



Exemption Policy re:

**Research Ethics Review
for Projects Involving
Digital Data Collection**

TABLE OF CONTENTS

I Definitions of Terms3

II Introduction5

 i Purpose and Scope of the policy5

 ii What are User Expectations of Privacy 5

 iii What is Digital Data Collection 6

III Ethical Issues of Digital Research.....6

 i Does list-mining, or like activities, qualify as research involving human subjects/participants?..... 6

 ii Is material posted online automatically in the public domain?..... 7

 iii Is list mining a form of naturalistic observation and isn't this type of research exempt from review? 7

 iv How can user expectations of privacy be assessed? 7

 v Is the use of archived content, even in discussion venues open to non-members, considered
 secondary use of data? 7

IV Summing Up.....8

V Criteria for Exemption from Ethics Review.....9

 i Exemption criteria relating to privacy 9

 ii Exemption criteria relating to confidentiality and anonymity 10

V1 Examples of online language relevant to ethics review exemption criteria
 for digital data collection 11

DEFINITIONS OF TERMS USED IN THIS DOCUMENT:

Chat Room: A virtual room where a chat session takes place. Technically, a chat room is really a channel, but the term room is used to promote the chat metaphor.

Content: Any material (e.g. text, images – still or moving, animation, sound) that appears on the internet in any capacity. In the context of research, specific content may be the object of a researcher’s analyses.

Digital Data Collection: Referring to the collection of data by means of digital technology including, but not limited to: internet websites, instant messaging (IM), voice-over internet protocol (VOIP), email, etc. by means of computer, phone, personal organizer or any other device that can be used to observe and/or communicate with other people.

Discussion Forum (plural = Fora): An online discussion group in which participants with common interests can exchange open messages. Forums are sometimes called newsgroups or conferences.

Email (Electronic Mail): Messages, usually text (sometimes images), sent from one person to another via computer. E-mail can also be sent automatically to a large number of addresses.

Instant Messaging: The act of instantly communicating between two or more people over a network such as the Internet. Instant messaging requires the use of a client program that hooks up an instant messaging service and differs from e-mail in that conversations happen in real time. There are many types of client programs such as MSN Instant Messenger, Yahoo Messenger and Skype.

Internet: The connection of two or more networks constitutes an internet - as in “inter-national” or “inter-state”. This encompasses local area networks (LAN’s). This term is also used to refer to the vast assortment of inter-connected networks that are linked via TCP/IP (Transmission Control Protocol/Internet Protocol).

List Mining: The practice of monitoring an online discussion venue for content relevant to a research project; and/or copying content (e.g. posted messages written by members of a discussion venue) into an offline document for analysis; and/or quoting material from an online discussion venue in the service of a research project, either in a publication, a presentation, or other form of information dissemination.

Listservs: A type of online discussion venue in which members who share a particular interest (e.g. parents of children with ADHD) can post and receive messages. A list owner/moderator manages submissions to the list and sends them out to list subscribers.

Network: A group of two or more computer systems linked together. There are many types of computer networks, including: local-area networks (LANs), wide-area networks (WANs), campus-area networks (CANs), home-area networks (HANs), etc.

Online: Connected to the internet, a LAN, the USENET, or any other type of information network.

Offline Identity: The real identity of someone who uses a pseudonym or nickname to identify themselves when they are online or the content they post.

Password: A code used to gain access (login) to a locked system.

DEFINITIONS OF TERMS USED IN THIS DOCUMENT:

Post, Posting, Posters: The act of entering a message into a network communications system. This term may also be used to refer to the message itself which may also be known as a “post”. Those providing the messages are sometimes referred to as “posters”.

Pseudonym/Nickname/Screen name: The name, or combination of letters and numbers, used by people to refer to themselves in the context of online communication.

Search Engine: A (usually web-based) system for searching information available on the Web

Terms of Use (ToU) or Terms of Service (ToS): A set of rules which are set up by the owner of an intellectual property or service to govern how they may be legally used.

TCPS: The Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. This document describes the policies of the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council of Canada (SSHRC). This joint Policy expresses the three Agencies' continuing commitment to the people of Canada to promote the ethical conduct of research involving human subjects and has been adopted as the minimum standard research ethics guideline by Queen's University.

USENET: A world-wide system of discussion groups, with comments passed among hundreds of thousands of machines. Not all USENET machines are on the Internet. USENET is completely decentralized, with over 10,000 discussion areas, called newsgroups.

User: Anyone making use of the internet to seek or provide information in the form of text, images, or sound. In the context of research, those who provide the content of interest to the researcher are referred to as users.

VOIP (Voice-over Internet Protocol): A method for communicating that converts analog voice signals to digital packets for transmission over the Internet or a private network. This term also refers to the various technologies, networks and software packages used to make telephone calls over IP networks, especially the Internet.

Website (or ‘site’): The entire collection of web pages and other information (such as images, sound, and video files, etc.) that are made available through what appears to users as a single web server. Typically all of the pages in a web site share the same basic URL, for example the following URLs are all for pages within the same web site:

INTRODUCTION

The General Research Ethics Board (GREB) and the Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB) are aware of the increasing frequency of research involving digital means of data collection and the potential for confusion over whether this type of research requires ethics review. Most digitally-based research that follows standard procedures of obtaining free and informed consent from individuals prior to collecting data, already fall under the category of human subjects/participants research. For example, a project involving an online questionnaire or a focus group in which all of the participants are aware that they are involved in research would certainly require ethics review. In contrast, a project involving content and frequency analysis of news articles about a particular topic (e.g. child poverty in Canada) would not require ethics review. But there are other types of digitally-based research that are not so easily defined.

Purpose and Scope of the Policy:

The purpose of this policy is to highlight the ethical issues involved in digital research and to specify appropriate criteria that researchers can use to assess whether their project requires human research ethics review.

This is an *exemption* policy in that the criteria specified on page 10 - 11 are to be used by researchers (and reviewers) to determine whether a project involving digital data collection is *exempt* from research ethics review. If a project does not meet all of these criteria, then it must be reviewed and approved by the appropriate REB at Queen's University before it can proceed.

If, according to exemption criteria, a project is not exempt from research ethics review this does *not* mean that the research project can not be done nor does it mean it will be subject to extra scrutiny. Rather, any project involving digital means of data collection under review will be treated like any other research project involving any other means of data collection.

This policy does not deal with issues pertaining to research involving digital data collection that are normally part of an ethics review, such as consent and debriefing procedures, privacy and confidentiality concerns, conflict of interest or methodological issues. This policy focuses exclusively on how to assess whether a digitally-based project is *exempt* from research ethics review. This assessment is based primarily on "User Expectations of Privacy"

What Are User Expectations of Privacy?

One point of view is that people who post anything online shouldn't have any expectation of privacy; that it is akin to shouting in the street. Another point of view is that individuals posting content online may have specific expectations about its transience, who may be viewing their messages, and how it may be used. In other words, they think of their posting as being in a private or semi-private space for a limited period of time.

A review of relevant literature and policies from other agencies revealed that there is general agreement that users do have some expectation of privacy, and, that these expectations vary according to specific characteristics of the site to which they are posting. For example, someone who posts a message in an on-line chat room that requires a formal registration process may expect that message to be viewed only by group members and not posted indefinitely. These individuals are not likely to expect their messages to be of particular interest to researchers for analyses or to be quoted in research presentations or papers. By the same token, users who post material on websites that are explicitly intended to be in the public domain, such as commercial news sites or weblogs (a.k.a. "blogs") can be considered to not have any expectations that their material should be considered private and/or confidential.

Although it can be argued that many users understand that their material may not be private, it is undisputed that there are also many users who are naïve to the possible uses of the information they post. These users can be described as "**technologically vulnerable**" and they are the people most in need of protection from the possibility of unwanted exposure in the service of research.

What is Digital Data Collection?

The solicitation or observation of digital information created or provided by humans is considered to be digital data collection. Digital information is text, voice, sounds, and/or still or moving images that have been created in a digital format or have been converted from another format to a digital format and are transmitted over various broadcast media (e.g. internet websites, private or public networks, email, text messages, voice messages, etc.). The transmission of digital information can take many forms and uses many different technologies. Here are some examples of research activities that are considered to be using digital data collection:

- Sending around a survey by email to be completed by selected participants who are asked to send it back either by email, regular mail, or by any other means of delivery;
- Running a focus group via an instant-messaging application and saving the text transcript of the session for later analysis;
- Recording a voice message conversation for use in a research project;
- Observing text messages posted in online fora (e.g. chat rooms) and copying and pasting users' messages into another application or transcribing them for use in a research project.

ETHICAL ISSUES INVOLVED IN DIGITAL DATA COLLECTION:

As the vast majority of digital data collection is typically done on the internet, the focus will be on the ethical issues of using material from this venue – with the understanding that similar issues apply to all forms of digital data collection.

Content posted on the internet may be viewed by some as an easily accessible database with no usage restrictions. The practice of using online material for research purposes is sometimes referred to as “list mining”. This term is commonly used to describe the process of looking for listservs (or other types of discussion venues) concerned with a particular topic or population, and then using the posted content on the site to pursue research questions. Typically, individual messages are copied from the online site and pasted into an offline document for text analysis or data categorization *without the message providers being informed that their material is being used for research*.

Researchers may believe that harvesting user content for their project without users' knowledge or consent poses no risk to anyone. But this may not be true. When identifying information is published outside of its original context, users may feel exposed or threatened and stop participating in sites they once enjoyed. The internet community from where the content was taken may also feel invaded or abused and possibly shut down, denying other users their service. Inadvertent exposure of an individual's offline identity from information published without their consent could have further consequences depending on the nature of the material (e.g. embarrassment, family upset, job loss, etc.).

Researchers do not set out to cause harm or distress to the people who may be involved in their projects. However, there is a lack of awareness in the research community about the potential harm that can be caused to internet content providers when their content is used and/or quoted without their consent. There is a sense that the internet is uncomplicated by the rules we apply to life offline. A brief review of the issues reveals that the internet is not free of social mores or expectations of privacy.

1. Does list-mining, or like activities, qualify as research involving human subjects/participants?

It is stated in Appendix I of the Tri-Council Guidelines that research requiring ethics review includes:

"... research conducted remotely [e.g. via mail, email, fax, phone] ... or from existing records not in the public domain ...".

Conducting research via the internet can be considered to fall under either or both of these categories.

2. Is material posted online automatically in the public domain?

Answers to this question vary depending on who is asked and the context of the material. This raises the issue of whose perceptions should dominate – the person who posted a message to a members-only listserv or the researcher who wants to harvest that message for analysis? In keeping with the spirit of the TCPS (i.e. respect for human dignity) we have determined that the user’s expectation of privacy should come first.

3. Is list mining a form of naturalistic observation and isn’t this type of research exempt from review?**TCPS Article 2.3:**

REB review is normally required for research involving naturalistic observation. However, research involving observation of participants in, for example, political rallies, demonstrations or public meetings, should not require REB review since it can be expected that the participants are seeking public visibility.

In Article 2.3 (above), it is clear that most research involving naturalistic observation requires ethics review but there are a few specific exceptions. One relevant criterion that determines whether naturalistic observation research needs ethics review is whether or not the people being observed are seeking to be observed. Similarly, when a user posts material on a site widely known to be viewable by the general public this may be taken as evidence that this user expects their material to have public visibility. In contrast, when a user posts material in a members-only chat room or discussion venue, or to a forum that is focused on a small sub-set of the population, this is to be taken as evidence that users do not anticipate that their material will be viewed by people outside this “community”.

4. How can user expectations of privacy be assessed?

Evidence of what user expectations of privacy may be can be deduced in several ways. For example, sites that have a members-only policy stated in their Terms of Use (ToU) or Terms of Service (ToS) are explicitly saying that the content is not in the public domain. Other websites may indicate that they consider all material on their site copyrighted and not to be used without permission, but allow some sharing of their content. For example www.cbc.ca states in their ToU that nothing on their site can be reproduced without their written permission. However, most of their news stories include an option to “email this story” to whomever the reader chooses. It is also acceptable to quote briefly from a news story or published work for the purposes of illustration or education as long as the source is given attribution.

Sites that require membership instill an expectation in their members that postings on that site will only be viewed by other members. Sites that explicitly describe their target membership (e.g. This site is for people grieving the loss of a pet) are setting out boundaries as to who should (and shouldn’t) become a member of the group and for what purpose. The title of a group (e.g. Sexual Abuse Survivors Forum) can be sufficient to implicitly state the membership criteria. However, the language describing group membership is not always so explicit.

5. Is use of archived content, even in discussion venues open to non-members, considered secondary use of data?

Secondary data is information that is used for a purpose other than that for which it was originally intended. Digital content is posted by users for a specific purpose: to provide or to seek new information, to communicate with members of their community, and/or possibly to seek general public attention. It would be rare for users to expect their content to be used in an unknown person’s research project. Thus, *all content used for research purposes without the user’s knowledge or expectation could be considered secondary data.*

In many cases, users’ content is archived indefinitely on the website to which it was originally posted. Users may or may not be aware that a comment they made in passing can be found in these archives for many years afterwards. In this respect, these archives of user content are also secondary data – if they are going to be used for a purpose other than originally intended. Use in an unknown person’s research would qualify as such a purpose.

Two articles in section 3 of the TCPS are relevant to the use of secondary data:

ETHICAL ISSUES INVOLVED IN DIGITAL DATA COLLECTION (continued):**TCPS Article 3.5:**

Researchers who wish to contact individuals to whom data refer shall seek the authorization of the REB prior to contact.

In the context of internet content, if researchers wish to contact the users who provided the content, whether it is for permission to quote their material or for additional information relevant to the research itself, they must seek ethics approval for this procedure.

TCPS Article 3.3:

If identifying information is involved, REB approval shall be sought for secondary uses of data. Researchers may gain access to identifying information if they have demonstrated to the satisfaction of the REB that:

- Identifying information is essential to the research;
- They will take appropriate measures to protect the privacy of the individuals, to ensure the confidentiality of the data, and to minimize harms to subjects; and
- Individuals to whom the data refer have not objected to secondary use.

Identifying information in the context of internet content could constitute a user's screen name or nick name which may be unique and well known (or readily known) to others. It could also take the form of information within the content that would allow others to discover the offline identity of the user. For example, descriptions (or images) of their home, their workplace, people they know, landmarks in their neighbourhood, etc.

SUMMING UP

Research involving digital data collection is a new frontier for researchers and research ethics boards. In this age of powerful search engines, indefinitely archived content, and fast-paced development of new technologies to collect and access information, it can be very challenging to protect the rights of individual human beings who are the source of digital data. But when material is to be used without free and informed consent the researcher is even more obligated to take all reasonable measures to ensure that their use of the material poses no risk of harm.

This policy is not meant to address all of the ethical issues involved in internet research. It is meant to give researchers a means of assessing whether their project can proceed without ethics review. On the following pages there is a list of eight criteria that, if met, will *exempt* a project from ethics review. That is, the project will be deemed to meet one of the exemptions from ethics review stipulated by the TCPS. For example, in Article 1.1 Research Requiring Ethics Review, research about an artist or a living person in public life based on publicly available information, archival materials or third-party interviews is not required to undergo ethics review. In Article 3.3 concerning secondary use of data, ethics review is required only when there is a risk that the data can be linked to individuals, and/or these individuals might possibly be identifiable in published reports or presentations.

The purpose of the criteria listed is to provide researchers, unit REBs, GREB, and HSREB with the means to evaluate whether a project needs ethics review. Projects that meet all eight criteria will be considered exempt from ethics review. Examples of research that meet (✓) and do not meet (✗) each criterion are provided below each of the criteria. Further examples relevant to the criteria can be found on page 11.

A project that does not meet each one of these criteria must be reviewed and approved by a research ethics board before it can proceed. Any doubt as to whether a project needs an ethics review can be addressed by contacting the Coordinator or Chair of GREB or HSREB (see: <http://www.queensu.ca/vpr/policies/committee.html> for contact information).

CRITERIA FOR EXEMPTION FROM ETHICS REVIEW RE: DIGITAL DATA COLLECTION

EXEMPTION CRITERIA 1 - 4: RELATING TO PRIVACY (✓ = meets criterion ✗ = does not meet criterion)

Criterion #1. The researcher can freely see the information online without having to register as a member of the discussion venue (e.g. chat room, discussion forum, listserv, etc.)

- ✓ Example: An open forum inviting anyone to voice an opinion of a particular issue. No membership is required to post or view a message. Nothing in the language of the homepage of ToU or ToS suggests that users may have any expectation that their messages will be afforded any degree of privacy

Here is a sample of unambiguous language to this effect from: <http://www.theatlantic.com/a/comments-policy.mhtml>

"If you do choose to post content on our Site, please do not do so unless you understand that your Postings will be available publicly. You should not assume that you are anonymous online and cannot be identified by your Postings."

- ✗ Example: A chatroom for university students interested in making friends. Membership is required to post messages.

Criterion #2. The discussion venue has no policy prohibiting research, or the reproduction, redistribution, and/or reprinting of its content.

- ✓ Example: The ToU or ToS as well as the homepage (or other explanatory information) reveals no statement restricting access to and/or the right to redistribute the content on the page and/or openly permits these rights Here is a sample of this type of language from: <http://creativecommons.org/licenses/by/3.0/>

"You are free: to Share — to copy, distribute and transmit the work"

- ✗ Example: Here is a sample of this type of statement from the ToU of "Facebook", a popular social networking site:

"You may not post, transmit, or share User Content on the Site or Service that you did not create or that you do not have permission to post."

Criterion #3. The owner/moderator of the discussion venue does not indicate in any way that messages posted on the site are private, confidential, or meant to be viewed only by people matching the characteristics of a particular interest group.

- ✓ Example: A forum called "BlogSite" invites the public to view and rate blogs listed by category
- ✗ Example: The moderator of a forum for people recovering from a particular illness grew tired of researchers attempting to recruit members of the forum for research. On the homepage and on the page for subscribing to the forum, the following message was prominently displayed:

"This forum is for people in recovery from X illness. It is not open to researchers for recruitment or observation purposes"

Criterion #4. The researcher will not be engaging in any deception of the users, such as 'posing' as a member of a group or suddenly becoming a regular contributor to the discussion venue. Similarly, the researcher, or anyone working on behalf of the researcher, was not already a member of the discussion venue for their own interest and is now making use of their membership for research purposes.

- ✓ Example: A researcher is analyzing opinion pieces on a particular topic posted to an open newsite.
- ✗ Example: A research assistant belongs to a chat room frequented by a particular interest group. This group is of interest to the researchers running a project she is involved in. At the request of the research team she becomes a regular contributor to the chat room, doing so in order to provoke responses relevant to the project.

EXEMPTION CRITERIA 5 - 8: RELATING TO CONFIDENTIALITY AND ANONYMITY

Criterion #5. The researcher will not directly contact any human sources of digital content.

- ✓ Example: A researcher has been collecting samples of Blogs (short for weblogs – a type of online diary or opinion page) for the purpose of analyzing the frequency and variety of acronym slang (e.g. LOL = laughing out loud). As he did not intend to use any identifying information in his analysis of the text he did not seek research ethics review.
- ✗ Example: The same researcher decides it would be interesting to see whether variations of acronym slang vary as a function of the writer’s birthplace, first language and gender. He intends to email each “blogger” to obtain this information. Now he must seek ethics review and approval before proceeding.

Criterion #6. Websites that are frequented by minors (persons under the age of 18) must not be part of the research project. Similarly, digital communication in any form that can be identified as being from a minor must not be included in the data collection.

- ✓ Example: A site that invites contributions of videos for public viewing by aspiring film makers includes age of the contributor in the biography information. A research project chronicling the phenomenon of online video creation avoids analysis or mention of any videos contributed by minors.
- ✗ Example: A site advertising toys for elementary school children invites young viewers to write reviews of the products. A researcher wants to analyze these reviews re: consumer knowledge of children. This project must come for research ethics review.

Criterion #7. The research plan must indicate that the identities of users (including nicknames and pseudonyms) will be disguised in any presentations or publications of the research works.

Criterion #8. The research plan must indicate that verbatim quotes will not be used in any presentations or publications. Paraphrasing and coded categorization of content are acceptable.

- ✗ Example re: #7 and #8: In a paper reporting an analysis of a listserv discussion about political activism, the researcher wishes to include samples of the conversations. One actual conversation looks like this:

SIgon12: ...almost got pepper-sprayed -- I smacked the cop with my sign and ran ...
AllyB08: I make my own spray – cayenne pepper and vinegar in a squirt gun. Try it!

- ✓ Example: The report of the conversation could be coded as in “Close to 50% of members advocated using pepper spray”. An acceptable anonymized and paraphrased version of individual comments could look like this:

“One member reported almost being sprayed with pepper by police during a protest. Another member gave out instructions on how to make pepper spray at home.”

EXAMPLES OF ONLINE LANGUAGE RELEVANT TO ETHICS REVIEW EXEMPTION CRITERIA FOR DIGITAL DATA COLLECTION

Note: The examples given here fall under the category of “fair use” in which short excerpts of materials may be quoted for educational purposes or in order to illustrate a point.

1. Indication that a discussion venue is not restricted to members only:

e.g. <http://www.caj.ca/services/list-serv-rules.html>

“Please note that the CAJ chooses to keep the list open to non- CAJ members, who are welcome as guests even as we encourage you to support Canada’s only national professional journalism organization by taking out (an inexpensive!) membership.”

2. Indication that a discussion venue is restricted to members only:

e.g. <http://www.ire.org/membership/listserv.html>

“The IRE-EDU Mailing List serves as a forum for journalism educators, particularly for those teaching public affairs, investigative or computer-assisted reporting. (This is a members-only listserv.)”

3. Indication that a discussion venue allows reproduction and/or redistribution of its content:

e.g. <http://www.mrm.org/about/you-may-freely-redistribute-our-content>

“Unless otherwise noted, you are permitted and encouraged to reproduce and distribute our articles in any format provided that you do not alter the wording in any way and you do not charge a fee beyond the cost of reproduction. For web posting, a link to our website is preferred. ”

4. Indication that a discussion venue restricts the reproduction and/or redistribution of its content:

e.g. http://www.peo.on.ca/Web_site_term_legal_restrictions.htm

“Professional Engineers Ontario (“PEO”) authorizes you to view and download the materials at this website (“site”) only for your personal, non-commercial use, provided you retain all copyright and other proprietary notices contained in the original materials on any copies of the materials. You may not modify the materials at this site in any way, or reproduce, copy, publicly display, distribute or otherwise use them for any public or commercial purpose. ”

5. Indication that a discussion venue is meant for a specific group of people (as defined by the list owner/moderator):

e.g. <http://acor.org/faqs/lists/listrules.html>

The ACOR lists are primarily for cancer patients, their supporters and loved ones, and health professionals who wish to lend their support and knowledge. If you are subscribing to the list for any other purpose, please contact the list managers PRIOR TO POSTING ANY MESSAGE or beginning any research project.

6. Indication that submitted content to a discussion venue is not confidential:

e.g. <http://www.itvs.org/beyondthefire/privacy.html>

“You agree that the material you post to this site shall not be confidential. Content that is posted to this site may be used and reprinted by ITVS for uses outside of this Website.”