

PUTTING PEOPLE ON THE MAP

PROTECTING CONFIDENTIALITY WITH LINKED SOCIAL-SPATIAL DATA

Panel on Confidentiality Issues Arising from the Integration of
Remotely Sensed and Self-Identifying Data

Myron P. Gutmann and Paul C. Stern, Editors

Committee on the Human Dimensions of Global Change
Division of Behavioral and Social Sciences and Education

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

PANEL ON CONFIDENTIALITY ISSUES ARISING FROM
THE INTEGRATION OF REMOTELY SENSED AND
SELF-IDENTIFYING DATA

- MYRON P. GUTMANN, *Chair*, Inter-university Consortium for
Political and Social Research, University of Michigan, Ann Arbor
- MARC P. ARMSTRONG, Department of Geography, University of Iowa
- DEBORAH BALK, School of Public Affairs, Baruch College, City
University of New York
- KATHLEEN O'NEILL GREEN, Alta Vista Company, Berkeley, CA
- FELICE J. LEVINE, American Educational Research Association,
Washington, DC
- HARLAN J. ONSRUD, Department of Spatial Information Science and
Engineering, University of Maine
- JEROME P. REITER, Institute of Statistics and Decision Science, Duke
University
- RONALD R. RINDFUSS, Department of Sociology and the Carolina
Population Center, University of North Carolina at Chapel Hill
- PAUL C. STERN, *Study Director*
- LINDA DEPUGH, *Administrative Assistant*

Contents

Executive Summary	1
1 Linked Social-Spatial Data: Promises and Challenges	7
2 Legal, Ethical, and Statistical Issues in Protecting Confidentiality	26
3 Meeting the Challenges	42
4 The Tradeoff: Confidentiality Versus Access	59
References	71
Appendixes	
A Privacy for Research Data <i>Robert Gellman</i>	81
B Ethical Issues Related to Linked Social-Spatial Data <i>Felice J. Levine and Joan E. Sieber</i>	123
Biographical Sketches for Panel Members and Staff	160

Appendix B

Ethical Issues Related to Linked Social-Spatial Data

Felice J. Levine and Joan E. Sieber

The ethics of research related to linking geographically explicit spatial data¹ and individual-, household-, or group-level social data is an issue of scientific and social significance. The capacity to measure location and context over time and with exact precision offers substantial opportunities to comprehend human, social, biological, and environment activities, interactions, and transformations at a level of sophistication that could not have been anticipated just a decade ago. The mesh of technological advances, computational capacity, multilevel statistical models, spatial analysis software, and robust data mining and management techniques makes it a ripe time for new explorations and applications to come to the fore using very precise locational information.² Along with these improved measurements and analytic methods come ethical issues regarding how best to use these new capabilities consonant with protecting the interests of research participants involved in such studies.

The most immediate ethical issue raised by linking different datasets or resources of any form is whether the integration of such information encroaches on the privacy of research subjects or compromises the confidentiality of information that otherwise is secure. Attention to issues of privacy of persons and confidentiality of data has increased over recent years.³ There is growing awareness of the scientific value of sharing data, the greater contributions made possible with microlevel data, and the potential uses from linking different datasets. Yet there is also mindfulness of the potential risks of confidentiality breaches due to intentional or inadvertent disclosure. In this current context, not unexpectedly, opportunities for link-

ing social and spatial data have also been accompanied by serious discussion of the confidentiality issues and policies involved in doing so (see, e.g., Rindfuss and Stern, 1998; VanWey et al., 2005; Golden, Downs, and Davis-Packard, 2005; Gutmann et al., 2005).

Whether in the biomedical or the social-behavioral sciences, new methodological capabilities or work at the frontiers of discovery invariably requires fresh consideration of ethical issues as an integral part of research. Especially in nascent areas of science in which practical experience is limited, grappling with ethical issues needs to go hand-in-hand with confronting theoretical, methodological, and operational considerations.⁴ Thus, it is notable that those attracted to or engaged in linking spatial and social data have already initiated the process of thinking reflectively and constructively about matters of confidentiality and reduction of the risk of information disclosure. The establishment of a National Research Council Panel on Confidentiality Issues Arising from the Integration of Remotely Sensed and Self-Identifying Data, with funding from the National Institutes of Health, the National Science Foundation, and the National Aeronautics and Space Administration, to address such confidentiality issues is a strong indicator of the salience of this topic to data producers, users, archivists, database managers, and those who review and support such work.

The purpose of this paper is to consider the ethical issues that come into play in research that links social and spatial data. Our aim is to present an overview of the ethical issues regarding the protection of human subjects, for researchers engaged in primary collection of social and spatial data, and for those engaged in secondary use of such data. First, we briefly highlight the ethical guidance available for researchers or research teams as they consider how best to undertake research on these data or provide such data to others. Second, we elaborate on and recommend as guidance the framework of ethical principles enunciated in the now classic 1979 Belmont Report, *Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). Third, we consider the range of ways ethical issues can manifest themselves in the course of collecting, providing, or using linked social-spatial data and how researchers might best advance ethically sound research and approach review by an institutional review board (IRB). Fourth, we examine such issues as consent, privacy and confidentiality, benefits and harm, and assessments of risk of harm and how to address them in research that either links or uses linked social-spatial data. Fifth, we specifically discuss the ethics of data dissemination, sharing, and access—emphasizing issues important to social-spatial research. Finally, we consider ethics education and training for those who collect, prepare, provide access to, use, or review research that links social and spatial data.

This focus on ethical considerations in social and spatial research is distinct from an analysis of the legal requirements that could apply depending on the data that are to be obtained. Use of extant information sources may be protected by privacy laws. Some of the most promising social and spatial research is addressed to issues in which privacy regulations are germane. Health research, for example, is a key area of inquiry in which access to confidential records, including precise locational information, could have tremendous scientific value and benefits to society. The Health Insurance Portability and Accountability Act of 1996 (HIPAA)⁵ protects individual privacy but allows for the use of health records for research without individual authorization. Such research needs to be evaluated as no more than minimal risk and needs to conform with a set of procedures and alternative methods to avert disclosure (e.g., meeting 18 specified criteria for deidentification, having a qualified expert determine what needs to be done to prepare the data for release).⁶ While researchers, data providers, and research analysts need to be mindful of legal requirements in planning their research, our purpose is directed to the ethical considerations that should guide collecting, gaining access to, analyzing, disseminating, or sharing such data irrespective of whether certain standards of privacy and confidentiality are required by law.

In emphasizing ethical considerations in research linking social and spatial data, we also do not intend to sidestep attention to the human research protection programs in place at academic or research institutions or the centrality of their IRBs for approval and oversight of research. Nor do we intend to minimize the challenge that can be involved in raising complex ethical issues to IRBs in areas in which the decision-making procedures are not yet developed. We do discuss the IRB review process directly. Our purpose in taking a broader approach to ethical decision making with social and spatial data is to focus attention on the research enterprise itself and how best to weigh factors in planning and executing research or in using or making accessible linked social-spatial data. We consider interaction with IRBs to be a key step in that process. While IRBs have direct institutional responsibility for the review of protocols and determinations about human research protection as stipulated in the Code of Federal Regulations for the Protection of Human Subjects (45 CFR 46),⁷ we see this interaction between researcher (producer/user) and IRB, and how to navigate it, as a part of the process of ethical decision making in human research, not as constituting that process in and of itself. Furthermore, many decisions having ethical implications are identifiable to the researcher not only prior to interacting with the IRB but also afterward; we regard these latter decision points to be integral to the overall process of ethical conduct.

ETHICAL GUIDANCE AND HUMAN RESEARCH PROTECTION

Given the social and behavioral science backgrounds of many of those engaged in social and spatial research, it might be expected that ethical norms would primarily derive from frameworks in these and adjacent fields. Although there are variations among codes of conduct (e.g., whether or not a code of ethics explicitly encourages data sharing), general standards in the social sciences have much in common regarding such issues as informed consent, intrusions on privacy, confidentiality and its limits, and benefits and harm. Whether the codes were promulgated in detail by the American Psychological Association (2003) or the American Sociological Association (1997) or in more summary fashion by the American Anthropological Association (1998), the Association of American Geographers (1998), the American Political Science Association (1998), the American Statistical Association (1999), or the American Association for Public Opinion Research (2003), there is on balance considerable consistency in their guidance.

One visible marker of specific interest in ethical considerations related to spatial data is the approval in 2003 of a geographic information systems (GIS) code of ethics by the Urban and Regional Information Systems Association (2003). By design, the code builds on a study of several dozen other codes. It states, among other guidance, that the GIS professional will protect individual privacy, especially about sensitive information; will encourage individual autonomy, including allowing individuals to withhold consent from being in a database, correct information, or remove themselves from a database; and will avoid undue intrusions into the lives of individuals (Urban and Regional Information Systems Association, 2003).

Exposure to research with human participants and related codes of conduct is by no means uniform among scientists and other specialists engaged in social and spatial research. Experts in remote sensing and other sophisticated locational measurements are typically not from the social and behavioral sciences or the health sciences, in which individuals or groups are the focus of inquiry and in which ethical guidance emphasizes the protection of human participants in research. Thus, in addition to the scientific richness of this interdisciplinary arena of study, there is also the challenge of fostering a deep appreciation among diverse researchers and research communities of the ethical issues at stake at each stage of the research process, from primary data collection through secondary use.

A second challenge flows from the fact that there is very limited research-based evidence about how ethical issues related to human research protection play out in the context of the collection or use of social and spatial research. In general, empirical study of ethical issues is far too scant across even well-established domains of inquiry, let alone new areas of research.⁸ The small body of literature addressed to linking social and

spatial data evidences an appreciation that this research area is a dynamic and fluid one and that expert knowledge can help produce research approaches that maximize advancing science consonant with human research protection principles. For example, Armstrong, Rushton, and Zimmerman (1999) do so by examining alternative methods of masking individual-level health data, testing the security of each approach for preserving confidentiality while permitting important uses. Similarly, Kwan, Casas, and Schmitz (2004) test three geographic masks with different perturbation radii to identify the optimum tradeoff between data confidentiality and accuracy of analytic results. These forms of empirical examination hold promise of producing useful guidance. Less directly, but also germane, Kwan and Lee (2004), using three-dimensional geovisualization methods and activity-travel diary data, found gender differences in time use, mobility, and travel patterns, but at the same time they cautioned that “individual-level activity-travel data geocoded to street addresses, given their reasonable degree of positional accuracy, may lead to considerable risk of privacy violation” (p. 63).⁹

THE BELMONT PRINCIPLES AS AN ETHICAL FRAMEWORK

In addition to drawing on ethics codes, recent national commissions, and relevant National Research Council panels, contemporary discussions of ethical considerations with social and spatial data (largely directed to issues of confidentiality) are taking place in the context of more than a 30-year history of ongoing attention to these issues in research and writing.¹⁰ More visible than any other, the Belmont Report articulated three overarching ethical principles that continue to offer a framework for responsible research conduct as well as form the basis of the Code of Federal Regulations for the Protection of Human Subjects (45 CFR 46). This report, issued by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, states the purpose of these principles as follows (p. 3):

Three principles, or general prescriptive judgments, that are relevant to research involving human subjects are identified in this statement.

Other principles may also be relevant. These three are comprehensive, however, and are stated at a level of generalization that should assist scientists, subjects, reviewers and interested citizens to understand the ethical issues inherent in research involving human subjects. These principles cannot always be applied so as to resolve beyond dispute particular ethical problems. The objective is to provide an analytical framework that will guide the resolution of ethical problems arising from research involving human subjects.

Thus, in setting forth these principles, the commission sought not to dictate but to create a culture of ethical decision making that could effectively serve researchers and IRBs alike.

The three ethical principles that are the foundation of the Belmont Report are respect for persons, beneficence, and justice. Depending on the complexities of a situation, the Belmont Report emphasizes that ethical decision making can—and often does—require balancing competing claims in order to accomplish the overall goals of the principles themselves. Briefly put, the principles are defined as:

1. Respect for Persons—Respect for persons incorporates at least two ethical convictions: first, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protection. . . . In most cases of research involving human subjects, respect for persons demands that subjects enter into the research voluntarily and with adequate information. . . .

2. Beneficence—Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being. . . . The obligations of beneficence affect both individual investigators and society at large, because they extend both to particular research projects and to the entire enterprise of research. . . .

3. Justice—Who ought to receive the benefits of research and bear its burdens? This is a question of justice, in the sense of “fairness in distribution” or “what is deserved.” An injustice occurs when some benefit to which a person is entitled is denied without good reason or when some burden is imposed unduly. . . .

It is the application of the principles of the Belmont Report that leads to considerations of informed consent, risk-benefit assessment, and the selection of subjects for research. As specified in the Belmont Report, respect for persons requires *informed consent* of research participants—meaning the provision of adequate information, participants’ comprehension of that information, and their voluntariness to be part of the research. *Assessment of risk and benefits* of research is closely related to beneficence—including an assessment of the probability of experiencing a harm, the magnitude of that harm (whether physical, psychological, legal, social, or economic), and the benefits that might derive to research participants or society from that research. The importance of risk reduction is also a concept emphasized in the Belmont ethical guidance. The third Belmont principle—justice—is embodied in the requirement that the *selection of subjects* needs to be appropriate to the research and ought not

to place an undue burden on certain populations or disadvantage them through omission as research participants.¹¹

Privacy and confidentiality are not explicitly mentioned in the Belmont Report, although they follow from the principles of respect for persons and beneficence and both are made explicit in 45 CFR 46.¹² *Privacy* refers to the interest that persons have in controlling others' access to them and private information about them. Individuals can vary in what they consider intrusive about themselves. In a research context, as long as human subjects willingly agree to participate in the research, can freely decide against providing certain forms of information, and can end their participation at any point, they have preserved their privacy right to control their information.

Confidentiality refers to how data will be handled by researchers, other data producers, and ultimately secondary analysts consonant with agreements with human subjects regarding private information.¹³ A corollary to participants' providing access to information in this trusting relationship is that researchers have the ethical responsibility to avoid intrusion on participants' privacy and to minimize the likelihood of harm from the disclosure of private information (both identity and attribute disclosure¹⁴). This commitment takes the form of a *confidentiality agreement* that provides assurances to research participants about what will be done with identifiable and private information about them. Except when data are collected *anonymously* (i.e., without identifying information) or the researcher is collecting only public information, the Belmont principles of respect for persons and beneficence lead researchers to consider confidentiality as part of the consent process and put into place data protection plans to reduce the likelihood of *personal identification*.

Like privacy and confidentiality, ethical guidance on *data sharing* can be deduced from the Belmont Report, but data sharing is not explicitly addressed in either this document or in 45 CFR 46. Much of ethical guidance in human research has focused on the intervention, interaction, and information acquisition processes. There has been far less attention to dissemination of results, access to data, or subsequent data use.¹⁵ The Belmont principle of beneficence emphasizes the value of addressing benefits that can accrue to participants, similarly situated others, and the larger society as well as to the entire research enterprise. Broad in its scope, this principle is particularly applicable to weighing gains that can come from data sharing—including the verification of results, consideration of competing hypotheses, and examination of new questions.

Overall the Belmont principles and derivative applications provide desiderata to help inform the ethical conduct of social and spatial research. Since the Belmont principles were developed primarily by physicians, they do reflect a conception of harm and benefit more appropriate to biomedical research than to social and behavioral science research. This emphasis is

problematic when the primary ethical concern is a possible invasion of privacy or a confidentiality breach due to using analytically precise coordinate data rather than when the concern is direct risk of physical harm. Similarly, the notion of autonomy set forth in the Belmont principles and operationalized via informed consent is much harder to understand when the choice is whether to participate in a survey linked to a complex set of locational measurements rather than when the choice is whether to participate in a treatment program that involves specific physical risks and benefits to the individual. Nevertheless, although the Belmont principles leave room for debate and uncertainty when applied to social and behavioral phenomena, the basic concerns of the principles and their emphasis on nuanced ethical decision making commend their use.

By design, the principles offer not answers, but expectations for balancing important considerations in undertaking ethically responsible research. The Belmont principles undergird the Federal Regulations for the Protection of Human Subjects and are also pervasively used across fields of human research. Their strength, however, lies in comprehending the flexibility that they were intended to foster, not in invoking them in a formulaic fashion. No ethical principles taken off the shelf can resolve dilemmas. Thus, in using the Belmont principles, researchers, data providers, and secondary analysts need to extrapolate from them to think through how they apply to social and spatial research.

ETHICAL CONSIDERATIONS, THE RESEARCH CONTEXT, AND RESEARCH PLANNING IN SOCIAL AND SPATIAL RESEARCH

Ethical Considerations

In general, the collection, use, and analysis of linked social-spatial data raise ethical issues that parallel those involved generally in handling identifiable, large-scale data sets on individuals or groups, whether the data are acquired directly or indirectly, and specifically when research involves linkages among microlevel data. Although not as powerful an individual identifier as DNA or other genetic material used in genetic studies, precise coordinate data in the social sciences is at once an identifier and a compelling social indicator that rivals most other forms of contextual measurement because it is location-specific and can be collected repeatedly, in multiple sites, and on a very large scale. It is rare, perhaps even unique, to have a single measure or indicator essentially serve as an exact identifier, either alone or in combination with only a few other variables.

The ethical principles and applications enunciated in the Belmont Report provide a framework for unraveling some of the complexities of social-spatial research. The ethical issues are at one level familiar ones: grappling

with how best to honor confidentiality agreements with research participants, minimize risk of disclosure of private information and potential harm, and maximize the benefits that can flow from research and access to the data. The potential identifiability of individuals and groups in studies involving linked social-spatial data makes it important for researchers to consider informed consent and the situations in which it can be waived; the nature of confidentiality agreements and protections; the risk of breaches of confidentiality and steps to ameliorate that risk; the magnitude of any potential harm from disclosure; and the benefits that can accrue to participants, their communities, or the larger society.

Attending to these considerations does not per se distinguish social and spatial research from other inquiries that cannot be undertaken anonymously or that involve identifiable and potentially sensitive personal information. With precise spatial data, the threshold for identifiability may be lower than in research in which analytic measures are not also personal identifiers, but the ethical principles shaping researchers' responsibilities are the same. Technological advances that can aid research can also contribute to increasing the probability of identification. For example, research using video recordings to study behavior in public places or that have research participants use wearable computers to monitor movement and interactions in work or social groups has considerable scientific potential, but it can also increase the risk of identifiability, even if the consequent harm is quite minimal. Similarly, spatial measurements are sufficiently precise in that they are at once invaluable to research and yet could make difficult protecting the identities of individuals and information about them from inadvertent or intrusive disclosure.

The very complexity of undertaking research of this genre does not mean that the work inherently involves more than minimal risk in terms of the type of harm or the likelihood of its occurrence. Also, research procedures can be put into place to reduce or ameliorate risk to a minimal level. Responsible conduct in research commends the use of advanced measurements and technologies to maximize scientific progress and the benefits of research while ensuring that any risk of harm for participants remains low.

Contexts of Research

In research involving the linkage of social and spatial data, there are a large number of persons who collect, use, or otherwise make decisions about how to maintain, preserve, and make such information available. Depending on the context, different individuals connected with the research may take on various roles in the development of a particular human research protection plan or the articulation of a strategy that will engender confidence in data sharing and use. The basic principles underlying ethical

decision making, whether by data producers or users, are no different from those in similarly complex, large-scale studies about people and their lives in which there can be data from multiple sites, multiple sources, and multiple time points. In all such research, there is an interest in and commitment to enhancing access and use in order to maximize addressing important issues while ensuring that confidentiality agreements are honored and the risk of personal identification is minimal.

Linkages between spatial and social data are being made by researchers at every point in the research enterprise, from primary to secondary use. For example, investigators are specifying designs that incorporate precise coordinate data in the research (e.g., home, workplace, school, recreation center; more than one location) or link to extant databases that provide precise coordinates. Secondary analysts, too, are examining individual-, household-, or group-level behaviors by using data that have those links or by enhancing those data through integrating additional resources. Even in the absence of precise spatial data, the merger of two deidentified databases or one set of public records and one or two deidentified databases raises the possibility of the reidentification of research participants. Identification is even more likely when highly refined locational data are in the mix and are intended to be used as analytic variables.

The data producer and user face particularly challenging circumstances when they generate new data or pursue data integration, analysis, dissemination of results, and sharing or transferring of these data to others. The archivist and the database manager also have responsibilities for how such data are to be preserved, stored, and potentially used.¹⁶ Finally the secondary analyst has the ethical responsibility to honor agreements for access, which include those agreements made with research participants as to use.

Purposive Planning

From the vantage of human research protection and review of research by an institutional review board, there are some immediate ethical questions for primary researchers and secondary users to consider. It is optimal, for example, to determine in advance whether data collection or linked analyses will be individually identifiable only by virtue of obtaining and using locational data; whether or not the consent of research participants will be obtained and, if so, in what form and with what assurances; and whether the likely benefits and the potential harms can be specified, and, in the case of potential harms, whether steps can be taken to ensure that they are low (e.g., embarrassment versus legal liability) and the risks of their occurrence are minimal (through strong data protection or access plans). A primary data producer and user can consider most of these issues in ad-

vance of initiating research or can specify them for follow-up review, up to and including strategies for data dissemination and sharing.

The secondary analyst does not create the data or the conditions for the research; nevertheless, she or he needs to develop a research plan consonant with confidentiality protections and needs to seek IRB review to the extent that the new work contemplates the integration of heretofore unlinked datasets or spatial measures.¹⁷ In the case of secondary data, the data archivist, data collector, or initial researcher can require licensing or other contractual arrangements with the secondary user or her or his institution, or the secondary user may need to work in a data enclave or other restricted-access environment in order to use the data. Each of these steps adds a level of review as a condition of access, controls the nature of that access, and includes the force of law to enhance confidentiality protections (see National Research Council, 2000, 2005).¹⁸ The extent to which such steps are necessary or appropriate depends on whether there is more than a minimal risk of disclosure and the probability of harm that any disclosure could entail.

Ethically responsible conduct in the collection or use of social and spatial data is sufficiently complex that it requires a planned, deliberative process. One useful way to think about the preparation of a protocol for review by an IRB, as well as the review process itself, is as a structured opportunity for primary researchers or secondary analysts to present to a group of peer scientists and community members a human research protection plan and approaches for undertaking sound and ethically responsible work. Because of the challenging issues involved in human research protection with social and spatial data, there are core ethical questions that need to be addressed: Is this human subjects research? Does the use of precise coordinate data add value to the topic under study? What is the process for gaining consent or the rationale underlying a request for a waiver of consent? How are issues of confidentiality to be addressed? What are the benefits of the research, and what are the risks of harm and strategies for amelioration? Each of these issues is considered in the next section.

THE BELMONT PRINCIPLES AND QUESTIONS TO GUIDE ETHICAL DECISION MAKING

The principles and standards specified in the Belmont Report provide a useful tool for the responsible planning and implementing of social and spatial research. For example, they can guide in assessing whether exact spatial data affect determinations of what constitutes human subjects research; judging the risks and benefits of certain research topics; and sorting out issues of confidentiality, data access, and data sharing. Fundamental to weighing how research can be done, how research data can be secured, and

how access to data can be provided are considerations of respect for the autonomy of human subjects, appreciation of their voluntariness, and assessment of the benefits for subjects and the research enterprise while assessing the risk of harm, the justness of inquiries, and the equitable distribution of benefits and burdens. These ethical principles help to frame questions that inform responsible decision making.

Human Subjects Research

Social and spatial research that otherwise involves no interaction or intervention can become human subjects research as defined in the Federal Regulations for the Protection of Human Subjects because precise coordinate data allows for *personal identification*. The Belmont principles are directed to the conduct of research *with* human subjects, and these principles shape the boundaries of what constitutes human subjects research in the federal regulations. There is considerable research in the social sciences using public records or other information that is publicly available or observable that is not human subjects research, even though it meets the research standard of contributing to generalizable knowledge.¹⁹ Information gathered without intervention or interaction with individuals or without *identifiable private information*²⁰ is considered to be outside the scope of human subjects research. Also, identifiable information about individuals that is publicly available is not identifiable *private* information, and hence it is also outside the scope of human subjects research.

Highly refined coordinate data can shift otherwise public information to the category of private identifiable information and thus human subjects research. For example, anonymous data on people's personal health habits becomes identifiable when linked to spatial data describing, with considerable accuracy, where a person lives. Such precise spatial data, coupled with other demographic descriptors of individuals, may enable an intruder to deductively identify individuals. This is a changed circumstance produced by major advances in observation technology and the capacity to record and store such information. Until recently, locational mapping, aerial photography, and other mechanisms to depict spatial relations were not sophisticated enough to yield private identifiable information and thus were outside the definition of human subjects research. The same transformation has occurred in the context of individual observation in public places where note-taking has been replaced by audio or video recording, and the potential identifiability of recorded data in public places can make research previously considered outside the definition fall under the scope of human subjects research.

It might be expected that the capacity to make refined measurements would lead data providers and secondary users to seek to have access to

these data. Secondary use of datasets that are not identifiable and are available for public use have less scientific potential but do not create the same concerns about disclosure risk. Secondary analysts need to weigh what forms of data meet their needs and what benefits to research may be lost without the use of more precise locational information.²¹

Many of the sophisticated techniques that have been employed to preserve and optimize the analytical value of data to secondary users can be generalized to social and spatial data. Data releases can vary depending on the needs of the secondary users. For highly qualified secondary users, the use of enclaves, licensing, and other related mechanisms, as described by Rodgers and Nolte (2006), can enable the secondary user to enjoy the same richness and usefulness that was available to the primary data user. Alternatively, judicious decisions by a disclosure review committee may result in the use of techniques, such as data swapping and suppression of geographic detail, and render the data appropriate for broader dissemination to secondary users (see, for example, O'Rourke et al., 2006).

Ethical responsibilities follow for researchers engaged in data collection or the analysis of data in which information is identifiable. Ethical research *with* known human subjects requires that they be aware of and informed about the research, that they agree to participate in it, that their information be treated in confidence, and that there be benefits to the work that outweigh the risks of harm. With known persons, researchers have fiduciary obligations to these individuals as part of the compact of their participation. If secondary analysts are studying data that are similarly identifiable, they also have the same obligation to honor agreements that have been previously made.

Topics of Research

Topics of inquiry vary in sensitivity and the likelihood that research participants may believe that they are sharing information that is highly personal and private. There are individual differences among participants as to their boundaries of privacy and what they are willing to share with researchers. These differences are exacerbated when it is not only the primary researcher but also others, later, who may gain access to individuals who are seeking to keep private their status, condition, or personal information. Individual differences in people's desire to control who has access to them and to information about them are likely to arise in some of the kinds of research that include spatial linkages to social data. For example, research on domestic violence, crime, stigmatized diseases, and natural disasters would be enhanced by geographic display of incidence data. Many persons in these circumstances are quite willing to participate in research and view quite favorably the opportunity to speak to a researcher or be part

of a larger enterprise. Others, however, who have been traumatized, perhaps repeatedly, are likely to be highly sensitive to and fearful of invasion of their privacy and any consequent, remotely possible intrusion on their person or their social circumstances.²²

Attention to the justness principle helps to assess whether the needs of certain populations commend the use of spatial data because of increased benefits that can derive from the research (e.g., vulnerability to toxic waste), or whether certain populations may be more vulnerable to being studied and to researchers seeking access to personal information (i.e., inequitable burden). Linked social-spatial data could add to knowledge on very personal, yet highly important topics (e.g., studying the relationship between health risks and access to health resources) that research participants and the larger society would value. Alternatively, such data could increase the vulnerability of already vulnerable populations to stigma or other forms of harm (e.g., studying drug use patterns proximal to high-crime “hot spots”). The key ethical questions include: To what extent does linking social and spatial data add to the importance of the research? To what extent does it add to the risk of disclosing personal information? How will the researcher or secondary analyst explain the benefits of the study and the value of social and spatial links to research participants and to the larger society?²³

Consent and Confidentiality Agreements

Informed consent of research participants is the standard ethical requirement for human subjects research. Researchers have an ethical responsibility to show respect for persons and earn their trust based on the assumption that people have agreed to participate on a voluntary basis and with sufficient information and understanding to make a decision. As specified in the Belmont principles, the standard is one of subjects’ having “sufficient” information. The principles allow for incomplete information in certain circumstances to ensure the validity of the research as long as the risk is minimal and the information being withheld does not pertain to risk. Typically, as part of the compact between researcher and research participant, consent to participate also includes an agreement to treat information as confidential and to ensure that no personal identifiers would disclose either subjects’ participation in the research or information about them.²⁴ The addition of fine-grained spatial data makes implementing this promise an additional challenge.²⁵

In making determinations about consent, the nature of consent, and whether to seek a waiver of consent under the federal regulations,²⁶ researchers and others collecting highly identifiable spatial data need to assess how they will approach the process of obtaining consent and whether and

under what circumstances they would seek a waiver of consent. For example, could a waiver be appropriate in a study of at-risk driving patterns using court records to identify drivers' license suspensions and home coordinate data to study car use, but in which no direct interaction with subjects is anticipated? Given the importance of ensuring persons' autonomy to participate in research and also of maximizing public trust in research, circumstances justifying waivers of consent require close scrutiny.

When spatial data are collected along with social data, it is important for researchers as well as IRBs to consider how the need for and use of coordinate data should be described in obtaining informed consent, what agreements of confidentiality should be made, and how explicit researchers should be about secondary or unanticipated use.²⁷ Without some explanation, it is not reasonable to expect that research subjects would understand either the potential risks or the benefits of social and spatial data. There is good general guidance in ethics codes and in recent reports on informing research participants about future data use that is equally applicable to primary researchers and data producers engaged in social and spatial research (see, e.g., National Research Council, 2005, Recommendation 14, pp. 80-81). Nevertheless, when new media and their conceivable risks are explained, it is all too easy for the researcher to assume that the potential research participant understands the terminology used to explain the technology and its risks, and it is likewise too easy for subjects to pretend to understand rather than appear uninformed. Moreover, such problems of miscommunication are likely to vary across different sectors of the subject population. Such techniques as cognitive interviewing can be usefully employed both to develop informed consent language that is understandable to the various relevant sectors of the population and as probes to evaluate comprehension by consenting individuals (Willis, 2006).

There are instances in which the consent of research subjects may not be possible for obtaining or using linked social-spatial data. Such instances are most likely to arise in the contexts of unanticipated or secondary use. Secondary analysts may seek to use social and spatial data for which there was no previous agreement about multiple research use during the original data collection. Also, primary or secondary researchers may identify a subsequent use for linked data for which recontacting research subjects to obtain consent may not be feasible—effectively making the research impossible if consent was required. In determining whether to seek waivers of consent, researchers need to weigh obligations to research subjects and to the scientific enterprise, as the Belmont principle of beneficence specifies. Under such circumstances, salient ethical questions include: Is the research of minimal risk and sufficient potential benefit to commend being pursued without consent? Will the researcher operate consistent with any prior confidentiality agreements, extrapolating to this circumstance? Can the

identity of research participants and information about them be protected in light of the privacy needs that people typically have?

Benefits

In general, the benefits of social science research typically accrue to society or to people with similar conditions or circumstances to the research subjects, rather than to individual subjects. Consonant again with the principle of beneficence, research participants should be made aware of this, and also of the fact that benefits derive from the accumulation of scientific knowledge based on information that they provide or make accessible and from having that information enhanced by linking to other information. Researchers can also communicate to participants the benefits that can derive from making the information they provide available to other qualified researchers who can reexamine findings or ask new questions using the same information.²⁸

The benefits of precise spatial measurements can best be understood in this context. More extensive measurement of contextual variables, such as location permits identifying and explaining patterns and differences on a group, community, or societal scale. Emphasizing these benefits does not mean that individuals do not themselves reap personal benefits commensurate with their time and engagement. Typically an aspect of ethically responsible research is to provide some tangible benefit to participants. In the case of a health survey with precise coordinate data, it could, for example, be a handout of proximal health clinics and routes of public transportation. With unsavory or undesirable human subjects, benefits may not accrue, but neither should direct harm due to their willingness to participate, assuming they are aware of ethical and legal limits.²⁹

Overall, in assessing and communicating the benefits of research, the salient ethical questions for researchers include: Are the research participants or their communities likely to benefit from more geographically explicit research? Are they likely to receive far fewer benefits if the use of geospatial data is severely restricted? Can researchers provide research participants or their communities with added benefit by adding the geographically specific information? Can the potential benefits of linking social and spatial data be reaped without research participants being exposed to undue risk of harm or disclosure? Can the researcher set forth the benefits of such data and not overpromise?

Risk, Harm, Risk Reduction and Confidentiality Protection

As with assessing benefits, the assessment of risk in social and spatial research needs to identify both general risks associated with the research

and any increased harm or risk of harm due to the use of precise locational information. For all studies, there are potential risks at each stage of the research enterprise, including in the identification or specification of the sample, data collection, database construction, data analysis, dissemination of results, and data sharing. Along with the benefits that derive from using exact coordinate data is some greater risk of disclosure made possible through the use of a readily identifiable variable.

The Belmont principles appropriately emphasize the distinction between risk of harm and severity or magnitude of harm and that the “benefits [of research] are properly contrasted with harms rather than the risk of harm.” Except at the extremes, determinations of level of risk and types of harm are frequently confused.³⁰ Social and spatial research on highly sensitive topics for which physical, psychological, legal, or economic harms are conceivable (e.g., a study of mobility patterns and self-protective behavior of abused spouses) place a higher burden on ensuring that preserved data have a very low (approaching zero) risk of disclosure and are protected by a very secure data protection plan.³¹ In some instances, researchers may wish to obtain certificates of confidentiality from federal agencies to protect, to the extent possible, some forms of data from forced disclosure.³² Much research of importance in social science is not on highly sensitive topics, and the primary risk of harm may take the form of transitory embarrassment, stress, or discomfort. Even under such circumstances, however, efforts to reduce the risk of disclosure remain important because of the ethical value placed on honoring agreements with research participants and the ethical principle of making information on participants accessible only if essential to addressing research issues.

With linked social-spatial data, there is an incremental risk of breaching confidentiality and the potential for disclosure due to the value of preserving and using precise locational information. There is an ethical obligation to minimize disclosure risk generally—even when it remains minimal. Precise coordinate data may continue to have analytic meaning for many years, but risks associated with its use may reduce over time as migration and other life course changes alter the identifiability of these data. Nevertheless, in implementing ethically responsible research and planning for access, issues for consideration include: What technical approaches can be used, and to what extent should they be used to reduce the identifiability of social and spatial data while still retaining their scientific and analytic value? What do researchers and others who produce, manage, or disseminate data need to know to minimize risk of disclosure? What forms of data protection plans and models of restricted access are most promising to maximize the use of data and to minimize dual use (that is, unanticipated and adverse use by an intruder)³³ or inadvertent disclosure? To what extent are different strategies or guidelines needed at different stages of research

(e.g., strengthened certificates of confidentiality,³⁴ guidelines for research reporting) in order to minimize disclosure risk?

Planning and implementing ethically responsible research, whether as primary researchers or secondary users, involve addressing this spectrum of issues. These questions serve to illustrate what needs to be asked by researchers, by data providers, by funders, and ultimately by review groups like IRBs in undertaking social and spatial research. The process is much more nuanced than a simple determination of how and at what level precise coordinate data must be masked to maximally reduce identifiability and potential breaches of confidentiality. If the data can be adequately protected from intruders, if inadvertent disclosure can be sufficiently reduced, if the risk of exposure is low, and if the harms from any exposure are only of minimal or transitory impact, then the core considerations to allow for ethical use have been met. Thus, the emphasis on strong data protection plans and conditions of responsible use is as important as masking data beyond a point at which its value would be substantially compromised.

ETHICS OF DISSEMINATION, SHARING, ACCESS, AND THE CONFIDENTIALITY NEXUS

As the foregoing discussion has emphasized, ethical decision making prominently includes attention to issues of confidentiality, but ethical considerations are larger and more comprehensive than confidentiality alone.³⁵ Because of the considerable scientific value of using precise coordinate data by primary researchers and secondary analysts, there is an inevitable tension between data dissemination or sharing and doing so consonant with the promises made to research participants not to disclose their identities or identifiable personal information about them. What are at once sound ethical standards—maximizing scientific gains from available information and ensuring that promises of confidentiality are kept to research participants—can conflict if the advancement of one compromises the other.

Dissemination

Ethical decision making in human subjects research typically focuses on issues that relate to identifying research populations and informing them about the study, gaining their agreement to participate, and minimizing the probability of any harm or risk of harm that might occur during the conduct of the research or with information gathered through it. Ethical responsibility as it relates to other steps in the research process, in particular research reporting and dissemination, is far rarer in discussion and decision making related to human research protection. Beyond confidentiality guarantees and cautions with respect to personal identification or the identifi-

ability of personal information, much is implied but little is elaborated on in the federal regulations. The Belmont principles and related standards provide useful guidance for greater attention to the dissemination phase.

First, the Belmont principles are explicit about the need to contribute to the larger body of knowledge. Also, they imply that fair treatment of those who participate in research includes the dissemination of results; beneficence depends in part on the dissemination of valid new knowledge. In addition, ethical standards related to reporting on research require that the data underlying results need to be presented in sufficient detail to permit readers to follow the logic of inquiry and assess the warrants underlying inferences.³⁶ These objectives need to be considered in the context of how information is publicly presented with linked social-spatial data.

The presentation of precise locational information can enhance contributions to knowledge, but, with locational data, the form of presentation of research results may require special measures or procedures to be as transparent as possible without risking disclosure of the identity of research participants. How will the data be presented or displayed to avert the likelihood of identifying research participants or the potential misuse of findings? How will the research methodology and design be described to allow for maximum transparency and the accumulation of knowledge but without risking inadvertent disclosure?

Depending on the precision of the locational data and the rarity of the social data that are linked, even a map display could reveal the identity of specific individuals without mention of any specific names. A study of drug users and their dispersion and density in a community may add immeasurably to knowledge of how social networks contribute to at-risk behaviors, but also published maps by household could be tantamount to published address books in certain neighborhoods. Thus, it may be necessary in presentations or published work to coarsen the displays, swap data, or extrapolate to similarly situated geographic spaces in the same or an equivalent neighborhood, or take other steps that allow for the reporting of results while preserving the confidentiality of linked social-spatial data.³⁷

Sharing

Given work of the scope, size, and significance of social and spatial research, the ethics of inquiry commends data sharing on the part of primary researchers and data collectors. Like the dissemination of results, data sharing also contributes to the important Belmont principle of contributing to the accumulation of knowledge. The Belmont Report emphasizes, as an element of beneficence, the improvement of knowledge and the benefits that can accrue to society in the form of knowledge to be gained from research. Data sharing in science can be seen as a means to that end:

fundamental to science is a commitment to openness of inquiry that enables the self-correcting and cumulative advancement of knowledge.³⁸

Prior to about 1975, openness regarding human subjects data was manifested by the way research methods were described and data were presented in publications; replication of results established the validity and generalizability of results.³⁹ In 1985, the National Research Council, under the leadership of its Committee on National Statistics, published the report *Sharing Research Data*, which was influential in its reach (National Research Council, 1985).⁴⁰ Almost immediately, the Division of Social and Economic Science at the National Science Foundation (NSF) took up the recommendations in that report and established a data archiving policy (National Science Foundation, Division of Social and Economic Science, 1986). By the late-1980s, some federal funding agencies, most notably NSF, began to encourage more formal sharing of documented data and materials in *all* areas of science.⁴¹ By the time the National Institutes of Health (NIH) fully elaborated its policy in 2003,⁴² the ethical underpinning and normative value of data sharing were quite evident both in official policy and related educative materials.⁴³

Ethical conduct in research involves not only attention to the value of data sharing but also doing so consonant with confidentiality agreements. Researchers and data producers need to plan for data sharing and the forms that data sharing can take. Especially with social and spatial data and other forms of information that may be readily identifiable, primary researchers and data producers need to ensure that research subjects are *sufficiently* informed about potential use of the data and to develop data sharing plans that can reasonably be expected to protect the identities of human subjects and personal information about them. As noted earlier, even with research in which potential harms are minimal, the broader commitment of honoring confidentiality agreements with research participants looms large even if the consequences of disclosure of personal identifiable information are small. Also, the reputation of human research with the general public will greatly influence the willingness of individuals to participate in research in the future.

From the vantage of ethically responsible research, the articulation of data sharing and data protection plans appropriate to the research go hand-in-hand. With large-scale social and spatial data (including that collected at many sites or over long-periods of time), there is the potential for considerable future use. Thus, gaining the consent of research participants could readily include noting that other researchers will have an opportunity to analyze the information. If a study is on quite personal or sensitive topics, primary researchers either could explain that information provided to others would be altered in such a way that identification would be virtually impossible, or they could indicate that other researchers can have access to

identifiable information only under restricted conditions in which they commit to honoring confidentiality agreements. If the researcher plans to share data through a third-party provider (e.g., a public archive), both the researcher and the provider need to anticipate whether the level and richness of the linked coordinate and social data allow for public use data files (typically limiting the detail that can be provided) or alternatively whether restricted access arrangements need to be made (e.g., licensing agreements or access at controlled sites).

There is good general guidance for investigators and for institutional review boards on specific ways to protect the privacy of human subjects and the confidentiality of the data.⁴⁴ There is need to develop and test approaches for providing access to precise coordinate data that can maximize the analytic potential of these measurements without risking the disclosure of identifiable information in primary or secondary use. Our purpose here is not to elaborate on the methodologies, the processes for sharing data (e.g., under the direct auspices of researchers, through a data archive or enclave), or even the timing of data sharing (e.g., released in waves for longitudinal study). Our aim is to underscore the ethical basis for data sharing and that data sharing and data protection can best be addressed together by researchers and by IRBs.

Access

Secondary users of public-use data or restricted data files have an ethical obligation to contribute to the advancement of knowledge in accordance with the agreements made to produce these resources. The ethical obligations of primary researchers extend to secondary analysts. Secondary analysts are reliant on the trust provided by research participants in the research enterprise, and thus the obligation of secondary analysts is not altered by the fact that they were not themselves party to any promise with the human subjects of research.

Public archives like the Inter-university Consortium for Political and Social Research (ICPSR) explicitly set forth the obligations of secondary analysts for responsible use.⁴⁵ With data that are either publicly available or available through limited or restricted forms of access, typically researchers have an ethical requirement to use data in their current form, without the integration of additional data or enhancements of other information, unless they take additional steps to assess the ethical issues related to an expansion or change. Except for data that are publicly available, this obligation also includes not otherwise sharing data with tertiary users. In social and spatial research, this guidance on secondary use is particularly important. Secondary analysts who seek to add precise coordinate information need to examine the ethical aspects as well as the feasibility of doing so

responsibly, and they need to consider issues of consent. In such instances, researchers may request waivers of consent, as noted previously, in seeking IRB approval. An IRB's ethical responsibility is to assess the reasonableness of waiver requests in light of the potential risk of harm and steps that will be taken to ameliorate that risk.

Contributing to the advancement of knowledge also obligates secondary analysts to acknowledge the sources of data as part of disseminating their results. To the extent that beneficence includes both an obligation to contribute to the well-being of research participants and the larger public good, acknowledgment of the connections between the new research and the initial research helps to ensure cumulative benefits. Most ethical guidance includes secondary analysts also acknowledging any assistance that they have received from primary researchers in gaining an understanding of or access to such data.

ETHICS EDUCATION AND TRAINING FOR SOCIAL AND SPATIAL RESEARCH

The scientific potential of linked social-spatial data and the complex issues involved in responsible social and spatial research raise questions about how best to prepare researchers, data managers, data stewards, and secondary analysts, among others, to engage in such work. Typically preparation for research of such complexity and sophistication focuses on issues of methods and measurement: At the data collection stage, research preparation tends to emphasize what information to collect and preserve and how best to ensure that different forms of data at different units of analysis can be meaningfully gathered and linked. Primary data collection includes obtaining the consent of research participants, but practices may vary widely as to whether data sharing or future use is noted as part of that process.⁴⁶ At the data management, analysis, and dissemination stages, research preparation focuses on how to store or provide access to data at varying levels of disclosure risk or turns to technical and statistical questions about how to retain scientific value without jeopardizing confidentiality agreements. These are all important issues for those engaged in producing or using linked social-spatial data, but, in these contexts, guidance is aimed at being more instructive about the requirements for use than educative about them. Attention to confidentiality, inadvertent disclosure, requirements with respect to any data enhancements or linkage are considered part of the process of providing access to data—with any heightened sensitivity to ethical issues at this stage being a secondary benefit.⁴⁷

Ethics education and training are not an explicit component of most graduate education programs. Ethical considerations across social and behavioral science fields and specialties are generally addressed sporadically

in courses or in the supervised training and mentoring that more experienced researchers provide. Scant materials are available for course or class unless one enters the specialized literature on human research ethics,⁴⁸ contributing to faculty's giving limited attention to ethical considerations as an aspect of methodology except among those with specialized interests or expertise in this subject. Added to this, the rapidly changing circumstances related to confidentiality issues due to advances in computer technology and fine-grained identifiable measurements (in areas like coordinate data or video recording) make for more questions than answers—a factor likely to further discourage curriculum development by the nonexpert. In this context, social and spatial research is no exception.⁴⁹

In recent years, IRBs are explicitly encouraging researchers to take courses⁵⁰ (typically available on the Internet requiring approximately one hour). In the case of research to be funded by NIH, since 2000, education is a requirement of receiving NIH support.⁵¹ IRB members are also required to undertake training to serve in this role. Further underscoring the importance of education and training, since 2005,⁵² as part of their assurance of compliance with the Public Health Service's Policies on Research Misconduct, institutions have a general responsibility to foster a research environment that promotes the responsible conduct of research and research training (with training responsibilities covering human subjects and data acquisition, management, sharing, and ownership among other issues).⁵³ The Office of Research Integrity⁵⁴ promotes educational activities and has oversight of institutional assurances. The current emphasis on responsible research conduct as part of the regulatory climate could support a shift in attention to ethical issues if it could be meaningfully encouraged by federal agencies and meaningfully implemented by researchers and their institutions.

Research societies in the social and behavioral sciences have sought to focus greater attention on human research ethics among their members and in departments that train in their fields (see, e.g., Iutovich, Kennedy, and Levine, 2003; Levine and Iutovich, 2003). Sessions at annual meetings, courses, and workshops are not uncommon—although attendance is variable. Over recent years, the American Sociological Association, the American Anthropological Association, the American Educational Research Association, and the American Historical Association, among others, have included human research protection issues on the agenda of the meetings of department chairs, directors of graduate programs, or, in the case of education research, graduate school deans. The American Statistical Association has a portion of its website dedicated to information and resources on confidentiality and privacy.⁵⁵ The Social and Behavioral Science Working Group on Human Research Protection, supported under a contractual agreement with the NIH Office of Social and Behavioral Science Research, has issued educational documents, prepared course material, and convened

courses at research society meetings since 2002.⁵⁶ Directed to issues of central concern to the social and behavioral sciences, these offerings focus on such core topics as consent, confidentiality, privacy, benefits and risks, and the use of public and restricted data files. Although these efforts are not specifically focused on social and spatial research or issues involved in the collection, management, and analysis of linked data by primary researchers and secondary analysts, they provide some basis for further targeted work.

As noted earlier, the goal of balancing access with the protection of confidentiality is set forth in written materials by data stewards to inform potential users.⁵⁷ Data providers (e.g., public archives or research teams providing their own direct access) also note their responsibility to train those engaged in data preparation, database management, and the review of requests for data access to avert inadvertent disclosure. Collaborative efforts across research societies and stewardship organizations could very well provide a framework for both offering high-quality education and further encouraging graduate departments to do so as an integral part of their training programs. Outreach should also include efforts directed to fields of science engaged in social and spatial research but with less experience in human research and related ethical issues.

That ethical requirements can be seen as hurdles by researchers and users is understandable given that what is required can be mechanistic in many instances or oblique as to its goals and intent. Attention is rarely focused on sensitizing researchers, database managers, or users to ethical considerations or how to weigh them in undertaking or being a part of social and spatial research. This situation is by no means unique to social and spatial research, however. Despite the expanded requirement that researchers take various online courses in human research protection to certify to IRBs that they are prepared to undertake research, there is little formal preparation in the undergraduate or graduate curriculum directed to the ethics of research and responsible research conduct.

Ethics education is often conceived as a top-down activity in which IRBs and IRB specialists educate IRB members, researchers, and students. A major deficiency of this approach is that it tends to present generalities and to overlook the commonly observed fact that the devil is in the details.⁵⁸ In the case of social and spatial research, this problem is accentuated by the fact that certain issues, such as the fineness/coarseness of the data, are a technical matter, as are various ways of intruding on the data set or protecting the data set from intrusion. Hence, it is particularly important that social-spatial data specialists are prominent in the development of ethics education in this realm, via textbook chapters, national and regional workshops, and journal articles.

FINAL THOUGHTS

This paper has sought to provide ethical guidance to those involved in producing, using, preserving, managing, and analyzing spatially linked data. Our aim was to present an overview of the ethical issues that come into play in the research process, from design through data sharing and dissemination. It is only recently that the capacity to collect precise coordinate data over many locations and points in time and to link them to social data has developed to a point that raises human subjects issues. Using the Belmont principles as a base and extrapolating from them, we have sought to examine ethical considerations and how they might be weighed here.

This paper seeks to raise issues, not only for those involved in social and spatial research, but also for those engaged in the review of it. In relation to IRBs, we recommend a highly proactive approach, since IRBs will be largely unaware of this complex new situation, and either naively overlook serious risks or, in the absence of good communication and a one-step-at-a-time approach, could introduce barriers that could unnecessarily limit such research. The opportunities for linking to important forms of data should not be avoided, for example, because they were not anticipated in advance when approval was initially sought or because the risks of harm could not be sufficiently assessed at a prior point. The IRB process allows for continuing review with provisional approval. Thus, under certain circumstances, researchers may want to provide a broad map of their work and follow up with subsequent review when it becomes germane. Documentation would grow with each new addition or use of the data, but the researcher would not need to anticipate all uses too far in advance.⁵⁹

Data producers and users who intend to undertake research involving linked social-spatial data will need to take time early in the planning stages to begin conversations with appropriate members of their IRBs. Of course, risks will emerge that are unforeseen; hence, the conversation must include some discussion of this possibility. There must be an understanding that these risks will be discussed openly and immediately with the IRB and incorporated into the data documentation. This rapport could deter naive risk taking by researchers or risk-averse actions by the IRB.

As implied by the observations above, research protection programs at academic or research institutions need to support and encourage IRBs to function as ethically effective decision makers. Institutional programs can be established and approved that allow IRBs to avoid mechanistic application of rules and to use the flexibility accorded to them. Openness to the coordination of multisite review or to preapproval for certain types of time-sensitive data collection are just two procedures that IRBs could introduce to facilitate review of social and spatial research consonant with human research protection. An emphasis on confidentiality and data protection

plans and a willingness to use waivers of consent could similarly facilitate discoveries related to geographic location while strictly adhering to minimal risk standards.

In social and spatial research and in human research more generally, IRBs and researchers, data producers, and secondary analysts would serve the advancement of knowledge and the ethical conduct of science through taking an unusually collaborative and collegial approach. Part of doing so requires essential expertise on IRBs or IRBs involving expert consultants related to the technical and ethical issues involved in social and spatial research. Representation on the IRB of scientists knowledgeable in spatial measurement, in data disclosure methods, and in approaches that can ameliorate risk would be optimal when there are sufficient numbers of relevant protocols. The promise of social and spatial research is so significant that it is incumbent on those who propose research and those reviewing it to proceed cognizant of the contribution of research participants and committed to *benefits for all*.

NOTES

1. As used here the term embraces all of the mechanisms that permit the identification of a location through latitude and longitude coordinates. The magnitude and speed of obtaining such information due to advances in remote sensing (from satellite images to high-resolution aerial photography) and global positioning systems (GPS), coupled with the growing sophistication of geographic information systems (GIS) to store and manipulate such data, have accelerated interest in research use and applications.
2. For an excellent overview of this rapidly emerging field, see National Research Council (1998).
3. To date, much of the attention on balancing data access and considerations of confidentiality has focused on federal statistical data collections, administrative records, and other public resources (see, for example, National Research Council, 2005; Lane, 2003; de Wolf, 2003). For earlier consideration of these issues, see Duncan (1993), National Research Council (1993), and U.S. General Accounting Office (2001).
4. Two important examples relate to research ethics in complex humanitarian emergencies (see National Research Council, 2002) and with victims of disasters (see Collogan, Tuma, Dolan-Sewell, Borja, and Fleischman, 2004; Collogan, Tuma, and Fleischman, 2004). For a general consideration of challenging research circumstances, see National Bioethics Advisory Commission (2001).
5. Public Law No. 104-191, 110 Stat. 1936 (1996).
6. For useful guidance, see de Wolf, Sieber, Steel, and Zarate (2006).
7. The Federal Regulations for the Protection of Human Subjects were adopted in 1991 and Subpart A (known as the Common Rule) was accepted by 17 federal agencies as policy. Only research funded by these agencies needs to be considered by an IRB at the relevant institution, but institutions under their assurance of compliance with the federal regulations (filed with the Office of Human Research Protections; available: http://www.hhs.gov/ohrp/assurances/assurances_index.html) gener-

- ally require that all human research receive IRB consideration, whether or not the work is extramurally funded, or whether it is funded by federal agencies (beyond the 17 signatories) or private foundations.
8. Calls for empirical research on human research ethics have increased in recent years. There is general awareness that human research considerations are shaped by too many assumptions about research participants (see, e.g., Levine and Skedsvold, 2007). Recent reports from the National Research Council addressed to issues of data access are strong in their calls for research (see, e.g., National Research Council, 2005, 2003a). In 2006, the *Journal of Empirical Research on Human Research Ethics*, published by the University of California Press, was established to serve as a forum for empirical research on such issues.
 9. Research, for example, that graphically displayed individual-level activity patterns—leaving from home to work but stopping to have coffee with friends rather than to arrive promptly for business meetings—could encroach on personal privacy and run employment risks if confidentiality were breached.
 10. For brief recent histories relating to the social and behavioral sciences generally, see National Research Council (2003a); also see the section on emergence of ethical considerations and related cites in Levine and Skedsvold (2007).
 11. *Respect for persons*, *risk-benefit*, and *justice* are key considerations as they relate to the autonomy of subject populations to participate in research and to ensure that their doing so is equitable in terms of inclusion as well as exclusion. For an important example of attention to ethical considerations in the conduct of research involving prisoners, see Institute of Medicine (2007). The committee undertaking this report sought to reexamine and address such important issues as what constitutes prisoner populations, whether review of research should shift from categories of research to a risk-benefit approach, and how justice might best be understood in the context of an ethical framework.
 12. Private information is one of the defining characteristics of research involving human subjects at 45 CFR 46.102(f); that is, information obtained in a context in which an individual might reasonably expect that no observation or recording is taking place or information that a person would reasonably expect will not be made public *and* is individually identifiable by the researcher. Subsequently, in setting forth the criteria for IRB approval of research at 45 CFR 46.111(a)(7), the need for provisions to protect the privacy of subjects and the confidentiality of data is emphasized. Confidentiality is also explicitly mentioned in the federal regulations at 46.116(a)(7) as an element of informed consent—that is, the need for informed consent to address the extent to which the confidentiality of records identifying research participants will be maintained.
 13. Privacy and confidentiality are distinct from anonymity, which generally refers to researchers retaining no record of the identity of research participants, either because unique identifiers are unknown to the researcher or they are not included as part of the data. For an accessible discussion of the distinction between privacy, confidentiality, and anonymity, see Sieber (1992:44-45). Some researchers and secondary analysts use the term “anonymization” to refer to the removal or alteration of identifiable information—although deidentification tends to be the preferred term to refer to eliminating or masking data to reduce the likelihood of potential disclosure (see National Research Council, 1993).
 14. Gutmann et al. (2005:2) made this useful distinction between the identity of subjects and information about them in the context of providing spatial data for secondary analysis. For a general discussion of identity disclosure and attribute disclosure, see National Research Council (2003:23-24, 143-144).

15. Ethical considerations in biomedical and behavioral research evolved first in the context of experimental research, including in clinical medicine, which put greater emphasis on subject recruitment, consent to participate, and benefits or risks of harm due to participation than on other phases of research—including data preservation, dissemination, access, or subsequent use. The National Research Council reports (1985, 1993, 2000, 2005) on data sharing and on access to research data—in particular public data and administrative files—are an exception to the dominant attention to the data collection stage.
16. Excellent suggestions are outlined in Gutmann et al. (2005).
17. IRBs at some institutions want to review research on extant data resources that include identifiable information even if the data are made available by third-party providers who have protocols and procedures in place for approving use. If additional data are to be linked by the secondary analyst, then IRB review is required because the additional data integration (whether or not there is new primary data collection) changes the conditions of research and potentially raises new ethical considerations in relation to research participants that need to be addressed.
18. For a recent description of ways in which data enclaves and other forms of limited access data sharing can be employed to permit qualified secondary users to analyze data with strict safeguards against disclosure of confidential information, see Rodgers and Nolte (2006).
19. The scope of this paper is directed to social and spatial research directed to producing and adding to generalizable knowledge. The definition of what constitutes research covered by the Code of Federal Regulations for the Protection of Human Subjects is set forth in 45 CFR 46.102(d), “*Research* means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. . . .”
20. According to 45 CFR 46.102(f), “Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subject.”
21. Increasingly IRBs at institutions are not doing additional review of protocols for research on public use files. For an excellent example, see the website of the University of Wisconsin, Madison, IRB at <http://www.grad.wisc.edu/research/compliance/humansubjects/7.existingdata.htm>. More generally, see the recommendation of the National Human Research Protections Advisory Committee on public use data files at <http://www.hhs.gov/ohrp/nhrpac/documents/dataltr.pdf>. Two NRC reports (National Research Council, 2003, Recommendations 5.2 and 5.3; 2005, Recommendation 6) urge the exemption of secondary analysis of public use files from additional IRB review based on certification of confidentiality protection from a data provider, including federal statistical agencies. The federal regulations at 45 CFR 46.101(b)(4) define as exempt “research involving the collection or study of existing data, documents, records . . . , if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.”
22. Empirical research on the complexity of undertaking research in traumatic circumstances or on traumatized populations is reviewed in Newman and Kaloupek (2004) and Newman, Risch, and Kassam-Adams (2006); see also, Griffin, Resick, Waldrop, and Mechanic (2003).
23. With certain topics of research or subject populations, researchers need to take special care to conceive of the research cognizant of the perceptions of human subjects about the study and the research procedures being used. There are many

- good examples related to ethnic and cultural populations and communities, including immigrant and refugee groups, in Trimble and Fisher (2006).
24. Since spatial data with precise coordinates by definition locate persons and their communities, community consultation about the consent process and informing communities about the research and its purposes may help to work out agreements (see, e.g., Melton et al., 1988; Marshall and Rotimi, 2001).
 25. Because precise spatial data are the equivalent of personal identifiers or close proxies for them, social and spatial research that includes such measures would typically require research participant consent. Unlike the completion and return of a survey, for example, that is completed online or received in the mail, for which executing the task can be presumed to be consent, collecting coordinate data at a person's home, workplace, or health clinic and recording or linking it to survey or social data would ethically require the knowledge and agreement of the persons potentially under study.
 26. The criteria for waivers of informed consent are set forth in 45 CFR 46.116(d).
 27. An IRB is likely to expect researchers to address what information will be conveyed to research participants about spatially explicit data and how they would be combined with other information collected in the study. An IRB is most likely to expect discussion of this linkage and any risk of disclosure when locational data are being obtained as part of a primary data collection, along with survey or other social data. The actual wording of such an informed consent process and how it is understood by potential subjects would, in accordance with ethical principles, be specified by the researcher, with explanation to the IRB as to why the information and the assurances are being presented in that format, the data protection plan to be put in place, and the level of risk of harm. Survey researchers know that some wordings of warnings raise undue alarm, erode willingness to participate in research, can skew the research sample, or may be misunderstood or not even be recognized, as when research participants sign a consent form without reading it.
 28. There is some evidence that people want their data shared if it is likely to benefit society and if risk to the research participant is minimal (see, e.g., Willison, 2003).
 29. The "idealized type" of human subject is a person of value in terms of community norms of decency and trustworthiness. Like other areas of inquiry, social and spatial research may focus on undesirable or unsavory persons (for example, a study of diffusion of fraudulent medical practices among physicians). The ethical obligation to be respectful of research participants and not to increase their vulnerability is part of the consent agreement. There are limitations to agreements relating in some instances to a duty to report (e.g., learning about identifiable child abuse) that need to be made clear to human subjects as part of gaining their informed consent (see the discussion of research populations in Levine and Skedsvold, 2007).
 30. "Risk" and "harm" are terms that are often conflated (see the Risk and Harm Report of the Social and Behavioral Sciences Working Group on Human Research Protections at <http://www.aera.net/aera.old/humansubjects/risk-harm.pdf>). "Harm" refers to potential adverse consequences and "risk" refers to the likelihood of their occurrences. There are standards for minimal risk implied in codes of ethics and enunciated explicated in 45 CFR 46.102(i) that set forth that the "probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests." While this definition offers rules of thumb, in no area does it provide the empirical clarity that would be useful (see also Wendler et al., 2005).

31. For useful general recommendations on confidentiality and research data protections, see the National Human Research Protections Advisory Committee (2002).
32. Certificates of confidentiality are issued by designated federal agencies to protect the privacy of research subjects by protecting investigators and institutions from being compelled to release information that could be used to identify subjects with a research project. They allow the investigator and others who have access to research records to refuse to disclose identifying information in any civil, criminal, administrative, legislative, or other proceeding, whether at the federal, state, or local level (see, e.g., the National Institutes of Health web site at <http://grants1.nih.gov/grants/policy/coc/background.htm>). For a compilation of federal research confidentiality statutes and codes prepared by the Social and Behavioral Sciences Working Group for the National Human Research Protections Advisory Committee, see http://www.aera.net/aera.old/humansubjects/NHRPAC_Final_Conf_Table.pdf.
33. Dual-use research is of major concern in the biological sciences. As defined in the National Security Advisory Board for Biosecurity Charter, dual use refers to “biological research with legitimate scientific purpose that may be misused to pose a biologic threat to public health and/or national security” (Shea, 2006: CRS-2).
34. Certificates of confidentiality vary in their reach and protection, and the need to strengthen or align them across federal agencies is generally recognized (see National Human Research Protections Advisory Committee, 2002).
35. Fienberg (2004) makes the point that protecting confidentiality is not synonymous with ethical behavior.
36. See section 7 on ethics in reporting in American Educational Research Association, (2006).
37. Ethical decision making can require consulting with expert peers to ensure that steps are taken in publications or presentation that do not compromise research participants but do so with a presumption that openness in research dissemination is optimal for transparent and well-warranted reporting. Other areas of science also face the challenge of how to maximize openness in research reporting while remaining sensitive to potential risks of harm. Some of the current discussion in the life sciences about the reporting of results consonant with concerns about security issues is a new domain deeply engaged in trying to understand how best to balance both ethical considerations (see, e.g., Vest, 2003; Somerville and Atlas, 2005).
38. For one of the earliest and most profound statements of the norms guiding science (originally published in 1942), see Merton (1973).
39. There was some early attention in the 1970s to issues of access to government data and the conditions for dissemination of microdata sets (including attention to linkages to survey data) in a report of the American Statistical Association (1977). See also the Bellagio principles, which were developed in 1977 at a conference of academic and government representatives from five countries (Canada, the United States, the Federal Republic of Germany, Sweden, and the United Kingdom) convened to consider privacy, confidentiality, and the use of government microdata for research and statistical purposes. The principles call for expanded access to the research and statistical community and also addressed issues of data linkage consonant with confidentiality protections (see Flaherty, 1978).
40. Also for an overview of the emergence of data sharing as a practice integral to the openness of science, see Sieber (1991). In recent years, the biological sciences have also been grappling with the principles underlying the sharing of data and software as well as materials related to publication. Based on discussion at a workshop, the National Research Council Committee on Responsibilities of Authorship in the Biological Sciences articulated recommendations for sharing publication-related prod-

- ucts that are familiar in tone and substance to those specified in the social and behavioral sciences (see National Research Council, 2003b).
41. The National Science Foundation first specified a data sharing requirement agency-wide in April 1989. The current statement of NSF policy on Dissemination and Sharing of Research Results (section 734) is in the Grant Policy Manual at http://www.nsf.gov/pubs/manuals/gpm05_131/gpm05_131.pdf.
 42. NIH issued Data Sharing Policy and Implementation Guidelines for grants of \$500,000 or more annually in direct costs, which is available: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
 43. See, e.g., Frequently Asked Questions at http://grants1.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm; Data Sharing Workbook at http://grants1.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf; Data Sharing Regulations/Policy/Guidance Chart for NIH Awards at http://grants1.nih.gov/grants/policy/data_sharing/data_sharing_chart%20.doc; Data Sharing Brochure at http://grants1.nih.gov/grants/policy/data_sharing/data_sharing_brochure.pdf.
 44. Helpful guidance is provided in Duncan (2003); see also O'Rourke et al. (2006). In addition, *Expanding Access to Research Data: Reconciling Risks and Opportunities* (National Research Council, 2005) specifically addresses a range of approaches to allowing greater access to federally collected data while strengthening confidentiality protections. The NIH documents also provide useful elaboration on considerations that can guide the development of data access and data sharing plans.
 45. See, e.g., the ICPSR Responsible Use Statement at <http://www.icpsr.umich.edu/org/policies/respuse.html>.
 46. Practices are changing as federal funding agencies like NIH are more explicit about data sharing and the need to address data sharing or future use as part of the process of obtaining informed consent. See the National Institutes of Health Data Sharing Policy and Implementation Guidelines at http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
 47. The National Longitudinal Study of Adolescent Health (Add Health) is a good example of a major nationally representative longitudinal study that provides potential users with straightforward information on available public-use data sets and restricted-use data sets, with spatial analysis data being available through restricted use. Access to restricted use data requires an IRB-approved security plan and agreement to a data-use contract (Requirements for access to Restricted-Use Contractual Data are described at <http://www.cpc.unc.edu/projects/addhealth/data/restricteduse>.) Educative guidance of steps to avert deductive disclosure is provided on the Add Health website at <http://www.cpc.unc.edu/projects/addhealth/data/dedisclosure>. The Project on Human Development in Chicago Neighborhoods, also a major longitudinal, multimethod study, has public-use files and restricted data available through the Inter-university Consortium for Political and Social Research. Precise locational data are considered sensitive information and obtainable through ICPSR's restricted-use agreement or secure data enclave (see <http://www.icpsr.umich.edu/PHDCN/about.html>).
 48. Exceptions include National Research Council (2003, 2005), which could be adopted in course and class. Also, for useful background texts, see Sieber (1992) and Fisher (2003).
 49. The Center for Spatially Integrated Social Science (CSISS) at the University of California, Santa Barbara undertakes a valuable range of activities to foster capacity building in researchers, including workshops, extensive bibliographic references, course syllabi, information on best practices, and so forth (see <http://www.csiss.org/>). The syllabi included on the website for courses taught on spatial analysis at different

institutions make no mention of ethical considerations. The CSISS also produced a very informative best practices volume in 2004 (Goodchild and Janelle, 2004). This book is directed to the potential value of thinking spatially and sets forth examples of spatial analysis, but there was no attention to ethical considerations for potential researchers or data analysts.

50. The University of Chicago Social and Behavioral Sciences IRB emphasizes education and provides useful educational resources (see <http://humansubjects.uchicago.edu/sbsirb/education.html>). Also, its *IRB & Investigator Manual* (see http://humansubjects.uchicago.edu/sbsirb/manual/sbsirb_manual.pdf) is a very helpful document for both those preparing research and reviewing protocols.
51. Effective October 2000, NIH requires education on the protection of human research participants for all investigators submitting applications for research involving human subjects under contracts or awards. See Required Education in the Protection of Human Research Participants at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-00-039.html>; also see Frequently Asked Questions for the Requirement for Education on the Protection of Human Subjects at http://grants.nih.gov/grants/policy/hs_educ_faq.htm. Although a good deal of information is offered on the website, the number and range of opportunities for training are quite limited, in particular for research grounded in the social and behavioral sciences.
52. See Public Health Service Policies on Research Misconduct, 42CFR Parts 50 and 93, at http://ori.dhhs.gov/documents/42_cfr_parts_50_and_93_2005.pdf.
53. Training in the responsible conduct of research was an element of National Research Service Award (NRSA) institutional research training grants (T32) prior to 2005, but attention to research conduct as part of institutional assurances heightened attention to this component: "Every predoctoral and postdoctoral NRSA trainee supported by an institutional research training grant must receive instruction in the responsible conduct of research. (For more information on this provision, see the NIH Guide for Grants and Contracts, Volume 21, Number 43, November 27, 1992, available: <http://grants.nih.gov/grants/guide/notice-files/not92-236.html>.) Applications must include a description of a program to provide formal or informal instruction in scientific integrity or the responsible conduct of research. . . ."
54. The mission of the Office of Research Integrity is to monitor institutions' investigations of research misconduct and promote the responsible conduct of research through education, prevention, and regulatory activities (see <http://ori.dhhs.gov/>).
55. This portion of the website is operated by the Committee on Privacy and Confidentiality of the American Statistical Association; see <http://www.amstat.org/comm/cmtepc/index.cfm?fuseaction=main>.
56. For further information on the working group and its educational activities, see <http://www.aera.net/Default.aspx?id=669>.
57. Organizations that serve as archives for data resources and stewards providing access for their use offer materials that serve to educate and inform researchers and secondary analysts about the ethical as well as technical issues involved in sharing and gaining access to data (see, e.g., ICPSR Responsible Use Statement at <http://www.icpsr.umich.edu/org/policies/respuse.html>). Also, the Henry A. Murray Research Archive of the Harvard-MIT Data Center is the repository for qualitative and quantitative research data at the Institute for Quantitative Social Science. It has materials on data archiving that offer brief guidance, from data collection through transfer to an archive, and on steps to facilitate data sharing (see <http://murray.harvard.edu/mra/service.jsp?id=55&bct=dData%252BPreservation.p5.s55>) or application for data use (see http://www.murray.harvard.edu/mra/showcontent.jsp?key=DATA_APPLICATION_FORM). The guidance sets forth conditions for use of various forms of data, including video and audio recordings.

58. One useful example of an accessible educative document is U.S. General Accounting Office (2001).
59. A similar approach was discussed by F.J. Levine regarding natural and humanitarian disasters and strategies for ongoing flexible review processes (see National Research Council, 2002).

REFERENCES

- American Anthropological Association
 1998 *Code of Ethics*. Arlington, VA: American Anthropological Association.
- American Association for Public Opinion Research
 2003 Protection of Human Participants in Survey Research: A Source Document for Institutional Review Boards. Available: http://www.aapor.org/default.asp?page=news_and_issues/aapor_statement_for_irb. [accessed November 25, 2005].
 2005 Code of Professional Ethics and Practices. Available: http://www.aapor.org/pdfs/AAPOR_Code_2005.pdf [accessed November 25, 2005].
- American Educational Research Association
 2006 Standards for reporting on empirical social science research in AERA publications. *Educational Researcher* 35(6):33-40.
- American Political Science Association
 1998 *A Guide to Professional Ethics in Political Science*. Washington, DC: American Political Science Association.
- American Psychological Association
 2003 Ethical principles of psychologists and code of conduct. *American Psychologist* 57:1060-1073.
- American Sociological Association
 1997 *Code of Ethics*. Washington, DC: American Sociological Association.
- American Statistical Association
 1977 Report of ad hoc committee on privacy and confidentiality. *The American Statistician* 31(2):59-78.
 1999 *Ethical Guidelines for Statistical Practice*. Alexandria, VA: Author.
- Armstrong, M.P., G. Rushton, and D.L. Zimmerman
 1999 Geographic masking health data to preserve confidentiality. *Statistics in Medicine* 18:497-525.
- Association of American Geographers
 1998 Statement on Professional Ethics. Available: <http://www.aag.org/publications/other%20pubs/ethicsstatement.html> [accessed November 25, 2005].
- Collogan, L.K., F. Tuma, R. Dolan-Sewell, S. Borja, and A.R. Fleischman
 2004 Ethical issues pertaining to research in the aftermath of disaster. *Journal of Traumatic Stress* 17:363-372.
- Collogan, L.K., F.K. Tuma, and A.R. Fleischman
 2004 Research with victims of disaster: Institutional review board considerations. *IRB: Ethics & Human Research* 26(July-August):9-11.
- de Wolf, V.A.
 2003 Issues in accessing and sharing confidential survey and social science data. *Data Science Journal* 2:66-74.
- de Wolf, V.A., J.E. Sieber, P.M. Steel, and A.O. Zarate
 2006 Part II: HIPAA and disclosure risk. *IRB: Ethics & Human Research*, 28(January-February), 6-11.

- Duncan, G.T.
 1993 Special issue on confidentiality and data access. *Journal of Official Statistics* 93:269-607.
 2003 Confidentiality and data access issues for institutional review boards. Pp. 235-247 in National Research Council, *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. C.F. Citro, D.R. Ilgen, and C. B. Marrett, eds. Panel on Institutional Review Boards, Surveys, and Social Science Research. Washington, DC: The National Academies Press.
- Fienberg, S.E.
 2004 Confidentiality in Geo-Spatially-Linked Data: How Robust Are the Solutions? Presentation at the Workshop on Confidentiality Issues in Linking Geographically Explicit and Self-Identifying Data, The National Academies, Washington, DC, December 9-10.
- Fisher, C.B.
 2003 *Decoding the Ethics Code: A Practical Guide for Psychologists*. Thousand Oaks, CA: Sage Publications.
- Flaherty, D.H.
 1978 Report of the Bellagio conference. *Journal of the Royal Statistical Society. Series A (General)* 141:401-405.
- Golden, M.L., R.R. Downs, and K. Davis-Packard
 2005 Confidentiality Issues and Policies Related to the Utilization and Dissemination of Geospatial Data for Public Health Applications. A report to the Public Health Applications of Earth Science Program, National Aeronautics and Space Administration, Science Mission Directorate, Applied Sciences Program. Prepared by the Socioeconomic Data and Applications Center, Center for International Earth Science Information Network, Columbia University, March 2005. Available: http://www.ciesin.columbia.edu/pdf/SEDAC_ConfidentialityReport.pdf [accessed November 25, 2005].
- Goodchild, M.F., and D.G. Janelle, eds.
 2004 *Spatially Integrated Social Science*. New York: Oxford University Press.
- Griffin, M.G., P.A. Resick, A.E. Waldrop, and M.B. Mechanic
 2003 Participation in trauma research: Is there evidence of harm? *Journal of Traumatic Stress* 16:221-227.
- Gutmann, M., K. Witkowski, C. Colyer, J.M. O'Rourke, and J. McNally
 2005 Providing Spatial Data for Secondary Analysis: Issues and Current Practices Relating to Confidentiality. Unpublished manuscript, Inter-university Consortium for Political and Social Research, University of Michigan (available from Myron Gutmann).
- Institute of Medicine
 2007 *Ethical Considerations for Research Involving Prisoners*. Committee on Ethical Considerations for Revisions to DHHS Regulations for Protection of Prisoners Involved in Research, L.O. Gostin, C. Vanchieri, and A. Pope eds. Washington, DC: The National Academies Press.
- Iutovich, J.M., J.M. Kennedy, and F.J. Levine
 2003 Establishing an ethical climate in support of research integrity: Efforts and activities of the American Sociological Association. *Science and Engineering Ethics* 9:201-205.
- Kwan, M.P., I. Casas, and B.C. Schmitz
 2004 Protection of geoprivacy and accuracy of spatial information: How effective are geographical masks? *Cartographica* 39:15-28.

- Kwan, M.R., and J. Lee
 2004 Geovisualization of human activity patterns using 3D GIS. Pp. 18-66 in M.F. Goodchild and D.G. Janelle, eds., *Spatially Integrated Social Science*. New York: Oxford University Press.
- Lane, J.
 2003 Key Issues in Confidentiality Research: Results of an NSF Workshop. Available: http://www.nsf.gov/sbe/ses/mms/nsfworkshop_summary1.pdf [accessed November 25, 2005].
- Levine, F.J., and J.M. Iutovich
 2003 Challenges in studying the effects of scientific societies on research integrity. *Science and Engineering Ethics* 9:257-268.
- Levine, F.J., and P.R. Skedsvold
 2007 Behavioral and social science research. In E.J. Emmanuel, R.A. Crouch, C. Grady, R. Lie, F. Miller, and D. Wendler (Eds.) *The Oxford Textbook of Clinical Research Ethics*. Oxford, England: Oxford University Press.
- Marshall, P.A., and C. Rotimi
 2001 Ethical challenges in community-based research. *American Journal of the Medical Sciences* 322(5):241-245.
- Melton, G.B., R.J. Levine, G.P., Kocher, R Rosenthal, and W.C. Thompson
 1988 Community consultation in socially sensitive research: Lessons from clinical trials on treatments for AIDS. *American Psychologist* 43:573-581.
- Merton, R.K.
 1973 The normative structure of science. In R.K. Merton, ed., *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.
- National Bioethics Advisory Commission
 2001 *Ethical and Policy Issues in Research Involving Human Participants: Vols I, II*. Bethesda, MD: National Bioethics Advisory Commission.
- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research
 1979 *Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. (GPO No. 887-809). Washington, DC: U.S. Government Printing Office. Also available: <http://ohsr.od.nih.gov/guidelines/belmont.html> [accessed November 13, 2005].
- National Human Research Protections Advisory Committee
 2002 Recommendations on Confidentiality and Research Data Protections. Available: http://www.aera.net/aera.old/humansubjects/NHRPAC_Final_Confidentiality.pdf.
- National Research Council
 1985 *Sharing Research Data*. Committee on National Statistics, S.E. Fienberg, M.M. Martin, and M.L. Straf, eds. Washington, DC: National Academy Press.
 1993 *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Committee on National Statistics, G.T. Duncan, T.B. Jabine, and V.A. de Wolf, eds. Washington, DC: National Academy Press.
 1998 *People and Pixels: Linking Remote Sensing and Social Science*. Committee on the Human Dimensions of Global Change, D. Liverman, E.F. Moran, R.R. Rindfuss, and P.C. Stern, eds. Washington, DC: National Academy Press.
 2000 *Improving Access to and Confidentiality of Research Data: Report of a Workshop*. Committee on National Statistics, C. Mackie and N. Bradburn, eds. Washington, DC: National Academy Press.

- 2002 *Research Ethics in Complex Humanitarian Emergencies: Summary of a Workshop*. H. Reed, Rapporteur, Roundtable on the Demography of Forced Migration, Committee on Population. Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- 2003a *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. Panel on Institutional Review Boards, Surveys, and Social Science Research, C.F. Citro, D.R. Ilgen, and C.B. Marrett, eds.. Washington, DC: The National Academies Press.
- 2003b *Sharing Publication-Related Data and Materials*. Committee on Responsibilities of Authorship in the Life Sciences. Washington, DC: The National Academies Press.
- 2005 *Expanding Access to Research Data; Reconciling Risks and Opportunities*. Panel on Data Access for Research Purposes, Committee on National Statistics. Washington, DC: The National Academies Press.
- National Science Foundation, Division of Social and Economic Science
- 1986 Data Archiving Policy and Implementation Guidance. Materials available from F.J. Levine at flevine@aera.net.
- Newman, E., and D.G. Kaloupek
- 2004 The risks and benefits of participating in trauma-focused research studies. *Journal of Traumatic Stress* 17:383-394.
- Newman, E., E. Risch, and N. Kassam-Adams
- 2006 Ethical issues in trauma-related research: A review. *Journal of Empirical Research on Human Research Ethics* 1(3):29-46.
- O'Rourke, J., S. Roehrig, S. Heeringa, B. Reed, W. Birdsall, M. Overcashier, M., and K. Zidar
- 2006 Solving problems of disclosure risk while retaining key analytic uses of publicly released microdata. *Journal of Empirical Research on Human Research Ethics* 1(3):63-84.
- Rindfuss, R.R., and P.C. Stern
- 1998 Linking remote sensing and social science: The need and the challenges. Pp. 1-27 in National Research Council, *People and Pixels: Linking Remote Sensing and Social Science*. Committee on the Human Dimensions of Global Change, D. Liverman, E.F. Moran, R.R. Rindfuss and P.C. Stern, eds. Washington, DC: National Academy Press.
- Rodgers, W., and M. Nolte
- 2006 Disclosure review procedures in an academic setting: Example of the Health and Retirement Study. *Journal of Empirical Research on Human Research Ethics* 1(3):85-98.
- Shea, D.A.
- 2006 *Oversight of Dual-Use Biological Research: The National Science Advisory Board for Biosecurity*. (CRS Report No. RL33342). Washington, DC: Congressional Research Service.
- Sieber, J.E.
- 1991 Introduction: Sharing social science data. In J.E. Sieber (Ed.), *Sharing Social Science Data: Advantages and Challenges*. Newbury Park, CA: Sage Publications.
- 1992 *Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards*. Newbury Park, CA: Sage Publications.
- Somerville, M.A., and R.M. Atlas
- 2005 Ethics: A weapon to counter bioterrorism. *Science* 307:1881-1882.

- Trimble, J.E., and C.B. Fisher
2006 *The Handbook of Ethical Research with Ethnocultural Populations and Communities*, Thousand Oaks, CA: Sage Publications.
- Urban and Regional Information Systems Association
2003 A GIS Code of Ethics. Available: http://www.urisa.org/ethics/code_of_ethics.htm [accessed August 9, 2005].
- U.S. General Accounting Office
2001 *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information*. (GAO-01-126SP). Washington, DC: U.S. General Accounting Office.
- VanWey, L.K., R.R. Rindfuss, M.P. Gutmann, B.E. Entwisle, and D.L. Balk
2005 Confidentiality and spatially explicit data: Concerns and challenges. *Proceedings of the National Academy of Sciences* 102:15337-15342.
- Vest, C.M.
2003 Balancing security and openness in research and education. *Academe*. Available: <http://www.aaup.org/publications/Academe/2003/03so/03sowest.htm> [accessed January 7, 2006].
- Wendler, D., L. Belsky, K.M. Thompson, and E.J. Emanuel
2005 Quantifying the federal minimal risk standard: Implications for pediatric research without a prospect of direct benefit. *Journal of the American Medical Association* 294:826-832.
- Willis, G.
2006 Cognitive interviewing as a tool for improving the informed consent process. *Journal of Empirical Research on Human Research Ethics* 1(1):9-23.
- Willison, D.J.
2003 Privacy and the secondary use of data for health research. Experience in Canada, and suggested directions forward. *Journal of Health Services Research and Policy* 8(Suppl 1):17-23.

Biographical Sketches for Panel Members and Staff

Myron P. Gutmann (*Chair*) is a professor of history and director of the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. Prior to joining the Michigan faculty in August of 2001, he was a professor of history and geography and director of the Population Research Center at the University of Texas at Austin. His research covers interdisciplinary historical topics, especially in health, population, economy, and the environment, and he has used a variety of approaches to study population-land use interactions. He also does research and writes on issues relating to data preservation and dissemination and about confidentiality protection in data used for secondary analysis. He is the author of *War and Rural Life in the Early Modern Low Countries*, and *Toward the Modern Economy, Early Industry in Europe, 1500-1800*, as well as more than 50 articles and chapters. He has been a member of the National Academies' Committee on the Human Dimensions of Global Change and its Panel on New Research in Population and Environment, as well as other national advisory committees and editorial boards. Gutmann received a B.A. from Columbia University, an M.A. from Princeton University, and a Ph.D. in history from Princeton University.

Marc P. Armstrong is a professor and chair of the Department of Geography at The University of Iowa, where he also holds an appointment in the Graduate Program in Applied Mathematical and Computational Sciences. A primary focus of his research is on the use of parallel processing to improve the performance of analysis methods used in spatial decision sup-

port systems. Other research interests are in mobile computing, privacy aspects of geospatial technologies, and evolutionary computation. He has served as North American editor of the *International Journal of Geographical Information Science* and on the editorial boards of many other journals. He has published more than 100 papers in a wide variety of academic journals, including the *Annals of the Association of American Geographers*, *Statistics and Medicine*, *Mathematical Geology*, and *Journal of the American Society for Information Science*. He received a Ph.D. from the University of Illinois at Urbana-Champaign.

Deborah Balk is associate professor at the Baruch School of Public Affairs and acting associate director of the Institute for Demographic Research at the City University of New York. Previously, she was a research scientist at the Center for International Earth Science Information Network at Columbia University, where she was also the lead project scientist for the Socio-economic Data and Applications Center, working on large-scale data integration of geographic, survey, and administrative data. Among her current projects, she is the principal investigator on two studies of urbanization and one on emerging infectious disease. She is a member of the Working Group on Urbanisation of the International Union for the Scientific Study of Population. She received a B.A. and M.A. from the University of Michigan-Ann Arbor and a Ph.D. in demography from the University of California at Berkeley.

Kathleen (Kass) O'Neill Green recently retired from her position as president of Space Imaging Solutions, a division of Space Imaging, LLC. While with Space Imaging, she directed programs that offered satellite imagery, remote sensing, and GIS (global information services) services to clients worldwide. She currently serves as an independent consultant and board member to public, private, and nonprofit natural resource and geospatial organizations. Her background includes 30 years of experience in natural resource policy, economics, GIS analysis, and remote sensing. She is the author of numerous articles on GIS and remote sensing and coauthored a book on the practical aspects of accuracy assessment. She is the recent past president of Management Association for Private Photogrammetric Surveyors (MAPPS), an organization of private mapping firms dedicated to advancing the mapping industry. She received a B.S. in forestry and resource management from the University of California at Berkeley and an M.S. in resource policy and management from the University of Michigan-Ann Arbor.

Felice J. Levine is executive director of the American Educational Research Association (AERA). Previously, she served as executive officer of the American Sociological Association, as a program director at the National

Science Foundation, and as senior research social scientist at the American Bar Foundation. Her work has concentrated on science policy issues, including research ethics, data access and sharing, and peer review; academic and scientific professions; and diversity in higher education. She was a member of the National Human Research Protections Advisory Committee of the U.S. Department of Health and Human Services and on the advisory committee for the decennial census. She is currently on the executive committee of the Consortium of Social Science Associations and served as its chair from 1997 to 2000 and is on the Board of Directors of the Council of Professional Associations on Federal Statistics. She is a fellow of the American Psychological Society and of the American Association for the Advancement of Science. She holds A.B., A.M. and Ph.D. degrees in sociology and psychology from the University of Chicago.

Harlan Onsrud is a professor in the Department of Spatial Information Science and Engineering at the University of Maine. His research focuses on the analysis of legal, ethical, and institutional issues that affect the creation and use of digital databases and the assessment of the social effects of spatial technologies. He is president elect of the Global Spatial Data Infrastructure Association (GSDI), past president of the University Consortium for Geographic Information Science (UCGIS), and past chair of the U.S. National Committee (USNC) on Data for Science and Technology (CODATA) of the National Research Council. He is a licensed engineer, land surveyor, and attorney. Current and past research projects have been funded by the National Science Foundation, the National Geospatial-Intelligence Agency, the Federal Geographic Data Committee, and the U.S. Department of Education. He holds B.S. and M.S. degrees in engineering and a J.D., all from the University of Wisconsin.

Jerome P. Reiter is an assistant professor at the Institute of Statistics and Decision Sciences at Duke University. His primary research areas include statistical methods for preserving data confidentiality, handling missing data, and making casual inference. He works extensively with the U.S. Census Bureau and the National Institute of Statistical Science on research in statistical disclosure limitation. He is a member of the committee on Privacy and Confidentiality of the American Statistical Association. He serves on the editorial board of the *Journal of the American Statistical Association*, the *Journal of Privacy and Confidentiality*, and *Survey Methodology*. He received a B.S. in mathematics from Duke University and a Ph.D. in statistics from Harvard University.

Ronald Rindfuss is the Robert Paul Ziff distinguished professor of sociology and a fellow at the Carolina Population Center (CPC) at the University

of North Carolina-Chapel Hill and a senior fellow at the East-West Center in Honolulu. He previously served as director of the CPC. As a social demographer, his work focuses on the timing and sequencing of cohabitation, marriage, childbearing, divorce, education, migration, and employment. He is currently working on the relationship between population processes and the environment, examining migration and social change in Thailand examining the consequences of child care patterns in Norway and examining changes in family processes in Japan. He is a past president of the Population Association of America and a fellow of the American Association for the Advancement of Science. He holds a B.A. from Fordham University and a Ph.D. from Princeton University, both in sociology.

Paul C. Stern (*Study Director*) is a senior staff officer at the National Research Council and study director of the Committee on the Human Dimensions of Global Change. His research interests include the determinants of environmentally significant behavior, particularly at the individual level; participatory processes for informing environmental decision making; and the governance of environmental resources and risks. He is the coeditor of numerous National Research Council publications, including *Population, Land Use, and Environment: Research Directions* (2005), *The Drama of the Commons* (2002), and *People and Pixels: Linking Remote Sensing and Social Science* (1998). He is a fellow of the American Association for the Advancement of Science and the American Psychological Association. He holds a B.A. from Amherst College and an M.A. and Ph.D. from Clark University, all in psychology.

