The Menlo Report

Ethical Principles Guiding Information and Communication Technology Research

August 2012

Homeland Security

Science and Technology

The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research

August 3, 2012

Executive Summary

This report proposes a framework for ethical guidelines for computer and information security research, based on the principles set forth in the 1979 Belmont Report, a seminal guide for ethical research in the biomedical and behavioral sciences. Despite its age, the Belmont Report's insightful abstraction renders it a valuable cornerstone for other domains. We describe how the three principles in the Belmont report can be usefully applied in fields related to research about or involving *information and communication technology*. ICT research raises new challenges resulting from interactions between humans and communications technologies. In particular, today's ICT research contexts contend with ubiquitously connected network environments, overlaid with varied, often discordant legal regimes and social norms. We illustrate the application of these principles to information systems security research – a critical infrastructure priority with broad impact and demonstrated potential for widespread harm – although we expect the proposed framework to be relevant to other disciplines, including those targeted by the Belmont report but now operating in more complex and interconnected contexts.

We first outline the scope and motivation for this document, including a historical summary of the conceptual framework for traditional human subjects research, and the landscape of ICT research stakeholders. We review four core ethical principles, the three from the Belmont Report (Respect for Persons, Beneficence, and Justice) and an additional principle *Respect for Law and Public Interest*. We propose standard methods to operationalize these principles in the domain of research involving information and communication technology: identification of stakeholders and informed consent; balancing risks and benefits; fairness and equity; and compliance, transparency and accountability, respectively. We also describe how these principles and applications can be supported through assistive external oversight by ethical review boards, and internal self-evaluation tools such as an Ethical Impact Assessment.

The intent of this report is to help clarify how the characteristics of ICT raise new potential for harm and to show how a reinterpretation of ethical principles and their application can lay the groundwork for ethically defensible research.

Working Group Participants

This report is the product of a series of workshops and meetings held over a period of sixteen months. The participants at these meetings are listed alphabetically below. In addition, the authors thank the dozen or so ICTR community members whose feedback was invaluable to assuring that this document reflects the ground truth sentiments of the professionals at the front lines of ICT research ethics.

- Michael Bailey, University of Michigan
- Aaron Burstein, University of California Berkeley
- KC Claffy, CAIDA, University of California San Diego
- Shari Clayman, DHS Science & Technology
- David Dittrich, Co-Lead Author, University of Washington
- John Heidemann, University of Southern California, ISI
- Erin Kenneally, Co-Lead Author, CAIDA, University of California San Diego
- Douglas Maughan, DHS Science & Technology
- Jenny McNeill, SRI International
- Peter Neumann, SRI International
- Charlotte Scheper, RTI International
- Lee Tien, Electronic Frontier Foundation
- Christos Papadopoulos, Colorado State University
- Wendy Visscher, RTI International
- Jody Westby, Global Cyber Risk, LLC

This Report is supported by funding from the U.S. Department of Homeland Security Science and Technology Directorate, Cyber Security Division. Points of view and opinions contained within this document are those of the authors and participants and do not necessarily represent the official position or policies of the U.S. Department of Homeland Security or the participants' respective employers. The content of this Report is intended to provide guidance, and it does not constitute legal advice nor should it be interpreted as conflicting with statutory mandates and other authoritative commitments governing actions by the Government.

Contents

A Introduction – Focus and Motivations

This report attempts to summarize a set of basic principles to guide the identification and resolution of ethical problems arising in research of or involving *information and communication technology* (ICT).[1] ICT is a general umbrella term that encompasses networks, hardware and software technologies that involve information communications pertaining to or impacting individuals and organizations. ICT has increasingly become integrated into our individual and collective daily lives, mediating our behaviors and communications and presenting new tensions that challenge the applications of these guiding principles.

ICT research (ICTR) involves the collection, use and disclosure of information and/or interaction with this ubiquitously connected network context which is overlaid with varied, often discordant legal regimes and social norms. The challenge of evaluating the ethical issues in ICTR stems in large part from the attributes of ICT: scale, speed, tight coupling, decentralization and wide distribution, and opacity. This environment complicates achieving ethically defensible research for several reasons. It results in interactions with humans that are often indirect, stemming from an increase in either logical or physical "distance" between researcher and humans to be protected over research involving direct intervention. The relative ease in engaging multitudes of distributed human subjects (or data about them) through intermediating systems speeds the potential for harms to arise, and extends the range of stakeholders who may be impacted. Also, legal restrictions and requirements have expanded considerably since the 1980s, and ICTR is unquestionably subject to a variety of laws and regulations that address data collection and use. While it is true that these individual complications are shared by traditional biomedical and behavioral research, this report seeks to manage the tension resulting from the simultaneous confluence of these complicating factors that occur with regularity in ICTR.

There is a need to interpret and extend the traditional ethical framework to enable ICT researchers and oversight entities to appropriately and consistently assess and render ethically defensible research.[2] Such a framework should also support current and potential institutional mechanisms that are well served to implement it, such as a research ethics board (REB). We build on the foundation set by the *Belmont Report*, which articulates three fundamental ethical principles and guiding applications of these principles for protecting human subjects of biomedical and behavioral research: respecting persons; balancing potential benefits and harms; and equitably apportioning benefits and burdens across research subjects and society.[3] The guidelines in this report are applicable to research that has the potential to harm humans, regardless of whether those humans are the direct research subjects or are indirectly at risk of harm from interactions with ICT. This report explains how the traditional framework fits within the context of the computer science sub-discipline of information security research. Specifically, this domain addresses ICT vulnerabilities, digital crime, and information assurance for critical infrastructure systems. These are areas where harms are not well understood yet are potentially significant in scope and impact. The framework proposed herein is germane to other disciplines that involve the use of ICT, including those targeted by the Belmont Report that now operate in ICT contexts.

A.1 Who is the Target Audience for this Report?

This report offers guidance primarily for ICT researchers (including academic, corporate, and independent researchers), professional societies, publication review committees, and funding agencies. Secondarily, this report aims to assist those who administer and apply these princi-

ples, such as oversight authorities (e.g., REBs), policy makers, attorneys, and others who shape and implement human subject protection policies and procedures.

This report does not recommend particular enforcement mechanisms. To the extent that enforcement of ethical practices is inconsistent across and within academic and non-academic ICTR, we intend this report to improve consistency in ethical analyses and self-regulation for both individuals and organizations striving toward ethically defensible research.

A.2 Historical Context

Despite a long history of well-publicized abuses, it took over a decade for the ethical standards prescribed in the Belmont Report to first be defined in the Code of Federal Regulations (CFR). Language from 45 CFR 46, which covers biomedical and behavioral research funded by the Department of Health and Human Services (HHS), was later adopted by all executive branch departments in what is known as the *Common Rule*.[4] It ushered in a government-wide requirement for REB oversight of research protocols to protect human research subjects. Prior to this point, there was no regulated oversight mechanism and biomedical and behavioral researchers relied on subjective, ad hoc, and inconsistent ethical compasses to guide their decision making.

In parallel during the 1970s, a U.S. Defense Advanced Research Projects Agency (DARPA) project was designing and implementing a communications architecture to support cooperative time-sharing of computational resources across large government-funded laboratories. Although this network architecture would eventually evolve into the global Internet, the community at the time was small, trusted, non-commercial, and research-oriented. This burgeoning Internet was not under constant attack from around the world. It did not provide access to numerous databases containing millions of personally-identifying records. It was not an integral part of providing and maintaining critical services or communications. A tiny number of people accessed the Internet during those early years compared to the billions of users who engage in this environment on a regular and almost unconscious basis today.

Early ICT research evolved without significant concern for human subjects, leading to instances where ethical considerations were either absent or misapplied because researchers failed to understand their relevance, or lacked any standards for assessment, accountability, or oversight. Cases include interactive studies of malicious software and platforms, engagement in active counterattack measures, exploitation and disclosure of systems vulnerabilities, and collection and sharing of sensitive information. The demonstrated potential for harm in ICTR illustrates the need to re-conceptualize the traditional *human subject protection* paradigm that underpins ethical oversight in other fields.

ICTR challenges us to re-conceptualize the traditional *human subject protection* paradigm that underpins ethical oversight. The foremost misunderstandings and disagreements about the applicability and scope of this protection in ICTR stem largely from how the Common Rule was written and has historically been interpreted. Specifically, *human subject* means, "a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information" (45 CFR 46.102(f)). Key terms here are "intervention" and "private information." Intervention does not just mean physical procedures, but also "manipulations of the [subject's] environment that are performed for research purposes," which could include manipulation of their computing devices, or automated appliances in the home. Private information is not just medical records, but "information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which

the individual can reasonably expect will not be made public." This could include electronic communications, or data captured by malicious actors recording online financial transactions in order to commit fraud. Taken as a whole, the intent of the Common Rule is to protect persons who might be harmed from involvement in research, not simply with whether humans are participating in research. Confusion starts because of the wording above and linkage of the terms *human* and *research subject*, and continues with the determination of *risk* and how to *protect humans* within a research study.

An evolved paradigm for applying ethical principles to protect humans who may be impacted by research considers activities having *human-harming* potential rather than simply looking at whether the research does or does not involve *human subjects*. Examples of potentially human-harming ICT artifacts that researchers may interact with include avatars in online virtual worlds, malware controlling compromised machines, embedded medical devices controlling biological functions, or process controllers for critical infrastructure. The significant changes brought about by ICT since the commencement of formal regulated research necessitates a reconceptualization of the application of ethical principles for research involving ICT.

B Restatement of Belmont Principles in the ICTR Context

In framing the principles and applications for evaluating and applying ethics in ICTR the Menlo Report explicitly adopts the Belmont principles and acknowledges the Common Rule regime which implemented that model. As such, this Report deliberately does not explore alternate ethical paradigms, and while not discounting that there may be novel implementations of the Belmont Report principles and applications that should be considered it makes no definitive recommendations in that regard. However, this Report does highlight areas within the Common Rule that are more consequential or problematic for ICTR.

The first three rows of Table 1 summarize the three core principles and their application as outlined in the Belmont Report.[5] We offer an additional principle to guide ethical considerations in ICTR research, listed in the fourth line of Table 1 We call this principle *Respect for Law and Public Interest* because it addresses the expansive and evolving, yet often varied and discordant, legal controls relevant to communication privacy and information assurance (i.e., the confidentiality, availability, and integrity of information and information systems). While respect for the law and public interest is implicit in Belmont's application of Beneficence, several challenging factors suggest these issues merit explicit consideration in the ICTR context: the myriad laws that may be germane to any given ICTR; conflicts and ambiguities among laws in different geo-political jurisdictions; the difficulty in identifying stakeholders, a necessary prerequisite to enforcing legal obligations; and possible incongruence between law and public interest.

C Application of the Principles

The challenges of ICTR risk assessment derive from three factors: the researcher-subject relationships, which tend to be disconnected, dispersed, and intermediated by technology; the proliferation of data sources and analytics, which can heighten risk incalculably; and the inherent overlap between research and operations. In order to properly apply any of the principles listed above in the complex setting of ICT research, it is first necessary to perform a systematic and comprehensive stakeholder analysis.

Principle	Application
Respect for Persons	Participation as a research subject is voluntary, and follows from informed consent; Treat individuals as autonomous agents and respect their right to determine their own best interests; Respect individuals who are not targets of research yet are impacted; Individuals with diminished autonomy, who are incapable of deciding for themselves, are entitled to protection.
Beneficence	Do not harm; Maximize probable benefits and minimize probable harms; Systematically assess both risk of harm and benefit.
Justice	Each person deserves equal consideration in how to be treated, and the benefits of research should be fairly distributed according to individual need, effort, societal contribution, and merit; Selection of subjects should be fair, and burdens should be allocated equitably across impacted subjects.
Respect for Law and Public Interest	*Engage in legal due diligence; Be transparent in methods and results; Be accountable for actions.*

Table 1: Proposed guidelines for ethical assessment of ICT Research.

C.1 Stakeholder Perspectives and Considerations

Stakeholder identification includes consideration of several factors: the degree to which information involved in the research identifies individuals (including their digital identities), groups and organizations and what behaviors, communications, or relationships are associated with such identification. Harms related to exposing the identity of research subjects engaging in sensitive behaviors, communications, or relationships, which they they assume to be private, can extend beyond the direct research subject to family, friends or other community relations. While this is also true of some research where the subject is the primary party at risk, in ICTR these harms may often be broader because ICT can amplify both the disclosure as well as the number of stakeholders impacted.

Further, ICTR often involves stakeholders that are non-research entities who rely on information and systems that are involved in the research and who may be harmed by its unavailability or corruption. Groups or organizations (e.g., companies or networks) may warrant different consideration from that of individuals, especially when applying the principles of Beneficence and Justice. Research involving ICT can be complex when the risks and benefits associated with multiple stakeholders require identification and balancing.

ICT Researchers In commercial, academic, and government sectors, ICT researchers have a vested interest in pursuing, sharing, and applying empirically grounded scientific knowledge. Research in economics, network science, security, and social behavior may inform operations, policies, and business models.

Human Subjects, Non-Subjects, and ICT Users Traditional biomedical and behavioral research requires protection of natural persons and certain data that identifies them. In ICTR,

the target of research may be an information system or associated data, which complicates the assessment of potential harm to users of that system or data. Primary considerations include the ability to interact with ICT without suffering harms such as disruption of access, loss of privacy, or unreasonable constraints on protected speech or activities. Victims of computer crimes are potential human non-subjects of research.

Malicious Actors A subset of ICTR involves criminal activity or potential exploitation of vulnerabilities in the design or implementation of ICT. The disclosures of some types of research results have a greater potential for misuse and thus greater value to malicious actors. This can provide a blueprint for widespread and wide-ranging harm by disclosing system vulnerability details of legitimate or malicious applications (the former by providing exploitation knowledge and the latter by illuminating countermeasures). Malicious actors avail themselves of published research results for nefarious purposes, which can result in harm that outweighs the intended research benefits. Consideration of this stakeholder's interest, therefore, involves understanding and avoiding or minimizing these potentially harmful impacts.

Network/Platform Owners and Providers Network owners or providers are typically commercial entities who are vested in safeguarding their physical and intellectual property, pursuing innovation and wealth, and building business and customer relationships. They are concerned about obligations associated with such representation. As intermediaries between a research and end users, they may be in a position of authority to serve as proxies for consent on behalf of their customers when it is otherwise impracticable for the researcher to individually obtain informed consent from end users.

Government: Law Enforcement Public law enforcement is mandated to advance criminal justice by protecting individuals and fostering public safety. Law enforcement also has an interest in research that improves its strategic, tactical, or operational efficacy in preventing, investigating, and responding to illegal activities. Examples include countering new and complex criminal ecosystems and instruments of crime such as botnets.

Government: Non-Law Enforcement Local, state, and federal government agencies are responsible for providing public services, protecting the rights of their citizens, and establishing law and policy governing social conduct. Research is an important vehicle through which the government can promote social good and innovation. For example, cybercrime research may enhance understanding of infrastructure risks, online social networks, or economic markets of criminal enterprises; influence the deployment of commercial countermeasure technologies; and inform the interpretation or reform of relevant laws and policies. Acknowledging the different scope of their mission, the military and Intelligence Community (IC) is another subset of this stakeholder group.

Society ICTR implicates the collective rights and interests of owners and users of networks and data to know, influence, and choose how and when to engage with information communications networks and systems. Society benefits from knowledge that improves policies, laws and the administration of justice, and the well-being of the lives of its citizens. Society may likewise be harmed through actions that negatively impact information systems infrastructures, or through the collection, use, or disclosure of information that may assist criminals as much if not more than ICT system developers and operators.

C.2 Respect for Persons

In the Belmont Report, the principle of Respect for Persons reflects two tenets: individuals should be treated as autonomous agents, and persons with diminished autonomy are entitled

to protection. This principle has been applied by involving as research subjects only those with sufficient understanding or awareness to provide *informed consent*, or by obtaining informed consent from legally authorized representatives (e.g., parents of minors, relatives of unconscious patients, or guardians of those incapable of deciding for themselves). In the ICTR context, the principle of Respect for Persons includes consideration of the computer systems and data that directly interface, integrate with, or otherwise impact persons who are typically not research subjects themselves.

C.2.1 Informed Consent

Informed consent is a process during which the researcher accurately describes the project and its risks to subjects and they accept the risks and agree to participate or decline. Subjects must be free to withdraw from research participation without negative consequences. Researchers obtain informed consent when research activity has the potential to harm individuals with whom a researcher interacts or about whom the researcher obtains identifiable private information. Research involving ICT also raises the potential for harms to secondary stakeholders who, while not the direct subjects of research, may have the right to autonomy.

Researchers should inform subjects that they may not benefit from the research, although society may benefit in the future. Researchers should be mindful that leveraging intended benefits to coerce or entice consent from subjects fails the voluntary participation element of informed consent. Examples include suggesting that research participants will receive improved or enhanced services, or that services will be degraded or withheld if a subject declines participation in or withdraws from a study. Informed consent for one research purpose or use should not be considered valid for other research purposes. When an individual is identified with a group or organization, individual consent does not imply consent from other members of the group. Finally, informed consent for one research purpose or use should not be considered valid for different research purposes.

The process of informed consent is intended to respect the autonomy of research subjects. The process involves three components: notice, comprehension, and voluntariness. Notice is typically achieved through a clearly written consent document that details the intended benefits of research activities and the risks to research subjects. The language level is kept to 8th grade or lower to improve the ability of subjects to comprehend the benefits and risks. The consent document stresses that participation is voluntary and that subjects are free to withdraw from research participation without negative consequences.

Research involving ICT also raises the potential for harms to secondary stakeholders who, while not the direct subjects of research, may also have the right to autonomy. When considering informed consent, we suggest researchers and REBs carefully explore the complex interconnected relationships between users and the myriad of organizations which provide ICT services. Decisions about mechanisms for obtaining informed consent, or requesting waivers of informed consent, may be impacted by whether entities have obtained valid authorization from their users – via explicit agreements or contractual terms of service – for participation in research activities. Such authorization, whether supportive or restrictive of research, should be appropriately balanced when considering informed consent.

When a researcher believes that obtaining informed consent makes the pursuit of research objectives impossible, the application process allows for researchers to seek waivers from an ethical review board. REBs make the determination of whether or not the Common Rule criteria of 45 CFR 46.116 and 45 CFR 46.117 allowing for alteration or elimination of informed consent have been met. These requirements ensure that: (1) The research involves no more

than minimal risk to the subjects; (2) The waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) The research could not practicably be carried out without the waiver or alteration; and (4) Whenever appropriate, the subjects will be provided with additional pertinent information after participation.

There are justifiable reasons why it may be impracticable for research to be carried out without a waiver or alteration of the informed consent process. Because of the difficulty in identifying all individuals from whom consent should be sought or in practicably obtaining consent, researchers or REBs may frequently conclude that seeking a waiver of informed consent or waiver of documentation of informed consent are the only options. For example, it may be infeasible to identify, or obtain consent from millions of users whose everyday communication generates traffic across a heavily aggregated backbone link in a traffic modeling study. Or it can be impossible to attempt to inform the owners of hundreds of thousands of compromised home computers that are being used as a single instrument of criminal activity (i.e., a botnet) under study. The Common Rule criteria for a waiver of documentation of informed consent in minimal or no-risk situations allows for less formal consent than a signed consent form, including verbal consent from a legally authorized representative rather than the research subjects themselves. REBs may also require some form of notification to research subjects, even if the REB does not require signed consent forms.

Some research involving retrospectively collected identifiable data may not be possible if consent must be obtained from all individuals identifiable in the data. In such situations, respect for persons is maintained by REBs instead focusing on data protections and/or removal of identifying information that is not germane to research as alternative means of minimizing potential harm and granting a waiver of informed consent for the research. Data that has already been de-identified and can be approved for exemption from REB review falls into a a special regulatory category of "pre-existing public data." REBs have some flexibility in how they define and interpret this class of data and some institutions maintain a list of pre-approved sources of such data that researchers may freely use. Data that is not on such pre-approved lists that contains fields that can identify individuals – even though it may be accessible to the general public – may not be considered "pre-existing public data." Researchers should therefor consult with their REB to discuss whether the data they wish to use falls under their institution's "pre-existing public data" exemption criteria, or whether they can qualify for a waiver of informed consent to re-use existing data in conformance with REB requirements. Prospective research is the more problematic case, where informed consent may be required by an REB unless it can be shown there is no risk what so ever.

As a contingency of granting a waiver of informed consent, REBs often require that the researcher notify subjects post hoc of their involvement in research, and demonstrate respect for autonomy by allowing subjects to direct the destruction of the data collected about them. Research involving deception may be performed by providing misleading data in the consent form, or with consent having been waived and no subject knowledge of the research activity at all. In either case, an REB may require debriefing in order to mitigate harm resulting from loss of trust in researchers by those subjects who were deceived. Research of criminal activity often involves deception or clandestine research activity, so requests for waivers of both informed consent and post hoc notification and debriefing may be relatively common as compared with research studies of non-criminal activity.

C.3 Beneficence

In the Belmont Report, the Beneficence principle reflects the concept of appropriately balancing probable harm and likelihood of enhanced welfare resulting from the research. Translating this principle to ICTR demands a framework for systematic identification of risks and benefits for a range of stakeholders, diligent analysis of how harms are minimized and benefits are maximized, preemptive planning to mitigate any realized harms, and implementing these evaluations into the research methodology.

C.3.1 Identification of Potential Benefits and Harms

Similar to traditional human-centered research, ICT researchers should identify benefits and potential harms from the research for all relevant stakeholders, including society as a whole, based on objective, generally accepted facts or studies. Since communication technologies intermediate so much of our lives, designing, conducting and evaluating ICTR may demand attention to potential societal benefits and harms related to: systems assurance (confidentiality, availability, integrity); individual and organizational privacy; reputation, emotional well-being, or financial sensitivities; and infringement of legal rights (derived from constitution, contract, regulation, or common law). Challenges identifying harms in ICTR environments stem from the scale and rapidity at which risk can manifest, the difficulty of attributing research risks to specific individuals and/or organizations, and our limited understanding of the causal dynamics between the physical and virtual worlds. As with all exploratory research, it can be challenging to articulate benefits such that subjects can make informed decisions. In ICTR our ability to qualitatively and quantitatively foresee the probable benefits is particularly immature.

One helpful approach to identifying harms is to review the laws and regulations that apply to an ICTR activity, and analyze the underlying individual and public interests that the research might negatively impact. While researchers are not expected to render legal conclusions or have legal subject matter expertise, they are obligated to respect what is written in the law and understand the underlying societal norms those laws represent. However, as the development of the law and technology occur at a different trajectory and pace, relying exclusively on the law may overlook important harms not expressly addressed by law. Similarly, it is not the role of researchers to judge guilt or innocence, but they should consider how malicious actors might avail themselves of published research results for nefarious purposes, and assess whether that potential harm might outweigh the intended research benefits.

C.3.2 Balancing Risks and Benefits

A simplistic interpretation of Beneficence is the maximization of benefits and minimization of harms. Beneficence does not require that all harm be completely eliminated and every possible benefit be identified and fully realized. Rather, researchers should systematically assess risks and benefits across all stakeholders. In so doing, researchers should be mindful that risks to individual subjects are weighed against the benefits to society, not to the benefit of individual researchers or research subjects themselves. Ideally, researcher actions are measured using the objective standard of a *reasonable researcher*, who exercises the knowledge, skills, attention, and judgment that the community requires of its members to protect their interests and the interests of others. As researchers gain a greater understanding of how to reason about and apply ethical principles, community norms and expectations about what is *reasonable* will evolve. From the subjective perspective of the researcher, especially in light of evolving community

standards, the elements of "integrity" are instructive: (1) discerning what is right and what is wrong, (2) acting on what you have discerned, even at personal cost; and (3) saying openly that you are acting on your understanding of right and wrong." [6]

When ICT is involved, burdens and risks can extend beyond "the human subject," making the quantification of potential harm more difficult than with direct intervention. It can be difficult to balance risks and benefits with novel research whose value may be speculative or delayed, or whose realized harm may be perceived differently across stakeholders. If there are plausible risks, researchers bear the burden of illuminating those risks and their consideration of how those risks will be managed, and not simply rely on outside reviewers or REBs to identify and oversee those risks.

In a direct intervention research scenario, balancing is partially addressed through the informed consent process. When a study involves minimal risk and a researcher can give valid scientific reasons for altering or eliminating the consent requirement, post-research debriefing may be required to respect individual autonomy. Balancing benefit and harm gets complicated when both deception and waiver of informed consent are involved, as may occur when studying social engineering using email (i.e., phishing). A researcher may seek to justify a waiver of the debriefing requirement under a *relative degree of harm* rationale, whereby deceived research subjects could suffer more harm from knowing researchers had deceived them than they would suffer from malicious actions. This in turn would be balanced by an REB against the knowledge developed through research intended to ameliorate the malicious harm. The process of comprehensive stakeholder analysis can assist both researchers and REBs to consider how best to balance benefit and harm in conformance with Common Rule waiver justification requirements (see 45 CFR 46.116 and .117).

While it is incumbent upon a researcher to identify and minimize potential harms, even with reasonable measures to detect and reduce them, harms may still occur. REBs must evaluate such risks in the context of what at-risk individuals actually experience in normal ICT usage, and in light of researchers' pursuit of generalizable knowledge that is vital to understanding the problem studied. For example, a researcher studying live malicious software may need to run the software on his own platform and observe its interactions with the criminals controlling it. Even with multiple layers of protection, the malicious software under study could still accidentally infect other computers. The risks posed by these accidental infections must be considered in light of everyday events that users encounter – programs crashing, malicious software accessing and infecting networked computers, and electronic communications being exposed – and must be balanced with potential benefits of understanding the behavior of the malicious software. Ethically defensible Beneficence lies on a spectrum between unequivocal adherence to averting all risk, which can have a chilling effect on beneficial research, and acting without regard to risk, which can be harmful to individuals and society.

C.3.3 Mitigation of Realized Harms

Some research involves greater than minimal risk, yet still has the potential to yield benefit to society and is allowed to be carried out. Despite appropriate precautions and attempts to balance risks and benefits in ICTR, such research may cause unintended side effects that harm stakeholders. Data breaches are one such form of harm, but others may exist from disruption of information systems. Research of greater than minimal risk that has been approved by an REB must undergo continuation review regularly in accordance with the period set for the study by the specific REB, but no less than annually. While reporting of adverse events is part of regular status reports, "serious adverse events" may need to be reported immediately to an REB for

possible actions. This can include the REB requiring a halt to research activities. For the same reasons that benefit is hard to calculate in ICTR, determining what could constitute a "serious adverse event" in the ICTR context is unclear.

In anticipation, researchers should consider preempting the escalation of realized harms by notifying affected parties or otherwise engaging mitigation actions. To that end, researchers should develop mitigation procedures and checklists, such as a contact list of parties to notify, if such unintended consequences ensue. Other potential harms that are reasonably foreseeable may have a low probability of occurring, but have a high impact. Researchers should anticipate such worst-case scenarios and make appropriate preparations to respond in a manner and scope that shows due diligence on the part of the researcher. It may be necessary and prudent to involve the researchers' own institutional risk management and oversight authorities and media relations in addition to the REB.

ICTR may involve records containing sensitive data about individuals, evidence of criminal activity, or that could potentially cause disruption to millions of computers around the world. ICT researchers must be aware of these harms as not only primary risks, but also secondary, collateral risks (e.g., to customers of primary data subjects or computer owners) and be prepared to responsibly inform affected stakeholders. In many cases, it is impracticable to notify all affected individuals, but it may be feasible to notify service providers or other entities who have the authority and capability – derived from their relationship with the affected stakeholders – to mitigate harm. A mitigation strategy should admit the variance in capacity and/or willingness of the notified entity to understand and act on the notification.

Research records that identify individuals pose a risk of disclosure as long as those records exist, and may fall under REB oversight because of the risk posed. Researchers should be prepared to continually protect these records for as long as those records exist and are under researchers' control. Upon completion or termination of approved research activities (allowing for a reasonable retention period approved by REBs in order to satisfy obligations of scientific reproducibility), the risky data should be destroyed. If records are maintained, the data should continue to be protected at the same level as was implemented during research under the same REB-approved mechanisms.

C.4 Justice: Fairness and Equity

In the Belmont Report, the principle of Justice is applied through fairness in the selection of research subjects, and equitable distribution of the burdens and benefits of research according to individual need, effort, societal contribution, and merit. Fairness should guide the initial selection of the subjects, as well as the apportionment of burdens to those who will most likely benefit from the research. Research design and implementation should consider all stakeholders' interests, although conflicting interests may render equal treatment impracticable. In the ICTR context, this principle implies that research should not arbitrarily target persons or groups based on attributes including (but not limited to): religion, political affiliation, sexual orientation, health, age, technical competency, national origin, race, or socioeconomic status. Neither should ICTR target specific populations for the sake of convenience or expediency.

It is important to distinguish between purposefully *excluding* groups based on prejudice or bias versus purposefully *including* entities who are willing to cooperate and consent, or who are better able to understand the technical issues raised by the researcher. The former raises Justice concerns, while the latter demonstrates efforts to apply the principles of Respect for Persons and Beneficence and still conduct meaningful research. All researchers have an obligation to not exclude/include individuals or groups from participation for reasons unrelated to the

research purpose. The arbitrary targeting of subjects in ways that are not germane to pursuing legitimate research questions violates this principle.

Challenges to obtaining informed consent from users might motivate a researcher to work with a service provider who has direct contractual relationships with its network's users. These may serve as legally authorized representatives as described in the Common Rule for situations of minimal risk and requests for waivers of documentation of consent through "short form" or verbal consent. Such decisions to engage entities who are willing and able to act as legally authorized representatives for obtaining consent and move forward with non-representative subject populations may raise fairness and equity concerns. Each provider with whom a researcher may interact will have varying levels of understanding and ability (or willingness) to act. If a researcher is required to get unanimous and uniform responses from all autonomous entities, it may be impossible to perform beneficial research. On the other hand, moving forward with risky research without the involvement, or at least awareness, of autonomous entities is undesirable as it may increase the potential for greater harm.

From an equity standpoint, open public disclosure of system vulnerabilities demands that researchers consider how the burdens and benefits of publicizing newly discovered vulnerability balance out. The burdens might be borne by the developers, yet actually might benefit malicious actors more in the short-term than developers or users of those systems. The calculation of benefits is actually a function of time, where malicious actors may act faster at exploiting vulnerability information than benevolent actors can act in mitigating the vulnerabilities.

C.5 Respect for Law and Public Interest

Respect for Law and Public Interest is implicit in the Belmont Report's application of Beneficence. In the context of ICTR, we include it as a separate principle with two applications – *Compliance* and *Transparency and Accountability*. The second application refers to transparency of methodologies and results, and accountability for actions. Transparency and accountability serve vital roles in many ICTR contexts where it is challenging or impossible to identify stakeholders (e.g., attribution of sources and intermediaries of information), to understand interactions between highly dynamic and globally distributed systems and technologies, and consequently to balance associated harms and benefits. A lack of transparency and accountability risks undermining the credibility of, trust and confidence in, and ultimately support for, ICT research.

There may be a conflict between simultaneously satisfying ethical review requirements and applicable legal protections. Even if a researcher obtains a waiver of informed consent due to impracticability reasons, this may not eliminate legal risk under laws that require consent or some other indication of authorization by rights holders in order to avoid liability. For example, information privacy and trespass statutes prohibit researchers from accessing, acquiring or disclosing communications or other protected information without the consent of the communicating parties or owner of the system. Until REBs can overcome limited ICT expertise on committees and in administrative staff positions, they may not be capable of recognizing that certain ICT research data actually presents greater than minimal risk and may erroneously consider it exempt from review or subject it to expedited review procedures that bypass full committee review. As long as there is a gap in the capacity of REBs to properly evaluate research proposals just entering the review process, researchers following the guidance provided in this report can help illuminate the risks and relevant laws so as to improve the REB oversight process.

C.5.1 Compliance

Researchers should engage in due diligence to identify laws, regulations, contracts, and other private agreements that are applicable to their research, and should design and implement ICTR that respects these restrictions. While legal controls that call for compliance can be numerous and wide-ranging, those that should inform ethical assessments cluster categorically around computer crime and information security, privacy and anonymity, intellectual property, computer system assurance, and civil rights and liberties. More specifically, ICT research may implicate rights and obligations related to: identity theft; unsolicited bulk electronic mail; privacy in electronic and wire communications; notification of security breaches; copyright and other intellectual property infringement; data security and destruction; child pornography; spyware and phishing; fraudulent deception; financial privacy; economic espionage; constitutional privacy; health information security and privacy; industry standards and best practices; and contractual privacy and acceptable use policies.

Respect for public interest can often be addressed by obeying relevant laws. If applicable laws conflict with each other or contravene the public interest, researchers should have ethically defensible justification and be prepared to accept responsibility for their actions and consequences.

C.5.2 Transparency and Accountability

Transparency is a mechanism to assess and implement accountability, which itself is necessary to ensure that researchers behave responsibly. These applications interact to ultimately generate trust in ICTR by the public. Transparency-based accountability helps researchers, oversight entities, and other stakeholders avoid guesswork and incorrect inferences about whether, where, and how ethical principles are addressed. Transparency entails clearly communicating the purposes of research – why data collection and/or direct interaction with ICT is required to fulfill those purposes – and how research results will be used. It also involves clear communication of risk assessment and harm minimization related to research activities.

Accountability demands that research methodology, ethical evaluations, data collected, and results generated should be documented and made available responsibly in accordance with balancing risks and benefits. Data should be available for legitimate research, policy-making, or public knowledge, subject to appropriate collection, use, and disclosure controls informed by the Beneficence principle. The appropriate format, scope and modality of the data exposure will vary with the circumstances, as informed by Beneficence determinations.

D Implementing the Principles and Applications

This document describes foundational ethical principles and their applications at a level intended to span a broad range of current and future research that will undoubtedly be affected by changes in ICT. For federally funded biomedical and behavioral research, the responsibility for evaluating whether a research project comports with these principles lies with REBs, which in the United States are known as Institutional Review Boards (IRBs). IRB review is a requirement for federally funded research, however researchers in the ICT field frequently either do not know of this requirement, or believe that they are not engaged in "human subjects research" and do not interact with their IRB at all. This report contends that ICTR will benefit from similar oversight, and the proposed guidelines will assist ICT researchers and oversight authorities identify, preempt and manage ethical risks. Current ICTR that does not fall under the purview of REBs would also benefit from community-derived self-regulation guided by

this report. Proactively and transparently engaging in ethical assessment of ICT research will help move the research community mindset in the direction of embedding ethics into ICTR design as productively and safely as possible, and more practically influence policy and governance at these crossroads.

Notes

[1]The term *information and communication technology* was coined by Denis Stevenson in a 1997 report to the United Kingdom government, *Information and Communication Technologies in the UK Schools: An Independent Inquiry* http://rubble.heppell.net/stevenson/ICT.pdf

[2]This report offers pragmatic guidance in the application of these fundamental principles to ICTR, and avoids taking a position in the philosophical debate about the uniqueness of computer ethics. For an overview of the philosophical debate, see Bynum, Terrell, "Computer and Information Ethics", The Stanford Encyclopedia of Philosophy (Winter 2008 Edition), Edward N. Zalta (ed.). http://plato.stanford.edu/archives/win2008/entries/ethics-computer/

[3]*The Belmont Report*, the touchstone document guiding human subjects research in the biomedical and behavioral research fields, was named after the conference center where it was drafted in 1976 (See http://ohsr.od.nih.gov/guidelines/belmont.html) This document similarly takes its name from the city where a substantial portion of the working group meetings that resulted in this document took place in 2009-2010.

[4]Fifteen government departments and agencies performing research involving human subjects adopted 45 CFR 46 Subpart A in what is known as the *Common Rule*. Each has its own guidance on the interpretation of their section of the CFR. Refer to guidance appropriate to the funding source.

[5]*See* http://ohsr.od.nih.gov/guidelines/belmont.html

[6]Stephen L. Carter Carter, Stephen L (1996). Integrity. New York: BasicBooks/HarperCollins. pp. 7, 10. ISBN 0-06-092807-7.