Born Archival: The Ebb and Flow of Digital Documents from the Field

Mark Turin

Facilitated by an infusion of funding from philanthropic sources, descriptive linguists have been galvanized to document the world's languages before they disappear without record. Linguists have responded to the "crisis of documentation" (Dobrin, L. M. & Berson, J. (2011), "Speakers and Language Documentation", in The Cambridge Handbook of Endangered Languages, P. K. Austin & J. Sallabank (eds), Cambridge University Press, Cambridge, pp. 187–211) by entering into increasingly collaborative partnerships with speech communities, producing "documents" that have both local relevance and academic integrity. The growth in access to digital recording technology has meant that contemporary research initiatives on endangered languages are not only born digital, but often birthed straight into an archive. Yet heritage collections of recordings made by ethnographers and linguists in the past are ever more endangered, becoming orphaned when their collectors die or fragmented into their component parts based on the medium of documentation when they are finally archived. Drawing on fieldwork in Nepal with a community speaking an endangered Tibeto–Burman language, and reflecting on the decade I have spent directing a digital humanities research initiative—the Digital Himalaya Project—I discuss how linguists and anthropologists are collecting, protecting and connecting their data, and how technology influences their relationship to documents.

Keywords: Archives; Endangered Languages; Documents; Digitization; Himalayas

Correspondence to: Mark Turin, Museum of Archaeology and Anthropology, University of Cambridge, Downing Street, Cambridge, CB2 3DZ, UK; South Asian Studies Council, MacMillan Center for International and Area Studies, Yale University, Henry R. Luce Hall, 34 Hillhouse Avenue, New Haven, CT 06520-8206, USA. Email: mt10003@cam.ac.uk

Introduction

This contribution has two interlinking parts, tied together by a prolonged relationship with a region: the Himalayas. The stage is set in the first section of the article, onto which two performances are then projected. To begin, I discuss the macro-politics and competing ethno-social agendas of contemporary "documentation" projects in linguistic anthropology. I then turn to the micro-politics of a particular field research project and my involvement in the production of documentary materials. Since digital documents are the primary products of contemporary research initiatives in field linguistics, the process of documentation naturally warrants closer attention. Finally, I reflect on the political aspects of an archive of digital documents that I co-established in 2000 and still direct—the Digital Himalaya Project—located at the universities of Cambridge and Yale. Through these various engagements, I address a recurring theme of this special issue, namely that "files are authoritative by virtue of their compilation" (Feldman 2008: 35) and consider what this means for documents that are not only born digital, but increasingly even born archival.

Data Production Practices in Field Linguistics

The recently rejuvenated sub-discipline of descriptive field linguistics is addicted to documents. Scholars generate them, funders demand them and members of some speech communities fetishize and deploy them for transparent political ends. The outputs of research projects that aim to describe hitherto poorly known and often endangered languages are increasingly measured by the volume of the documents that they produce. Quality is not everything, quantity also counts, as demonstrated by a strong focus on data curation and future-proofing strategies in online language archives, such as the DoBeS archive in Nijmegen[1] and the Endangered Languages Archive maintained at the School of Oriental and African Studies in London.[2]

That linguistic documentation projects should be so explicitly absorbed with the production and distribution of documents is worthy of study in itself, but of particular interest to this issue is the degree to which these modern languages collectors pause to reflect on the political implications of their work, and the extent to which they are being asked to consider the social and community consequences of their documentation. In this section, I argue that field linguists have considerable experience of dealing with the tension between analogue, tangible documents (printed word lists and grammatical sketches) versus those whose genesis and distribution strategy is largely digital and intangible (database records, digital entries into online archives or logged audio-video content).

Megill and Schantz rather provocatively suggested that a "document becomes a verb for the archivist" (1999: 21). Six years later, in his *Corporate Memory: Records and Information Management in the Knowledge Age*, Megill fine-tuned his earlier position with an entire chapter on "The Document as Verb", proposing that while the International Organization for Standardization defines a document as an "object", "in the Knowledge Age, a document is more often a verb than a noun" (2005: 33). I

would extend Megill's proposition further and suggest that in my work, and also for other ethnographers, field linguists and museum curators, "document" has always been both a verb and a noun. While an explicit focus on the agency of the fieldwork endeavour may be a more recent concern, field linguistics has been motivated by the calling of primary documentation since its origins. Researchers have collected data on the diversity of human linguistic expressions, with the field linguist positioned as both "documenter" of a specific language and subsequently producing "documents" (as nouns and objects) that are quantifiable products. Present-day linguists are being asked for more information on the pragmatics and ethics of their collection methods at the point of archival accession and deposit, and students are being trained to document (verb) their documents (noun) by providing data about their data (meta-data).

In this context, Lise M. Dobrin and Josh Berson write of the "crisis of documentation" (2011: 205) affecting field linguistics. The nod to anthropology's great moment of introspection, popularly framed as "the crisis of representation", is unambiguous. Language documentation, they argue, is an increasingly "social activity" (2011: 197), with documentary linguists going to great lengths to "establish more equitable power relations with speakers through the use of participatory, community-based research protocols" (2011: 207). This engagement is surely to be celebrated, but it has not been an equally comfortable transition for all. The recognition of endangered language speakers "as persons, as opposed to mere sources of data" (2011: 189) combined with the dawning realization that fieldworkers often inadvertently document an idiolect (a variety of a language unique to an individual) or at best a sociolect (spoken by only one socio-economic segment of the community), rather than anything that can be called a language as a whole, has generated a raft of new questions that linguists need to address.

Research projects in field linguistics are increasingly framed as cooperative, community-based, participatory collaborations for social good—at least to funders and in public declarations—a kind of social movement "that has brought academic linguists out of their offices and libraries and into a shared space" (Dobrin & Berson 2011: 187). In the minds of many contemporary linguists working on endangered languages, both the creation of linguistic documents and the exercise of documentation itself have explicit activist components. While states may continue to discriminate against minority languages and their speakers through punitive legislation and enforced structural invisibility, some linguists have taken on the roles of advocate, supporter and—in all senses—"documenter" of speech communities. Dobrin and Berson note that the "thematicization of collaboration" (2011: 203) that has emerged as a central methodological issue in documentary linguistics knows no precedent in the discipline.

Linguistics, then, has made the transition from documentation as salvage or rescue work—which set out to reclaim from the debris of modernity the last vestiges of indigenous linguistic purity—to viewing documentation as a participatory and even community-led process, through which differently prioritized and variously weighted "documents" are produced for diverse groups of stakeholders. To be clear, linguists have not become political activists or community mobilizers overnight, but have

rather acknowledged that the same data can be retooled and retasked in different documentary packages that can satisfy both their need for career advancement and publication, as well as the requirements of funding agencies *and* the speech communities whose languages they have had the privilege of researching. Far from shying away from potential confrontations, or viewing this reconfigured research landscape as a systemic challenge to their knowledge system, many linguists have embraced the new idiom of collaboration with speakers of minority languages as an exciting opportunity.

But as anthropologists have discovered through many decades of fraught and contested engagements, at the heart of this "new" linguistics lies a growing awareness that the "power imbalance in the documentary encounter . . . is at odds with the motivations for conducting the research in the first place" (Dobrin & Berson 2011: 189). Two dilemmas lie at the heart of these new linguistic partnerships. First, the act of creating and disseminating traditional linguistic objects and documents—such as grammars, texts, dictionaries and corpora (real-world examples of natural spoken language)—may unwittingly reproduce the very same "suspect power hierarchy that linguistics-in-recognition-of-indigenous-rights" so proudly set out to dismantle (Dobrin & Berson 2011: 202). Second, and even more troubling for the heuristically minded language fieldworker, is that communities—like their languages—are rarely as bounded as first hoped. Even though linguists have introduced into their descriptive and scholarly lexicon a range of terminologies to add nuance to the continuum of articulations that lie along the accent-dialect-language spectrum, few have been equally reflexive about the porosity and fluidity of ethnic boundaries. Linguistic monographs are still built on meaty chapters devoted to complex verbal morphology and clause structure, with discussions of the cultural history of a speech community usually relegated to a slim section entitled "the People". What happens when these speakers turn out to have as many political positions as they have dialects and voices, with various groups laying claim to different agendas and aspirations? And then, what if the subjects (no longer objects) of study—still commonly referred to by linguists as "informants" or the altogether more World Bankish term "consultants"—willingly participate and collude in their own objectification and "documentization" for their own ethno-political ends?[3] Linguists are noticing that collaboration works both ways—they are being "collaborated with" as well as "collaborating with"—and the documents of their research, as well as the documenters themselves, are being harnessed by speech communities in creative and often unexpected ways. As a case in point, in the section below, I reflect on my own work on Thangmi, an unwritten Tibeto–Burman language spoken by around 30,000 people in Northern-central Nepal and in the Darjeeling district of the state of West Bengal in India.[4]

Thangmi: Activists in Search of a Linguist

After almost a decade of research on the Thangmi language, I finally produced a document in 2006: a grammar of two dialects of the language, incorporating a number of shamanic oral texts, some ethnography and a trilingual lexicon (Turin 2011). The

manuscript—while much anticipated by some in the Thangmi community (perhaps because it was so long overdue)—was not equally well received by all of the speakers with whom I had worked. The 1990s and early years of the twenty-first century were a period of massive political and social upheaval for Nepal, with a violent civil war and a level of instability that the country had not seen in its recent history. Through this period of unrest, which coincided with my research, many members of the Thangmi community—one of Nepal's most traditionally marginalized and economically impoverished peoples—were beginning to assert themselves in a loud chorus proclaiming their ethnic pride. They claimed autochthony in a traditional homeland and a unique language, and activists positioned the group as deserving attention from the national administration. While some of the Thangmi community were still interested in the more abstract idea—and then the product—of a descriptive grammar of their language, others were beginning to ask what it was for, who owned it, why it was in English and how it was going to help them.

As a partial and anticipatory response to these substantive questions, two years before my dissertation was completed, I had already compiled a *Nepali–Thangmi– English Dictionary* together with my long-time Thangmi research assistant (Turin & Thami, 2004). Published in Kathmandu with a printing subvention from the British Embassy in Nepal to make it more affordable, I had somewhat naively thought that this trilingual lexical booklet would forestall some of the criticism. In fact, the production of our dictionary simply provided a timely spark to an outpouring of local lexicography that accompanied a critique of my endeavour. Our perfunctory word list served to catalyse two further dictionaries in response, both solely compiled by native speakers without foreign intervention or funding, and both also larger, heavier and more complete as documentary products. Rather like an archetypal fairly tale, the dictionaries just kept on getting bigger in line with the dream of a comprehensive, "complete" Thangmi dictionary. Our humble undertaking had been just an early appetizer. The gold standard and ultimate documents for comparison were large, heavy monolingual Nepali dictionaries; and indigenous lexicographers were working towards such a monograph by indigenizing Nepali words and including every possible verbal conjugation in their lists to bolster the number of pages and thus engorge the lexicon. After centuries of orthographical invisibility, dictionaries—as political documents—were becoming a new unit of value, compiled and deployed in a competitive display of local lexicography to garner the favour and attention of a newly inclusive state that was taking stock of its linguistic minorities and considering offering them tangible benefits calibrated to their perceived level of indigeneity.

In comparing the contents of the two documents that I had produced, a question had surfaced among my Thangmi interlocutors: how could the massive differential between the size of my dissertation (900 pages) and the size of the tiny collaborative dictionary (116 pages) be explained? What was in the English (foreign) book that was not in the Nepali (local) one? What was I leaving out? Was the community being short-changed? For some Thangmi speakers, my dissertation had achieved a positive symbolic status (large heavy book), but had no practical role (impenetrable

linguistic annotation and in English), while the shorter collaborative dictionary had a noted practical effect (accessible and affordable), but was lacking in symbolic impact (on account of being locally published and so small). Thangmi language activists wanted both practical impact and symbolic capital for their documents—and who could blame them—and regarded my endeavour as a useful proof of concept that could be improved upon and developed further. As Riles has noted, documents can change "social and material form" (2006: 14) as they move from one setting (a dissertation to be evaluated within a university) to another (a resource for a community of practice). The journey of my monograph may be said to have been even more fundamental, changing not only in form, but also in meaning: from distilling a decade of research, to effectively distancing a community from their speech form, in its embodiment as an objectified and impenetrable "document" about their language. When creating my linguistic documents, I felt that I had not anticipated how they would be received, circulated, instrumentalized and taken apart (Riles 2006: 18), even though I felt that I had worked collaboratively and consulted community members at each step.

Without dwelling on the process of indigenous lexicography, there are two issues worth drawing out. First, as anthropologists know full well and as linguists are just beginning to discover, partnerships and collaborations are always contested. The simple refrain of "giving back to the community", to be found in so many contemporary grant proposals, is more problematic than many imagine. Documents are powerful and potent, and as the editors of this issue have shown in their introduction, the authority and authorship of documents is seldom singular. Second, I suggest that in these contestations lie interesting research questions that only begin to emerge when the researcher engages with community demands. This process of negotiation and arbitration can be enormously intellectually fulfilling, as it helps to challenge and transcend mechanical discussions of cultural repatriation to arrive at a richer and more finely graded space where deeper trust is established through the production and comparison of documents. In all of this, we must not lose sight of the verb–noun correspondence: I set out to "document" a language, and in the process created a number of different "documents" whose interpretations I could no longer control.

Through all of this, the Thangmi community—and I as their partial agent and sometime linguistic advocate—were journeying on the well-trodden path from orality to literacy, leveraging their verbal expressions into written form through documents: sometimes digital, mostly physical and always political. There is no space to dwell on the wider issues of orality and textuality here (see Ong 1982; Pollock 2006; Finnegan 2008; and Gaenszle 2010) for helpful discussions), and the sometimes surprising directionality of the process in Nepal (from textual to oral, as well as the other way around), but rather we may reflect on the question of how "one can write about a culture that has already been written by its native spokesmen" (Burghart 1996: 17). How do partnerships actually work, and how are technologies changing, both the experience of documentation and the document itself? The apparent paradox of globalization is located in that final question: the very processes that are portrayed as eroding cultural and linguistic diversity are at once bringing individuals

into closer contact with one another and providing affordable and appropriate tools to document these expressions in non-linear form. As John Miles Foley has noted, oral tradition and the internet can be seen as "homologous technologies of communication",[5] a convergence that is not lost on members of the Thangmi community who perceive new digital recording technologies as opening up not only documentary, but also representational, space for their own expressions which had previously been precluded by the long dominance of textual modes of production and dissemination.

In summary, then, linguists have transitioned from extractive models of research (parachuting into fieldwork zones, collecting words, analysing the data and publishing in the West) to methods that are more engaged, appropriate and at times, even reflexive. In so doing, they have also embarked on a search for new audiences, including speech communities themselves, and have embraced web-based technologies to disseminate their research findings digitally, with online data stores and corpora being included in cases for tenure. One such online ethno-linguistic repository, the Digital Himalaya Project, is the focus of the final part of this paper.

Collect, Protect, Connect: Digital Documents from the High Himalayas

I have become intrigued by the steady repositioning of (particularly Western) institutions as custodians of collections held in trust, rather than as the owners of culture and history, and the concomitant democratization of archival space. Online repositories are becoming adept in what Greg Rawlings refers to in this issue (in a different context) as the "aesthetics of the subtle", negotiating the competing interests of preservation and access that can pull in opposite directions. Archives have been successfully rebranded: they are no longer musty places where documents go to die, but sites of interaction and energy, connection and outreach. As a case in point, we may note that the British National Archives, which now hosts 1,000,000,000 documents online, used the phrase "Bringing history to life through UK government records" as its strapline for much of 2010.[6]

This shift in orientation was just beginning in 2000, when together with the historical anthropologist Alan Macfarlane, I established the Digital Himalaya Project (http://www.digitalhimalaya.org) to develop digital collection, storage and distribution strategies for multimedia anthropological information from the Himalayan region. The plan was simple enough and the timing appeared to be right: many archival ethnographic materials, such as 16 mm films, still photographs, videos, sound recordings, field notes, maps and rare journals were fast degenerating in their current formats and were in urgent need of rehabilitation. As anthropologists who worked in Nepal, it was logical that we focus our attention on the Himalayas and the unique archives and impressive collections to which we had access in Cambridge.

As we established the project, we noticed an incongruity. Even though anthropologists were becoming ever more concerned about cultural endangerment and the far-reaching impact of rapid socio-economic change, and funding was available for scholars to document indigenous cultures and languages that were at risk of disappearing without record, very few social scientists (with notable exceptions) were working to

ensure that ethnographic collections from previous generations were maintained, refreshed and made accessible, both to the research community and to the descendants of the people from whom the materials were collected. To this end, with a grant from the Anthropologists Fund for Urgent Anthropological Research,[7] the Digital Himalaya Project was established as a strategy for digitally archiving, digitizing and disseminating legacy ethnographic materials from and about the Himalayan region.

Alongside the preservation aspect mentioned above, we had two other primary aims: to make our digital resources available over broadband Internet connections for researchers and students, and to return copies to source communities in the countries of origin—such as Nepal, Bhutan, the Tibetan Autonomous Region of China and the Himalayan regions of India. Echoing Noel Lobley's discussion of Hugh Tracey's African recordings in this issue, we were already engaged in the complex web of ethical and practical questions surrounding the circulation of archival content and historical documents to the communities and contexts in which they were recorded. When setting our project in motion, we had naively imagined that users in the West would be downloading the content from the Internet while "the Rest" would be accessing the collections on DVDs and CD-ROMs. This paper is written as a partial reflection on that innocence.

While each cultural object or artefact has its own life history, a unique trajectory through time and space, the narratives of individual documents have a tendency to be flattened by the anonymity of large archives even as they spill out of its sides. In some cases, the sheer number of analogue or digital documents can reduce a unique image, audio file or artefact to one in a series, a spec of unaffiliated data. Similarly, the individuals involved in a document's life may be left by the wayside when a document becomes assimilated into an archive: subject, collector/artist and interpreter collapsed into notes in the margin. But at the same time, a well-structured and multimodal archive can help to reveal a document in relation to its peers, involving new actors in its future life as an artefact of study: researcher, commentator, community member and public.

Ethics and Responsibilities, and the Limitations of Informed Consent

Whether online or on DVD, issues of confidentiality and consent remain central to the construction of a digital archive. Anthropologists need no reminding of the need to obtain consent from their informants—whether Wall Street businessmen or Tibetan nomads—especially when representing them in writing. But the potential pitfalls are even more acute when dealing with the immediacy and lack of anonymity inherent in visual representations, or, in the case of much of the material that made its way online through Digital Himalaya, when the images were taken generations ago when the mass distribution of visual content as we now experience it was inconceivable.

Even though early anthropologists may have believed that the people they filmed and photographed consented to these activities (there is no space here to discuss coercion and the unequal power dynamics of the anthropological endeavour itself), the

advent of the digital age fractures the very basis of that "informed consent". When Christoph von Fürer-Haimendorf started his fieldwork in Nepal in the 1950s, the country had just opened to the outside world for the first time in a century. How could Haimendorf's informants have consented to having their images broadcast over broadband Internet fifty years later? How could they have anticipated that the words they uttered (gossip about their neighbours, political criticism of the monarchy) might be available to millions of faceless viewers across the world? Although many of the individuals represented in Haimendorf's films have long since died, what happens when their descendants search through a digital archive and happen upon their grandparents taking part in some compromising activity or making statements still embarrassing to the family? How can anyone know how their image will be manipulated in a week's time, let alone over the next ten years? As we were beginning to see through our work on the Digital Himalaya Project, old film does not die, it simply gets clipped into smaller pieces and Tweeted, further removed from the moment of its original registration, and used for ever-more unexpected purposes.[8] These were some of the questions that provided the intellectual frame for our project.

It was also clear that archives become more complex when the "documents" in question are representations of human "subjects", as was the case for the ethnographic archives in which we were interested, including photographs, films, sound recordings and field notes on people's lives, their cultures and their practices. The challenge we faced was how to structure, build and then maintain a public digital archive that contained hundreds of hours of film and thousands of photographs, representing numerous individuals spanning a 100-year time period in a part of the world where digital media was just beginning to take hold in 2000. How could we ensure that the representations of long deceased people who we were about to make available online were ethical and appropriate? How might we integrate into an online portal the often-conflicting intentions of all the individuals involved in the life of an ethnographic document—subject, collector and archivist, as well as the descendants of the subject and collector—and at the same time, honour the culturally embedded history of each document and catalogue it in a consistent manner? Riles describes documents as "paradigmatic artifacts of modern knowledge practices" (2006: 2), but I wonder how strictly modern they are. What of the palm leaf manuscripts produced, distributed and consumed across large tracts of Asia as far back as the fifteenth century BC—examples of early mobile archives—that recorded historical, as well as mythical occurrences in the ancient world? Perhaps the documents that Digital Himalaya was hosting could be thought of as "cultural texts" (Riles 2006: 12), in the sense of their cultural embodiment in a physical landscape where they had been produced and into which they were being re-introduced, as well as the cultural context provided by our disciplinary frame of reference.

Multimodal Documents and the Case for the Web

Archivists specializing in the curation of moving images have used the phrase "nitrate won't wait" to describe the urgency of migrating silver nitrate film to more durable digital formats. Not only are anthropological collections dating from the early

twentieth century fast degrading, but they are also becoming orphaned, as the technology needed to view them is obsolete and the people needed to make sense of them and contextualize the contents are even harder to find. While it has "long been a central canon of museum philosophy that fossilized collections die" (Mayhew 2006: 87), there has not been an equal recognition that many heritage collections of recordings made by fieldworkers in previous generations are now critically endangered, and that they risk becoming unusable when their collectors die or when they are fragmented into their component parts based on recording medium. The politically motivated destruction of documents is touched upon at points throughout this special issue, but what concerned us more in the context of the Digital Himalaya Project was something more pervasive and apparently more passive: the wholesale destruction of documents through neglect and inaction. While a physical document might endure—a single negative or snippet of audio recording—it would be situationally barren and devoid of ethnographic resonance without context and frame. Moore's law on the exponential rate of technological change in computing provided a powerful if brutal lesson in impermanence and non-attachment to external form: it remains entirely possible to read a book that is 500 years old (as many scholars of classical languages and cultures regularly do), but it is close to impossible to find a computer anywhere within the University of Cambridge that can read an "old" 8 or 5/14 inch floppy disc dating back to the 1980s. The pace of innovation and thus of obsolescence was moving ever faster, and few fieldworkers had paused to reflect on issues such as the longevity and persistence of their recordings before they embarked on fieldwork. The curious challenge of online collections was that they appeared to exist everywhere and nowhere, at once infinitely replicable (and thus seemingly safe) yet tragically fragile if not backed up or regularly migrated.

While "audio-visual" had become a technology buzzword in the 1990s, ethnographic fieldwork had been "multimedia" or "multimodal" for about 100 years, with early anthropologists using still cameras, wax or plastic cylinder record phonographs alongside pen and paper to document their experiences. When these scholars returned home, though, they were expected to write books in which precious little of the material that they had recorded could be accommodated. And when anthropologists retired, and later passed away, their collections of recordings and photographs might be left in shoeboxes in their attic, only to be donated to university libraries and archives that often did not really want them and very rarely had time to catalogue them, let alone use them for teaching and research.

So while fieldwork was inherently immersive, and had long conscripted all manner of equipment as documentary apparatus, an anthropologist's holistic collection risked being split apart when the researcher returned home, with its location to be determined by the format of the recording medium: text to the library; sound to the audio archives; photographs to the image collections; and cine film regularly left stranded and unclaimed as it straddled categories and was hard to curate. We saw the fast-developing web as a natural site for these diverse materials to be recombined and integrated, served up in a searchable and retrievable multimedia format, recreating—perhaps for the first time—something of the holistic fieldwork experience,

for both the observer and the observed. Returning to Riles, who reflects on Heimer's assumption that "documents strip away context" (2006: 9), we had found that the online archive provided a way to enrich and recontextualize, rediscovering context that in non-digital documents was more difficult. Through our digital repository of ethnographic recordings, the project team were becoming actors in the rehabilitation of often inaccessible and uncatalogued "documents", and agents in their subsequent dissemination to a wider public. With some users perceiving us to be digital publishers, we were discovering that through the process of digitization, archival "objects" such as old film were being transformed into "documents", taking on levels of authenticity and orthodoxy which they may not have had in their analogue and less downloadable forms.

The two most extensive collections that we initially digitized were 100 hours of 16 mm film from various parts of the central and Eastern Himalayas, filmed between 1936 and 1980 by Christoph von Fürer-Haimendorf, Professor of Asian Anthropology at the School of Oriental and African Studies in London and one of the greatest ethnographers of his time, supplemented by his detailed field diaries and photographs;[9] and the Williamson Photographic Archive: 1,700 photographs taken between 1930 and 1935 by Frederick Williamson, British Political Officer, in Tibet, Sikkim and Bhutan in addition to twenty-three reels of his little known 16 mm film. Williamson's collection was held in the Museum of Archaeology and Anthropology at the University of Cambridge, and included a number of rare historic images.

For budgetary reasons, we digitized Fürer-Haimendorf's films ourselves, projecting the footage and then filming the output through a box of mirrors known as a film chain adaptor, and then hosting compressed video clips on our website. These snippets caught the attention of the British Universities Film and Video Council who then funded the professional digitization of most of the footage, using telecine projection. From this, we learned another lesson: digitization was a continuous and ongoing process, not a one off; and we began to think of digitizing a subset first before committing to undertake the digitization of an entire collection, for reasons of time and cost.

Through the Fürer-Haimendorf and Williamson film collections, unexpected collaborations began to take place. The custodians of such collections back in the UK often had only limited knowledge about the footage that they held, based on a few quickly scribbled notes on a film canister or from an ancient accession form. Back in the Himalayas, however, descendants of the individuals who featured in these films could provide a great deal of additional information, and their insights added enormous value to the collections. Returning to source communities with DVDs and hard discs, then, became far more than a routine process of cultural return in digital form, but rather an exciting opportunity for partnership by which collections were enriched and better understood. As part of the Digital Himalaya project, we travelled to Sikkim, Bhutan and Tibet with digitized footage, providing copies to institutes, universities and colleges in the region, as well as to the descendants of the people we could identify from the recordings. Through the project, digital documents

were being distributed to communities who had a stake in the content, who then reincorporated these materials into their deepening understanding of their own past. As noted by Robson, Treadwell and Gosden, historical documents such as these are actually living "cultural items" of considerable importance to local people, and we were rapidly discovering that "artefacts from the distant past are key to cultivating a sense of self and community in the present" (2006: xv).

Through such interactions and partnerships—online and in the field—we were also becoming aware of the surprising valence of different kinds of documents. On one occasion, what appeared to us to be an extremely sensitive recording taken by Williamson of the Tibetan Army involved in a training exercise in 1932, proudly flying the flag of an independent Tibet, did not have the power that those of us managing the archive had imagined. To our surprise, access to this footage is still not restricted from within the Great Chinese Firewall, even though the authorities were perfectly aware of the project, and users across the Tibetan Autonomous Region have continued to have untrammelled access to view and download our films, while the BBC website remains blocked to all web users in China. How and why does the passage of history depoliticize some documents, but repoliticize others?

Online Host or Digital Publisher? JSTOR for the Himalayas

By 2002, Digital Himalaya had tested the water and established that an appetite existed for online collections of cultural content, and we were in a position to develop the project further. As scholars, we regularly used digitized academic journals available through services such as JSTOR,[10] but were surprised to discover that no publications originating in Himalayan countries could be found through such online repositories, severely restricting the access, impact and visibility of the research that these publications contained. With the agreement of editorial boards and publishers, we started sourcing and scanning back issues of a large number of journals, magazines and imprints of Himalayan studies from Nepal, Bhutan, India and Tibet.

After several years of scanning, and thanks to creative partnerships at the universities of Cornell, Virginia and Alabama, we now host back issues of almost forty important publications for free download from our website, comprising well over 200,000 pages of digitized text.[11] The aim of this journal repository was to stimulate sales and subscriptions by digitizing and hosting back issues of publications (at no cost to the publisher), some of which were even out of print, and thus to provide a web presence for collections that might otherwise not have made it online. This PDF archive of journals and magazines turned out to be surprisingly popular, especially within the Himalayan region where Internet access was widening, but access to good libraries and full collections of printed matter remained poor.

While a satisfied user community, illustrated by over 300 visits a day to the Digital Himalaya website, would be gratifying enough, our journal digitization programme also raised interesting intellectual and ethical issues. Did we have the right, without approaching individual authors, to digitize and host their work, albeit for free, with no registration requirements? Having secured agreements from the publishers and

editorial boards, none of whom—it should be noted—had ever required their contributors to sign a copyright release or waiver, we trod very gingerly to begin with, hosting one issue at a time and patiently waiting for the backlash. Once again, we were surprised at the supportive response of the scholarly community and authors in particular, many of whom wrote to us to express their delight at finding their own articles published 30 years ago freely available online; some had even misplaced their original copies.

But perhaps we should not have been so surprised. After all, one of the world's most august institutions, and one not prone to skittish infringements of intellectual property or copyright—the British Library—had already opened the envelope far wider than expected with a progressive "Notice and Takedown Policy" laid out on its "Terms of Use" webpage. By hosting a wide range of online content, including archival sound recordings, the British Library believed that it was acting in the public interest. By default, their access tap had been set at open and full stream, and could be slowly tightened or closed down, but only with due cause:

> If you are a rights holder and are concerned that you have found material on our website, for which you have not given permission, or is not covered by a limitation or exception in national law, please contact us in writing ... [British Library Notice and Takedown policy, http://www.bl.uk/aboutus/terms/index.html#notice]

As John Willinsky has convincingly shown in his book *The Access Principle*, the key to copyright is the right of authors to profit from their work (2005: 49). Journal publishers have not made their editors or authors financial partners in the process, in the way that book publishers have, with the result that individual authors do not lose out with journal digitization, but rather reach new audiences and readerships for free. The infrequent requests that we do receive from authors to take down PDFs that we host either relate to the publisher releasing an incorrect version of a file to us for online hosting, or the author wanting to amend the text of their original document. While in such cases, we encourage authors to contact the journal editors and ask for a correction to be printed in the next issue, their approach to us reflects a revealing conflation of the role of publisher, printer and online host. In another instance, we were approached by a scholar asking that we not digitize the next issue of a certain journal as the wider dissemination of its contents would, he believed, be against the public interest. Fortunately, we had no plans to digitize this issue so we did not have to decide whether to rescind a file or restrict access in any way. Through these requests, it became clear that we were involved in what Riles refers to as "document making" (2006: 18), the production and distribution of documents in new arenas. While an article's first incarnation may have been in a print journal, bound to its physical medium and constrained in distributive potential, the Digital Himalaya website was offering the potential of digital rebirth and association with thousands of other similar documents.

While the hierarchically flat nature of our website is in some ways a great information equalizer, as an unpublished Master's dissertation submitted by a student for online hosting occupies the same nodal level as a peer-reviewed article by a

senior professor, the digital circulation of such documents has implications for power relations between authors and their readers. Through tools such as Google Analytics and other web tracking services, we can now log the number of downloads of each digital document in any given country, and in some cases, within a particular institution or region. Once downloaded from our website, however, the secondary distribution of digital objects is harder to track, and we know that PDF scans that we produced are circulating freely and being co-hosted around the world, in line with the Creative Commons license that governs the use of all of our web resources.[12]

It has also become considerably easier for a disgruntled author to effect change by asking for the removal of a single file from a digital archive than it was to seek and destroy every offending copy of a published journal. As Digital Himalaya became perceived as a *de facto* online publisher, I return to Feldman's suggestion that files can become "authoritative by virtue of their compilation" (2008: 35). Our project, through the documents that we hosted, had developed considerable if unexpected power. The process of scanning, digitizing and hosting functioned as a kind of rehabilitation, through which documents were becoming more true, more pervasive and more assertive.

Conclusion

If we can reclaim analogue recordings and field documents from the fate of obsolescence by assuring that they are properly digitized, catalogued, contextualized and valued as meaningful representations of individual linguistic expressions, life histories and social memory, and most importantly involve those who featured or published in such collections in the process, we will be on the way to creating an appropriate ethnographic and linguistic archive for the digital age. Meeting these goals necessitates a constant process of engagement with the form of our documents and the communities from which they originate: speech communities, fieldworkers, publishers and authors. Such partnerships require us to think of an open, non-linear archival structure that offers a range of access points and paths for different audiences to meet their diverse expectations. Each step along the way requires careful consideration of the central questions that this article raises, to ensure that initiatives such as Digital Himalaya, and the many language documentation projects active around the world today, continue to both remember and reflect the past, yet remain culturally responsive resources for the future.

For Marilyn Strathern, "the artifact is something one treats *as if* . . . it were simply a found object in the world" (Riles 2006: 16–17, emphasis in original). As a corollary, might a document then be something one treats *as if* it were simply a found object in the archive? Archiving documents that were previously free-floating, either online or in physical collections, can effect change not only by impact, but also by association. Whereas Jacques LeGoff argues powerfully against the neutrality and objectivity of documents, suggesting that "the document is what remains" (1992: xvii), in this paper, I hope to have shown that the document is just the beginning.

Notes

[1] Funded by the Volkswagen Foundation, the Dokumentation Bedrohter Sprachen (DoBeS) archive can be found online here http://corpus1.mpi.nl/, last accessed on Thursday, 27 January 2011.
[2] See http://elar.soas.ac.uk/, last accessed on Thursday, 27 January 2011.
[3] For a discussion of this process in Nepal and Northern India, see Shneiderman (2010).
[4] The ethnonym of choice for the ethnic group and their language is Thangmi, but the nation state of Nepal recognises them by the exonym Thami, which is also the official spelling of their last name. Both terms—Thangmi and Thami—are in use, although I have more recently opted for the former as it more accurately reflects the desires of the majority of the speech community.
[5] See http://www.pathwaysproject.org, last accessed on Monday, 31 January 2011.
[6] On its website http://www.nationalarchives.gov.uk/.
[7] The grant was routed through the Royal Anthropological Institute in the United Kingdom, to whom we remain very grateful.
[8] See Nordstrom (1993) for a discussion of the pornographic overtones of some early ethnographic photography.
[9] See Macfarlane (2010) for a discussion of the unique breadth and depth of von Fürer-Haimendorf's films.
[10] JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content on a trusted digital archive of more than one thousand academic journals.
[11] Online at http://www.digitalhimalaya.com/collections/journals.
[12] Creative Commons license (CC BY-NC-ND 3.0) allows users to copy, distribute and transmit the work, but only under the following conditions: users must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse the work); they may not use this work for commercial purposes; and they may not alter, transform, or build upon this work.

References

Burghart, R. (1996), "The Formation of the Concept of Nation-State in Nepal", in *The Conditions of Listening: Essays on Religion History and Politics in South Asia*, C. J. Fuller, & J. Spencer (eds), Oxford University Press, Delhi, pp. 226–260.
Dobrin, L. M. & Berson, J. (2011), "Speakers and Language Documentation", in *The Cambridge Handbook of Endangered Languages*, P. K. Austin, & J. Sallabank (eds), Cambridge University Press, Cambridge, pp. 187–211.
Feldman, I. (2008), *Governing Gaza: Bureaucracy, Authority, and the Work of Rule, 1917–1967*, Duke University Press, Durham, NC.
Finnegan, R. (2008), "Data—But data from what?", *Language Documentation and Description*, vol. 5, pp. 13–28.
Gaenszle, M. (2010), "Documenting ceremonial dialogues: An *in vitro* performance and the problem of textualization", *Language Documentation and Description*, vol. 8, pp. 66–82.
LeGoff, J. (1992), *History and Memory*, Columbia University Press, New York.
Macfarlane, A. (2010), "Early ethnographic film in Britain: A reflection on the work of Christoph von Fürer-Haimendorf", *Visual Anthropology*, vol. 23, no. 5, pp. 375–397.
Mayhew, N. (2006), "Cultural Property: A Contribution to the Debate", in *Who Own Objects? The Ethics and Politics of Collecting Cultural Artefacts*, E. Robson, L. Treadwell, & C. Gosden (eds), Oxbow Books, Oxford, pp. 83–89.

Megill, K. A. (2005), *Corporate Memory: Records and Information Management in the Knowledge Age*, K.G. Saur, Munich.

Megill, K. A. & Schantz, H. F. (1999), *Document Management: New Technologies for the Information Services Manager*, Bowker-Saur, London.

Nordstrom, A. D. (1993), "Persistent images: Photographic archives in ethnographic collections", *Continuum: Journal of Media & Cultural Studies, 1469−3666*, vol. 6, no. 2, pp. 207–219.

Ong, W. (1982), *Orality and Literacy: The Technologizing of the Word*, Methuen, London and New York.

Pollock, S. (2006), *The Language of the Gods in the World of Men: Sanskrit, Culture, and Power in Premodern India*, University of California Press, Berkeley, CA.

Riles, A. (ed) (2006), *Documents: Artifacts of Modern Knowledge*, University of Michigan Press, Ann Arbour.

Robson, E., Treadwell, L. & Gosden, C. (2006), *Who Own Objects? The Ethics and Politics of Collecting Cultural Artefacts*, Oxbow Books, Oxford.

Shneiderman, S. (2010), "'Producing' Thangmi ritual texts: Practice, performance and collaboration", *Language Documentation and Description*, vol. 8, pp. 159–174.

Turin, M. (2011), *A Grammar of the Thangmi Language with an Ethnolinguistic Introduction to the Speakers and Their Culture*, 2 Volumes, Brill, Leiden.

Turin, M. & Thami, B. B. (2004), *Nepali−Thami−English Dictionary*, Martin Chautari, Kathmandu.

Willinsky, J. (2005), *The Access Principle: The Case for Open Access to Research and Scholarship*, MIT Press, Cambridge.