Free/Libre and Open Source Software: Policy Support

FLOSSPOLS

Deliverable D05
ETHICAL GUIDELINES

Revised: October 20, 2004

Ethical Guidelines for Online Research

(Bernhard Krieger)

About this document: purpose, development and validation

Much of the research undertaken for the FLOSSPOLS project will take place online. As outlined in the proposal for this project 'specially the online developer survey and the ethnographic studies would form an intrusion into the private sphere of the respondants'. However '[n]either European and national directives nor the ethnographic disciplinary tradition inform on how one should undertake research on human beings in online contexts, where much of the information used is technically publicly available, but may become sensitive in the context in which it is collected and presented'.

In this document we will describe 'ethical guidelines, which will be considered in the phase of data collection, data analysis and dissemination of the findings'. These guidelines have been developed as rules of conduct during the whole project as well as a generic 'recommendation for researchers undertaking investigations on human beings in online contexts'. This paper aims to look at the specific situation of Internet research and how established ethical principles can be applied to it.

Since most of the ethically critical research for the FLOSSPOLS project is undertaken in the ethnographic studies these guidelines are based upon principles established in the anthropological discipline. The deliverable has been written by one and supervised by other researchers of the University of Cambridge. It thereafter has been discussed with the concerned consortium member (University of Maastricht) and adjusted to their corresponding requirements. Furthermore we consulted several members of the subject community to comment on the guidelines. In reference to the work package W3 (developer survey methodology)and deliverable D6 (developer survey design) the relevant ethical aspects will underlie additional evaluation in the pretests of the survey among further community members.

Introduction

The "Association of Social Anthropologists of the UK and the Commonwealth" (ASA) identifies

five groups towards which a researcher has an ethical responsibility. There are the participants of a study, the people supporting the research (funding organizations, local institutions, etc.), the colleagues within the discipline, the governments of the researchers and the host country, and the wider society (ASA 1999). Research of the free software / open source community is quite comparable to more 'traditional' research projects in terms of ethical responsibility towards the last four groups. However, there are some new ethical challenges with regard to the first of these groups, who are the participants of the community.

The predominant ethical aspect of the research on which we will reflect in this document relates to a specific characteristic of the phenomenon of investigation: The production of open source software does not take place in one specific location, but is dispersed throughout the world. There are of course regional concentrations. People interested in the operating system Linux, for instance, regularly meet in so called Linux User Groups (LUGs), often in bars or coffee shops in order to discuss technical issues or to enjoy the company of other people interested in Linux. However, the existence of the open source phenomenon cannot be thought of without recent developments in information and communication technologies, specifically technologies related to the Internet. Community member use various technical channels such as websites, emails, mailing lists, chats or newsgroups in order to stay in contact and cooperate with each other. These technical channels actually are foundational for the community. It is this context of virtuality which deserves special consideration during the design and conduct of online research as well as in the dissemination of the results.

Whereas the literature of general ethical problems in the research of the free software phenomenon is very limited (Emam 2001; Singer n.d.), a discussion of research ethics in virtual environments in general is under way. Alluding to the principle of minimizing possible harm and maximizing potential benefits of research for its participants, the American Association for Advancement of Science (AAAS) identifies two major ethical aspects that should be particularly considered during design and conduct of the research on human subjects on the Internet, as well as for the dissemination of the results. There are informed consent and protection of privacy and confidentiality (Frankel and Siang 1999). Let me start with the former.

Informed Consent

 If we look at the history of research ethics, the need for an informed consent for research on human beings became dramatically clear in the context of the massive violation of human rights by

German scientists in the concentration camps of Nazi-Germany and Eastern Europe. In a reaction to medical experiments on women and men the "Nuremberg Code" stated in its first paragraph the necessity of consent of human participants to medical research:

> "The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be situated as to be able to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress, over-reaching, or other ulterior form of constraint or coercion, and should have sufficient knowledge and comprehension of the elements of the subject matter involved as to enable him to make an understanding and enlightened decision." (Nuremberg Trial 1949 p. 181)

Facing the shock of crimes against humanity of nazi scientists this statement seems self-evident. In medical research the individual subject must be informed about the risk of the research to her/his health so that s/he can thereafter decide whether or not to participate. This procedure is based on the assumption that during the research the personal domain of human subjects (e.g. his/her body or psyche) is intruded upon. However in social sciences one can argue that some methods do not intrude on the personal domain of the participating subject. For instance by observing interaction on a Dutch flower market researchers can learn about a specific type of auctioning model without doing any harm to any participating actor. It is therefore not necessary, one could argue, to get the consent of the participating actors. The underlying argument then would be that one would only need consent for the conduct and dissemination of research if harm to the involved subjects were anticipated and research in the public domain can be carried out without the consent of the participating human beings.

Apart from the question of whether such a claim could legitimately be employed for research in general, in the case of the free software / open source community it opens up the problem of how one conceptualises the Internet. It is probably people's most obstinate assumptions that the Internet is freely available and this is the source of the claim that everything which happens 'in' it, belongs to the public domain. This is of course not the case. Like in the offline world, access to many resources is only conditionally permitted. But even if access is possible to everyone, at any time; even if we agreed that the Internet is a public space such as a park or a metro station, this does not mean that everything that happens 'in' it would belong to the public domain. For this would imply – to proceed with the offline analogy – that a conversation in a tramway is part of the public domain, simply because it is audible in a publicly accessible space and therefore would not require consent by the speakers to be included in research.

However this is exactly the position of some researchers who investigate online phenomena. The report of the Association of Internet Researchers' ethical committee mentions the example of chatrooms and states the opinion of some of its members:

> "[I]n contexts such as chatrooms which are always open to anyone and thus are "public" in a strong sense, and in which (i) user names are already pseudonymous and (ii) in light of their option to always "go private" if they wish, users thus choose to participate in the public areas of the chatroom and may thereby be understood to implicitly give consent to observation." (Ess 2001)

Other members of the committee disagreed on this issue however. They made the argument that "participants in chatrooms have a reasonable expectation that their communication is ephemeral and is not being recorded or studied without their consent" (ibid)..

Cavanagh (1999) argues that informed consent for research on participants of online interaction is not necessary. She refers to Goffman (1971), arguing that the self has no individual foundation but is a "reflexive constitution by and of the social world." (Cavanagh 1999). Thus as in offline contexts, the online self is "produced through ritual, through the practices and relations which constitute the intersubjective fabric of the online social world." (Cavanagh 1999). The self becomes thereby the product of its social environment. What then follows is a denial of personal agency:

> "In seeing textual production online as a form of self presentation and production which occurs within co-present, co-ordinated spaces of interaction we divorce the text from the subjectivity of the "author", aligning it instead as interactive ritual. Thus we are considering, not the expression of individual personalities, but the strategic means and forms of interaction within the media. The data is therefore, by implication, a product not of individual agency but of social ritual" (Ibid)

Consequently, she argues that "the data takes the form of an insight which is not peculiar to any specific individual and therefore does not attach a need to obtain informed consent from the participants" (ibid). However, even if we considered conversation online as a ritual, this would not imply that either the group, or the participants dispense on any right against its appropriation. Harrison, for instance, demonstrated how individuals and groups compete over access to rituals as a resource (Harrison 1992). Many would certainly not agree on the principle appropriability of rituals by researchers or others.

Cavanagh's approach does not seem to provide a solution to these ethical problems. However the wider questions she raises are still important: Is material which is freely accessible over the Internet open for unlimited use, including research? Does the fact that this material is exposed in the public space mean, that one would not need the informed consent of the observed participants? These theoretical questions regarding informed consent relate on an operational level to a further, more practical problem in online research: Even if the researcher has the consent of a possible subject for his/her investigation s/he faces the problem, whether s/he can be sure that the person did understand purpose, content and possible consequences of the research. It is common practise not to undertake research on mentally impaired people or children without the consent of their custodians or parents. It may be difficult to verify the mental condition and the actual age of a possible informant in an

online context. However, since operating a computer and participating in online communication requires a certain amount of intellectual capability, we expect to encounter relatively few free software participants who will not be able to comprehend possible harm arising from research. Concerning the question of age, one could think of meeting young people, who participate in the community. However the F/LOSS survey showed that less than one percent of the programmers were under the age of 16, which we consider to be old enough to make an informed decision.

Privacy

 The second issue we will discuss in the context of research in virtual environments is linked to, but still different from the question of informed consent. Privacy is certainly an issue Internet users are much concerned about. However, researchers assuming that everything that happens 'in' the Internet belongs to a public domain and is therefore open to research activity, often disregard the "right to confidentiality and anonymity" (ASA 1999 p. 4)[1]. There are two aspects of privacy on which we want to focus in the following paragraphs. They concern different stages of the online research. The first regards the ASA guideline 5a: "Care should be taken not to infringe uninvited upon the 'private space' (as locally defined) of an individual or group" (ASA 1999 p. 4). Some Internet users in general and some free software participants in particular may not want to be subject of research at all because they appreciate the value of privacy in their online environment. This is especially important to consider before and during the conduct of investigation. The second aspect is especially important during the phase of the dissemination of the research results. It concerns the anonymity of people who participate in the research, and decide that their identity should be disguised. The ASA guidelines propose means such as "the removal of identifiers, the use of pseudonyms and other technical solutions to the problems of privacy in field records and in oral and written forms of data dissemination." (ibid). We shall discuss some of these means, which could ensure the anonymity for possible participants.

But first we will report on examples in which researchers ignored the desire of their subjects in both regards. By violating basic rules for privacy they missed the goal of minimizing possible harm for the researched community.

King describes the discouraging effects of researchers undertaking investigation of Internet communities, which are set up to discuss sensitive and personal information. He reports on reactions of a support group finding out that researchers targeted their community. He quotes one of its members:

> "The reason I will not say anymore about myself is that since, I have become a member of this list, I have been seeing more and more postings from students doing research papers or working on their advanced degrees or journalists looking for interviews for articles. When I joined this, I thought it would be a *support* group, not a fishbowl for a bunch of guinea pigs. I certainly don't feel at this point that it is a "safe" environment, as a support group is supposed to be, and I will not open myself up to be dissected by students or scientists." (King 1996 p. 122)

A second example goes even further. In 1994, Finn and Lavitt published an article on an online self-help group on sexually abused people (Finn and Lavitt 1994). Without the consent of any of the group's members the authors observed, downloaded, analysed and published postings made by the group. Statements of the groups' members were quoted verbatim including the exact date and time of the posting. By disguising only the names of the group's members the researchers did not do enough to protect their anonymity. Verbatim quotes are traceable by search engines such as 'Google' or 'Dejanews' even by not especially computer-literate people. This includes not only the content of websites, but also of other electronic channels such as email lists or newsgroups. Therefore an identification of individual participants is easily possible. In any case, the authors did not even try to protect the anonymity of the group. By naming it, they violated their privacy, which caused not only psychological harm to the members, but posed a threat to the very existence of the community itself. Due to the loss of trust in the confidentiality of the group, the participants hesitated to openly discuss their very private issues, which was the purpose of the community. Privacy as defined locally in the context of the Internet does not concern a certain place, but a technical matrix. The research and publication strategy of these researchers would therefore actively break the above-mentioned rule of conduct (see ASA 1999 paragraph 5a).

This example also demonstrates problems some researchers seem to have with the correct use of pseudonyms. This concerns especially the dissemination of the research results. Finn and Lavitt argue that their participants' identities were disguised because they only quote their pseudonyms in their publication. However what they mean by pseudonyms are the names their researched subjects chose during their online activities. The protection of their participants' anonymity is to a large extent active only for their offline identities, since most people could not link these used names with the names in the passports of the participants. However in the context of computer mediated communication people often are identified more with these names than with the names they use in offline contexts. Much research points to the fact that the construction of an online self is perceived as no less important for many people's identity than their offline self (see e.g. Turkle 1995; Markham 1998). The fact that people separate these two parts of their identity does not imply that only one part has the right to be protected during research and in its dissemination. Therefore if people use nicknames for the participation in an online forum, researchers should not misunderstand them as pseudonyms, but treat them with the same confidentiality as they do with

names people use in offline contexts.

Conclusion

It is clear that any intrusion of researchers into a community is certainly disturbing for its members. This is a phenomenon also experienced in offline communities. So what are the ethical consequences of all this for research practises? Can online environments be legitimately used as a source of data, or should one rather limit oneself to employ them only as a tool to communicate with possible informants that one meets offline? We think the basic distinction between public and private can be used as a tool which can help to decide how to deal with data accessible on the Internet. Thereby it is important to consider that what is private and what is public does not depend on legal requirements or on the researcher's own opinion, but should be seen from the perspective of the members of the respective group. However, we are convinced that not all material that is available online requires informed consent. For example, it is the intention of the creator of a website to present content to the World Wide Web. There are technical means to protect parts or the whole of a website from unauthorised usage. Therefore websites can certainly be regarded as a kind of publication and are therefore subject only to legal requirements of intellectual property right protection.

More difficult decisions rise in cases when the person that provides the content is not the same as the creator of the forum on which it is provided. As shown above, the content of these forums may not be regarded as belonging to the public sphere. There is certainly scope for the interpretation of the situation. However there are some signs, which give users and researchers a hint towards an anticipated attitude of the forum's handling of privacy issues. King proposes that the degree of accessibility to the data is such an indication (King 1996 p. 122). People posting on temporarily limited online chats of news websites probably assume less privacy than members of an Internet community to which one has to be registered in order to participate. Also the size of an Internet forum provides an indication of the perceived degree of privacy. A group such as the Linux kernel developers' mailing list with several hundred developers and many thousands of postings per year is less likely to count as a private domain than the email distribution list of a small local Linux User Group (LUG) with only a few participants that probably share also other social contexts. Furthermore the topic of an Internet forum gives some ideas about the perceived privacy of its members. Forums set up to deal with mainly technical questions are certainly thought to be more for the public than groups set up to discuss private issues (see example described by King above).

Certainly most free software communities are not set up to discuss sensitive personal information. However, what may be considered as confidential information that is shared only with people of a specific forum is not necessarily obvious for the researcher. One has to reflect on possible implications of data gathering, analysis and publication.

Concerning the question of confidentiality one will have to take care to efficiently protect these participants who do not want to be credited for their participation and wish to remain unidentified. As discussed above, pseudonyms have to be chosen carefully in order to protect individual members' privacy. Furthermore groups may be protected as a whole against their identification. Bruckman proposes a couple of strategies that can be employed, depending on the degree of confidentiality. She thereby mentions means such as "The group is not named", "Names, pseudonyms and other identifying details are changed", "Verbatim quotes are not used if a search mechanism could link those quotes to the person in question" (Bruckman Forthcoming p. 35). These or other techniques to disguise the identity of single individuals and whole groups who do not want to be identified will be employed in cooperation with the participants during and after the research. Of course credit has to be given to all individuals and groups whose participation in the research helped to achieve the results.

Finally there are certainly several degrees of intrusion during the different stages of the research. For anthropologists, participation is by definition a method of data collection. The simple fact of his/her presence in the field – be it a technical matrix used in online communication or a physical setting offline – already implies an intrusion. The researcher will make notes, take screenshots, save emails and other forms of communication. This is much comparable to offline situations in which a researcher is using a notebook to keep record of discussions or a camera to take pictures and film sequences. Therefore also in online situations researchers should make clear to the people s/he is working with that s/he is a researcher.

Obviously an anthropologist will make use of the information to digest it into fieldnotes. These fieldnotes are actually written by the researcher and are one of the main sources for later analysis. However, this does not automatically imply the right to actually publish the primary material s/he was collecting without the explicit consent of the persons involved (see above). In any case, all material which makes single persons as such identifiable, be it a photograph, an email or a quote from an online chat needs permission for actual publication.

This paper has been written to make researchers aware of ethical problems that occur in the context of research in online environments. It proposes that the ethical questions are comparable to those in

conventional research settings among human beings in social contexts. However the circumstances of online research make it more difficult for the researcher to evaluate the situation and to adjust his/her behaviour appropriately. S/he therefore has to more closely reflect on the respective situations and take over more responsibility for his/her own action.

Regarding the FLOSSPOLS surveys, the consequence of the ethical measure we postulate here is that we make as sure as possible that the persons we address with this survey get to know the purpose and background of the survey in order to allow an informed consent. Moreover, we have to assure our respondents that we do not intend to violate their privacy and that all information they give us will be kept strictly confidential. This warranty includes the assurance that the survey is not for commercial purpose and that we will not publish any result that would allow identifying our respondents. A critical aspect with regard to the FLOSSPOLS developer survey is that we cannot exclude that underage persons will be surveyed, since the average age of the FLOSS developer community members is quite low. We will therefore explain that we do not intend to survey underage persons systematically and that we are aware of the problems that are aligned with it, but that we acknowledge the fact that young persons are active in the community and that we ask only questions that are directly related to these activities within the community or that are necessary to interpret the responses to these questions properly.

1. See also the "Statements on Ethics – Principles of Professional Responsibility" issued by the American Anthropological Association (AAA 1971 (amended 1986)): "Informants have a right to remain anonymous. This right should be respected both where it has been promised explicitly and where no clear understanding to the contrary has been reached."

References:

AAA. 1971 (amended 1986). Statements on Ethics - Principles of Professional Responsibility. http://www.aaanet.org/stmts/ethstmnt.htm: American Anthropological Association.

ASA. 1999. Ethical Guidelines for Good Research Practice. http://les1.man.ac.uk/asa/Ethics/Ethical%20Guidelines.pdf: Association of Social Anthropologists of the UK & the Commonwealth.

Bruckman, Amy. Forthcoming. Studying the Amateur Artist: A Perspective on Disguising Data Collected in Human Subjects Research on the Internet. *Ethics and Information Technology* Draft version available online on http://www.cc.gatech.edu/classes/AY2003/cs6470fall/bruckman-names.pdf:1-38.

Cavanagh, Allison. 1999. Behaviour in Public? Ethics in Online Ethnography. *Cybersociology online on http://www.socio.demon.co.uk/magazine/6/cavanagh.html (last visited 3.4.2003)* (6).

Emam, K. El. 2001. Ethics and Open Source. *Empirical Studies in Software Engineering* 6 (4):291-2.

Ess, Charles. 2001. AoIR ethics working committee – a preliminary report. http://aoir.org/reports/ethics.html: Association of Internet Researchers.

Finn, J. , and M. Lavitt. 1994. Computer-based self-help groups for sexual abuse survivors. *Social Work with Groups* 17:21-46.

Frankel, Mark S., and Sanying Siang. 1999. Ethical and legal aspects of human subjects research on the Internet. New York: American Association for the Advancement in Science.

Ghosh, Rishab, Ruediger Glott, Bernhard Krieger, and Gregorio Robles. 2002. F/LOSS. Maastricht: International Institute of Infonomics.

Goffman, Ervin. 1971. *Relations in public: microstudies of the public order.* London: Penguin.

Harrison, Simon J. 1992. Ritual as Intellectual Property. *Man* 27 (2):225-244.

King, Storm A. 1996. Researching Internet communities: Proposed ethical guidelines for the reporting of results. *The Information Society* 12 (2):119-127.

Markham, Annette. 1998. *Life Online: Researching Real Experience in Virtual Space*. Walnut Creek: Alta Mira.

Nuremberg Trial. 1949. *The Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2*. Washington: US Government Printing Office.

Singer, Janice. Ethics In Situ. *http://cip.umd.edu/singer.htm (last visited 3.4.2003).*

Turkle, Sherry. 1995. *Life on Screen: Identity in the age of the Internet*. New York: Simon & Schuster.