Curating the Ethnographic Moment

Andrew D Asher, Indiana University - Bloomington
Lori M Jahnke, Emory University

Available at: https://works.bepress.com/andrew_asher/42/
Curating the Ethnographic Moment

Andrew Asher
Lori M. Jahnke

Abstract

Digital technologies have vastly increased the volume of materials that ethnographic researchers are able to collect. They have also enhanced possibilities for rapidly sharing data not only with other researchers, but also with research subjects and the public at large. However, the relative ease of collecting digital materials can quickly outpace the data management skills of researchers, who usually have little or no training in recordkeeping, preservation, or curation strategies. Making digital data available online can also compound ethical dilemmas faced by ethnographers, who often find themselves torn between contradictory imperatives to share data publically and to protect the privacy and confidentiality of the individuals and communities they study.

Using interviews conducted with researchers at five universities, this article describes particular problems associated with preserving and curating ethnographic data. We emphasize the everyday practices of ethnographers by examining the practical aspects of working with diverse ethnographic materials, the unique status of ethnographic fieldnotes—which are at once intensely personal and private but also the basis for public authority—and the negotiation of ethical questions raised by varying or conflicting cultural standards for research consent and data use. Based on these observations, this article explores the implications of data management, preservation, and stewardship practices for the processes of ethnographic research and publishing. We conclude by making recommendations for data scientists who work with ethnographers to curate their research data.

Introduction

By vastly increasing the volume and types of data that field researchers can collect, digital technologies have reshaped ethnographers’ relationship with their accumulated research data and materials. Whereas paper-based fieldnotes and other physical ethnographic records are often deposited in repositories and made accessible only after a scholar’s death, the emergence of born-digital materials has created and enhanced possibilities for rapidly sharing data, not only with other researchers but also with research subjects and the public at large via online repositories. While many researchers support the principal of public access to research data, making this data available online can compound the ethical dilemmas ethnographers face. Ethnographers often find themselves torn between contradictory imperatives to share data publically and to protect the privacy and confidentiality of the individuals and communities they study.¹

Unlike experimental data, ethnographic data, by definition, are not reproducible.² Ethnographic fieldwork is inextricably linked to a particular time, place, and society, making the primary data of ethnographers unique and unrecoverable. For this reason, today’s ethnographic fieldnotes are particularly important to preserve as tomorrow’s historical records; they document observations
of places and cultures that—because they are in a constant state of change—will never exist in the same configuration again. Since ethnographic fieldnotes provide the empirical basis for researchers’ public authority, it would likewise seem logical for fieldnotes and other ethnographic materials to be made available for verification purposes in online repositories as soon as possible after the publication of research reports, articles, or monographs.

However, several characteristics of ethnographic materials cause them to be particularly difficult to make publically accessible. First, many ethnographic researchers are reluctant to share their data for personal reasons. Ethnographic fieldnotes are intensely personal documents for many researchers, who often view them as extremely valuable for their own work, but doubt that they would be useful to other researchers without the accompanying contexts and experiences in which the materials were collected. Secondly, ethnographic materials also routinely contain private, confidential, or culturally sensitive information about both research subjects and the researcher, which often requires time-consuming or costly procedures to redact, and may limit or prevent public dissemination and secondary analyses of the data. Finally, ethnographic materials are regularly collected in contexts with varying or conflicting cultural standards for consent and data use, raising ethical questions about the efficacy of archiving these materials. For these reasons, balancing the multiple imperatives of data sharing, informed consent, and confidentiality for research participants is a fundamental issue for ethnographers preparing to deposit their materials in an archives, and the ethics surrounding these decisions was one of the most common and recurring themes in the interviews conducted for the study we present here.

Using data gathered during the 2011-2012 academic year in a qualitative study of researchers’ data management and curation practices, this article examines the practical and ethical dilemmas ethnographers face as they work with diverse research materials across varying, and sometimes conflicting, cultural standards for consent and data use. By emphasizing the everyday practices of researchers, this study sought to gain a holistic understanding of the workflows involved in the creation, management, and preservation of research data in order to better comprehend researchers’ unmet needs within data curation processes. Twenty-three researchers from varying ranks working in a variety of social science disciplines at five universities participated in interviews for this study (see Table 1), which focused on how researchers collect and analyze data, how they manage and preserve these data, and what training they have had in data-curation practices.

Although scholars across many disciplines were grappling with the ethical and philosophical problems of data sharing—often in the absence of coherent policies for data archiving and release from funding agencies, professional organizations, or their employing institutions—this article will focus on the experiences of nine of these researchers who conducted ethnography as part of their data collection. Because ethnographic research involves a complex social interplay between researchers and respondents, often from differing cultural contexts, the management of these data presents unique difficulties for researchers who usually have little or no training in preservation, curation, or data-stewardship strategies. This article will examine how researchers navigate these difficulties when dealing with the two principal types of ethnographic data: fieldnotes—the diverse array of notes, writings, and other materials upon which researchers base
Table 1. Rank and academic discipline of the study participants. *The nine starred participants conducted significant amounts of ethnographic research.

<table>
<thead>
<tr>
<th>Participant #</th>
<th>Rank/title</th>
<th>PhD Discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-03-100511</td>
<td>Assistant Professor</td>
<td>Anthropology (Biological)</td>
</tr>
<tr>
<td>1-17-121211</td>
<td>Assistant Professor</td>
<td>Education</td>
</tr>
<tr>
<td>*2-12-111011</td>
<td>Assistant Professor</td>
<td>Environmental Science</td>
</tr>
<tr>
<td>*2-16-1202011</td>
<td>Assistant Professor</td>
<td>Anthropology</td>
</tr>
<tr>
<td>3-05-102111</td>
<td>Assistant Professor</td>
<td>Developmental Psychology</td>
</tr>
<tr>
<td>3-07-102111</td>
<td>Assistant Professor</td>
<td>Political Science</td>
</tr>
<tr>
<td>3-08-102111</td>
<td>Assistant professor</td>
<td>Political Science</td>
</tr>
<tr>
<td>2-22-021512</td>
<td>Professor</td>
<td>Environmental Studies</td>
</tr>
<tr>
<td>2-15-120211</td>
<td>Associate Professor</td>
<td>International Relations</td>
</tr>
<tr>
<td>5-09-103111</td>
<td>Data Scientist</td>
<td>Geology</td>
</tr>
<tr>
<td>1-01-72911</td>
<td>Digital Curator</td>
<td>Slavic Languages</td>
</tr>
<tr>
<td>*5-10-103111</td>
<td>Graduate Student</td>
<td>Sociology</td>
</tr>
<tr>
<td>*5-20-020212</td>
<td>Graduate Student</td>
<td>Environmental engineering</td>
</tr>
<tr>
<td>*5-21-02032011</td>
<td>Graduate Student</td>
<td>Anthropology</td>
</tr>
<tr>
<td>3-06-101111</td>
<td>Grant Coordinator of a research center</td>
<td></td>
</tr>
<tr>
<td>*1-04-100511</td>
<td>Postdoctoral Fellow</td>
<td>Sociology (Demography)</td>
</tr>
<tr>
<td>4-25-120511</td>
<td>Postdoctoral Fellow</td>
<td>Education (Learning Sciences)</td>
</tr>
<tr>
<td>3-14-113011</td>
<td>Postdoctoral Fellow</td>
<td>Architectural history</td>
</tr>
<tr>
<td>*4-19-012012</td>
<td>Postdoctoral Fellow</td>
<td>History</td>
</tr>
<tr>
<td>*2-13-111411</td>
<td>Professor</td>
<td>Sociology</td>
</tr>
<tr>
<td>*5-11-103111</td>
<td>Professor</td>
<td>Sociology</td>
</tr>
<tr>
<td>1-02-100511</td>
<td>Professor</td>
<td>Anthropology (Archaeology)</td>
</tr>
<tr>
<td>4-18-121911</td>
<td>Researcher</td>
<td>Anthropology (Archaeology)</td>
</tr>
</tbody>
</table>
their ethnographies—and the research artifacts, such as transcripts and audio or video recordings, produced from more formal, direct interactions with respondents.

The Data Sharing Imperative

Funding agencies—especially agencies administering public funds—are increasingly requiring qualitative research data to be made publically available. Since 1995, the Economic and Social Research Council (ESRC)—a major source of funding for qualitative research in the United Kingdom—has required its awardees to archive their data as a condition of funding. Similarly, the National Science Foundation (NSF) and other US funding agencies (e.g. National Institutes of Health and the National Endowment for the Humanities) have begun to require formal data-management plans detailing how researchers will meet data-sharing expectations. Our assumption is that this trend will continue as both public and private funders seek maximum returns on their investment in research. While not always explicitly required, the goal of many of these policies is for research data to be deposited in an appropriate repository with online access for both other researchers and the public. For ethnographers, who are often accustomed to retaining exclusive rights over their fieldnotes and other materials, this can represent an uncomfortable shift in research practice, even when these repositories take into account the particular confidentiality and consent demands inherent to ethnographic research (discussed further below) and implement appropriate safeguards such as embargoing or redacting sensitive materials.

In principal, the ethnographic researchers we interviewed for this study saw data preservation and sharing as positive values, even if they were not sure about how to preserve and share data or who would potentially use the materials. When asked about what parts of her research data she would like to preserve, one anthropologist explained:

That’s the thing—If you keep anything long enough it becomes interesting. . . . For myself . . . I would love to archive everything and maybe put some kind of time stamp on personal information where it can’t be released for some amount of time. It’s so hard to know what people in the future would be interested in. . . . I try to keep data in such a way that hopefully it would be comprehensible to someone else if they were willing to work through my crazy software. . . . If I had somebody’s fieldnotes from 100 years ago from my field site, I’d be totally fascinated by that, so you never know what people are going to be interested in.

(2-16-120211)

The scholarly societies to which many of these researchers belong also broadly support data sharing and potential reuse. The most recent 2012 revision on the American Anthropological Association’s (AAA) ethics statement explicitly states that it is the obligation of the researcher to preserve records and make results accessible, “Anthropologists have an ethical responsibility for ensuring the integrity, preservation, and protection of their work” and, the “results of anthropological research should be disseminated in a timely fashion” provided it is not “at the expense of protecting confidentiality.” The previous 1998 version also addressed data access and preservation, but not as systematically or forcefully. Likewise, the American Sociological Association sets similar expectations in the data sharing section of its Code of Ethics.
The benefits of data archiving and sharing for scholarly disciplines and the advancement of knowledge are straightforward: preserved materials can be subjected to secondary analysis, unnecessary duplication of research can be minimized, historical records can be maintained, and, importantly, scholars can independently evaluate the validity of published results by consulting the primary data. Furthermore, online repositories might enable new forms of scholarship and writing. For the first time, materials preserved and made accessible online have made it feasible to make publically available the corpus of primary texts that form the basis of support for ethnographic argumentation. While a non-digitized, paper-based collection could also serve this purpose, it is unlikely that more than a handful of interested readers of a particular ethnography would physically travel to the repository to examine the documents. Repositories that make resources available online fundamentally change this dynamic. In *Ethnography as Commentary: Writing from the Virtual Archive*, Johannes Fabian explores these possibilities by conducting what he describes as an experiment with “writing ethnography ‘in the presence of texts’” located in a publically accessible online repository. Fabian terms this genre of writing “commentary,” a text-centered form of ethnography that assumes his reader can simultaneously consult a primary text for her or himself via an online archive.

The text Fabian utilizes is a conversation he and Kahenga Mukonkwa Michel had about a ceremony Kahenga conducted at Fabian’s home in Zaire in 1974. The transcription, translation, and an audio excerpt of the discussion are now housed in the Archives of Popular Swahili on the website of the Language and Popular Culture in Africa project. Fabian argues that the repository of documents serves an important epistemological purpose by “mediat[ing] between . . . events, the presentation of evidence, and our aims and claims to produce anthropological knowledge.” By allowing the reader of an ethnography to interrogate and evaluate the primary data on which it is based, the archive reveals how ethnographies are composed through the confrontation between the researcher, his or her respondents, and the texts they co-create.

Fabian’s instructive experiment notwithstanding, for many researchers the practice of making ethnographic materials available online is fraught. Ethnography requires researchers to occupy a methodological position that mediates and translates between multiple cultural contexts. This positioning leads to a number of difficulties for ethnographers as they attempt to meet their obligations to multiple constituencies—including funders, cultural institutions, and the researcher’s employing university, in addition to the individuals and communities the researcher works with—while negotiating the processes of placing materials in archives and repositories. When working with human subjects, ethical pitfalls abound, and often ethnographers must resolve contradictions between these groups.

At a very basic level, policies requiring access to research data may “directly [contradict] some ethnic groups’ cultural traditions concerning secrecy and controlled access to information.” For example, an anthropologist might gain access to knowledge that is secret, or restricted for cultural or social reasons. During his fieldwork with the Saramaka, a group in Suriname who descended from escaped African slaves, Richard Price gained knowledge about the “Old-Time People” who experienced the war for liberation from roughly 1680 to 1762. The Saramaka consider knowledge about this time and these people powerful, dangerous, and restricted. Price struggled with the ethical implications of publishing a book within a different cultural system.
that uses this knowledge as its primary subject. Price notes that by putting the information in a public document, he potentially “deprives [the] author of control (except perhaps via the language in which [the book] appears) over [his or her] audience.” This problem is compounded exponentially once information is made available online (as Price’s book now is, at least in part, in Google Books) by removing even the physical restrictions of a bound monograph.

Information need not be “exotic” to create publishing difficulties for ethnographers. For example, an ethnographer working in a corporate, government, or other institutional setting in the United States might be privy to proprietary or confidential information about a product or service that must be ethically and legally protected. A researcher in environmental engineering explained her process when conducting interviews:

Another big concern is confidentiality because I record those conversations, and then there’s kind of an extended period of time in which I allow those who I’ve talked with to redact parts of our conversation if they’ve talked about things they would prefer to remain private. . . I’m talking to several companies about the development and production of their technologies and so there’s some proprietary information that comes up and that’s really the main concern.

Indeed, accidental release of this type of information might put the researcher at as much legal risk as her respondents.

When preparing to preserve ethnographic data, issues of ownership and authorship are also not always straightforward, and it is imperative that researchers communicate and address these potential problems with archivists, data curators, and repository and digital collection managers before depositing ethnographic materials. For example, consent forms for interviews might not clearly state who holds the copyright or who has a license to use the recording or transcripts, potentially leading to confusion over the legal status of these materials and making it more difficult for a researcher to transfer materials to an archives. Intellectual property and copyright are themselves culturally specific concepts, and a legal right to dispose of or deposit materials as a researcher sees fit does not equal an ethical right to do so. The people providing information, the funders, the institution that employs the researcher, and the community in which the researcher works all have a moral interest in the disposition of the data.

Parry and Mauthner go so far as to say, “[B]ecause the construction of qualitative data is a joint endeavor between respondent and researcher, both parties should retain authorship/ownership rights over the data,” which creates “practical, legal and ethical implications for archiving and re-use.”

Some researchers, including Fabian, argue that since ethnographers and their respondents are joint agents in creating ethnographic materials, both parties should be named to acknowledge their contributions. However, naming respondents as authors or contributors contradicts an ethnographer’s imperative to protect the confidentiality of their research participants, which is a basic standard required by most Institutional Review Board (IRB)-approved research and by professional-ethics statements. While David Zeitlyn observes that “a default assumption that notes will be anonymized conflicts with an individual’s moral right to be recognized as the author of his or her words,” many, if not most, ethnographers would likely be uncomfortable
naming their participants in almost any context, but especially in publically accessible online repositories. This is likely to hold true even if a researcher philosophically accepts that his or her authority to determine the level of sensitivity that the data require privileges the researcher’s position vis-à-vis his or her respondents.26

Many ethnographers thus understand themselves not so much as owners of their data, but rather as stewards or custodians of data that are produced by a complex set of relationships between researchers and respondents.27 As the stewards of these data, researchers are often the de facto interpreters of the terms by which data are made available, choosing what and when to publish or deposit materials in repositories or archives. Despite his assertion that correspondents should be named, Fabian notes that he exercises this stewardship by attempting not to publish anything that might be dangerous or damaging to the people with whom he worked—a position that is probably made easier by the fact that the texts he is working with were already 30 years old when he transcribed and posted them on the web.28 Many researchers we interviewed adopted a similar “attitude of moral responsibility towards research participants, expressed in the way in which data are handled,”29 citing ethical concerns about appropriate use of the data as well as issues of confidentiality and consent. Even for researchers who support public access to data, these concerns cause them to question funders’ data-sharing requirements. A researcher in education who was preparing to apply for additional funding to create a longitudinal dataset by re-interviewing her previous respondents noted her uncertainty:

I know that a lot of places when you apply for funding they want eventually for [data] to become public. I haven’t had to deal with it, but I’ve definitely thought about it . . . . I want to make sure to preserve the qualitative data because it’s very important to me . . . so it’s a concern for me personally to save it . . . . [But] once it’s public, or shared between other people, I would like to get help eventually on making the data itself more confidential because they are just interview transcripts which, you know, there are a lot of [confidentiality] issues with that, and that I would have to [make the data itself more confidential] before I shared it . . . . I don’t know if there are limits to the extent to which I can make it public or not.

(1-04-100511)

In the context of making their data accessible, these confidentiality and consent negotiations are some of the most difficult issues for scholars to resolve before depositing their ethnographic materials in data repositories or archives. These issues are also most effectively addressed as early in a research project as possible, so that researchers can implement the necessary practices and protocols to insure that confidentiality and consent can be assured in the long term prior to data collection. As Charles Humphrey and his colleagues discuss, the integral involvement of archivists and data-curation specialists in the planning stages and throughout qualitative research projects greatly increases the quality of the data products and decreases ethical problems associated with deposition in repositories and archives.30 As they point out, this collaboration is beneficial to both researchers and archivists, since good archival preparation practices are often synonymous with good research practices.31

Despite efforts by organizations such as the Council for the Preservation of Anthropological Records (CoPAR) to foster this type of collaboration32 and educate ethnographers in data
curation and archival preparation practices, it appears that relatively few ethnographers seek out the assistance of archivists or data management specialists until after data collection is completed, unless funding requirements compelled them to do so. Of the ethnographers we interviewed, only two had consulted with data management services prior to the data collection phase of their research, both as a result of the NSF’s requirement that data-management plans be included in new funding proposals. Only one project had a systematic plan for depositing its ethnographic data in an archives, principally because this research was conducted as part of a very large and well-funded research initiative. This project had amassed some 40,000 pages of ethnographic fieldnotes, which were being preserved but not yet been shared publically. The researcher leading this project explained:

It’s news to anthropologists that data should be made public. It’s not something that’s in their mindset. In fact they are appalled by it. . . . However, that [ethnographic] data was paid for using a million dollars of federal money and I’m using it with some of my grad students who are essentially doing secondary analysis of fieldnotes. I think that in the long run that data should be made public, and could be, but very carefully. . . . I don’t think that’s a problem . . . my guess is we’re going to get there. . . . A very new area now is “what do you do with electronic fieldnotes?” and “can those be made available for researchers?” If nothing else, I think this archive of fieldnotes . . . in 20 or 30 years could be a great historical source and should be preserved on that basis. So, I would like to see us work out ways of carefully having anthropological and ethnographic data made public too. (5-11-103011)

In working out the procedures for sharing these materials, this researcher went on to emphasize the paramount importance of protecting research participants’ confidentiality. Indeed, all of the ethnographers cited confidentiality concerns as one of the foremost issues that they would need to discuss and resolve with an archivist or repository manager before depositing their materials, making it perhaps the most critical area for developing collaborative practices between researchers, data curators, and archivists.

The Confidentiality Imperative

The ethics statements of both the American Sociological Association and the American Anthropological Association require researchers to protect the confidentiality of their respondents (although the AAA statement asserts that expectations of credit should also be addressed). Protecting research participants’ confidentiality is routinely an explicit promise within social science research, and is an almost universally assumed requirement of IRBs when reviewing research proposals. Setting aside the research participant’s potential moral right to authorship as discussed above, the nature of ethnographic research often renders the promise of confidentiality problematic, if not outright impossible. As Zachary Schrag observes, maintaining the confidentiality of fieldwork sites and research participants discussed in publications has been a perennial problem for social science researchers, as details necessary to make ethnographic and other fieldwork-based writing meaningful often simultaneously make these places and people identifiable, especially to people familiar with the location or topic. Increasing confidentiality protections almost always come at the cost of decreasing the amount of contextual information
included for the ethnography’s audience, and can effectively prevent data from being deposited in an archives or repository (or sometimes even collected). A sociologist recounted the difficulties he encountered on one project:

I wanted to do life histories with priests [in central Pennsylvania], and part of the problem was . . . we got into a situation where people might tell me things about their personal lives that are sort of not confidential in the IRB sense but that might be upsetting to their congregations—like I talked to one priest who had been married three times, where if the congregation had known about that they would have been very upset. There’s nothing illegal about it; this person’s not shy about telling that, but it could have been damaging. But I had the sense that . . . if I went around and did these interviews, I didn’t want to destroy the data because I thought they had some historical importance. . . . [The IRB said,] “and when you get done with this study if you’re not going to destroy the data you have to store it in an archive for 50 years, and I checked with [university] archives and they said, “we’re not going to store anything like that.” . . . I wanted to preserve the data, but there was no way to do it. . . . I actually quit the project [because the IRB made it impossible to do]. (2-13-111411)

While protecting research participants’ privacy can be difficult, even in carefully prepared and edited publications, the nature of ethnographic materials makes them especially difficult to confidentially preserve over time. Audio and video recordings present obvious confidentiality problems by capturing a respondents’ voice or image, but even interview transcripts can be problematic. A common practice in archiving ethnographic materials is to make attempts to remove potentially identifying information from archived transcripts. However, even these “anonymized” transcripts often still contain personal details that allow the possible identification of research participants. For example, in the interviews conducted for this study, participants regularly discussed their discipline and research interests. Coupled with the names of the universities where the research was conducted—which are contained in all of our research proposals and publications—this seemingly innocuous information would probably render any of our participants identifiable with a simple web search if the full interview transcripts were released. This might not seem like a significant risk for scholars describing projects that they routinely discuss publically, but depending on the other details contained in the transcript it might also be embarrassing or potentially damaging (for example, comments about university administrators, other colleagues, etc.). As is often the case with qualitative research, removing these contextual details would diminish, if not destroy, the usefulness of these data to future researchers—the ostensible audience for these materials. Unfortunately, each piece of contextual detail retained about a research participant to aid in data interpretation simultaneously creates more potential for identification.

Given technological advances, removing all conceivably identifying details might still be insufficient to guarantee that transcripts are rendered anonymous. In fact, anonymizing transcripts may soon no longer be technically possible. Arvind Narayana and his colleagues have demonstrated that it is possible to accurately identify anonymous online authors up to 20 percent of the time based on writing style alone, given a sufficient corpus of example texts. As more and more content moves online, one can only assume that these techniques will become more

robust, so that identifying a participant from the vocabulary or speech patterns in a transcript may eventually become a matter of applying the correct software. These realities set up even the most well-intentioned researchers to fail to meet the obligations required by IRBs and the agreements made with their respondents.

When archiving ethnographic materials, confidentiality can also be equally important for protecting the researcher. This is especially true for the archiving of fieldnotes, which are profoundly personal for many ethnographic researchers. Researchers are often quite ambivalent about these data, regularly describing feelings of anxiety, inadequacy, or embarrassment about their work and what it reveals about them as scholars. When talking about her data collection process, an anthropologist in our study observed:

That’s the thing with fieldnotes, you never show them to anyone . . . as long as it works for you . . . it’s not like you have to show it to someone else and have them make sense of it, which is kind of a shame because it would be nice to have data in formats where people could, you know, sort of, archive that information. (2-16-120211)

In his study of fieldnote practices, Jean Jackson observes, “Many respondents point out that the highly personal nature of fieldnotes influences the extent of one’s willingness to share them: ‘Fieldnotes can reveal how worthless your work was, the lacunae, your linguistic incompetence, your not being made a blood brother, your childish temper.’” Anthropologist Simon Ottenberg describes his fieldnotes similarly, “[W]hen I was younger, I would have felt uncomfortable at the thought of someone else using my notes, whether I was alive or dead—they are so much a private thing, so much an aspect of personal field experience, so much a private language, so much part of my ego, my childhood [as an anthropologist], and my personal maturity.” This attachment and ambivalence may make researchers reluctant to turn over control of their notes to an archive—especially one that has the purpose of making materials available publically on the internet. This professional practice not only makes it very difficult to substantiate or verify ethnographers’ data, suggesting a need for ethnographers to develop more proactive and sensitive data-sharing procedures, but also creates a situation where ethnographic materials are saved, but with inadequate plans for preservation. Zeitlyn notes, “Paradoxically most anthropologists want neither to destroy their field material nor archive it.” So, too, with our study: almost no one had made plans for their data beyond the short term, let alone the final dispensation of their materials at the end of their career or after their death.

While archivists regularly assure scholars that the papers of “ordinary” or not-famous researchers are sometimes the most valuable, researchers themselves are often unconvinced. Many of our respondents had trouble imagining the potential audience for their data and primary materials, although one Environmental Studies researcher shared data on graffiti images she collected during fieldwork following a particular ethnic conflict event. The data were not directly related to her primary research questions, but another scholar used them in a Master’s thesis (and possibly in a future dissertation). Nevertheless, doubts about future usefulness contribute to researchers’ reluctance to undertake sometimes arduous and time-consuming data preservation and annotation tasks, such as redacting confidential information, and to neglect advance planning for the eventual archiving of their materials, such as assuring adequate consent procedures. This
is potentially a key area of advocacy for data curators and archivists, who are well positioned to both educate researchers in the importance of preserving primary data for future generations of researchers, even if its immediate value is not apparent, and to provide tools such as cross-indexes and finding aids that make these data more useful and comparable.43

The Consent Imperative

Both IRBs and disciplinary societies’ ethics statements emphasize the importance and necessity of obtaining informed consent from research participants. Given the possibility of archiving materials indefinitely, and especially for web publication, obtaining fully informed consent may be a futile endeavor. Libby Bishop points out that because researchers cannot adequately speculate on all the uses in which data might be employed, explicit consent is “logically impossible.”44 However, Bishop also observes that this should not paralyze researchers from depositing their materials in archives, since consent for the primary research activity is in itself partial and no researcher can fully explain all information about any research project.45 Zeitlyn makes this point in another way, noting that cultural differences may also limit the ability of a researcher to obtain consent in a way that is recognizable to IRBs: “Once material is archived, it may be consulted by unknown others who might use it in novel ways. . . . Can a researcher obtain meaningful prior informed consent if the uses to which the material might be put in the future cannot be explained?”46

While Bishop suggests responding to this problem by providing examples of how archived research might be reused or by obtaining open-ended consent,47 others suggest a more comprehensive approach that understands consent as a process rather than a one-time event.48 This approach presents significant difficulties when applied to the lifecycle of an entire ethnographic research project, especially once data and other fieldwork materials are deposited in archives for the purpose of long-term access and preservation. If appropriate consent for archiving is not obtained at the outset of a project, or is sought when new uses are discovered for materials, contacting participants to authorize this usage can be extremely difficult and costly for ethnographers, if not outright impossible. Furthermore, a consent model that requires re-contacting participants also requires confidentiality or anonymization procedures to be reversible, meaning that the researcher must retain records that can link individuals to the materials about them. Retaining contact information or other identifying records is not best practice for assuring participant confidentiality, and in the case of truly sensitive research, doing so unnecessarily maintains records that could be subject to subpoena or otherwise compromised. For example, in 2012 the US Department of Justice issued a subpoena in connection with a murder investigation on behalf of Great Britain for access to oral history interviews conducted with Irish Republican Army and Loyalist paramilitary members as part of “The Belfast Project,” archived at Boston College’s Burns Library. While the project’s confidentiality agreement promised interviewees that these materials would be embargoed until their death, this guarantee was subject to exceptions required by American law; a point that may have been unclear to the research participants. Both Boston College and the researchers are presently contesting the subpoenas, but the initial rulings by a federal district court and the First Circuit Court of Appeals have found that the subpoenas are legal.49 As this case demonstrates, whenever identifying information exists, there is a possibility that its disclosure can be compelled. From the point of
view of research-participant confidentiality, it is therefore preferable to destroy materials that enable identification or re-contact as soon as possible after research is complete.

Since anonymization is an all-or-nothing decision that must be irreversible, even for the researcher, if it is to be effective, in order to make consent an ongoing process, one must also accept that the data cannot be made fully anonymous. The question then becomes how researchers can meaningfully extend consent decisions to include archives without rendering them unmanageable for their administrators or unnecessarily diminishing their analytical value. Balancing the need to maintain appropriate levels of confidentiality with the administrative costs is a critical area for collaboration between researchers, archivists, and data specialists. Fully anonymizing datasets is probably neither necessary, nor cost effective, for relatively low-risk materials, but both the researcher and the archivists must fully understand the potential risk scenarios before these decisions can be made, particularly in an environment where privacy protections are an increasing concern for research participants. While discussing the challenges he faces in complying with funder’s public access requirements, a sociologist observed:

> Clearly the biggest problem is the problem of protection of privacy. The concern of privacy has been ramped up tremendously over the last 10 or 15 years, and the process of getting permission to analyze data can be difficult, but a trend in social science data is to include more and more information that's sensitive. . . . I think the field is trying to now establish appropriate levels of protection for particular kinds of data and is trying to balance this problem of privacy with public access. And it is a challenge and it’s going to require some new modes of doing things. (5-11-103111)

> Developing these modes of work and processes of consent requires not only the participation of researchers and scholarly societies, but also intensive cooperation with data management specialists, technologists, and archivists.

Conclusions and Recommendations

Researchers and archivists alike face tremendous challenges in effectively preserving ethnographic research data and providing meaningful access for researchers and other stakeholders while simultaneously protecting sensitive data. In order to be effective, data sharing, confidentiality, and consent imperatives all require the active participation and investment of researchers themselves. Decisions about access, issues of authorship attribution, adequate consent, and the level of confidentially required cannot be made without the input of someone who is deeply knowledgeable about the contexts in which the research took place. However, as we have seen, none of these decision processes necessarily end at the moment when the materials are deposited in an archives or repository, and a researcher’s ethical obligation to his or her research participants does not end once he or she has transferred responsibility for their preservation. This sense of responsibility makes many researchers reluctant to relinquish control over their ethnographic materials, lest changing political, social, or cultural conditions necessitates a more restrictive approach to data access than was initially required.
For example, an Environmental Studies researcher noted that currently her data does not present much risk to her respondents:

Most of the data I’ve collected really have limited privacy issues or any kind of harm that could come to respondents. There are some topics that I’ve looked at, that depending on, when—you know, timeliness—at what point these were published, that identifying information could be problematic. In [central Asian country] it’s a lot easier—at least currently—the situation for my topics isn’t dangerous. That doesn’t mean it can’t be in the future and if I extended the research to neighboring countries then I would probably have some more serious confidentiality issues. (2-12-111011)

If the political situation for her respondents changed, however, this researcher might be ethically obligated to change the access conditions of any materials that she had archived, even if her consent procedures had addressed this eventuality. Repositories for ethnographic materials must be careful to include procedures to respond to these types of situations in their donor agreements, and researchers should discuss the unique qualities of ethnographic materials prior to their deposit. As Parezo points out, “it is unethical and unreasonable to expect the archival community to guess which fieldnotes contain sensitive information and what are reasonable use restrictions.”

Unfortunately, preparing materials for deposit into an archives is often a low priority for ethnographers for a variety of reasons, and the archiving of ethnographic data is further hindered by a persistent lack of training. While comprehensive historical review is beyond the scope of this article, Silverman already noted the problem of a lack of formal training in data collection and management practices among anthropologists almost 20 years ago, before the rise of many of the complications that the expansion of digital data collection created. Almost none of the scholars we interviewed reported that data curation or stewardship training was part of their graduate curriculum, and data management was usually only addressed narrowly in relation to methodological approaches and problems. The ethical and practical difficulties involved in managing and preserving large amounts of research data were rarely addressed, with most researchers acquiring the skills they needed via trial and error. None of the scholars interviewed during this study expressed satisfaction with their level of expertise in data management, and few had access to individuals who could provide knowledgeable guidance. While it would be preferable for graduate curricula to include this type of training, it seems unlikely that this change will occur in the near term. It is, therefore, likely that archivists and data specialists will need to provide this training, preferably by developing relationships with ethnographic researchers as early as possible in the research process.

The lack of confidence in their data-management training compounds researchers’ apprehension about the ethical issues contained in their data. Researchers are naturally—and rightfully—conservative about the potential risks in sharing data, especially if they are not convinced they can protect their respondents’ confidentiality, or that they have obtained appropriate consent—issues that the difficulties of cross-cultural communication make more complex. Without such assurances that these ethical imperatives can be addressed, many researchers are unlikely to
invest in archiving their materials. “In many cases, their desire to avoid the ethical risks of inappropriate data release may outweigh the costs of potential data loss.”

Because of these interrelated concerns, it is critical that the archival and data curation systems available to ethnographic researchers have multiple and fine-grained levels of access and privacy controls. Tools must be developed that can appropriately manage confidential data and that allow researchers, working in conjunction with their respondents, to exercise control over their data, administer who has access to it, and account for culturally specific restrictions on data use. One example of a system attempting to address these requirements is the Murkutu content management system (http://www.mukurtu.org/), which makes cultural protocols a fundamental part of access to materials. This allows for the easy creation of fine-grained and culturally appropriate user groups with specifically defined usage rights, and it is being developed in cooperation with indigenous communities.

Because the nature of ethnographic materials makes them potentially subject to ongoing and evolving dilemmas of confidentiality and consent, archivists and data managers should understand that these materials may require ongoing curation, preferably through processes and procedures that incorporate researchers in a collaborative and meaningful way. Although many researchers would welcome the ability to more easily share their data, particularly in collaborative projects and in response to data-sharing requirements, they are almost universally resistant to any arrangement in which they would relinquish control over access to the data. As one anthropologist described:

People don’t hesitate, at all, to share data with collaborators that they trust. … If you provide a mechanism for collaboration, even if it’s just Google Docs or something, you know, people share data easily and freely. It’s when it becomes an anonymous process that they seem to get balky. (1-02-100511)

The goal of making materials publically available online, then, will require archival systems and processes that are cognizant of the complex social relationships embedded within ethnographic materials, not only between the researcher and his or her respondents, but also between and among the many stakeholders in ethnographic data. The development of additional policies, ethical guidelines, and best practices by disciplinary organizations, funding agencies, universities, and cultural institutions could further aid researchers and archivists in navigating the ethical dilemmas contained in the archival process.

Because much ethnographic data is collected by individuals in relatively small and often idiosyncratic datasets, it is especially vulnerable to loss due to benign neglect or lack or resources, even if the data are historically unique. A sociologist described one such vulnerable dataset:

We did life histories of 45 people in a [church] congregation in [central Pennsylvania town]. . . . It’s probably the only body of material on a single congregation of that depth and complexity that exists, and I’m just too lazy to have figured out what to do with it and write it up and everything, but you could imagine somewhere somebody who wants to do research on congregations would
find this incredibly valuable. But what’s going to happen to it [in the long term]? And what’s going to happen to it is, the copies I have will end up in a trash can, and my computer hard drive will eventually crash or get outmoded or get erased, and that’ll be the end of it [if there isn’t a place to archive it]. (2-13-111411)

In order to prevent these losses, data management specialists and archivists must specifically curate these materials. They must also build relationships with ethnographers to resolve the unique ethical dilemmas these data present and the uneasy tensions between researchers’ desire to share data publically and obligations to protect the privacy and confidentiality of the individuals and communities they study.

Finally, despite the acknowledged importance and value of archival materials to the disciplines and future researchers, academic rewards systems offer very little payoff for the work that goes into making ethnographic materials available, and ethnographic researchers are often under-resourced in the funding, time, and personnel support needed for preparing materials for archiving. Archivists must therefore work to make archival systems and the development of archival processes and policies as straightforward for researchers as possible, ideally assisting in integrating the development of metadata and other secondary materials (for example, codebooks, explanatory materials, finding aids, and ontologies) necessary to archiving and ensuring the long-term usefulness of primary data as early in the research process as possible. Likewise, funding agencies should improve funding for long-term data management, preservation, and curatorial services, either directly or via guidelines to institutions for recouping indirect costs, especially when the grant mandates archiving or public data accessibility. By providing these supports, archivists and funders would greatly assist researchers in responding to the imperatives of providing access to data and assuring the confidentiality and consent of their respondents, as well as resolving the ethical dilemmas contained in their data, thereby helping assure the preservation of irreplaceable ethnographic information.

3 This study was funded by the Alfred P. Sloan Foundation and managed by the Council on Library and Information Resources (CLIR). See Jahnke, Asher, and Keralis, The Problem of Data.
4 Penn State University, Lehigh University, Bucknell University, Johns Hopkins University, and University of Pennsylvania.
7 The number in parentheses is the participant identification number; please refer to Table 1.

This version was slightly revised in 2009 to make minor changes to some of the statement’s language. Since this revision did not alter the structure or main points of the statement, we continue to refer to the 1998 version (as does the AAA’s ethics website). American Anthropological Association, Code of Ethics of the American Anthropological Association, Statement, 1998, B.4-5, C1.

American Sociological Association, Code of Ethics, 1999, 13.05


Ibid., 3.


Fabian, “Magic,” 123

Fabian, *Ethnography*, 54. One does wonder, however, why Fabian considered a traditionally published monograph necessary to this project, rather than simply including it in the archive. Given that his book annotates the online texts using paragraph numbers that assume a specific online layout, it would seem more effective to have created hyperlinks between the commentary and the primary materials. In this way, the “experiment” seems incomplete.


Ibid., 23


Parry and Mauthner, “Whose Data,” 141.

Fabian, *Ethnography*, 16-17.


Fabian, *Ethnography*, 16-17

Annamaria Carusi and Marina Jirotka, “From Data Archive to Ethical Labyrinth,” *Qualitative Research* 9, no. 3 (2009): 292.
32 Parezo, “Preserving Anthropology’s Heritage.”
33 CoPAR has released a series of fourteen bulletins outlining specific issues involved in the preservation of anthropological materials and addressed to practicing anthropologists. These bulletins are available at http://copar.org/bulletins.htm. Unfortunately, since most of these bulletins were written in the late 1990s, many of the technology discussions are out of date. However, the conceptual discussions remain highly relevant. 
38 Jackson, “I Am a Fieldnote,” 10-12.
39 Ibid., 22.
41 Zeitlyn, “Anthropology in and of the Archives,” 471.
44 Bishop, “Ethical Sharing,” 263.
46 Zeitlyn, “Anthropology in and of the Archives,” 471; see also Parry and Mauthner, “Whose Data,” 147.
47 Bishop, “Ethical Sharing,” 263.
49 The Society of American Archivists Oral History Section maintains a webpage summarizing the case and providing links to relevant documents at http://www2.archivists.org/groups/oral-history-section/the-belfast-case-information-for-saa-members.
50 Zeitlyn, “Anthropology in and of the Archives,” 471.
See Carusi and Jirotka, “From Data Archive,” 294-95.

See Corti, Day, and Backhouse, "Confidentiality and Informed Consent," 4.4, on the expense of creating anonymized archives.

ICPSR’s “Guide to Social Science Data Preparation and Archiving Phase 5: Preparing Data for Sharing” provides a model for these processes. See http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter5.html.


Ibid., 17.

Ibid., 17-18.

A list of community groups and organizations participating in the Mukurtu project is available at http://www.mukurtu.org/community.html.

The NSF’s recent decision to list datasets as one of the research products to be included in the section of its grant proposals where researchers describe their qualifications does suggest, however, that the current rewards system may gradually be changing (see http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_sigchanges.jsp).