Ethical Issues in Empirical Studies of Software Engineering

Janice Singer and Norman G. Vinson

Abstract—The popularity of empirical methods in software engineering research is on the rise. Surveys, experiments, metrics, case studies, and field studies are examples of empirical methods used to investigate both software engineering processes and products. The increased application of empirical methods has also brought about an increase in discussions about adapting these methods to the peculiarities of software engineering. In contrast, the ethical issues raised by empirical methods have received little, if any, attention in the software engineering literature. This article is intended to introduce the ethical issues raised by empirical research to the software engineering research community and to stimulate discussion of how best to deal with these ethical issues. Through a review of the ethical codes of several fields that commonly employ humans and artifacts as research subjects, we have identified major ethical issues relevant to empirical studies of software engineering. These issues are illustrated with real empirical studies of software engineering.

Index Terms—Ethics, empirical studies, software engineering, legal issues.

1 Introduction

Dr. Jonas is a professor at a well-known university in the newly created software engineering program. He recently embarked on a research project to determine how collaboration styles influence software quality. His hypothesis is that software engineers who work well together produce better software. Dr. Jonas collects data by observing SE teams at local companies. He then categorizes the teams according to their success at collaboration. He also collects metrics for software components previously developed by the same teams. Dr. Jonas plans to correlate the collaboration quality measures with the metrics to determine whether teams that collaborate better produce higher quality code. A few weeks into the research program, a manager asks to see Dr. Jonas' field notes and wishes to know how his company compares to the other companies regarding the metrics assessment. What should Dr. Jonas do? To whom is he obligated?

Researchers conducting Empirical Studies of Software Engineering (ESSE) often face such ethical dilemmas. Yet, to date, little has been written about ethics as it applies to ESSE. Moreover, the community has yet to produce or adopt a common set of guidelines geared towards the ethical issues faced by empirical software engineering researchers [18]. In this paper, we take the first step towards the creation of a set of ESSE guidelines by familiarizing the major stakeholder groups in ESSE research (researchers, sponsors, and potential subjects) with the basic elements and standards of research ethics. We begin with a discussion of how and why research ethics is important and relevant to those involved in ESSE research. We then point out that existing codes of ethics, though relevant, are not directly applicable to ESSE research. Therefore, we use the codes to identify four core research ethics principles and subsequently illustrate these principles with real ESSE research projects.

Our discussion is most relevant to research projects that employ human subjects or that involve the collection of information that can lead to the identification of individuals. Such identifiable information can be collected through the observation of humans or through the examination of artifacts (i.e., source code or documents). Note that this covers much of empirical software engineering research whether it involves metrics, workplace studies, or process studies. Our discussion also applies to SE *practice* when it involves human subjects or identifiable information (usability testing for example), though our main focus is on ESSE research. Our paper does not apply to other areas of software engineering practice or research, such as the development or application of standards, or components research. For a discussion of ethical issues arising in these areas, there are a number of excellent books (e.g., [19], [31], [38], [39]). The interested reader is also directed to the IEEE-CS/ACM Software Engineering Code of Ethics and Professional Practice [17], [15] and the ACM Code of Ethics and Professional Conduct [1].

There are several reasons why researchers, corporate sponsors, and potential subjects of ESSE should be concerned with research ethics. Researchers who upset their subjects risk losing their cooperation or honesty [29]. Researchers who upset the subjects' employers or managers risk losing access to the subjects, to funding, or to other resources. Canadian, Australian, and American researchers from academia or those receiving government funding risk losing their funding if they do not follow mandated ethical guidelines [22], [28], [30], [34], [35]. Sponsors of ESSE research must also understand how research ethics guides the behavior of the researchers, and how unethical

• *The authors are with the National Research Council, Canada, Institute for Information Technology, Montreal Rd., Bldg. M-50, Ottawa, ON K1A 0R6. E-mail: {janice.singer, norm.vinson}@nrc.ca.*

Manuscript received 9 Dec. 1999; revised 17 Sept. 2001; accepted 23 July 2002.
Recommended for acceptance by S. Pfleeger.
For information on obtaining reprints of this article, please send e-mail to: tse@computer.org, and reference IEEECS Log Number 111101.

behavior, on the part of management or the researchers, can jeopardize a project [21]. For example, coercing employees to participate as research subjects can often lead to invalid data. Ethical issues should also be of great concern to potential subjects. Subjects must understand their rights in order to ensure that they are appropriately shielded from harm, such as loss of employment. Different forms of harm are discussed in the cases presented below. Researchers, sponsors, and potential subjects may have all different reasons to be concerned about ethics, but they should all be concerned nonetheless.

In searching for relevant literature, we found that only a small part of the vast body of literature on engineering and science ethics directly relates to ESSE research involving human subjects or artifacts. For example, a large portion of the literature in research ethics deals with issues that apply broadly across research disciplines, such as authorship, the relationship between graduate students and their advisors, and scientific fraud [9], [10]. This literature is relevant to ESSE to the same extent that it is relevant to any other scientific discipline, but it does not deal with humans as subjects or software engineering per se. The literature on computing and engineering ethics [19], [31], [38], [39], [40] primarily discusses ethical issues raised by engineering and computing practices, such as following recognized engineering standards, or ensuring the security of computer systems. Similarly, the codes of ethics promulgated by several professional computing associations such as the ACM and IEEE-CS [1], [4], [15] focus on ethical issues arising from computing *practice*. This literature does not cover the ethical issues arising from the human subjects research occurring in ESSE. Nonetheless, the ACM and IEEE-CS/ACM SE reflect many of the same fundamental ethical principles as the codes that do cover human subject research. Consequently, we refer to the ACM and IEEE-CS/ACM SE codes in our ESSE case illustrations, though often in a context that differs from those for which the codes were written (see [5]).

Human subjects research ethics is primarily covered in guidelines authored by government funding bodies and social science organizations. Note, however, that ESSE has many peculiarities not well-represented in these research domains, such as organizations serving as research subjects, long-term classroom studies, and the use of artifacts (e.g., source code). Even when ESSE researchers' practices resemble those of social scientists, the social science codes may not deal with the ethical issues raised by these practices. For example, many anthropologists in the private sector conduct workplace studies whose results are potentially damaging to their subjects by, for instance, providing information for restructuring [13], [14]. However, the new version of the American Anthropological Association's code of ethics does not directly address ethical issues raised by such workplace studies [2], [14]. Ignoring the reality of workplace studies, the AAA code even seems to prohibit such work, stating that the anthropologists' primary responsibility is to protect their subjects from any harm ([2], Sections 3.1 and 5). Because of the peculiarities of ESSE and such lapses in professional codes of ethics, we found that, while the human subjects codes of ethics were relevant to ESSE, their application to actual situations faced by ESSE researchers was far from obvious.

Consequently, rather than attempting to apply existing codes directly to ESSE examples, we reviewed the codes and abstracted four representative high-level ethical principles: informed consent, scientific value, beneficence, and confidentiality.[1] These principles constitute the accepted standards of ethical research practice. One can of course take issue with these standards, but, in this regard, it is important to understand that *our aim in this paper is simply to present the standards—not to debate them.* Our presentation proceeds by illustrating each of the four principles with examples taken from the ESSE literature or related to us by ESSE researchers. Such an exposition allows readers to perceive how these principles can be applied in the situations they face.

Moreover, because the codes of ethics were not designed with common ESSE research practices in mind, they tend to prohibit those practices rather than accommodate them. This is particularly problematic for researchers in Australia, Canada, and the US, who are regulated by governmental codes of research ethics. Consequently, in addition to illustrating the ethical principles, we provide some procedural suggestions to increase the compliance of ESSE research practices with the codes of ethics.

2 Examples of Ethical Issues in ESSE

The following examples illustrate the ethical principles identified in our review. Each example is first described and then discussed with respect to ethical concerns. Note that all examples are based on real cases of ESSE research, some published in the literature, some related to us in conversation. However, the names and contexts have been changed to maintain the anonymity of the parties involved.

2.1 Informed Consent: The Case of the Student Subjects

Dr. Gauthier is on the faculty of a large research university. She is interested in how different views of source code influence program understanding and has therefore built a tool that offers a data flow view, a control flow view, and an architectural view of a system. She wants to see which of the different views help software engineers design and maintain source code more effectively. Unfortunately, Dr. Gauthier does not have access to industrial software engineers to test her tool. Consequently, she decides to use the students in her software engineering class as test subjects. She divides the students into four sections. Each of three sections is given one of Dr. Gauthier's tools with a different view. The fourth section uses the standard tools provided by the university programming environment. Dr. Gauthier gives all four sections the same midterm project. She finds that some of views offer modest gains in productivity.

Perhaps the primary ethical principle in human subject research is that of full informed consent on the part of the subject to participate in the research project [12], [20], [34]. Ethicists do not fully agree on the necessary components of informed consent, but it is clear that it must contain at least some of the following elements: disclosure, comprehension

1. The Appendix provides the full list of codes we reviewed.

and competence, voluntariness, the actual consent or decision, and the right to withdraw from the experiment. Disclosure refers to the information that the researcher must provide the subjects before they decide to participate in the experiment. This information usually includes, but is not limited to: the purpose of the research, the research procedure, the risks to the subjects, the anticipated benefits for the subjects and the world at large, and a statement offering to answer the subjects' questions. The intent is to provide the subjects with all the information they need to understand how the research affects them. The need for comprehension compels the researcher to present the information in a way the subjects can understand, e.g., eschewing technical jargon that is outside the subjects' repertoire. The element of competence refers to the subjects' ability to make a rational informed choice. This criterion is intended to protect vulnerable subjects who may not understand the nature of the research, such as children or the mentally disabled. Voluntariness specifies that informed consent must be obtained under conditions free of coercion and undue influence, and that the consent must be intentional. Finally, subjects are given the right to terminate their participation at any time without having to provide any explanation [12], [23], [33], [34]. In some cases, subjects may be able to exclude their data, but this is not always true—especially in instances where withdrawal of one subject's data harms the other subjects in some way (e.g., by making the experiment invalid) [34].

In the above example, Dr. Gauthier violated the principle of informed consent in that she simply did not obtain consent from the students involved. Even if the students had been given the opportunity to refuse participation, their consent would have been vitiated by Dr. Gauthier's potential influence on their decision. Valid informed consent in classroom research is extremely difficult to obtain because of the professor's power over the student's grades. There is always the potential for a student to fear a reprisal in the form of a lower grade for refusing to participate, or to anticipate a benefit in the form of a higher grade for participating [34], [36]. This applies whether the professor intends to reward participants, punish nonparticipants, or not. The ethical difficulty arises not from the professor's intent but from her power.

Thomas Puglisi, former head of the human subjects division of the United States Office for Protection from Research Risks[2] (OPRR), makes an even stronger statement [25].

> I must conclude that recruiting subjects in class, with the instructor present, is inherently coercive and clearly violates 45 CFR 46.116: "An investigator shall seek such consent only under circumstances that . . . minimize the possibility of coercion or undue influence." The unequal power relationship between instructor and student is one that should not be viewed casually. Active, meaningful protections are needed to avoid coercion when instructors appear to endorse recruitment of subjects into the research of others. Additionally, it is my view that the power relationship simply cannot be equalized when instructors attempt to recruit their own students into their own research, and such recruitment should never be permitted, no matter how (seemingly) benign the research.

Methods that eliminate the professor's power over the students' grades relieve the pressure on students to consent, thus ensuring the validity of their consent. For example, a professor can use another professor's students granted that the recruitment occurs in a noncoercive manner (i.e., the professor is not present during recruitment and the other professor does not endorse the research project). A professor can also conduct the research after the semester has ended, when the participants are no longer his students. Alternatively, if the research component is part of the regular curriculum, it might be possible to gain consent to use the students' data after the semester is over, e.g., by mailing them a consent form. Maintaining the anonymity of the participants also mitigates the classroom consent problem. For example, a graduate student can administer a questionnaire instead of the professor [24].

Similar consent difficulties arise in SE field research. As with the students, an employee's decision to participate could be unduly influenced by the perception of possible benefits or reprisals ensuing from the decision. For employees, the benefits or reprisals would be thought to come from management, who approved the study, rather than from the researchers. Additionally, researchers must guard against management coercing the employees into participation. Not only is such coercion unethical, it is also dangerous from a pragmatic perspective, in that coerced subjects may be uncooperative and/or provide invalid data [21]. To guard against such coercion, the researcher should emphasize to management the importance of voluntariness. The researcher should then reiterate this when meeting with the subjects themselves and assure them that they will not be reported if they decide not to participate. As with the students, protecting the anonymity of the participating employees will mitigate the undue influence on the consent process.

Note that offering remuneration can also constitute undue influence [34], [36]. The problem with inducements in general is that they can encourage potential subjects to undertake risks that they would not normally accept. Accordingly, inducements are less problematic when the risks to subjects are minimal [36]. Nonetheless, even then, inducements should not be so great that they constitute undue influence. This begs the question: How much is too much? Many university departments where experiments with humans are regularly conducted have standard rates of remuneration. This can present a problem to ESSE researchers, as it is unlikely that a software engineer would be willing to participate in an experiment for such a low rate of pay. Where remuneration rates become a problem, it is important to educate the board reviewing research proposals for compliance with ethics guidelines and regulations. Such boards are often comprised of individuals with no software engineering knowledge and will not know what is standard for this community. Cogent reasons in support of a proposed rate of remuneration will increase the likelihood that the review board will accept the proposed rate. An alternative to remunerating every subject is to hold a drawing for the latest electronic gadget. To be eligible for the drawing, an SE must participate in the experiment. This

2. The former United States Federal Office whose mandate was to oversee the American Institutional Review Boards (IRBs) that ensure compliance to US research ethics regulations.

technique is less expensive than remunerating every SE, yet should still appeal to SEs, and may be more acceptable to a review board.

Another common practice in ESSE is to examine data. Here too, consent is required when identifiable information is being collected. The IEEE-CS/ACM software engineering code [17] requires consent from companies before their property can be used. The ACM code goes further in also requiring consent from organizations and individuals before their systems are accessed. In addition, the ACM code ([1] ACM Article 2.8) provides examples of the types of access requiring consent: access to "communication networks and computer systems, or accounts and/or files associated with those systems [...]." Thus, a researcher wishing to log user commands would first have to obtain consent from the individual software engineers, from the manager of the software engineers, and from the company itself. Only when an individual cannot be identified from the data collected is it possible to dispense with obtaining the consent of individuals (see Section 2.7, Exceptions).

A final aspect of informed consent relates to how the data is used. Article 1.7 of the ACM [1] code states that information gathered for a specific purpose should not be used for another purpose without further informed consent. In an ESSE context for example, data collected to describe a software process should not be used to evaluate group members without their explicit informed consent to this additional use of the data. In general, it is better to specify in the informed consent document all anticipated future uses of the data. If a researcher wishes to use data for a purpose that was not specified in the informed consent document, she/he must generally consult a review board to ascertain how to gain the informed consent for the current intended use. An exception of this rule exists for data sets in which all identifiers have been stripped and, therefore, the original contributors of the data cannot be identified from the current data set [34].

In sum, software engineering researchers should generally obtain the informed consent of their research subjects and/or host organizations. Unfortunately for ESSE researchers, two of the main subject populations, students, and company employees, are considered vulnerable populations by the codes promulgated by Canadian and Australian government funding agencies [22], [28], [34] and the US federal government [36]. Moreover, social science disciplines that often conduct research with students or employees have yet to devise procedures to protect subjects that could easily be applied to the type of long-term classroom studies common in ESSE or to ESSE field studies [13], [14]. Finally, in some cases, it is not clear whether employees have the right to refuse to participate in a research project sanctioned by management (see Sections 2.3 and 2.6 below). The ESSE community has an opportunity to examine these issues and to develop solutions that allow research to proceed while protecting their potential subjects' rights.

2.2 Scientific Value: The Survey Says...

Chuck Amaro is an associate at a research firm. He just completed his PhD degree, and is now consulting on a project on the use of design reviews in industry. One of Chuck's tasks is to determine how design reviews are really conducted in the real world and to what ends. Chuck has never done this kind of research before, but he feels confident that he knows what to do. He develops a "common sense approach, as opposed to a specific, rigorously defined social science approach."[3] Chuck interviews 50 software engineers on three different continents, each for two to 11 hours. The engineers were selected based on their proximity (to reduce travel costs), notoriety, and similarity to his target audience.

Scientific value has two components: the importance of the research topic and the validity of the experimental results. The importance of the research topic is usually evaluated in the context of potential risks and benefits to both subjects and society at large. Many codes reflect the principle of beneficence by stating that the research must provide the greatest possible balance of benefits to risks [23], [34], [35]. Research that is of little importance will have difficulty meeting the test of beneficence.

The second component of scientific value is the validity of the results. If the results are not valid, they do not reliably or faithfully represent reality. Consequently, any conclusions drawn on the basis of those results will be incorrect, and will also have no value. A study producing invalid results has no scientific value. If a study has no value, then its benefits cannot outweigh its risks to the subjects, and it cannot pass the beneficence test. The ethical standard is that a study must be expected to produce valid results in order to be undertaken [22], [36], [41].

Typically, invalid results are produced by the misapplication of reputed methodologies, methodological oversights, or the use of disproved methods. For example, consider the finding that source code written in Java is debugged more quickly than source code written in C++. Consider further that the defect samples were biased such that all the Java defects in the experiment were considerably easier to fix than the C++ defects. We would then expect Java debugging to be quicker just because the Java defects were easier to fix. Therefore, results showing a difference in debugging time would at least partly, and perhaps wholly, be due to this difference in the defect samples. This sampling bias would also prevent us from generalizing the results beyond the study, where this methodological artifact would not come into play. Consequently, this experiment would provide no information about whether the *inherent* differences between Java and C++ could produce a reliable, *general* difference in debugging time. Due to a methodological error (a sampling bias), the results, and therefore the study, are of no value.

In practice, researchers face a variety of tradeoffs in selecting research methodologies. These tradeoffs involve the degree of validity of the results, their importance, their breadth, the potential harm to the subjects, the ease with which the data are collected, and so on. In selecting a method, researchers must weigh these tradeoffs and be prepared to justify their choice to their ethics review board.

In the professional codes, the issue of validity is dealt with by articles discussing competence. These articles instruct members not to conduct any work for which they

3. This phrase is quoted from the article itself. However, we do not provide the reference in order to maintain the anonymity of the parties involved.

have no competence (e.g., [1], [3], [16], [26], [27]). Furthermore, the American Psychological Association explicitly prohibits members from using techniques that have been shown invalid [3]. The IEEE-CS/ACM SE code also directs members to seek other professionals' council when needed [17]. Moreover, the information technology codes encourage members to apply, and to be familiar with, all the relevant standards [1], [17]. In the context of ESSE, this could be interpreted as familiarity with, and application of, standard research and statistical methodologies.

Dr. Amaro's study, our example described above, had little scientific value because his "common sense approach" which he himself describes as lacking rigor, is invalid. Standard social science methodologies were developed expressly because such "common sense approaches" were shown to provide unreliable and invalid data. So, although the research topic is important, the validity of the study is questionable. Even if there had existed little information on how design reviews are (really) conducted, and it therefore would have been acceptable to use an exploratory methodology (interviews), this methodology should have been used correctly.

The issue of scientific value is critical for ESSE in that some researchers are not completely familiar with the methodologies they use. The ethical codes suggest that using improper methodologies in ESSE studies involving humans or companies is unethical. The reasoning is that, since the methods are flawed, the results will be invalid so the merit of the study is nil. Weighed against possible harm to the subjects, a study without merit should not be undertaken.

As this discussion has shown, the issue of scientific value is intimately tied to the issues of harm to, and benefit for the research participants. This is the topic of our next two cases.

2.3 Beneficence—Human: The Case of the Reengineered Engineers

Dr. Brandt conducts research on source code reengineering and automated translation. To carry out his work, he needs access to programs with several million lines of source code. He obtains access from his industrial partners. Upper management has always been happy to have its source code updated by Dr. Brandt, but the software engineers who maintain the source code have not been so appreciative. Consequently, Dr. Brandt has implemented procedures to minimize the impact of the source changes on the software engineers. First, he involves the software engineers in all of the issues surrounding the project's schedule and the new source code's integration into the existing system. He also arranges for the software engineers to receive training in the new source code's language. Moreover, he insists that management allot the software engineers time to simply explore the new source code. These procedures give the software engineers control over the whole translation process, thus reducing their stress. They also allow the software engineers to more easily transfer at least some of their expertise (e.g., knowledge of source code/domain relationships) to the new source code.

The concept of beneficence requires a favorable balance of benefits to harms [23], [29]. This means that researchers must maximize the benefits to society and the subjects while minimizing the possible harms that can result from the research—this is often referred to as the risk/benefit ratio. In theory, then, greater benefit can justify greater risk of harm. For example, in biomedical research, it may be possible to subject patients to harmful radiation in the search for a cure for cancer.

When weighing beneficence, it is important to consider how the benefits and risks affect each stakeholder involved in the project. Beneficence should be maximized, as much as possible, for each stakeholder group. However, tradeoffs that can adversely affect one stakeholder group will sometimes need to be made. As the preceding case shows, research that benefits the employer (reengineered code) can also harm the employees. In such cases, it is important to offset the harm to the greatest extent possible.

Dr. Brandt's code translation harms the software engineers in several ways. For example, it greatly disrupts their work; if they are unfamiliar with the new language, the translation can place their employment at risk; and the loss of control over the code creates a great deal of stress. These consequences are clearly encompassed by the definition of harm specified by governmental organizations [22], [36]. Though the ACM code [1] does not specify such consequences as harmful, its recommendations are quite clear regarding cases such as Dr. Brandt's. "Organizational leaders [management in this case] are responsible for ensuring that computer systems enhance, not degrade, the quality of working life." [Article 3.2, ACM Code]. Moreover, the needs of those affected by the implementation of a system must be included in the system requirements and the system must be shown to meet those needs [Article 3.4, ACM Code]. The IEEE-CS/ACM SE code has a similar provision for user requirements. Thus, both Dr. Brandt and management were ethically responsible to minimize the harm to the software engineers, even if it could not be completely prevented. Consequently, Dr. Brandt acted properly in ensuring that there were procedures in place to help the software engineers learn about the new system.

Dr. Brandt's involvement of the affected SEs in the project itself parallels recommendations for archaeological research, which can sometimes cause upheaval. "For field projects, archaeologists should consult with appropriate representatives of the local community during the planning stage, invite local participation in the project, and regularly inform the local community about the results of the research [Archaeological Institute of America, Code of Professional Standards, Article II.4.] [6]. Similar advice, also applicable to ESSE, is provided in Section 6.B of the Tri-Council's statement [34].

Consider a parenthetical note regarding the problems with consent in cases such as Dr. Brandt's. In such a case, could the consent of the software engineers be obtained? If so, could it be deemed truly voluntary given that upper management had already decided to translate the code? Given that upper management had made this decision, what consent was left for the software engineers to give? In other words, the decision to translate the code was not the software engineers' to make; it was a management decision. Consequently, the software engineers had no choice but to accept the translation. Was there anything left to which the software engineers could refuse to consent? These issues are

further explored in the context of metrics research, in Section 2.6.

2.4 Beneficence—Organizational: The Case of the Process Modeler

Dr. Johns works in a software engineering research center. Her research deals with process improvement. Dr. Johns is quite excited by a newly published process model. Consequently, she collects process data from a software development team working for a large government contractor. Using the model to analyze her data, Dr. Johns finds five major flaws in the contractor's software process, including the contractor's over-reliance on one team leader. Dr. Johns is very impressed with the new model's usefulness and publishes her results in a publicly available conference proceedings.

The ethical issue presented here is the minimization of harm, but at the organizational, rather than the individual level. By publicly disclosing the flaws in the contractor's processes, Dr. Johns has put their government contracts at risk. Even if the company name is not published, it is quite possible that a reader will be able to identify the company based on the description of their processes. If this information is brought to the government's attention, the government may well terminate the contracts. Moreover, her publication might impede the company's efforts to obtain other contracts.

Minimization of harm at the organizational level applies quite broadly in ESSE. For example, a researcher may evaluate source code from several different companies and name these companies in an appendix to a published article (a real case). Negative evaluations could lead prospective clients to choose competing products, even if the relationships between the evaluations and companies are equivocal. None of the IT codes we reviewed discuss this type of financial harm. However, they do have provisions protecting confidentiality, privacy, and consent that would protect companies and individuals from such harm [1], [17].

On the other hand, if protecting an organization places the public at risk, EEs and SEs are asked to whistle-blow, that is reveal information damaging to a company in order to protect the public [15], [16]. The Tri-Council Statement also recognizes that public-policy research may legitimately harm organizations that are discovered to be acting inappropriately [34]. Such conflicts of interest between the hosting organization and the public at large complicate the implementation of beneficence.

Similar complications arise from conflicts between the interests of the employee-subjects and their employer. For instance, if researchers uncover problematic processes in a company, how should they attempt to minimize harm? To minimize harm to the company, the researchers should inform management of process problems that could harm the company through increased costs and reduced product quality. However, this could result in dismissals, thus harming individuals. While the codes of ethics do not provide much guidance, these issues are discussed by Becker-Kornstaedt [7].

2.5 Confidentiality: The Case of the Novice Programmer

Dr. Smith was interested in how novice programmers gain expertise. He contacted a personnel manager at a local company who was also interested in this research topic as the company was rapidly expanding and was therefore spending a great deal of money and effort training new employees. Dr. Smith signed an agreement with the local company. The company would provide him with access to experts (gurus) and novices, and he would help the company improve its training procedures. Dr. Smith spent the next several months interviewing the experts and novices. Because it was a small company, however, he had access to only a very small subject population. In the end, he interviewed two experts and followed 10 novices' work over several months. In the final report, Dr. Smith included a table showing the number of languages in which each of his subjects could program and their success in training. Subjects were not named but instead were identified by numbers. When the research was complete, Dr. Smith made the report available to the personnel manager as he had promised.

The ethical issue discussed here concerns confidentiality. In general, confidentiality has two components: anonymity and confidentiality of the data [29]. Anonymity is preserved if no one can identify the participants of an experiment. The ideal protection of anonymity involves not collecting any data that can be used to identify subjects; not even names. However, in many cases, researchers will be required to collect a signed consent form from the subjects, which will constitute a record of their participation. Beyond that, it is preferable not to collect any personal information. Anonymity also involves severing the subject's identity from his data set so that he cannot be identified through an examination of his data set. For example, the subjects' names should not be linked with their data. To identify individual data sets, subject numbers or aliases can be used instead of subject names. Researchers should also try to ensure their research subjects' anonymity by not letting any coworkers witness researcher-subject interactions.

Confidentiality involves the privacy of the data collected. The informed consent document should describe exactly who will access the raw data and for what purposes. Typically, raw data are stored under lock and key. In written or oral reports, confidentiality and anonymity can be protected by aggregating the data. (For example, one can report cross-subject averages, medians, standard deviations, or standard errors instead of raw data).

Dr. Smith's experiment raises three difficulties for confidentiality: he conducted his study in the workplace, he had few subjects, and he included in his report information that could be used to identify individuals (i.e., the number of known programming languages). In workplace studies, such as Dr. Smith's, it is often difficult to maintain anonymity since coworkers can often witness the interactions between the researchers and subjects. Workplace studies also increase the likelihood that subjects will be identified from reports of individual subjects' characteristics. Because coworkers know each other, they will be more successful at identifying the subjects than strangers would. For example, Dr. Smith reported the number of programming languages each subject knew, making it

easier for coworkers to determine who participated in the study. Once the subjects have been identified, it then becomes possible to link their identity to portions of the data. In the above example, it was possible for coworkers to determine each subject's success at training. It is preferable to report aggregated data (e.g., cross-subject averages) instead of raw data, as this makes it much more difficult to identify individual subjects and their data. Unfortunately, such aggregation is less effective at preserving confidentiality and anonymity when there are few subjects, as is common in ESSE. Because ESSE studies often have features (few subjects, occur in the workplace) that make the preservation of anonymity and confidentiality more difficult, ESSE researchers should disclose the limits of confidentiality and the implications thereof to the subjects as part of the informed consent procedure.

2.6 Consent, Confidentiality, and Beneficence: Taking the Measure of Artifacts

Dr. Foot is a metrics researcher. He examines the relationships between metrics and defect rates in order to determine which metrics relate to software quality. In order to carry out his work, he must have access to the source code of very large systems. This access is provided by the companies that developed the systems. In conducting his work, he discovered that programmer identity accounted for a large amount of variance in the defect rate. In other words, some programmers' source code had many more defects than other programmers'. Dr. Foot mentioned the importance of programmer identity in his report, but refrained from revealing individual programmers' identities or presenting any individuals' defect data. Nonetheless, this did not preclude him from reporting a more general relationship between class complexity and defect rate.

It is not uncommon to examine artifacts in ESSE. For example, in metrics research, source code and class design documents are examined. Process modelers also examine documents describing a company's processes [7]. Reporting flaws in such documents can harm the company, as described in Section 2.4 above. However, additional ethical issues arise when the artifacts identify the individuals who created them, as in the example given above. This raises the issues of consent, confidentiality, and beneficence in regard to those individuals.

Other fields exist in which artifacts are the objects of investigation: archaeology, artistic and literary criticism, and public policy research. The issues raised by archaeological research are more similar to those already discussed in Section 2.3, The Case of the Reengineered Engineers. Indeed, in that section, we did refer to ethical archaeological research practices. However, in archaeological research, the individuals who created the artifacts are no longer living. Consequently, the issues of consent, confidentiality, and to a large degree, harm do not arise. In contrast, artistic and literary criticisms can often harm the artist, and similarly, public policy research can harm an organization. It is recognized that a beneficence requirement would undermine those fields of work [8], [34]. For example, such a requirement would prevent art critics from producing any negative reviews! In contrast, in both metrics and process research it should be possible to avoid harming both individuals and companies by maintaining the confidentiality of some of the data and the anonymity of the subjects. Accordingly, Dr. Foot was able to protect the individual programmers from harm by maintaining their anonymity and the confidentiality of their defect data. As noted in Section 2.4 above, however, the protection of employees can conflict with the interests of the company.

Consent is another issue faced by ESSE researchers analyzing artifacts. Clearly, the consent of an organization is required when obtaining access to their source code or documents [17]. However, one can wonder whether the consent of the creators (authors, programmers) is also required. This is one of those peculiarities of ESSE that has not been dealt with elsewhere. However, we can extrapolate from the US [35] and Canadian [34] government guidelines. Both the US and Canadian guidelines specify that consent is not required to use information in the public domain. Moreover, the US regulations specify that only the collection of identifiable information that is of a private nature (e.g., a medical record) requires consent. It seems that the critical issue regarding consent is the subject's expectation that the information will remain private.

In considering how this applies to ESSE, we will only examine the implications for metrics research. The conclusions will generalize to other forms of ESSE with the same characteristics. In the case of metrics, the subjects (the programmers) do not expect their code to remain private within the company. In a sense, their source code is placed in a limited public domain constituted by the company. Consequently, the need to obtain the consent of the programmers is reduced if not eliminated. The need for consent is also reduced when the subjects are protected from any possible harm. In the case of metrics, it is important to note that the company will usually already have found the defects in its source code, and, if it so desired, could determine the number of defects introduced by each programmer. Consequently, by simply analyzing the defects, and, as did Dr. Foot, by refraining from reporting individuals' defect rates, the researcher does not create any risk for the programmers beyond what they face from the company's quality assurance programs. In short, because programmers expect their work to be reviewed by the company and are not placed at any additional risk by metrics research, it should not be necessary, from an ethical perspective, to obtain their consent. However, to maintain a good relationship with research subjects it might be advisable to involve the programmers as described in Section 2.3 The Case of the Reengineered Engineers (see also [6]).

2.7 Exceptions

Here, we briefly discuss general exceptions to the requirements of obtaining informed consent, respecting confidentiality, and beneficence. One type of exception occurs when there is no information in the raw data that could allow a particular individual to be identified. Here, informed consent and confidentiality are generally not required. This situation would occur most commonly in ESSE in the analysis of artifacts, such as source code or documentation, *in which the authors could not be identified* [34], [35]. Of course,

the consent and confidentiality of the company are still required [1].

Consent and confidentiality are also not required when examining public records or public activities where the expectation of privacy does not exist, if the data collected contain no personal identifiers and no harm comes to the subjects [34], [35]. Note the relevant criterion is the expectation of privacy, not whether there truly is privacy. This distinction is important in situations people mistakenly assume to be private, such as online newsgroups, or Web browsing. In such cases, consent and confidentiality are required [37]. In contrast, it is acceptable to measure client-server traffic without obtaining consent from each user of the system if none of the file names or client identities are collected [36].

Another exception occurs when more harm results from maintaining confidentiality than from breaching it. For example, where mandated or permitted by law, confidentiality can be breached to protect an individual from harm [3]. The IEEE code and the IEEE/ACM Software Engineering code also assert that members should make decisions consistent with the safety, health and welfare of the public, and to disclose promptly factors that might endanger the public or the environment. This is an interesting provision potentially allowing an engineer to harm his employer by publicly revealing product deficiencies that could harm members of the public [16], [17]. (This is known as "whistle-blowing.") In the context of ESSE, a researcher could uncover similar information creating a dilemma between harming the company hosting the project and allowing harm to come to the public.

Certain research designs may also present exceptions to the above general principles. For instance, at some institutions in the US, some classroom research is exempt from ethics review. Some interview or survey studies may also be exempt from review [35]. In general, these exceptions depend on the rules in force at a particular institution. For example, some US ethics review boards review research that is nonetheless exempt from federal regulations [36]. Therefore, researchers should ensure that they understand all the ethical guidelines and procedures with which they are required to comply (e.g., granting council guidelines, university guidelines, IP guidelines within companies, NDAs, etc.)

In summary, exceptions to any set of guidelines can occur. Researchers suspecting their project to be exempt from guidelines or regulations should consult their local ethics review board, or at the very least, their colleagues.

3 Conclusion

Let us revisit Dr. Jonas from the introduction. After collecting field data at some local companies, one manager asked Dr. Jonas for copies of his notes and wanted to know how his company compared to others regarding the metrics assessment. He was faced with the unenviable choice of breaching the confidentiality of his subjects or angering a host organization. Unfortunately, poor Dr. Jonas finds himself in a difficult situation. Either choice will probably lead him to lose access to his subject population. How could Dr. Jonas have avoided this ethical predicament? The answer is clear: He should have considered the ethical implications of his research at the planning stage, and ensured that all subjects and hosting organizations understood their rights and responsibilities before they consented to participate.

It is important for ESSE researchers to consider the ethical issues raised by their project while it is still in the planning stage in order to avoid Dr. Jonas' fate. It is incumbent upon each of us, as researchers, to be aware of the ethical issues that we will face in our research, and to act ethically towards our research population. "Responsible and ethical research is not a matter of codes, policy, or procedure. Rather, responsible and ethical research centers on a commitment to protect the participants of one's study from potential harm." [37, p. 130].

However, for the field as a whole, this is insufficient. The ESSE community must develop its own code of research ethics adapted to the peculiarities of ESSE, such as the use of organizational artifacts (like source code). We have taken the first small step towards the goal of developing a set of ethical guidelines for ESSE research. By reviewing ethical guidelines from other fields, we have identified four core ethical concepts upon which these guidelines are based: informed consent, scientific value, beneficence, and confidentiality. We illustrated these issues with ESSE case examples so that ESSE researchers will more readily see how ethics apply to their work. In addition, we have suggested some procedures to increase the compliance of ESSE research to the existing ethical guidelines.

Should ESSE researchers fail to define their own code of ethics, an unworkable code may be imposed upon them. For instance, in Canada, the three government granting agencies have created a common code of ethics applicable to all grant recipients [34]. Initial drafts of the proposed code were based primarily on biomedical codes of ethics. Some of the procedures detailed in this draft posed great problems for social science and humanities researchers. For instance, the code would not have allowed any negative art and literary critiques [8]. The code's authors have taken these concerns into account such that the final draft includes guidelines and procedures that seem appropriate for many of the affected disciplines [34]. Nonetheless, this was a very difficult and controversial exercise [8]. Similar problems have occurred in the US with respect to anthropological research [11]. These experiences should serve as a lesson to the ESSE community. The ESSE community should develop a workable set of guidelines and procedures to ensure that ESSE research can proceed ethically.

Appendix

Ethical Codes

Below is a list of the ethical codes we reviewed in writing this paper.

A. ACM Executive Council, "ACM Code of Ethics and Professional Conduct," *Communications of the ACM*, Vol. 36, No. 2, 1993, pp. 99-105, http://www.acm.org/constitution/code.html.

B. American Anthropological Association, "Code of Ethics of the American Anthropological Association," 1998. See also http://www.aaanet.org/committees/ethics/ethcode.htm.
C. American Educational Research Association, "Ethical Standards of AERA," http://www.aera.net/about/policy/ethics.htm.
D. American Psychological Association, "Ethical Principles of Psychologists and Code of Conduct," *American Psychologist*, vol. 47, No. 12, 1992, pp. 1597-1611. See also http://www.apa.org/ethics/code.html.
E. American Sociological Association, "ASA Code of Ethics," 1997, http://www.asanet.org/members/ecoderev.html.
F. American Statistical Association, "Ethical Guidelines for Statistical Practice," http://www.amstat.org/profession/ethicalstatistics.html.
G. Archaeological Institute of America, "Code of Processional Standards," http://www.archaeological.org/pdfs/AIA_Code_of_Professional_StandardsA5S.pdf.
H. Canadian Psychological Association, "Companion Manual to the Canadian Code of Ethics for Psychologists," 1991, Old Chelsea, QC: Canadian Psychological Association, 1992.
I. IEEE Ethics Committee, "IEEE Code of Ethics (IEEE Policy 7.8)," 1990, http://www.ieee.org/committee/ethics/#Code.
J. IEEE-CS/ACM Joint Task Force on Software Engineering Ethics and Professional Practices, "Software Engineering Code of Ethics and Professional Practice," 1998. See also http://www.acm.org/serving/se/code.htm.
K. National Health and Medical Research Council (NHMRC). "National Statement on Ethical Conduct in Research Involving Humans," Canberra: AusInfo, 1999. See also http://www.health.gov.au/nhmrc/publications/synopses/e35syn.htm.
L. NRC Human Subjects Research Ethics Committee, "Research Involving Human Subjects: Guidelines for Institutes," National Research Council Canada, 1995.
M. M. "Nuremberg Code," In US Department of Health and Human Services (DHHS), Public Health Service, National Institutes of Health, Office for Protection from Research Risks. Protecting Human Research Subjects: Institutional Review Board Guidebook. Washington, DC: Government Printing Office, 1993. See also http://ohrp.osophs.dhhs.gov/irb/irb_guidebook.htm.
N. Register of Professional Archaeologists (RoPA). "Code of Conduct and Standards of Research Performance," see also http://www.rpanet.org.
O. Society for American Archaeology, "Principles of Archaeological Ethics," http://www.saa.org/AboutSAA/Ethics/prethic.html.
P. Society for Applied Anthropology, "Ethical and Professional Responsibilities," http://www.sfaa.net/sfaaethic.html.
Q. The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, "The Belmont Report: Ethical Principles and Guidelines for the Protection of Humans Subjects of Research," Washington DC: US Government Printing Office, 1978. Department of Health, Education and Welfare, Publication No. (OS) 78-0012. See also http://ohrp.osophs.dhhs.gov/humansubjects/guidance/belmont.htm.
R. Tri-Council, "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans," Ottawa, Canada: Public Works and Government Services Canada 1998, Catalogue No: MR21-18/1998 E. See also http://www.nserc.ca/programs/ethics/english/policy.htm.
S. U.S. Department of Health and Human Services, "Protection of Human Subjects," Code of Federal Regulations, Title 45, Part 46, 1990.
T. UNESCO, "Ethical Guidelines for International Comparative Social Science Research in the Framework of MOST," http://www.unesco.org/most/ethical.htm.
U. UNESCO, "Universal Declaration of Human Rights," http://www.unesco.org/human_rights/dba.htm.
V. World Medical Association, "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects," as amended by the 52nd WMA General Assembly, Edinburgh, Scotland, 2000. See also http://www.wma.net/e/policy/17-c_e.html.

REFERENCES

[1] ACM Executive Council, "ACM Code of Ethics and Professional Conduct," *Comm. ACM,* vol. 36, no. 2, pp. 99-105, 1993, see also http://www.acm.org/constitution/code.html.
[2] American Anthropological Association, Code of Ethics of the American Anthropological Association, 1998, http://www.aaanet.org/committees/ethics/ethcode.htm.
[3] American Psychological Association, "Ethical Principles of Psychologists and Code of Conduct," *American Psychologist,* vol. 47, no. 12, pp. 1597-1611, 1992, see also http://www.apa.org/ethics/code.html.
[4] R. Anderson, "Social Impacts of Computing: Codes of Professional Ethics," *Social Science Computing Review,* vol. 10, no. 2, pp. 453-469, 1992.
[5] R.E. Anderson, D.G. Johnson, D. Gotterbarn, and J. Perrolle, "Using the New ACM Code of Ethics in Decision Making," *Comm. ACM,* vol. 36, no. 2, pp. 98-107, 1993.
[6] Archaeological Institute of America, "Code of Professional Standards," http://www.archaeological.org/pdfs/AIA_Code_of_Professional_StandardsA5S.pdf.
[7] U. Becker-Kornstaedt, "Descriptive Software Process Modeling—How to Deal with Sensitive Process Information," *Empirical Software Eng.,* vol. 6, no. 4, 2001.
[8] Canadian Association of University Teachers (CAUT), "CAUT Responds to Tri-Council Code," *CAUT Bull.,* Nov. 1997, see also http://www.caut.ca/english/bulletin/97_oct/tricouncil.htm.
[9] Committee on Science, Engineering and Public Policy of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *Responsible Science: Ensuring the Integrity of the Research Process,* vols. 1 and 2. Washington, D.C.: Nat'l Academy Press, 1993.
[10] Committee on Science, Engineering and Public Policy of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, *On Being a Scientist,* second ed. Washington DC: Nat'l Academy Press, 1995.
[11] B. du Toit, "Ethics, Informed Consent, and Fieldwork," *J. Anthropological Research,* vol. 36, no. 3, pp. 274-286, 1980.

[12] R.R. Faden, T.L. Beauchamp, and N.M.P. King, *A History and Theory of Informed Consent.* Oxford Univ. Press, 1986.
[13] C.E. Garza, "Studying the Natives on the Shop Floor," *Business Week,* pp. 74-78, Sept. 1991.
[14] C.E. Garza, "The Touchy Ethics of Corporate Anthropology," *Business Week,* p. 78, Sept. 1991.
[15] D. Gotterbarn, K. Miller, and S. Rogerson, "Software Engineering Code of Ethics is Approved," *Comm. ACM,* vol. 42, no. 10, pp. 102-107, 1999.
[16] IEEE Board of Directors, *IEEE Code of Ethics,* IEEE 1990, http://www.ieee.org/about/whatis/code.html.
[17] IEEE-CS/ACM Joint Task Force on Software Engineering Ethics and Professional Practices, *Software Engineering Code of Ethics and Professional Practice,* 1998, see also http://www.acm.org/serving/se/code.htm.
[18] D.R. Jeffries and L. Votta, "Empirical Software Engineering: Guest Editor's Special Section Introduction," *IEEE Trans. Software Eng.,* vol. 25, no. 4, pp. 435-437, 1999.
[19] D. Johnson, *Computer Ethics.* Englewood Cliffs, N.J.: Prentice Hall, 1993.
[20] J. Katz, *Experimentation with Human Beings.* New York: Russell Sage Foundation, 1972.
[21] P.H. Mervis and S.E. Seashore, "Creating Ethical Relationships in Organizational Research," *The Ethics of Social Research,* J. Sieber, ed., pp. 79-104, 1982.
[22] National Health and Medical Research Council (NHMRC), National Statement on Ethical Conduct in Research Involving Humans,Canberra: AusInfo, 1999, see also http://www.health.gov.au/nhmrc/publications/synopses/e35syn.htm.
[23] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Humans Subjects of Research,* Washington DC: US Government Printing Office, 1978, Department of Health, Education and Welfare, Publication No. (OS) 78-0012, see also http://ohrp.osophs.dhhs.gov/humansubjects/guidance/belmont.htm.
[24] R. Penslar, *Personal Communication.* 2000.
[25] T. Puglisi Quoted with Permission from MCWIRB, an IRB listserv, June 2001.
[26] Register of Professional Archaeologists, *Code of Conduct and Standards of Research Performance,* 2000, http://www.rpanet.org.
[27] Society for American Archaeology Ethics in Archaeology Committee, *Principles of Archaeological Ethics,* Soc. for Am. Archaeology, 1996, http://www.saa.org/Aboutsaa/Ethics/prethic.html.
[28] S. Sedgley, Personal Communication, Director, Policy & Planning, Australian Research Council, 2000.
[29] J.E. Sieber, *Planning Ethically Responsible Research.* Newbury Park, Conn.: Sage Publications, 1992.
[30] J.E. Sieber, "Protecting Research Subjects, Employees and Researchers: Implications for Software Engineering," *Empirical Software Eng.,* vol. 6, no. 4, 2001.
[31] R. Spinello, *Case Studies in Information and Computer Ethics.* Englewood Cliffs, N.J.: Prentice Hall, 1997.
[32] Society for American Archaeology, "Principles of Archaeological Ethics," 1996, http://www.saa.org/AboutSAA/Ethics/prethic.html.
[33] J. Thomas, "Introduction: A Debate about the Ethics of Fair Practices for Collecting Social Science Data in Cyberspace," *The Information Soc.,* vol. 12, no. 2, pp. 107-117, 1996.
[34] Tri-Council, "Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans," Ottawa, Canada: Public Works and Government Services Canada, 1998, Catalogue No: MR21-18/1998 E, see also http://www.nserc.ca/programs/ethics/english/policy.htm.
[35] U.S. Department of Health and Human Services (DHHS), "Protection of Human Subjects," *Code of Federal Regulations,* Title 45, Part 46, 1990.
[36] U.S. Department of Health and Human Services (DHHS), Public Health Service, National Institutes of Health, Office for Protection from Research Risks, *Protecting Human Research Subjects: Institutional Review Board Guidebook,* Washington, DC: Government Printing Office, 1993, see also http://ohrp.osophs.dhhs.gov/irb/irb_guidebook.htm.
[37] D. Waskul and M. Douglass, "Considering the Electronic Participant: Some Polemical Observations on the Ethics of On-Line Research," *The Information Soc.,* vol. 12, no. 2, pp. 129-139, 1996.
[38] J. Weckert and D. Adeney, *Computer and Information Ethics.* Westport, Conn.: Greenwood Publishing, 1997.
[39] C. Whitbeck, "Teaching Ethics to Scientists and Engineers: Moral Agents and Moral Problems," *Science and Eng. Ethics,* vol. 1, no. 3, pp. 299-307, 1995.
[40] C. Whitbeck, *Ethics in Engineering Practice and Research.* New York: Cambridge Univ. Press, 1998.
[41] World Medical Association, "World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects,"as amended by the 52nd WMA General Assembly, Edinburgh, Scotland, 2000, see also http://www.wma.net/e/policy/17-c_e.html.

Janice Singer is a cognitive psychologist working in the Software Engineering Group of the National Research Council Canada. Dr. Singer conducts studies into the work practices of software engineers with the goal of improving both software engineering tools and processes. Her current methdological approaches are primarily ethnographic, such as observation and interviewing. In the past, however, she has had a great deal of experimental experience. Dr. Singer is also interested in the application of ethics to research in software engineering. She currently serves on the Research Ethics Board of the National Research Council Canada, which reviews research proposals ranging in scope from human computer interaction studies to complex biomedical experiments.

Norman G. Vinson is a cognitive psychologist whose research focuses on visual perception and spatial cognition as they relate to human-computer interaction. He is also interested in the psychological processes involved in data visualization and analysis. His current project is to devise and test design principles supporting human navigation in virtual or augmented reality. Drs. Vinson and Singer are examining the ethical issues involved in such research to ensure that human subjects research practices at the IIT are both ethical and reasonable.