Running head: Data Security

Data Security in the Information Age:

Important considerations for protecting electronic data and participants while conducting research

Jamison E. Judd

Matthew J. Ross

University of Connecticut, Neag School of Education

Abstract

This paper presents a technical view of the ethical concerns researchers must consider when working with electronic data. It explores the various steps necessary to protect the fruits of research, in a best-practices manner. Topics covered include proper collection, storage, protection, erasure of data, and security, as well as participant identity protection. Security, in particular, has become an issue in the collection, storage, sharing, and destroying of data. Issues related to electronic data are now becoming a hot topic for many Institutional Review Boards. The intent is to leave the reader with a general knowledge of the complex problems inherent in ethically handling electronic data and provide an outline of solutions for these problems.

Data Security in the Information Age:
Important considerations for protecting electronic data and participants while conducting research

What responsibilities does a researcher have in regards to securing data? The answer to this question can be quite complex. The data should be secured first and foremost in accordance with the federal, state and local mandates. It should also match the any institutional policy that is currently in place, which will often be stated in the IRB proposal. Securing data is a serious matter, as much research that takes place contains personal information. Electronic formats for data pose special problems, and it is my goal to tackle some of them in a helpful and informative manner.

Anyone receiving federal funding, or at an institution, which receives federal funding, must comply with the policies set forth by the Office of Research Integrity. The program set up by the ORI requires institutions to create their own internal review process that monitors research activities. The current code in effect dates back to 1989, and does not include any specifics on electronic data (Office of Research Integrity, 1989). It is assumed to be handled like paper-based data in terms of security. The APA also lacks any specific rules or regulations regarding electronic data in their latest code of professional ethics (APA, 2002). They have however, developed a "Statement by the Ethics Committee on Services by Telephone, Teleconferencing, and Internet" (APA, 1997). This statement, briefly summarized, says that for any medium used for research that is not specifically noted in the code of ethics, the psychologist has an obligation to use their best judgment in making sure that research is done in an ethical manner. Final consideration of data security should be addressed by your institutional review board, and they can help you in complying with the necessary regulations.

A survey is a document that provides a researcher with a “detailed critical inspection” of a particular subject or event (WordNet, 2004). Research surveys have traditionally been administered orally or in written format. With the continued growth of availability to the Internet in all levels of society, and the innovation of newer and easier programming solutions, an increasing number of surveys are being administered online.

Web-based surveys contain the same elements as their written/oral predecessors. Questions can be formed using multiple choice, quantitative responses, and qualitative responses. Answers are entered using text boxes, text fields, radio buttons, check boxes, and list boxes/menus. The many methods with which to phrase and answer questions give the researcher a multitude of ways to ask key questions and in some cases multiple times.

Surveys are typically designed with the following components: an opening page containing information about the research being done, a consent form, and the survey questions. In the case of some web-based surveys, an identifier page (establishing the identity of the participant) is also included. Before a participant begins the survey, he/she must consent to participating in the research.

Early forms of web-based surveys used email as the means of data collection. Survey results were returned to the researcher in the body of an email. This presented a very time consuming task for the researcher, as the data was then transferred to an analysis tool. The chance of error in the transfer was great, and the chance for lost data due to transmission errors was greater. Today, many web-based surveys are designed as a front-end to a database. A database is a collection of tables relating to a particular function. The tables contain rows (also known as records or cases) and columns (also known as fields or variables). By entering the data

directly into a table within a database file, the researcher can port the data set directly to an analysis tool (SPSS, Excel, etc.) without error.

Obtaining Consent

In conducting any research, the first element is the consent form. Obtaining informed consent is essential to successful research. "[Consent] is more than simply having a potential research participant sign a consent form; it is a process by which necessary information is communicated to the participant by the researcher." (Sales, Folkman, 2000) Obtaining consent can be difficult when administering web-based surveys, dependant on the participant pool. Obtaining the consent of a minor can be very challenging when done over the web. A parent or guardian must also consent to the child participating. Many commercial websites avoid underage populations by requiring a valid credit card number be entered. Although this assumes that every person over the age of 18 has a valid credit card. It also does not embrace the growing trend for ATM/Debit cards and the growing number of under-18 year olds with these cards. Disclosure statements are also another effective way of dealing with underage issues.

On the other hand, obtaining consent of an adult by means of the Internet is much easier than the inconvenience of arranging a meeting, accommodating travel time, and other interference. A researcher can place a consent form for any study on the web, and ask participants to review and accept/decline far in advance of any research activity. The researcher might also have a brief pre-screening survey that can be administered prior to any visits, for time saving analysis.

A person's identity is protected by the type of consent that person gives. When designing web-based surveys, one method of obtaining consent is more challenging than the other. The easier of the two methods of obtaining consent is consent with confidentiality. This method

allows the programmer to attach a true, unique identifier to the data collected. The agreement between the participants and the researcher bars the researcher from releasing that identifier to any other source. It merely acts as a means for follow-up. This unique identifier might be the participants' social security number, or piece of that number, a unique id associated with the participants employer, university, or school, or some other pre-arranged identifier. This unique identifier can be used to prevent the participant from returning to participate a second or third time, especially in web-survey situations where data is collected only once. This construct must be programmed carefully, as will be discussed later.

Anonymity is the harder of the two methods of consent to accommodate as a programmer. The identity of the participant must be kept separate from the data. This involves a few additional lines of code in order to ensure that the being entered by the participant is kept together without tying the response to a personal identifier of the participant in any way. A pseudo-identifier must be developed by the programmer to accomplish this. The most common problem with data collection under the constraint of anonymity is the ability to discern that a participant has not previously participated, and/or is not misrepresenting him/herself.

One design, used by the staff at the Neag School of Education, for informed consent can be applied to either confidential or anonymous consent. In this design, a unique pseudo-id is set for the participant. In a confidential and some anonymous situations a unique personal identifier is also collected; however, it is not stored until the data has been completely collected. The method of informed consent determines how that identifier is stored. In a model where confidentiality is the method of consent, the identifier is stored with the data after the survey has been administered (Diagram 1).

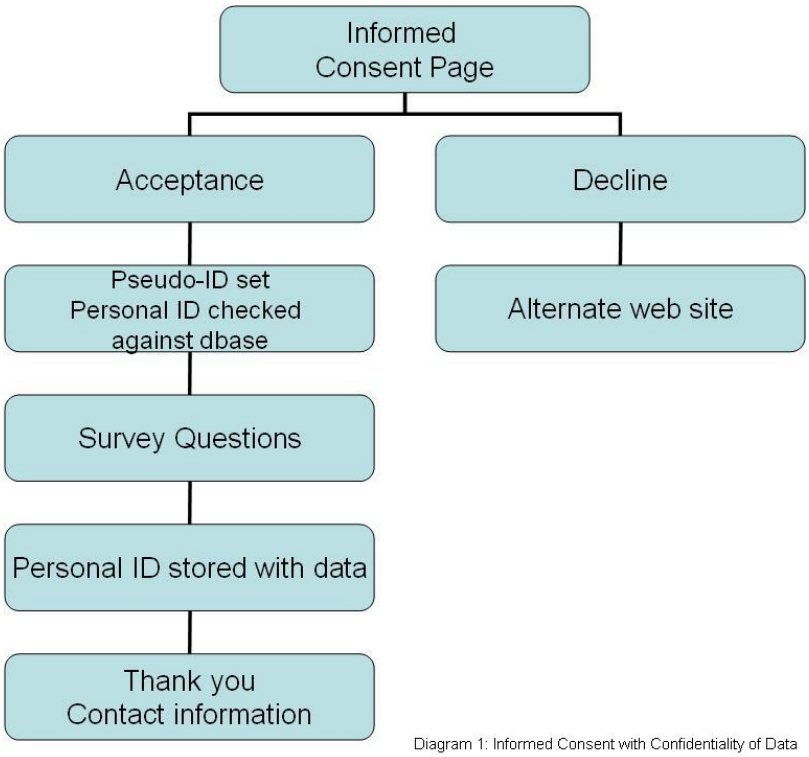


Diagram 1: Informed Consent with Confidentiality of Data

The personal identifier is needed at the beginning to determine whether the participant has already taken the survey. A participant taking a survey more than once can pose a threat to the integrity of the data.

In the case of collecting data anonymously, the data collection can operate in two ways. As mentioned previously, one of the key problems with anonymous data collection is ensuring data integrity. If identity is completely anonymous, the researcher can not check the data to make sure a participant did not complete the survey multiple times.

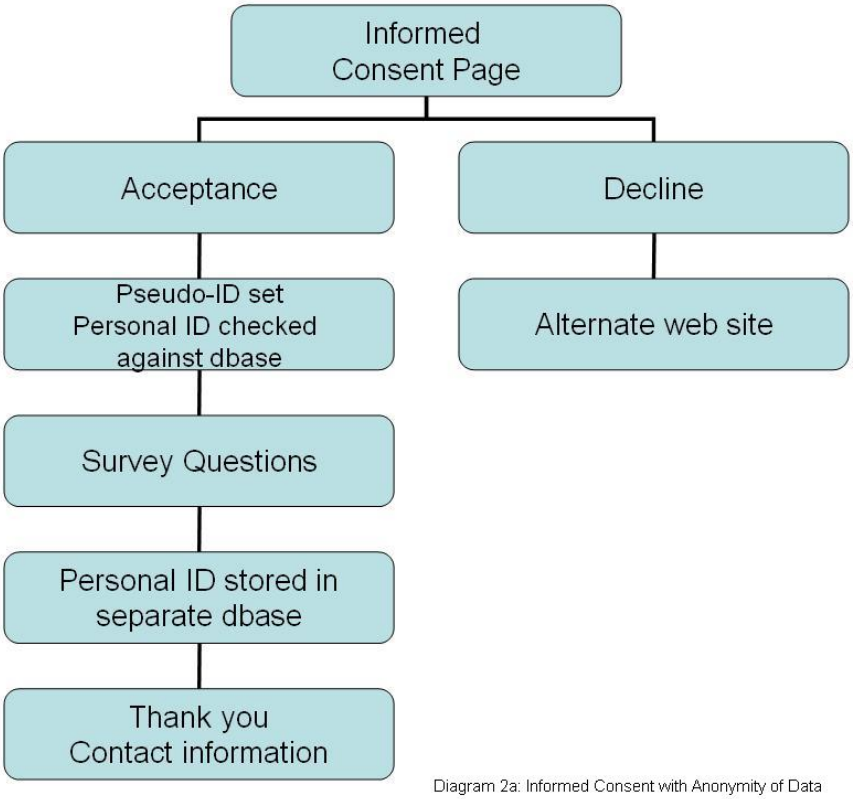


Diagram 2a: Informed Consent with Anonymity of Data

Diagram 2a shows the personal identifier being collected and checked against a database table, similar to the method used for confidential data collection. However, later in the process, the personal identifier is stored in a separate database table, rather than with the data. This identifier is also never stored with a time/date stamp. This new table is that which the personal identifier is check against earlier in the process. Because the identifier is stored in a separate table, the researcher can not connect the data to the participant. The personal identifier table can also be destroyed after the data collection has occurred, further protecting the participant.

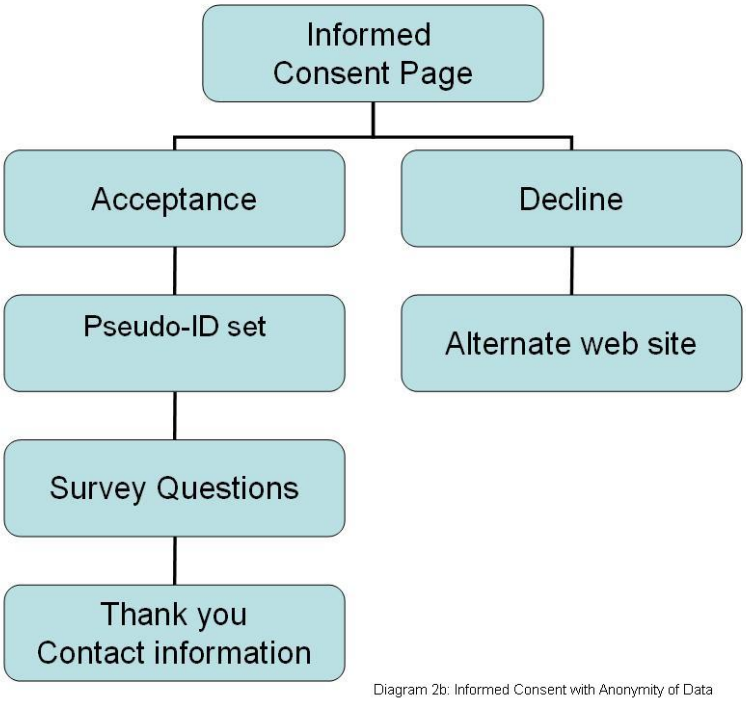


Diagram 2b: Informed Consent with Anonymity of Data

Diagram 2b shows the flow of data collection when no personal identifier is being collected. This method should be reserved only for extreme cases of anonymity, as data integrity can not be determined unless the participant is taking the survey in a controlled situation.

The code for these designs is rather simple. A sample of code used for a Professional Development School Teacher Evaluation is given in Appendix A. This example is for confidential data applications. It shows the personal identifier being collected along with the program of study. In the action, the code checks that personal identifier against a separate table containing all the personal identifiers of completed surveys (table PDS_Teacher_ID). If the identifier already exists, the participant is directed to an error page, and then to an alternate website. If the identifier does not exist in the second table the participant continues with the rest of the survey.

This design is important to prevent participant misrepresentation, intentionally bad data, and possible data manipulation. By eliminating the possibility of a person completing the survey more than once or intentionally taking the survey multiple times to skew the data, the survey instrument is more reliable and valid. It also protects the participant by making the data truly confidential or anonymous, depending on the application. LAN administrators and programmers can create databases that encrypt the personal identifier making it impossible to read the identity even if the database table is shown to the researcher or another person. The personal identifier would show as a series of *******.

Reliability and Validity

The next critical piece to the survey instrument design is the validity of data entered or collected throughout the body of the survey. The continuous development of new client-side programming languages (those incorporated into the user's browser software) such as JavaScript provides for the validation of participant data without a need for additional server transactions. This technology provides a vehicle for the minimization of incomplete or erroneous data sets. Server-side languages (those hosted on the web server) such as Perl and CGI provide for the manipulation of data to such a degree that files can be prepared for analysis with little to no researcher intervention, eliminating time consuming and error-prone pre-analysis processing by hand. (White, Carey, Dailey, 2001)

The common mistake made by many researchers when creating web-based surveys is the use of client-side scripts programmed to force the participant's completion of all data fields. If all the fields are not completed the form can not be submitted. In the USF study, this resulted in two reflections. "First of all, there is an ethical issue. All participants have the right to withdraw from participation in a research study. On paper surveys, participants often exercise the right to

leave specific data fields blank. These same privileges should exist for web surveys." The second issue brought to light was also with regard to the completion of data fields. "From a practical perspective, failure to allow the participant to omit a field leaves only the option of total withdrawal from the study. On the chance that partial data may still be useful for some purposes, most researchers will probably prefer a partial data set to none at all, reserving the option to eliminate data sets that are not useful." (White, Carey, Dailey, 2001)

Instead, researchers and programmers should validate the type of data entered. For example, in a field requiring a participant's date of birth, the researcher would not want text. Typically the date of birth is given in MMDDYY format. The programmer should validate that the information being entered is that format and is numeric rather than textual.

Cost-Benefit Analysis

The benefits of administering a survey via a web-based forum far outweigh the cost to the participant and the researcher. The benefit to the participant is the ability to participate in web-based research at his/her own leisure. In the case of the researcher, the initial cost of the research development and design, and the survey instrument designs are still apparent. However, the need for additional help to collect, aggregate, and analyze the data once collected is no longer needed. Instead, the data is stored to a concise database, which can then be ported to a statistical analysis program. If configured properly, data is much safer stored on a hardware device than being stored in paper format, portable media, etc. The time savings is also a great factor in reducing the cost.

Overall, the cost of hardware is significantly down compared to costs 10 years ago. In fact, most institutions already have data equipment capable of hosting such web-based surveys for small to medium size research projects. Furthermore, the researcher can reach a much

broader audience by means of the web, than what was typically available in a community, university setting, etc. Research shows that about 51 million households currently have internet access, an increase from approximately 11 million in 1994. (UVA, 2004) A report by the Strategis Group predicts "more than 90 million households to be on the Internet in five years. Internet penetration of businesses reached 6.3 million in 1999 and will rise to 8.3 million by 2004. The Strategis Group foresees a combined total of 171 million Internet users in 2004." (Pastore, 2000) As the numbers grow, so does the potential participant pool for researchers.

The benefits of storing data electronically are numerous. The data can easily be stored in a small physical space, and quickly copied, manipulated, and analyzed. Unfortunately with this convenience comes new issues of security. In the past, electronic data was stored on removable media such as magnetic tape or floppy disks. Such materials could easily be secured by removing them from the computer and locking them in a file cabinet. In modern times most computers are connected to some sort of a network, and this connectivity creates security risks. The steps which must be taken to properly secure research data vary with the sensitivity of that data.

I will break down research data into three main categories: low risk, medium risk, and high risk. Low risk data is that which cannot be attributed to an individual because of a lack of personally identifiable information, a.k.a. anonymous. Medium risk data contains personally identifiable information, but nothing of a sensitive nature. For example, a subject completes a Likert scale evaluation on their feelings about work. High risk data is anything which contains personally identifiable information and is of a sensitive nature. Any data which pertains to an individual's physical or mental health, and well being may be protected under the Health Insurance Portability and Accountability Act of 1996, also known as HIPAA (National Institutes

of Health, 1996). Check with your IRB to see if HIPAA applies to your research. The level of security that you enforce should be at a minimum level consistent with the risk of the data.

In order to understand the risks involved with storing data in an electronic format, it is necessary to understand the basics of electronic data. Computers store data in binary format. Binary is the base-2 number system, in which the only numbers are 0 and 1. This is necessary because computers are built with devices called transistors. A transistor can be represented as a type of switch which works in two states, either on or off. Each digit of information, either a one or zero is called a bit. These bits are grouped into strings which have different schemes of encoding. A commonly used one is called ASCII (American Standard Code for Information Interchange) (ASCII Table). Eight bits are called a byte, which can represent 2^8 different combinations, or 256 different characters. For example, a capital "A" is encoded in ASCII as 01000001. The technical specifications of how data are encoded on each system are not important to understand, as long as it is known that all data is broken down into binary for digital storage.

For any level of risk, the first issue to address is the security of the medium the data is being stored on. The best option is to store data on a removable medium, and always keep a backup copy or two depending on the value of the data. This allows for the data to be physically secured. There are many secure options for removable media, such as CD, DVD, tape, USB drives and external hard drives. All of these media are lightweight and easily portable. Most are fairly inexpensive as well.

Table 1: Costs of removable media

Media Type	Storage Capacity	Average Cost per Media	Cost / megabyte
CD-RW	700 MB	$2.00	$0.0028
DVD-RW	4700 MB	$5.00	$0.0011
DAT	12000 MB	$10.00	$0.0083
HDD	120000 MB	$100.00	$0.0008
USB Flash	128 MB	$35.00	$0.2734

Source: Froogle, Pricewatch

If data must be stored on a non-removable medium such as an internal hard drive, the next best option is to make sure the computer is not connected to a network. Recently there have been a rash of virus and worm outbreaks which take advantage of vulnerabilities in Microsoft Windows operating systems. Viruses are not limited to Windows, they are also written for the Mac OS, and Unix and Linux. Windows is by far the most popular operating system, and therefore the one targeted by the majority of virus writers. According to Symantec, “a computer virus is a small program written to alter the way a computer operates, without the permission or knowledge of the user” (Symantec). A worm is a type of virus that spreads through networks infecting other computers without any user intervention. These types of viruses are particularly worrisome as computers can get infected simply by being connected to the network. The best

way to avoid being infected by viruses and worms, and to avoid being hacked is to keep any computer with sensitive data disconnected from the network.

In some large research projects, there is a need to share data in real-time across large geographic areas. This is the case when researchers from multiple institutions collaborate on a project. In this instance it may be necessary to store data on a network and have it be internet accessible. Another need for networked data is online data collection, which is growing in popularity. In either case, it is important to take steps to ensure that the data is not easily accessed by outsiders. There are several methods of securing data which can be used individually or in conjunction.

The first is password protection. It is important when choosing a password, to make sure that it cannot easily be guessed. Often people choose obvious passwords, such as their birth date, names of family members or pets, or even their own name. Such a choice compromises security as an outsider could easily determine what the password is. Creating a strong password is essential to good security. Here are some tips from Microsoft on choosing a strong password (Microsoft, 2004):

- Make the password long, at least 7 characters
- Mix together upper case, lower case, numbers, and special characters
- Don’t use real words
- Don’t use adjacent numbers or letters on the keyboard

Another important method which can be used to prevent unauthorized viewing of sensitive data is encryption. Encryption is a mathematical process by which data is scrambled using a key. The key is either based on a randomly assigned number or on a password. The strength of the

encryption is based on the length of the key. You may have seen encryption in use if you have ever visited a secured website (which begins in https:// instead of the regular http://). The extra s stands for secured, and if you locate the lock icon in your browser you can determine the level of encryption. Windows 2000, XP, and 2003 all have file encryption built into the operating system. This allows users to encrypt files on the hard drive so that if someone were to break into the computer they would have to know your user ID key to decrypt the file. The user ID is encrypted using your login password and your unique system ID. By default XP and 2003 use 256-bit encryption (Microsoft, 2002), which means there are 2^{256} different possible combinations (approx 1.1×10^{77}). File encryption can be set up to work transparently to users, so anyone who is authorized will not even notice the encryption/decryption happening in the background. This is a must for sensitive data.

A final option that is very useful for protecting network-connected machines is to utilize a hardware firewall. A firewall is a system which prevents computers external to the firewall from accessing your computer. Network communications break up information into packets. These can be expressed as a sort of electronic envelope which has the sender address, receiver address, and data all included. Intelligent firewalls are able to look at where packets originate from and block any network traffic that did not originate from the internal side of the network. This keeps anyone outside from gaining access to your computer and data. Another less expensive option is a software firewall which performs the same functions as a hardware firewall, except it is a program instead of a physical box.

Hopefully this has been a helpful introduction on how to secure electronic data. To summarize, make sure physical access to the computer is secured. If possible keep the computer disconnected from the internet. If internet connectivity is required, use either a hardware or

software firewall. Encrypt data on the system using built-in file system encryption, or using third party software such as PGP. Finally, make sure that any passwords protecting the file are strong to prevent hacking of your account. Follow these guidelines and you will have more secure data.

Eventually at the end of a research project there will come a need to destroy some of the original data. Paper data and VHS tapes will often be incinerated since reuse could lead to problems with recovery of data off of the media. Electronic data is no different. The best thing to do is destroy the media if possible. Any type of removable media is easily destroyed. Floppy disks, CDs, and DVDs can all be shredded to prevent anyone reading the data contained on them. They are all inexpensive and the cost of the media does not usually justify trying to reuse them. What about other media such as hard drives and USB drives?

Erasing a file in the operating system or formatting a disk may lead you to believe that you have somehow gotten rid of the data that was stored on it. Usually that is not the case. Most media have a special reserved area of the disk that is an analogue of the table of contents in a book. This is where the computer looks to find out where physically on the disk a file is stored. The special area where this information is stored is called the File Allocation Table (FAT). When you erase a file all that usually happens is that the portion of the FAT that tells where the file is gets erased. If we return to the book analogy, this would be like erasing “Chapter 5” from the table of contents, without actually removing all the information in Chapter 5. An entire industry has developed which specializes in data recovery. Software exists which can scan the entire area of a medium looking for files. Once again to return to the book analogy, the program would scan through all the pages looking for information even if it’s not listed in the table of contents. It is therefore possible to find the file even though the user is under the impression they deleted it.

Storage on a medium varies depending on the type of technology used. There are basically two types in use today: magnetic and optical. Magnetic media utilizes some sort of disc or tape which can be magnetized. A tiny magnetic head moves over the surface and either magnetizes spots to be positive or negative, representing a binary one or zero. When the disc is quick formatted, zeros are written over the FAT. When a full format is done on the disc, zeros are written over the whole surface. Erasing magnetic tape is a time-consuming process, and should only be done if it is necessary to reuse the media. In general any CD or DVD media is now at the point where taking the time to erase it fully for reuse would cost more than just destroying and replacing it. Even after being formatted data can often be recovered using special software.

Optical discs use a reflective metallic layer to store data. CD-R media, which is writeable only once, has a dye layer which gets burned by the higher power write mode of the laser. CD-RW discs have a layer made of an alloy which has some very special properties. Depending on the temperature it's heated to, when it cools it will crystallize in a different manner that changes the reflective nature of the surface (Tyson). A reflective surface is read as a zero while a non reflective surface is read as a one. DVDs work in a similar manner. Factory pressed DVDs have multiple layers to store data on, but until the next generation of DVD, writers can only burn a single layer. Similar to magnetic media, a format of a CD-RW or DVD-RW will often erase only the FAT and not the entire disc. Certain programs such as Nero allow for a full format, erasing all data on the disc. Data recovery can also be performed on optical media if it has not been fully formatted. Forensic methods can be used to recover data from supposedly formatted discs, so if the material is sensitive the sensible approach is to destroy the media.

Magnetic media is much more resistant to full erasure of the disc than optical media. In fact, data can sometimes be recovered from a disc that has been formatted completely or overwritten with other data. While time consuming and costly, methods have been developed using scanning electron microscopes to read previous magnetic patterns imprinted in the disc. In order to overcome this problem of possible recovery of deleted data, a system has been developed to ensure that it is completely erased. In its essence, it writes random bits over the whole surface of the area where the file was, or the whole disc. This process is called wiping and is necessary for ensuring that no data can be recovered from your magnetic media.

Wiping has become a controversial topic amongst the data security community. There are various opinions on how many times you should wipe a disc to make sure it is unrecoverable by any means (Garfinkel, 2003). A single pass wipe is the standard and a good preventative measure. The Department of Defense 5220.22-M directive for secure deletion of data calls for a 3-pass wipe (United States Department of Defense, 1995). The National Security Agency requires a 7-pass wipe. In 1996 Peter Gutmann of the University of Auckland wrote a paper on "Secure Deletion of Data from Magnetic and Solid-State Memory." In his research he used various methods to recover data from supposedly wiped drives, and his paper has become a sort of bible for protection against data recovery techniques. His recommendation for secure deletion of data involves a series of 35 consecutive wipes, some writing random data and some writing common characters (Gutmann, 1996). While this method is excessive for everyday deletion of files, if a drive were being disposed of or transferred to another user, the Gutmann method is preferred.

Conclusion

As technology advances, the research academy will begin to tackle a new forum for the same critical issues dealt with decades ago. Survey administration is not a new subject; web-based surveys are a different medium for administration. How will participants' rights be protected? How will research be conducted in a fair, reliable way, while not hindering the overall scholarship of the researcher? How can a researcher protect his/her electronic data in as safe (or safer) a manner as paper-based studies? Institutional Review Boards must confront the reality of advancing technology and all it has to offer the world of research. Hopefully in the near future the APA will adopt some guidelines for properly protecting research. In the meantime, it is the researcher's responsibility to conduct studies using his/her best judgement and make ethical decisions where guidelines currently do not exist. Continuing advances in technology are not only the problems, but the answers to the problems.

References

American Psychological Association. *APA Statement on Services by Telephone, Teleconferencing, and Internet*. (1997, November 5). Retrieved April 20, 2004 from http://www.apa.org/ethics/stmnt01.html

American Psychological Association. *Ethical Principles of Psychologists and Code Of Conduct 2002*. (2002, August 21). Retrieved April 20, 2004 from http://www.apa.org/ethics/code2002.pdf

ASCII Table. Retreived April 28, 2004 from http://www.asciitable.com/

Froogle. Retreived May 2, 2004 from http://www.froogle.com

Garfinkel, S.L., Shelat, A. *Remembrance of Data Passed: A Study of Disk Sanitization Practices*. (2003 Jan/Feb). Retreived May 1, 2004 from http://www.computer.org/security/garfinkel.pdf

Gutmann, P. Secure *Deletion of Data from Magnetic and Solid-State Memory*. (1996, July 22). Retrieved May 1, 2004 from http://www.cs.auckland.ac.nz/~pgut001/pubs/secure_del.html

Microsoft Corporation. *Encrypting File System in Windows XP and Windows Server 2003*. (2002, August 1). Retrieved May 1, 2004 from http://www.microsoft.com/technet/prodtechnol/winxppro/deploy/cryptfs.mspx

Microsoft Corporation. *Creating Stronger Passwords*. (2004, March 9). Retrieved May 2, 2004 from http://www.microsoft.com/security/articles/password.asp

National Institutes of Health. *Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule*. (2003, September 25). Retrieved April 21, 2004 from http://privacyruleandresearch.nih.gov/pr_02.asp

Office of Research Integrity. *Programs: Assurance/Compliance*. (2002, February 15). Retrieved April 23, 2004 from http://ori.dhhs.gov/html/programs/assurance.asp

Pastore, M. (2000, February 16). US Households Continue to Go Online. *ClickZ Stats.* Retrieved April 28, 2004 from http://www.clickz.com/stats/big_picture/geographics/article.php/5911_305261

PriceWatch. Retrieved May 2, 2004 from http://www.pricewatch.com

Sales, B.D. & Flokman, S. (Eds.) (2000). *Ethics in research with human participants.* Washington, DC: American Psychological Association.

Symantec Inc. *What is the difference between viruses, worms, and Trojans?* (1999, April 12). Retrieved May 2, 2004 from http://service1.symantec.com/SUPPORT/nav.nsf/docid/1999041209131106

Tyson, Jeff. *How Removable Storage Works: Optical: CD-R/CD-RW*. Retrieved April 28, 2004 from http://computer.howstuffworks.com/removable-storage8.htm

White, J., Carey, L., & Dailey, K. (2001). Web-based instrumentation in educational survey research. *WebNet Journal: Internet Technologies, Applications and Issues. (In Press).* Retrieved April 28, 2004 from http://fcit.coedu.usf.edu/surveydemo/surveypaper.html

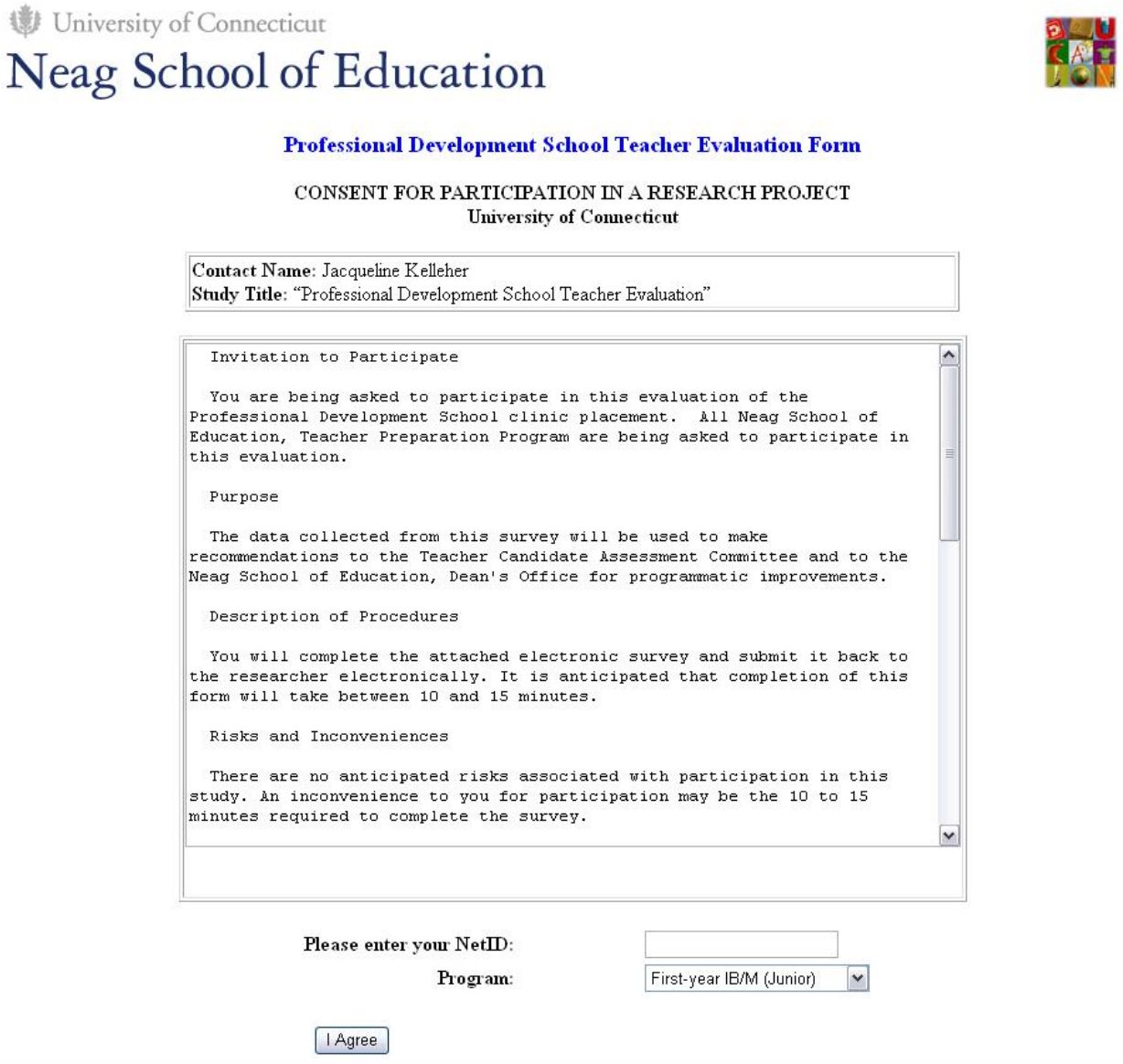
WordNet 2.0 Search (2004) Retrieved April 26, 2004 from http://www.cogsci.princeton.edu/cgi-bin/webwn

United States Department of Defense. *National Industrial Security Program Operating Manual*. (1995, January). Retrieved May 1, 2004 from http://www.dss.mil/isec/nispom_0195.htm

University of Virginia. (2004). *1994-2004 Ten Years of the Web at U.Va.* Retrieved April 28, 2004, from http://www.virginia.edu/HP2004/webstats.html

Appendix A

The following pages follow the structure in Diagram 1 (p. 6). The consent page is given in view format, followed by the HTML/CFML code that interprets the data to determine whether the participant has responded previously.


University of Connecticut
Neag School of Education

Professional Development School Teacher Evaluation Form

CONSENT FOR PARTICIPATION IN A RESEARCH PROJECT
University of Connecticut

Contact Name: Jacqueline Kelleher
Study Title: "Professional Development School Teacher Evaluation"

Invitation to Participate

You are being asked to participate in this evaluation of the Professional Development School clinic placement. All Neag School of Education, Teacher Preparation Program are being asked to participate in this evaluation.

Purpose

The data collected from this survey will be used to make recommendations to the Teacher Candidate Assessment Committee and to the Neag School of Education, Dean's Office for programmatic improvements.

Description of Procedures

You will complete the attached electronic survey and submit it back to the researcher electronically. It is anticipated that completion of this form will take between 10 and 15 minutes.

Risks and Inconveniences

There are no anticipated risks associated with participation in this study. An inconvenience to you for participation may be the 10 to 15 minutes required to complete the survey.

Please enter your NetID:

Program: First-year IB/M (Junior)

I Agree

```
<cfparam name="FORM.STU_ID" default="1">
<cfquery name="qGetMasterID" datasource="evaluations">
SELECT *
FROM PDS_Teacher_ID
WHERE STU_ID = '#FORM.STU_ID#'
</cfquery>
<html>
<head>
<title>Processing...</title>
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1">
</head>
<CFSET #form.YEAR# = "2004">

<cfquery name="StartRecord" datasource="evaluations">
        INSERT INTO PDS_Teacher
        (STU_ID,PROGRAM,YEAR)
        VALUES
        ('#Form.STU_ID#', #Form.PROGRAM#, '#Form.YEAR#')
        </cfquery>

<body>
<CFIF qGetMasterID.RecordCount EQ "0">
        <cflocation url="page1.cfm?STU_ID=#FORM.STU_ID#">
        <CFELSE>
        <cflocation url="error.cfm?STU_ID=#FORM.STU_ID#">
</CFIF>
</body>
</html>
```