Editors’ introduction (LDD12)

DAVID NATHAN, PETER K. AUSTIN

Cite this article: David Nathan, Peter K. Austin (2014). Editors’ introduction (LDD12). In David Nathan & Peter K. Austin (eds) Language Documentation and Description, vol 12: Special Issue on Language Documentation and Archiving. London: SOAS. pp. 4-16

Link to this article: http://www.elpublishing.org/PID/134

This electronic version first published: July 2014

This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See http://creativecommons.org/licenses/by-nc/4.0/

EL Publishing

For more EL Publishing articles and services:

Website: http://www.elpublishing.org
Terms of use: http://www.elpublishing.org/terms
Submissions: http://www.elpublishing.org/submissions
Editors’ introduction

David Nathan, Oxford University
Peter K. Austin, SOAS University of London

All but two of the papers in this volume of Language Documentation and Description are written up versions of presentations given at the Language Documentation and Archiving Workshop held at SOAS on 17th November 2011 in conjunction with the third Language Documentation and Linguistic Theory conference. There are two additions: the first is a chapter on research and creative arts practice across cultures and disciplines by John Wynne that describes work he carried out over several years in collaboration with Robert Munro and David Nathan (SOAS) on interactions between archival materials and sound art. The second addition is David Nathan’s account of the model of language archiving that he and his colleagues developed for the Endangered Languages Archive (SOAS) inspired by social networking. It is a revised version of a book chapter published in 2013 which covers and complements many issues relevant to the other papers in this volume.1

1. Language documentation and archiving: a fickle relationship

Woodbury (2011: 163) points out that a recognisable form of language documentation has been practised for well over a century, at least since the work of Boas and Sapir. However, although recordings, field notes and documents from Boas, Sapir and other early documenters have been physically preserved (Johnson 2005: 140), it is only recently that archiving has come to be seen as a distinguishing mark and systematic component of documentation practice.

For those involved with endangered languages today, whether of a theoretical or applied orientation, the terms ‘language documentation’ and ‘archiving’ slip off the tongue together as if they have always been connected. But they have been systematically linked only since the late 1990s, when Nikolaus Himmelmann, in his seminal paper for documentary linguistics (Himmelmann 1998: 168), stated that:

1 The remainder of this Introduction is based on David Nathan’s introduction to the Proceedings of the Workshop on Language Documentation and Archiving (Nathan 2011b).
Language Documentation … is concerned with compiling, commenting on, and *archiving* language documents (emphasis added – editors).

and foresaw many of the issues that continue to occupy us (Himmelmann 1998: 191):

- technical problems … such as the choice of an appropriate recording and presentation technology (sound recording, video, multi-media applications, etc.), the problem of archiving and maintaining documentations, and the problem of providing and controlling access to documentations

In an influential paper, Bird and Simons (2003) described the same issues in terms of ‘portability’, the sustainability of digital documentation across different computing environments and over time.

The pairing of documentation and archiving also appears in several other contexts, including ethics, access and training. United via ethics we find, for example, Dwyer (2006: 40) emphasising that ‘properly archiving collected data is far more respectful to a speaker community than piling it in the back of a closet’. Dwyer (2006: 35) also identifies archiving as a ‘phase’ of documentation that carries forward and fulfils language speakers’ preferences:

> [d]uring the archiving phase, the researcher must carry though the wishes of the consultants in terms of anonymity and recognition … [and] on user access to the materials (community, scientific researchers, general public)

In a recent chapter ‘Archiving and language documentation’, Conathan (2011) interweaves documentation and archiving through considering access and intellectual property issues, where documentation and archiving intrude on and affect each other’s practices. Nathan (2013), drawing on an analogy with libraries, describes archivists and depositors as ‘joint librarians’ for endangered languages materials, but where depositors play the major role because they are the ones who understand the materials’ context and content. Archiving and documentation regularly appear as a duo in the curriculum of training courses such as those run by the 3L consortium, CoLang (formerly InField), LingDy, and others.

This *prima facie* relationship between documentation and archiving has not, however, received anything like the same scrutiny as that between documentation and description. The latter pair have been theorised by Himmelmann and discussed and debated by many since in a large number of conferences and publications (cf. Austin and Grenoble 2007).
It is surprising, therefore, that this volume appears to be the first journal publication symmetrically targeted at both language documentation and archiving. We hope that the papers presented here challenge the directions of our disciplines; they offer a glimpse into the future, not only of endangered languages archiving and language documentation, but perhaps even the survival of particular languages.

2. The papers

The papers in this volume deal with three themes: new models for archiving, enhancing archive usage and effectiveness, and new methods for creating and structuring archive content.

The first theme – new models for archiving – is addressed in Anthony Woodbury’s paper. He is concerned about the apparent under-usage of archives by their potential audiences and looks at the nature of these audiences to provide desiderata for what ought to count as ‘good’ documentations, or, as the title of his paper puts it, documentations that people ‘understand and admire’. Paradoxically, he starts out considering paper documents and paper archives in order to remind us, perhaps, of the dictum that in using new technologies we should not forget what previous technologies did very well. The body of his paper consists of a timely and much-needed set of proposals for rethinking the genres, content, and arrangement of language documentation. Bringing the discussion full-circle back to audiences, he suggests that audiences can most fruitfully be ‘critics’ of documentation, for example as reviewers, thus reinforcing calls for, e.g., journal-style reviews of documentation to be added to the ecology of language documentation. This could be taken as another reminder not to readily abandon familiar and effective genres. Peer reviews, as part of an evaluation and feedback loop, are an indispensable component of the scientific process and the evolution of ideas (see, e.g., Allen et al. 2009).

Woodbury has also quietly exposed the issue of archivists’ contributions to the presentation of materials, a question that most endangered languages archives have barely grappled with. Archivists’ contributions relate not only to contextualisation of materials but also to the software and design issues involved in creating screen interfaces that appropriately delineate whose ‘voice’ the audience is reading/hearing. Although the production and presentation of contextualising matter and finding-aids and the preparation of exhibition materials are standard fare for archivists in ‘traditional’ archives, in the language documentation and digital archiving fields they have largely been covertly re-delegated to documenters (or left un-done). This may perhaps be due to most present-day endangered languages archivists having
backgrounds in linguistics or information technology rather than archiving (or museum studies, librarianship, or related fields).

Gary Holton’s paper looks at how archives can serve as mediators between the materials they contain and the audiences they intend to serve. He notes that even when access to materials is being provided to language communities by an archive, through the internet or in person, it may still be the case that materials are not being used. He argues that archives must work for something more than providing access and that this will involve presenting collections in a way that is pertinent to potential users by enriching metadata descriptions with relevant fields, or highlighting particularly useful resources within large collections. To achieve this, the archive must know and work closely with its user community.

Community links are also the focus of Mary Linn’s paper which proposes a new approach she calls CBLA (community-based language archiving). She draws interesting parallels between models for participatory linguistic fieldwork and new models for archiving. She raises a number of innovative, indeed challenging, ideas such as ‘decentralised curation … [where] there is no need for an archivist at all’. This might resonate with some archives who delegate much of the curation process to depositors through issuance of guidelines and software that governs structures and formats. However, Linn actually takes curation to go far beyond checking formal properties of data. She considers ‘radical user orientation’, where the archivists’ primary curatorial task, namely contextualisation, is centred on the context of the users themselves, because it is they (especially as community members welcomed into the archive ecology) who ultimately determine the success of archives in meeting their goals. Such departures from classical archiving approaches form the basis of Linn’s proposal for CBLA, in which the language community is involved in every step, from documentation planning to curating to dissemination. Linn provides a case study showing how such approaches have had a positive social impact on communities and revitalisation through increased archive usage and resultant language activities.

The paper by Edward Garrett focuses on individual language speakers and communities who participate in language documentation projects. He proposes an archiving component, complementing the approach of Linn, called PDLA (participant-driven language archiving) which assigns role-appropriate rights and responsibilities to the individuals and speaker

---

2 For a discussion of the need for archives to consider pertinence of their collections as part of preservation strategy, see Nathan 2006b.
communities involved in language documentation, including them in the curation, preservation and dissemination of materials. They can add further materials, comments, or contextualisation. Or they can identify themselves or their relatives and claim moral rights\(^3\) in recordings and other materials. They can also make corrections to erroneous data, interpretations, and attributions. PDLA aims to facilitate cooperation and collaboration, and promote the aims of community curation as described by Linn and Christen (2011). It could result in a less ‘commodifying’ approach to language documentation and archiving (Dobrin et al. 2009). Garrett’s proposals expose the weakness of frequently-heard claims about supporting ‘communities’ by recasting the general problem as one of supplying concrete services to individual persons.

By enabling speakers to assert their moral authority over materials that they participated in creating, and connecting them to other archive users, PDLA raises new possibilities and challenges for transforming the direction of language documentation and archiving.

The second theme – enhancing archive usage and effectiveness – begins with Joshua Wilbur’s unique, semi-biographical account of his interactions with various archives in Saami country, Sweden, where he conducted his documentary fieldwork. With the explicit goal of making his documentary materials more accessible to Saami people, he negotiated with three different archives, and presents here his experiences and observations. He finds that smaller, local archives, although having the greatest potential to reach community members, have very particularised (in some cases, limited) capacities, resources, skills, policies and preferences. These mean that in order to reach local communities through such archives, the documentary linguist may need to invest considerable time and effort not only in negotiating with archive management and technical staff, but also in acting as a technical consultant to them. Wilbur also makes an important distinction between ‘discovery’ and ‘promotion’. Discoverability is the ability of potential users to identify a relevant resource, typically through metadata-based search; it is the oft-stated rationale for certain types of (standardised) metadata schemes and data aggregating portals (Bird and Simons 2001). However, Wilbur finds that ‘usage of materials … is not guaranteed by their mere presence in an archive’, regardless of metadata, and that archives need to ‘actively promote the language materials they have been … trusted with’.

Gabrielle Gardiner and Kristen Thorpe are also concerned with community access and particularly with establishing protocols for Indigenous Australian materials in the Aboriginal and Torres Strait Islander Data Archive (ATSIDA). They describe the development of ATSIDA, its vision, partners and key stakeholders, and outline the importance of archives working closely with researchers, their data and the communities to which the data relates. They describe in detail the protocols that have been developed to guide preservation and ongoing sustainable use of Indigenous Australian research datasets, embracing principles of trust, respect and engagement with Indigenous communities.

John Wynne’s paper – which should be read in conjunction with the interactive multimedia app Hearing Voices: Speakers / Languages4 – looks at how creative arts practices can dovetail with documentation and linguistic research in drawing attention to the importance of language survival and revitalization. Wynne’s fieldwork with Khoi and San languages in Botswana, and with Gitxsan in Canada, led to the creation of sound art installations which make creative use of the recorded and archived documentary materials. The sound environments he creates aim to draw the visitor into a space in which to consider the aesthetic, linguistic and socio-political dimensions of the languages, and language endangerment more generally. Contemplative gallery installations are accompanied by archival materials and contextual information in various formats including interactive multimedia. The paper addresses the tension that arises between the collected materials and the uncluttered, contemplative work the artist hopes to make.

The final paper on this theme, by Paul Trilsbeek and Alexander König, addresses the widely-felt concern that materials in digital endangered languages archives are under-used. They identify two main audience categories, academics and speaker communities, and turn to the latter as new, so far untapped, providers of language documentation. While many communities are keen to participate in the new media landscape characterised by YouTube, the authors examine some of the problems that archives would face in handling community-sourced uploads to public portals, whether to sites such as YouTube or to extensions of archives such as DoBeS. In particular, they doubt that other sectors of the audience would be confident of the provenance, veracity and ethical conduct associated with resources

4 Hearing Voices: Speakers / Languages (Wynne and Munro 2014) is included with this issue LDD12 and is available for free download - see the volume catalogue for access details.
‘contributed by unknown depositors’. The authors hint at an interesting reversal in the properties of community-resourced versus researcher-sourced materials: while a YouTube-fired zest for public exposure could give rise to masses of freely available but (presumably) less authoritative material from community members, many academic researchers tend to over-apply access restrictions on their (presumably) more authoritative materials. The authors therefore call for a greater willingness among researchers to share access to their materials, while respecting source community wishes.

The third theme – new methods for creating and structuring archive content – is the focus of the paper by Sebastian Nordhoff and Harald Hammarström. They argue that in addition to documentary materials, language archives should also hold grammars of languages, and in a form that reflects the logical structures of grammars. Their chief interest is in reinterpreting grammars firstly as ‘grammatical descriptions’ and then as ‘granular annotations’, in order to widen the applications and usages of grammatical information, especially via the semantic web, where information can be searched and processed in terms of its logical and ontological structures. After surveying the structure of typical printed grammars, and dispensing with their printed-page-only properties, they propose a 21st century approach to digital grammar-writing as ‘nonlinear database[s] of micropublications’.

The final paper by David Nathan presents a new model of language archiving developed at the Endangered Languages Archive (ELAR) at SOAS that draws upon concepts and practices developed for web-based social networking. Nathan argues that language archives have typically provided a narrow, one-way access strategy, where academic documenters provide materials labelled with access restrictions (where required), which archives disseminate to users (who are usually also academics) whilst enforcing such restrictions. ELAR aims to offer more flexible and equitable access to various types of users rather than use this traditional model which privileges researchers. The model reconceives the archive as a platform for building and conducting relationships between information providers and information users. ELAR implements a nuanced ‘access protocol’ system which dynamically and securely manages access to materials by providing channels for providers and users to negotiate access. ELAR takes the user interface seriously, paying attention to its design, layout, interactivity, controls and navigation – all factors which contribute to usability and accessibility in the wider sense.
3. Discussion

A number of issues recur in the papers of this volume, and several papers offer challenges to the current paradigm for archiving language documentation. We have selected some of these issues and challenges for brief further discussion.

Community curation

Several authors write about what could be called, following Christen (2011), ‘community curation’. The sharing and participatory approaches proposed by Linn, Garrett, Nathan and Woodbury (and appearing to some degree in other papers too) are innovative models that challenge current practice. Linn countenances archivists (in the conventional sense) being dispensed with altogether, while Garrett and Woodbury see language speaker audiences as correctors and critics. We are presented with a radical inversion: the archive concept of ‘context’ is no longer that of the materials, or their (supposed) provenance, but of the users. And these users are principally the language speakers and their descendants, who are also, in a virtuous circle, central participants in the documentation and archiving processes.

Not forgetting researchers, Woodbury, and Nordhoff and Hammarström propose new models and structures for information and documents, opening up documentation to the possibilities enabled by bringing together disparate sources of related information through the semantic web, and including annotations and other contributions from researchers other than the original creator(s).

Synthesising these very different types of participation will pose a challenge. A major tendency so far has been to polarise them: Trilsbeek and König focus on their archive’s distinction between their depositors, who are authoritative and ‘internal’, and community-sourced and crowd-sourced materials that are fraught with unknowns. Wilbur injects a rarely considered class of participants: small, local archives, who may be the best option for reaching communities, after all.

Promotion

The papers by Holton, Wilbur and Woodbury explore the idea that archives need to do more than acquire, curate, preserve and disseminate materials. Woodbury, for example, would like to see materials become more visible and better publicised through exhibitions and reviews, so that archive holdings ‘could interest and intrigue many more people’. Taken together, the three papers suggest that to reach target audiences, archives need to actively
promote and mediate the materials they hold. This also suggests to us that archives need to develop relationships with their audiences that are not based purely on file-level access to language materials, because the success of archive outreach may depend on developing contact, relationships and trust in order to encourage usage and other interaction with the materials. And, independently of dissemination of materials, archives’ promotion of language resources in the public sphere would help ‘valorise’ and thus sustain the very languages represented in collections.

**Contextualisation**

Contextualisation of materials is at the heart of archiving, but in many of our contemporary archives the art of contextualisation has given way to the science of software development. However, communities may well wish to play a role in framing the interpretation of their materials to others (cf. Christen 2011: 197). Similarly, community access to materials is not reducible to file transfer, but in reality entails access to meaning (Christen 2011: 194; Nathan 2013), and meaning that is relevant, as Holton argues. The most radical proposal is Linn’s CBLA, which is not about negotiation or mediation of metadata, or even about promotion. It is baldly about sharing or handing over management of the archives themselves, since ‘when communities and families know what’s in archives and how they work, the collections get used more.’

**The form of documentation**

Despite extensive theorisation of documentation in previous work, there has been little discussion of the form of documentation: its granularity, structure, organisation, links, and how it is to be navigated. Several papers here do address the issue. Nordhoff and Hammarström give a detailed alternative view of the shape of grammars for a networked age. Woodbury’s paper describes ‘a Noah’s Archive, a one-time sampling of the uses of a language for a grammar, dictionary, or thumbnail linguistic ethnography’, with detailed proposals for content. He also describes materials by Knut Bergsland from 1959 which, although printed, are essentially models in hypertext and text retrieval techniques that today’s digital environment can easily deliver but language documentation has not yet conceived for itself.

**Publishing**

A corollary of Woodbury’s many proposals, including increased archivist contributions, attention to genre, exhibitions, promotion, and reviews of documentations, is that what archives do is expressed better as ‘publishing’
rather than ‘dissemination’. The idea that ‘[a]rchiving can be considered a form of publishing’ first appeared in Johnson (2005: 143), and is explored in Nathan’s paper in this volume (see also Nathan 2011b).

General observations

More generally, there are interesting tensions which hint at future possible divergences. For example, some authors advocate a corpus-based framework with its emphasis on a ‘representative sample’ for ‘scientific study’, while Linn’s CBLA model and Garrett’s PDLA evoke a more participatory-organic-evolutionary framework (for a close examination of the resonances and dissonances between corpus and documentary approaches, see Cox 2011). Similarly, some believe that further codification and implementation of standards is a key to greater sharing, while others are wary of the constraints and biases introduced by the imposition of standards (see also Christie 2005). Finally, there is the question: are language documentation and archiving a cross-disciplinary affair, or two parts of the same discipline? The papers here seem to range along this one-or-two disciplines axis, so the question remains open, at least for now.

4. Conclusion

All the papers in this volume share the goal of exploring how the linguistic and archive communities can provide effective and ethical responses to language endangerment. Documenters and archives all wish to raise the quality of documentary materials and the effectiveness of technologies for disseminating them. What the papers, taken together, tell us clearly is that quality and effectiveness are not meaningfully measurable without considering audiences and the nature of their interactions with archive materials. Engagement with language speakers will be necessary for progress to be made. Here is an example: current web trends are towards mash-up pages, mobile apps, and aggregating portals. These gather resources based on a particular user’s preferences, and display them according to topic, geographic location, or language. But what happens when a user wants to see a page of information regarding a specific person such as language speaker X? Unless speaker X is truly a participating element of the system, as a member, owner, or curator, rather than a mere metadata-point, then such a page will be

---

5 These will be even more useful for endangered languages community members with the increasing use of mobile devices.
an incomplete and insipid representation with poorly-implemented access management because nobody except speaker X can properly understand their own original communicative intentions or decide who information should be shared with. We are fortunate, therefore, to see the maturation\(^6\) and continued rise of online social networking and innovative apps which personalise individuals’ interactions with a variety of resource providers and provide further exemplars for implementing the participatory models suggested by several authors here. Whether dedicated language-resource platforms are going to be effective is yet to be confirmed; when it comes to matters of rights, communications and sociality it is likely that well-designed systems that work for everyone will be the best ones for language speakers too.

Another thing also seems clear: for too long we have proffered (and accepted) glib statements about the myriad advantages of the internet, for example that it solves our major problems by reaching everyone. Such statements might have struck a chord in 1996 but today are digital prehistory. Thomas Friedman, celebrated journalist and author, recently observed that less than ten years ago ‘Facebook didn’t exist, Twitter was a sound, and Skype … was a typo.’\(^7\) All of these are, of course, platforms for social interaction. Our future successes will come not from focusing on technologies but on communities of interaction.

As Woodbury says, there is ‘much, much to be done’. Ours is still a work in progress, perhaps barely begun.

---

\(^6\) It is only in the last few years that dynamics for online privacy and sharing have become deeply and widely debated, and as a result are evolving to become more nuanced and conventionalised. The release of Google+ (Google’s platform for social networking), as the first real competitor to Facebook, precipitated a true ‘battle for ideas’ on the territory of ethical practices for sharing and privacy.

\(^7\) Thomas Friedman, ‘What went wrong with America?’ Highlights of an address given at the Melbourne Town Hall on 29 July 2011. Big Ideas, ABC Radio National, 8 September 2011.
References


Acknowledgements

All the papers in this volume were anonymously reviewed by at least two external reviewers and then revised for publication in light of the reviewers’ and editors’ comments. We are grateful to colleagues for their help with the reviewing process. Thanks also for publication assistance to Souung-U Kim, who did an excellent job of copy-editing, proof-reading and checking references and web links and to Mary Chambers for proof checking. As always, we are grateful to Tom Castle for design and layout.
# List of Contributors

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Location</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabrielle Gardiner</td>
<td>UTS Library</td>
<td>NSW, Australia</td>
<td><a href="mailto:Gabrielle.Gardiner@uts.edu.au">Gabrielle.Gardiner@uts.edu.au</a></td>
</tr>
<tr>
<td>Alexander König</td>
<td>Max Planck Institute for Psycholinguistics</td>
<td>The Netherlands</td>
<td><a href="mailto:Alexander.Koenig@mpi.nl">Alexander.Koenig@mpi.nl</a></td>
</tr>
<tr>
<td>Edward Garrett</td>
<td>Department of the Study of Religions</td>
<td>UK</td>
<td><a href="mailto:heacu.mcintine@gmail.com">heacu.mcintine@gmail.com</a></td>
</tr>
<tr>
<td>Mary S. Linn</td>
<td>Sam Noble Museum of Natural History</td>
<td>University of Oklahoma</td>
<td><a href="mailto:mslinn@ou.edu">mslinn@ou.edu</a></td>
</tr>
<tr>
<td>Harald Hammarström</td>
<td>Max Planck Institute for Psycholinguistics</td>
<td>The Netherlands</td>
<td><a href="mailto:Harald.Hammarstroem@mpi.nl">Harald.Hammarstroem@mpi.nl</a></td>
</tr>
<tr>
<td>David Nathan</td>
<td>Zoology Department</td>
<td>UK</td>
<td><a href="mailto:david.j.nathan@gmail.com">david.j.nathan@gmail.com</a></td>
</tr>
<tr>
<td>Gary Holton</td>
<td>Alaska Native Language Center</td>
<td>Alaska, USA</td>
<td><a href="mailto:gmholton@alaska.edu">gmholton@alaska.edu</a></td>
</tr>
<tr>
<td>Sebastian Nordhoff</td>
<td>Freie Universität Berlin</td>
<td>Berlin</td>
<td><a href="mailto:sebastian.nordhoff@langsci-press.org">sebastian.nordhoff@langsci-press.org</a></td>
</tr>
</tbody>
</table>
Kirsten Thorpe
Indigenous Unit
State Library of New South Wales
Sydney
NSW, Australia
kthorpe@sl.nsw.gov.au

Anthony C. Woodbury
Department of Linguistics
University of Texas at Austin
Texas,
USA
woodbury@austin.utexas.edu

Paul Trilsbeek
Max Planck Institute for Psycholinguistics
Nijmegen
The Netherlands
Paul.Trilsbeek@mpi.nl

John Wynne
London College of Communication
University of the Arts
London
UK
jw@sensitivebrigade.com

Joshua Wilbur
Skandinavisches Seminar
Albert-Ludwigs-Universität Freiburg
Freiburg
Germany
joshua.wilbur@skandinavistik.uni-freiburg.de