) [6].

The Internet has introduced complexities into human research that were unforeseen in earlier ethics guides [2], and has also led to several more guides, usually arising from

specific disciplines [7]. An area of particular concern for medical informatics is the research of online discussion forums, or bulletin boards.

There are two broad models that inform the way in which researchers view postings in online forums. The first is a "human subject" model, and the second is a "textual object" model.

THE HUMAN SUBJECT MODEL

The human subject model has its origins in the medical field, and the traditions that began with considerations drawn from the Nuremberg Trials [1]. The human subject model views online forum postings as expressions of humans, and emphasises that the ethics guiding various issues, such as privacy, disclosure of information, and informed consent, should be treated in much the same way that any researcher would when conducting studies of humans. This is of particular importance when sensitive medical information is the focus of forum discussion [3]. Guidelines grounded in this model warn the researcher to tread sensitively and carefully when conducting research into these forums [3; 8], as the ethical issues need to be treated properly.

The first issue of concern is informed consent, and the

advice in the guidelines hinges primarily on the amount of accessibility and publicity the contributors in the forums are generally aware of, and desire, versus the amount of privacy that appears to govern the rules of registration and participation in the forum [2-4; 8]. When uncertainty occurs, caution is usually advised. One obvious reason for caution is that, even if consent is given, these online groups are fluid [3; 4]. Consent may be given today, and a new individual may join the group tomorrow, and be unaware of the ongoing research. In addition, other problems such as actually obtaining the consent and the use of pseudonyms pose particular practical problems [2; 3; 9].

A second issue raised in the human subject model, is that of securing privacy. The first and obvious step to securing participants' privacy is ensuring that participants are not named in the published research. Secondly, however, care must be taken about quoting qualitative data, because these data can be used to search for identifying information, and exposure of the identifying information runs the risk of great psychological distress [3; 4; 8; 10]. Even in a publicly-visible site, many researchers feel that there is some expectation of privacy [3].

Protecting individual participants on the research site can become difficult when one wishes to obtain research data from hundreds or even tens of thousands of participants. To increase protection, while it might not be the instruction from an institutional review board (IRB) or ethics' committee to do so, some researchers in medical and medically-related fields either disguise or do not name the research sites at all [9-12]. This approach is in line with what Bruckman classifies as "heavy disguise" [9]. Even outside of medical fields, disguising or hiding the name of the site has the advantage of being able to protect the "regulars" [9].

Finally, there is the relationship between the researcher and the group of people being studied; researchers are advised to take great care to not be viewed as spies or intruders, [3], and disruptions of group processes are to be avoided [4].

TEXTUAL OBJECT MODEL

Contrasting with the human subject model is the argument that postings in online forums and other areas of web sites are not humans, but are textual objects, and should be treated as such.

For a start, as pointed out by Bruckman [9], in the USA, a "Human subject means a living individual about whom an investigator (whether professional or student) conducting

research obtains

(1) data through intervention or interaction with the individual, or

(2) identifiable private information." [9; 13]

This definition certainly does not refer to texts that are posted in an easily-accessible forum if no personal identifiable information is given, and, therefore, according to this model, the conflation of humans with texts is inappropriate. As many researchers point out, the textual object model is supported by much 20-century literary theory (such as the work by Wimsatt & Beardsley [14] and Roland Barthes [15]) which clearly separates any discussion of text from the discussion of the author or even the author's intention [9; 16].

In addition, in the US, all work on the Internet is considered copyrighted, so the only real ethical issues under consideration revolve around fair use, proper citations, and protection of copyright [7; 16]. In this sense, even the domain name is regarded as a textual object and should be afforded the same protection, but no more [16].

Finally, the model argues, because the Internet is a public area, any expectation of privacy is misplaced – Walther uses the offline analogy of conversations in a public park to argue that "people do not expect to be recorded or observed although they understand that the potential to do so exists" [17]. The analogy is especially appropriate if the researcher does not specifically link the data to a human, in which case the data-gathering is much the same as gathering information from old newspapers or any archival data [16; 17].

THE TEXTUAL OBJECT IN PERSPECTIVE

While the textual object model certainly does have strong support, three further points must be considered.

The first is that arguing from a literary theoretical viewpoint is fraught with danger – literary theory is not a single coherent theoretical approach, but rather covers a vast range of theories, frequently grounded in other disciplines (e.g. Anthropology, Sociology and Psychology) and philosophical approaches (e.g. Structuralism, Marxism and Feminism). Just as one may find a literary theory supporting a point of view, one may find any number that oppose it. The approaches most often cited in support of the text as object model are chiefly from the theoretical viewpoints called New Criticism and deconstruction [16]; if one uses these as a basis on which justify disclosure, then one should also argue

why any number of respectable opposing theories were not chosen.

Secondly, many of these arguments in literary theory discuss the relationship between text and authors' intention [14]; ethics concerns in studying online forums does not hinge on the author's intention, but rather on the author's identification. Even if the piece of text is an object, in a case where a quotation from the piece of text leads to the identification of the author, then the text cannot be viewed as an isolated object only.

Thirdly, in medical research, there is a long tradition of not disclosing any identifying information about objects, and this may be for a range of reasons. An illustrative and typical example is an article by Ehara and Marumo [18] which reports on forensics' methods for examining lipstick. Lipstick is clearly an object, and the article describes 174 lipsticks from 11 different manufacturers, yet does not mention any of the lipsticks or manufacturers by name. In this case, even if all 11 manufacturers were mentioned, listed in alphabetical order, their names could not be connected to any of the specific objects, and yet the authors chose not to identify the manufacturers. Significant is the fact that nowhere do they justify their decision – there is no need, as this is a commonly-accepted practice.

More recently, and perhaps far more in the public interests, Leung et al. [19], reporting on the accuracy of iodine reported by multivitamin companies, do not disclose the names of any the companies whose products they study, neither do they justify their decision. Again, there is no reason to justify a commonly-accepted practice.

THE STANCE OF THE RESEARCHER

The researcher of online forums is caught between the two models. Notwithstanding the fact that many of the arguments are grounded in national laws, the individual researcher finds very little definitive that applies across the globe. Commenting on the Association of Internet Researchers (AoIR) Ethics Working Committee 2002 Report, Jankowski and van Selm make the point that “the main implicit message of the report is that no definitive, single set of ethical guidelines is possible for a field as diverse as internet research” [2].

In addition, even Walther, who argues that postings into forums cannot be classified as, and need not be given the same protection as, human subjects, notes that researchers “must make their own individual ethical decisions with

regard to activities such as quoting or reflecting names or pseudonyms in their ultimate publications, and should indeed do so in mind of some of the points that the Report raises” [17].

A possible solution that respects the diversity of ethical views is suggested by Ess [7]; one that takes the legal requirements as a minimum, and then researchers are free to impose extra restrictions on themselves, if they so wish [7]. Given that laws change according to a society's needs and perceived needs, for many researchers, using the law as a minimum may be the only viable option. Medical researchers are painfully aware of a track record in which deferring ethics entirely to law is what led to many of the inhumane experiments in Nazi-occupied Europe, and formed much of the motivation for independent research ethics in the first place. Many other countries, from the USA to South Africa, have bleak areas in their history where medical professionals too easily deferred their ethics to the prevailing law, or to those norms that the dominant group in the society regarded as acceptable behaviour [20-23].

HOW MUCH TO DISCLOSE ABOUT AN ONLINE WEB SITE?

The pointed issue of relevance to this paper is the issue of information disclosure. Frankel asks an all-important question: “How much description of an online community should a researcher provide?” [3]. Given the impact of the human subject model, many researchers may wish to err on the side of caution – and even advise their students to do so [9]. There is an opposing tension, however. For the research to have value, readers need to know something about the site so that the context of the research can be understood.

It is this opposing tension that leads to the central issue to be examined in this paper. While issues explored deeply in most of the research cited above consider the risks associated with disclosing too much information about a site, this paper explores the risks associated with disclosing what some may see as too little information about the site.

This paper will discuss some of these risks. In doing so, it will use the case of a paper published by the author, and the discussions on the Internet that followed its publication.

METHODS

In 2009, the author published a paper [24] that dealt with online forum postings on a medical web site. Based on the guidelines given by Eysenbach and Till [8], it was established that informed consent was not required. To

justify this, it was necessary to give some broad information about the site (e.g. that it was not password-protected), and some aggregated data (e.g. number of users). In addition, the names of some of the forums were given. Giving this information also enabled readers to understand the context of research.

Nevertheless, because that paper dealt with forums, the author elected not to disclose the identity of the site (neither its name nor its domain), nor the geographical location of the hosting server. In addition, no participant names (including pseudonyms) or any other identifying information or qualitative quotations from the site were given in the publication.

This approach is closer to the human subject model than to the text as object model. Because the site was aimed primarily at medical professionals, and the journal at which the paper was aimed is a medical informatics journal, the human subject model was deemed appropriate.

After the paper's publication, the author used general search engines to search the Internet for responses to the paper, and the data from the responses were gathered.

In the case of publicly-visible blogs and Facebook, again using Eysenbach and Till [8] as a guide, permission to quote was not required. In case of postings to small discussion lists and private emails, consent was obtained to identify and quote from participants' postings.

The responses were themed using NVivo Version 7. Because this paper is concerned only with the issue of information disclosure, only the discussions referring to that issue will be given and discussed here.

RESULTS

In her work on disclosure, Amy Bruckman [9] describes work done by her and other researchers, and presents the impact of the online ethics debate in a series of useful "lessons." This paper will similarly present the results of this case, ending each sub-section with a lesson for the researcher who elects to withhold information.

DISCUSSED ON THE INTERNET

Within two-three weeks of publication, there was reaction to the research on a range of blogs, discussion lists, other Internet sites, and private emails sent to the author. In the words of one blogger, news of the publication "lit up feeds and emails" [25]. While this might be an over-statement, it is true that a Google search on the paper's title barely a month

after publication identified over 700 links. This excluded the unknown number of articles and news sites (e.g. [26-28]) that discussed the publication, but did not refer to the actual title of the paper.

Although the topics ranged widely, one that appeared to attract a great deal of attention was the fact that the researcher had withheld the identity of the site under investigation.

Lesson: Any article published today may be discussed on a large number of websites. The act of non-disclosure of some information about the researched site sparks interest, and researchers should expect their work to be debated in the light of the non-disclosure, whatever their reasons may be.

REQUESTS FOR INFORMATION

Following the publication, the author received 12 private emails requesting the identification of the site.

Nine of the emails were from publishers and editors of journals who were being requested to participate in a follow-up survey of journal editors [29]. One was from a reporter for one of the journals. In all these cases, the requesters were notified that, for ethical reasons regarding research into websites hosting discussion forums, this information could not be given. These requesters did not pursue the subject any further.

The 11th e-mail was from a researcher requesting the identity of the website for research purposes. The identity, and some other information, was supplied on the understanding that the information was provided purely for the purposes of research, and should not be made public. The researcher accepted the information under those conditions.

The 12th e-mail was from Kent Anderson, a blogger and Executive Director, Product Development, The New England Journal of Medicine and Journal Watch who requested the identity of the site, or reasons for the non-disclosure. As this was suggested as an acceptable alternative, he was sent an explanation similar to the explanation sent to the journal editors and publishers.

Lesson: Non-disclosing or disguising the information in the published paper will be the first step only. After publication, researchers may be approached by a range of other researchers, journalists or members of the public requesting further information about the study, including the information that the researcher has intentionally not disclosed. Some of these may be polite requests, while

others may be more challenging in tone.

A REJECTION OF THE ETHICS ARGUMENT

Although the ethics argument for non-disclosure appears to have been accepted by the journal editors and publishers who contacted the author, it was rejected by some bloggers and discussion list participants as grounds on which to withhold the site name. On his blog, Anderson argued that the site's name "isn't personal information about people but a web address or domain" [25]. In addition, it "was publicly available," and the author had not made "any promises of confidentiality." The reason for non-disclosure was deemed to be a "vague discomfort wrapped up as 'ethics'" [25].

In a private email to the author, Anderson wrote that he "appreciates" the position, but that it is "nonsense" [30]. He described the withholding of the information as "weird," and argued further: "For me, this boils down to research integrity. I can't think of any study that would conceal the subject of its findings, especially if the subject is an inanimate object" [31]. In his mail, Anderson compared the approach to a geologist who finds a rock with particular characteristics, but refuses to identify the rock so that others can confirm or refute the findings [31].

In response to Anderson's blog, Eric Hellman raised a possible ethical problem of posting a link to an unethical site, but agrees that "it's still important to 'show your work'" [25].

Lesson: While many fellow-researchers will accept the explanation of non-disclosure on ethical grounds, the human subject model does not have unanimous acceptance in the case of revealing textual (or "inanimate") objects, including domain names. This will exist even if the content of that text is not known (In this case, there was the assumption that the domain name did not contain personal details), or if knowledge of the site's details would naturally allow the forums' posters to be identified. More, the researcher's use of ethics to justify the non-disclosure may be viewed with scepticism, if not disdain.

THE CREDIBILITY AND VALIDITY OF THE STUDY WILL BE CALLED INTO QUESTION

In his blog, Anderson considered the research paper "incomplete" as it withheld information that he deemed important for the sake of clarity. He went on to ask the rhetorical question of how anyone else could "confirm, refute, or re-analyze the research if such a vital link is missing and actively concealed by the researcher?" [25]. The

missing information was regarded as "key," omitted "for no good reason" [25].

Further in his mail, Anderson argues "But, unless there's something unseemly to the reasons you are withholding this information, I guess we'll just move on" [31].

In a similar vein, in a publicly-accessible discussion list where the publication was being discussed, Thomas Krichel wrote "I would not pay much attention to a paper that cites an unidentified web site. They could have made up that data" [32].

Lesson: Not disclosing information in a publication may be interpreted as a concealment of vital or key information, or it may be suggested that the data were falsified, that the author has "unseemly" reasons for the non-disclosure, and the validity of the entire study may be called into question.

NON-DISCLOSURE AS PART OF A GAME

Anderson reported that he had attempted to find the site by "digging" for it, but had failed. He extracted pieces of information taken from the paper, and presented them as "clues" on his blog, requesting readers to use that information to discover the identity of the site [25].

Unbeknown to Anderson, on the same discussion list cited above, Mark Funk had asked "What, no detectives on this list?" [33]. He then described a process by which he had searched for and found a site that he believed to be the research site. He identified the site as "www.smso.net." Funk was not able to provide a link to the site, as it appeared to be no longer in operation, so he placed a link to an archive site, and to other related information. Unfortunately, the archive site does not list pages that were visible at the time discussed in the original research. Funk did, however, end his posting with: "I doubt very much that the data were made up" [33].

Funk's entry was found by Philip Davis, who then responded to Anderson by posting a link to Funk's posting, along with the comment that "we shouldn't have to speculate on the source nor validity of the data" [25]. Anderson updated his blog, announcing "We have a winner!" [25]. He described Funk's "detective work" [25], and also placed a link to the archive site. As already mentioned, however, the archive site does not list the site's pages covered at the time of the original research, so it is uncertain as to how those pages could be used to verify the data in the paper, one of the prime reasons that Anderson had given for wanting to

know the name of the site. In his blog, Anderson makes no reference to Funk's ending comment.

Lesson: The act of not disclosing information will be seen as part of a detective game, a riddle to be solved, with the original research providing the clues. If available data are used as clues, identifying information may be discovered and publicly displayed.

MEMBERS (REAL OR NOT) DISCLOSE THE INFORMATION

A person in a Facebook group announced on Facebook that it was their site that had been studied in the paper [34]. The announcement was not regretful, but was rather openly advertising.

On the discussion list, Mark Funk posted a note informing the list of this Facebook posting, and placed a link to the Facebook announcement. He commented that, "And in case there were any doubts, on this page is a proud link to the ISPUB journal article about the site" [35].

Funk's posting was picked up by Anderson, who displayed the information in his blog.

Lesson: Individuals, even if others on the site do not wish to have it disclosed, may take it upon themselves to identify themselves as being part of the research site. This may happen irrespective of whether they are correct or not.

PRESSURE TO CONFIRM OR DENY THE SITE

In follow-up mails, Anderson asked the author to confirm the identity of the site as the one displayed on Anderson's blog [31; 36]. The argument was that the researcher was alone in his belief that it was necessary to maintain the non-disclosure. The author declined to confirm the site. In his blog, Anderson argued that the author "is alone in his unwillingness to acknowledge that SMSO.net is the subject of his research" [25].

Responding to Anderson's blog, a reader identified as "Dr. Gunn" praised Anderson's work as "fantastic" and displaying a " 'go for the jugular' instinct that so many pro reporters seem to have lost" [25].

Lesson: Once a suitable site has been identified, pressure in the form of follow-up mails and further comments may test a researcher, irrespective of whether or not the site is correctly identified. The implication will be that the information is now known publicly, and that to continue to deny it is an oddity.

QUICK DISSIPATION

Within a week of the initial posting, most of the commentators had moved on to other topics. It appears that either the issue was not that important after all, or that they felt that the riddle had been solved, so there was no need to dwell on it any longer.

Lesson: The game of finding the information quickly becomes more important than the reason for finding it. It is quite possible that, when a suitable site is identified, the discussion will end there.

DISCUSSION

This paper has presented the results of comments in various Internet sites and emails in response to the authors' non-disclosure of information in published research. This section discusses some of these results in the light of the literature and the broader context of research. It will use the same sub-headings to refer to the sub-sections in the Results.

DISCUSSED ON THE INTERNET

An argument supporting publishing in OA journals is that OA articles have more citations than NOA articles [37-40]. (Even where this argument has been countered, the research has shown that OA articles are accessed more than NOA articles [41], and a greater number of accesses is a useful predictor of later citations [42]).

In this instance, it would be too soon to measure citations. What has been demonstrated, however, is that open-access does allow for immediate commentary on the Internet, and this commentary may occur across many different sites. In this study, how much of this interest was due to the nature of the open access publishing model and how much due to the issues researched remains to be seen.

REQUESTS FOR INFORMATION

The researcher needs to be very sure about which ethical model is being used to guide and justify the non-disclosure of information. If the human subject model is being applied, then consistency requires that it will have to be applied in responding to requests also. This may also be a requirement of one's IRB or ethics committee. In any case, the researcher should decide whether or not it would be permissible to disclose the information to some parties, and under which terms.

A REJECTION OF THE ETHICS ARGUMENT

It is noteworthy that all the correspondents from journals and publishers who requested information accepted the ethical

argument for non-disclosure. Of the 10 emails from journals and publishers, only five were from medical or medically-related journals and publishers, indicating that the ethical argument does have some acceptance beyond the medical fields.

Part of Anderson's argument is that the non-disclosed information "isn't personal information about people but a web address or domain. It was publicly available" [25]. The relationship between identifying the domain and the people using that domain has already been covered in this paper. The weakness of the publicly-accessible argument is further apparent when one considers email addresses: for the most part, almost everybody's name and email address are publicly available – but that does not mean that people participating in research should have their names and email addresses disclosed in publications dealing with that research.

There is also the argument that, in the course of the research, the author had not made promises of confidentiality [25]. It is quite true that, if promises of confidentiality have been made, then these need to be upheld. It does not necessarily follow, however, that if promises of confidentiality have not been made, then one is obliged or encouraged or even free to disclose this information. Simply because somebody has not promised not to harm me, does not give them the right to do so.

THE CREDIBILITY AND VALIDITY OF THE STUDY WILL BE CALLED INTO QUESTION

No researcher wishes to publish incomplete or inadequate work. Doing so obviously impacts on one's standing in the academic community. For that reason, comments questioning the validity of the results if the required information is not disclosed, or "concealed," are bound to make researchers second-guess their ethical ground for non-disclosure.

Unfortunately, this is a risk that almost all research dealing with human subjects takes when the identity of human subjects is withheld. As Bruckman notes, in "an open scientific community, individuals ideally publish results sufficiently detailed for others to attempt to duplicate those results and affirm or question the findings. This idealized model from the physical sciences is always hard to replicate in social sciences, but even harder when the act of protecting subjects adds substantial new barriers to follow-up inquiry by others" [9]. Bruckman notes that "The better you protect your subjects, the more you may reduce the accuracy and

replicability of your study" [9].

In addition, the non-validity argument ignores the thousands of survey results that are published every year. In these, figures are cited, but the raw data are never disclosed, and replicability with the same participants is almost impossible. In much the same way, as was noted in the description of the research by Ehara and Marumo [18] and Leung et al. [19], a non-disclosure of identifying information is common in studies of objects also.

A final issue that might be considered by the researcher is the impact on the research site of releasing this information. The survey of editors referred to earlier indicated that up to a third of journals accessed through such a file-sharing site would consider taking legal action against such a site [29]. Although this reaction is specific to this site, it is not difficult to believe that similar situations might exist with other sites. In these instances, if the researcher intentionally identified the site, he would no longer be an observer of a phenomenon, but rather an agent impacting directly on the phenomenon. This would be more in the line of investigative journalism with a view to exposure of a particular group, not the work of an academic researcher.

NON-DISCLOSURE AS PART OF A GAME

The activity of people hunting for and then disclosing undisclosed information has been found by Bruckman also [9]. In the case cited by Bruckman, however, the disclosure originated primarily from people who had been studied, and was not seen as detective work by outsiders.

Researchers need to be aware that the rulings of IRBs, and ethics' committees, and journal guidelines are binding on the researchers, but are not necessarily binding on the general public, nor on those working with different ethical models and guidelines. (It is important to note that the author is not criticising the ethics of any of the commentators referenced – the point is made that they are working from a different ethical viewpoint, one that appears to be more strongly guided by the text as object model).

MEMBERS (REAL OR NOT) DISCLOSE THE INFORMATION

Similar to the lesson given by Bruckman, these results have shown that "anonymity may be hampered by the subjects themselves" [9]. Just as the ethics used to guide to researcher are not binding on outsiders, they are also not binding on people whose site has been studied, or on people who believe or claim that their site has been studied.

In this case, though, given that the activities on the research site were probably infringing on copyright laws, and US government agencies are known to be monitoring sites like Facebook for copyright infringement [43], an announcement like this one on Facebook was probably not the best thing that a member of such a site could do.

PRESSURE TO CONFIRM OR DENY THE SITE

Pressure to confirm or deny the identified site will be exerted on the researcher irrespective of whether the identified site is the research site.

Arguably, more pressure will be exerted on the researcher if the site presented as the “winner” [25] is not the research site. If the site is incorrectly identified, the researcher may be tempted to say so, and may feel obligated to say so. Logic would argue that that it would be acceptable to announce that the identified site is not the research site, because denial does not disclose anything about the research site. The problem with this approach is that the researcher may then be probed with a range of possible sites, and will feel compelled to respond with denial for every site that is not the research site. He will find himself an unwitting participant in the game. Then, when the correct site is named, if the researcher remains silent, he will, in effect, have affirmed the site as the research site.

If the researcher has supplied only limited or disguised information, then there is the risk that a different site may match that information. From these results, it appears that those seeking the site will stop only when they have found a seeming match.

The researcher should take great care in responding to such requests.

QUICK DISSIPATION

The quick dissipation of the discussion may raise questions about the motivation for needing the site information in the first place. In this case, the initial motivations were given as the need for clarity, and needing “to confirm, refute, or re-analyze the research” [25]. It is noteworthy that none of the later postings were then able to link the discovered information to their goals. Assuming that the site identified was the correct site, there was no discussion about how this new-found data added clarity to the research. More importantly, there was no reference to how this information was being used to “to confirm, refute, or re-analyze the research.”

A concept that exists in many countries’ legal systems on disclosure revolves around the benefit of disclosure versus non-disclosure in terms of the public good. The question of the benefit of this disclosure (correct or not) to the public good remains unexplored.

ENHANCING THE DISGUISE WITH ‘MARYUT SITES’

In her research in online discussions, Turkle does not disclose the name of all the online sites she studies, and she disguises names and events in the lives of her subjects [10-12]. As mentioned, this approach is part of what Bruckman classifies as “heavy disguise” [9]. Given the widespread detective work described in the results of this study, this level of disguise may not be enough to completely protect the identity of the site and the participants.

If the researcher wishes to increase the protection of the site, the disguise may be expanded to the use of what I call a “Maryut site.” The term “Maryut site” is a reference to the story of the creation of a decoy site at Maryut Lake to prevent Alexandria Harbour’s being bombed during World War II. The process of using a Maryut site would be the following:

The researcher creates a fake (or “Maryut”) web site that has a structure similar to the research site.

The researcher then populates the Maryut site with plausible information. This would include structures (e.g. names of forums) that are found in the research site, plus additional forums. The new information would need to be of such a nature that it does not detract from the validity of the research. An example would be the name of a forum that one might expect to find on such a site but that does not exist in the research site.

In the research paper, amongst the real information listed, the researcher lists the fake information that is found only in the Maryut site. This information must not materially affect the research, in much the same way that alterations to patients’ experiences in Turkle’s research does not materially affect her research.

The Maryut site is taken off-line, and archiving sites may list only top-level pages that will hold the Maryut site’s general information (e.g. forum names and number of participants).

If the information from the published research is then identified as “clues,” this information will more closely

match the Maryut site than it will match the research site. If this information is used to search for the research site, the information on the Maryut site will divert “detectives” away from the research site, and will point them either to the discontinued Maryut site or to the archived site.

The researcher might even post information into blogs, discussion lists, online forums and social networking sites, under pseudonyms, directing detectives to the Maryut site.

Naturally, it is possible to create more than one Maryut site, all similar, but with small differences, to be found under different but similar search strategies. Some may never be found.

An essential part of the disguise may be for the researcher never to disclose whether or not Maryut sites were used during the research.

The ethics of making a Maryut sites, in the interests of safeguarding the non-disclosure, may be a point of discussion by IRBs and ethics’ committees.

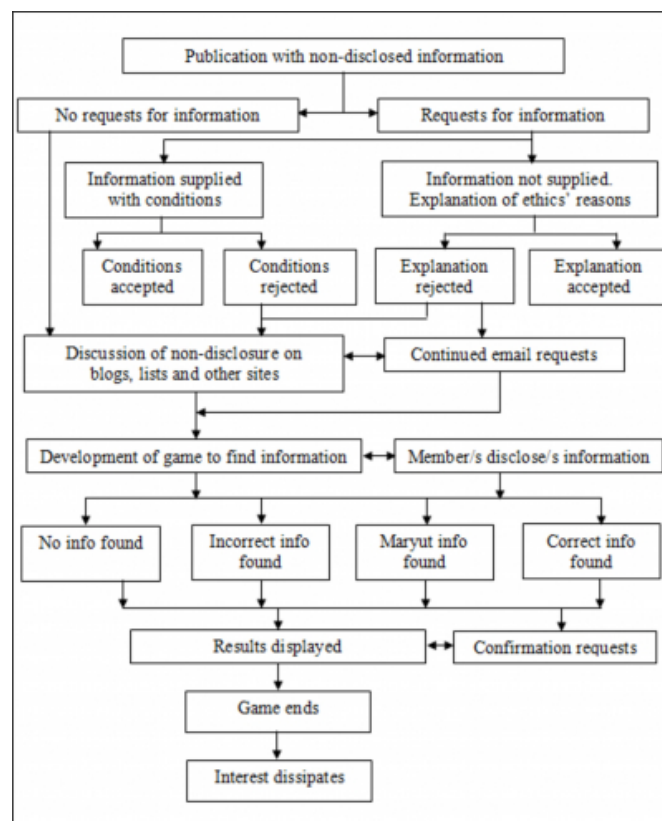
A danger with this method is that the Maryut site may inadvertently point the detectives to a valid secondary site that has nothing at all to do with the research.

A MODEL OF EVENTS FOLLOWING NON-DISCLOSURE

It might seem premature to develop a model from one case. In this instance, however, the data presented in the results come from a range of different commentators. In addition, the work of Bruckman supports several of the results presented here.

Figure 1

Figure 1: Model showing possible events following non-disclosure in a publication



IN THE BROADER CONTEXT OF RESEARCH

The broader applicability of the model is still to be tested. Specifically, a second paper, dealing with the same site, and guided by the same ethical principles is to be published in this journal [44]. Responses to that paper may serve to supply information leading to an expansion of the model described above.

Naturally, the non-disclosure of information is not only part of Internet research, and lessons can be learnt from this model with possible application elsewhere. Returning to the article on lipstick by Ehara and Marumo [18], from the information supplied in that paper, it would be possible to run tests and identify the manufacturers and even the lipstick. It would then be a simple matter of making the information public. Similarly one could attempt to replicate the research of Leung et al. [19], and identify the companies openly.

In neither case, however, does this appear to have happened. There might be several reasons. One of these may be that, to perform that testing, one requires specialised and expensive equipment, and considerably more effort and skill than a

simple Internet search on a few phrases. Future applications of the model in Figure 1 may lead to adjustments to reflect the ease with which the “detective work” can be performed.

CONCLUSION

In the field of medical informatics, the study of online forums is becoming increasingly important, and the disclosure of information in the course of publication raises ethical issues. This paper has shown that the ethics guiding disclosure of information from online forums has a complex theoretical background, and also has complexities in practical application. Guidelines do exist, but, in many instances, researchers must decide on issues by themselves. A recommended guide is one that respects a diversity of viewpoints, and takes the law as a minimum, and then researchers are free to impose extra restrictions on themselves, if they wish.

Through the case, this paper has indicated that, while disclosing too much information carries its own dangers, disclosing what some may see as too little information runs risks also. These risks include requests from third parties for the information, criticisms of the researchers’ reasons for non-disclosure, questioning of the researcher’s motives and the validity of the research, and finally, the development of a ‘game’ designed at disclosing the non-disclosed information.

In addition, the paper has indicated that researchers who wish to maintain the non-disclosure may enhance the site’s disguise or non-disclosure by creating a “Maryut site” which may prevent the research site from being discovered.

Finally, based on the lessons learnt from this case, the paper has presented a model showing the possible events that will follow non-disclosure of information on the research site. The model may better prepare researchers in the future.

References

1. National Institutes of Health Office of Human Subjects Research [USA]: Nuremberg Code; <http://ohsr.od.nih.gov/guidelines/nuremberg.html>; n.d. (Accessed 23/11/2009).
2. Jankowski NW, van Selm ML Research ethics in a virtual world. Guidelines and illustrations. In: Carpentier N, Pruulmann-Vengerfeldt P, et al., (eds). Media Technologies and Democracy in an Enlarged Europe; Tartu University Press, 2007.
3. Frankel MS: Ethical and legal aspects of human subjects research on the Internet. Washington; American Association for the Advancement of Science; 1999.
4. Eysenbach G, Wyatt J: Using the Internet for Surveys and Health Research. *J Med Internet Res*; 2002; 4(2):e13.
5. National Committees for Research Ethics in Norway: Guidelines For Research Ethics in the Social Sciences, Law and the Humanities; Oslo; 2006.
6. World Medical Association: World Medical Association Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects; <http://www.wma.net/en/30publications/10policies/b3/index.html>; 2008 (Accessed 12/01/2010).
7. Ess C: Introduction. *Ethics and Information Technology*; 2002; 4:177-88.
8. Eysenbach G, Till JE: Ethical issues in qualitative research on internet communities. *BMJ*; 2001; 323:1103-5.
9. Bruckman A: Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology*; 2002; 4:217-31.
10. Turkle S: *Life on the Screen: Identity in the Age of the Internet*; New York; Simon & Schuster; 1995.
11. Turkle S: Constructions and Reconstructions of Self in Virtual Reality. *Mind, Culture, and Activity*; 1994; 1(3):158-67.
12. Turkle S: Multiple subjectivity and virtual community at the end of the Freudian century. *Sociological Inquiry*; 1997; 67(1):72-84.
13. National Institutes of Health Office of Human Subjects Research [USA]: Code of Federal Regulations, Title 45, Public Welfare, Part 46: Protection of Human Subjects; Bethesda; 2005.
14. Wimsatt W, Beardsley M: The Intentional Fallacy; In *The Verbal Icon: Studies in the Meaning of Poetry*; Lexington; University of Kentucky Press; 1954; 3-18.
15. Barthes R: *Image Music Text*; New York; Hill and Wang; 1977.
16. Bassett EH, O’Riordan K: Ethics of Internet research: Contesting the human subjects research model. *Ethics and Information Technology*; 2002; 4:233-47.
17. Walther JB: Research ethics in Internet-enabled research: Human subjects issues and methodological myopia. *Ethics and Information Technology*; 2002; 4:205-16.
18. Ehara Y, Marumo Y: Identification of lipstick smears by fluorescence observation and purge-and-trap gas chromatography. *Forensic Science International*; 1998; 96:1-10.
19. Leung AM, Pearce EN, Braverman LE: Iodine Content of Prenatal Multivitamins in the United States. *NEJM*; 2009; 360(9):939-40.
20. Brown JG: Institutional Review Boards: A Time for Reform; Department of Health and Human Services, Office of Inspector General [USA]; 1998.
21. Mclean G, Jenkins T: The Steve Biko affair: A case study in medical ethics. *Developing World Bioethics*; 2003; 3(1):77-95.
22. Steinbrook R: Improving protection for research subject. *NEJM*; 2002; 346(18):1425-30.
23. Silove D: Doctors and the state: Lessons from the Biko case. *Soc Sci Med*; 1990; 30(4):417-29.
24. Masters K: Opening the non-open access medical journals: Internet-based sharing of journal articles on a medical web site. *Internet Journal of Medical Informatics*; 2009; 5(1):<http://tinyurl.com/kmoajournals> (Accessed 28/10/2009).
25. Anderson K: Breaking the Chain of Inquiry - When Journals and Journalists Fall Short <http://scholarlykitchen.sspnet.org/2009/11/12/breaking-the-chain-of-inquiry/>; 2009; (Accessed 18/11/2009).
26. Timmer J: Med Students Hoist P2P Jolly Roger to Get Access to Papers; <http://arstechnica.com/science/news/2009/10/med-students-hoist-p2p-jolly-roger-to-get-access-to-papers.ars>; 2009; (Accessed 23/11/2009).
27. Jordan S: The Latest File-Sharing Piracy: Academic

Journals. TeleRead

<http://www.teleread.org/2009/11/03/the-latest-file-sharing-piracy-academic-journals/>; 2009; (Accessed 23/11/2009).

28. Terris B: The Latest File-Sharing Piracy: Academic Journals. The Chronicle of Higher Education <http://chronicle.com/blogPost/The-Latest-File-Sharing/8662>; 2009; (Accessed 23/11/2009).

29. Masters K: Articles shared on a medical web site - an international survey of non-open access journal editors. Internet Journal of Medical Informatics; In press.

30. Anderson K: <(Private email to author)>; ISPUB Paper; 2009; 11/11; (Accessed 11/11/2009).

31. Anderson K: <(Private email to author)>; ISPUB Paper; 2009; 14/11; (Accessed 14/11/2009).

32. Krichel T: Academic Journal File-Sharing (CHE); liblicense-l@lists.yale.edu; 2009, 09/11; (Accessed 10/11).

33. Funk M: Academic Journal File-Sharing (CHE); liblicense-l@lists.yale.edu; 2009, 11/11; (Accessed 12/11).

34. "Visnja," An Article About SMSO and SMSO Members



))) Special Regards to All Active SMSO Members!;

Facebook Posting;

http://www.facebook.com/posted.php?id=171602826061&share_id=168693639775&comments=1#s168693639775; 2009, 03/11, (Accessed 07/11/2009).

35. Funk M: Academic Journal File-Sharing (CHE); liblicense-l@lists.yale.edu; 2009, 12/11; (Accessed 13/11).

36. Anderson K: <(Private email to author)>; ISPUB Paper; 2009, 13/11; (Accessed 13/11/2009).

37. Antelman K: Do open access articles have a greater research impact? College & Research Libraries News; 2004; 65(5):372-82.

38. Eysenbach G: The open access advantage. J Med Internet Res; 2006; 8(2):e8.

39. Eysenbach G: Citation advantage of open access articles. PLoS Biology; 2006; 4(5):e157.

40. Hajjem C, Harnard S, Gingras Y: Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. IEEE Data Engineering Bulletin; 2005; 28(4):39-47.

41. Davis P, Lewenstein BV, Simon DH, Booth JG, Connolly MJ: Open access publishing, article downloads, and citations: Randomised controlled trial. BMJ; 2008; 337:a568.

42. Brody T, Harnard S, Carr L: Earlier web usage statistics as predictors of later citation impact. Journal of the American Association for Information Science and Technology (JASIST); 2006; 57(8):1060-72.

43. Editor: Twitter Tapping. New York Times (Online Edition); (Editorial);

<http://www.nytimes.com/2009/12/13/opinion/13sun2.html?th&emc=th>; 2009; New York; (Accessed 13/12/2009).

44. Masters K: Opening the closed-access medical journals: Internet-based sharing of institutions' access codes on a medical web site. Internet Journal of Medical Informatics; In press.

Author Information

Ken Masters, PhD
ITHealthEd, Austria