

Archives 2.0 for Endangered Languages: from Disk Space to MySpace

David Nathan
Endangered Languages Archive
School of Oriental and African Studies, London
djn@soas.ac.uk
www.hrelp.org

1. Introduction

The Endangered Languages Archive (ELAR) at SOAS¹ has been set up to preserve and disseminate digital documentations of endangered languages, especially those created through the funding activities of our sister programme, ELDP². ELAR is developing its online catalogue to take advantage of web-based social networking in order to address the inherent complexities of access and distribution in the domain of language documentation. Many of these documentation materials are associated with sensitivities and access restrictions because endangered language communities and their speakers are under various pressures and deprivations, which are, in many cases, the causes of the decline of their languages. The need for care is amplified by the fact that language documentation practices favour recording of spontaneous, naturalistic speech, which can easily include content that might cause embarrassment, or worse, for the speakers.³

Under our new approach to addressing this situation, the archive is no longer essentially defined by its data repository function, but is reconceived as a forum for conducting relationships between information providers (usually the depositors) and information users (language speakers, linguists and others), using the now-familiar idioms of the Facebook or eBay websites, which in turn are part of the recent phenomenon known as “Web 2.0”. This paper shows how a Web 2.0 approach neatly addresses two defining characteristics of endangered languages documentation materials - sensitivity, and diversity. Rather than a disseminating archive having to continually interpret and broker complex and changing access conditions, parties can negotiate directly with each other to achieve more flexible and creative outcomes. Through this approach, we aim to enhance the distribution of materials, to foster more usage of them in research and mobilisation for support of language communities (Nathan 2006), and to encourage a more critical approach to the nature of the materials themselves.

2. Protocol

During two decades of recognition of the threats to the sustainability of the world’s languages, there has been increasing attention to documentation of endangered languages. However, by most criteria, the increasing amount of documentation has in itself provided few positive outcomes for communities that want to maintain their

¹ ELAR is one component of the Hans Rausing Endangered Languages Project (HRELPP), hosted at the School of Oriental and African Studies, University of London, and funded by Arcadia. ELAR’s web presence is at elar.soas.ac.uk and www.hrelp.org. For Arcadia, see <http://www.arcadiafund.org.uk>.

² Another component of the HRELPP project, Endangered Languages Documentation Programme. See www.hrelp.org/grants.

³ There are direct threats, for example where languages are spoken in war zones, or where recorded conversations reveal illegal activities. Less dramatically, but no less important, recordings of conversation in small communities can easily contain damaging statements about others within the community.

languages, or for the evolution of a linguistics discipline that could help them to do so. One way of increasing the effectiveness of language documentations is to build systems that take into account their unique nature as records of spontaneous, communicative acts in authentic social contexts (Himmelmann 1998). Documentation materials can have personal, social, cultural and pedagogical potencies, rather than being merely value-free “data” commodities that are grist to the mill of linguistic glossing and grammatical and typological distillation (Dobrin et al 2009). The processing of, access to, and subsequent use of such materials all need to be performed sensitively and ethically. At ELAR, we label these areas of sensitivity and restrictions with the term “protocol”, reflecting the tension between on the one hand formulating, implementing and maintaining access restrictions, and, on the other hand, making materials accessible to the right people for the right purposes.

A modern, effective archive for such materials needs to formulate and implement an appropriate protocol scheme to govern the operation of its access system. Traditional approaches to online access control will not effectively achieve the right balance between control and access for endangered languages. As a result of input from depositors we believe that:

- depositors prefer formulations of access restrictions that are more fine-grained than binary open/closed systems. They prefer a range of choices such as the ‘graded access’ system of AILLA⁴, or even more nuanced categories of control
- binary open/closed schemes minimise access because depositors with specific requirements are forced to “fall back” to safety by making materials fully closed
- archives need to demonstrate understanding of the sensitivities associated with materials, so that depositors are inspired to have confidence and trust in those archives and are likely to allow the most liberal possible access restrictions.

When ELAR opened in 2005, we developed its deposit form to take these factors into account. We gave much consideration to the design of the protocol section. We surveyed about thirty other similar forms for related facilities/institutions, and workshopped our proposals at meetings at SOAS and at the DELAMAN archives group meeting at the University of Texas in 2005.⁵ The result was the scheme shown in Figure 1, which shows the main component of the protocol section. Since then, the form has been filled in by all depositors, and we have received no adverse feedback about the protocol scheme. This paper takes up the topic of the implementation of this scheme in Section 4.

3. Diversity

Specialised archives require specialised solutions. As well as protocol, an archive for endangered languages documentation has to deal with the great diversity of its materials, clients, and stakeholders.

⁴ The Archive of the Indigenous Languages of Latin America, at the University of Texas. See <http://www.ailla.utexas.org/site/gas.html>.

⁵ Digital Endangered Languages and Musics Archive Network. See <http://www.delaman.org>.

The field of language documentation is an emergent and evolving one, with few current conventions for what counts as a language documentation. Language materials have quite different data semantics from library, business and other documents. Business data, for example, is anchored in well-defined concepts such as quantities, costs, and product codes, which have widely shared and stable interpretations that also correspond in uncontroversial ways to real-world object and properties. Libraries enjoy a special conventionality of their objects' attributes (author, title etc.) which are not only formalised by publishers but also made available to them by shared authority file sources.

By contrast, the world of language data is a distinctly varied - if not chaotic - one, with its categories, rather than being predetermined and centrally provided, needing to be derived bottom-up from its wide-ranging (and possibly yet undiscovered) data and methodologies. Linguistic nomenclatures do exist, but language *data* as symbolic representations consists of speculative and contestable interpretations⁶ rather than measurements or standard attributes. We know that the majority of human languages are not yet documented. Accruing work in language documentation suggests that languages can differ from each other in arbitrarily complex ways. Thus, we have the paradoxical situation that linguistic study seems to *guarantee* non-interoperability of its data because that data is already metadata, i.e. we do not have agreed-upon *facts* that will “ground out” its metadata semantics.

Endangered languages documentation has been characterised by diversity since its inception. Its seminal description (Himmelman 1998) saw its methods and outputs as inherently heterogeneous, in order to capture

a multipurpose and comprehensive record of the linguistic practices characteristic of a speech community ... [where] the emphasis is on the collection and representation of primary data rather than theory and analysis

(Himmelman 1998:166).

It ought to be no surprise, then, that documentary materials do not confirm to a single template. Indeed, we at HRELP encourage creative approaches to formulating and conducting funded documentation projects. The contexts of projects range from recording the “whistled language” of a tiny Amazonian community⁷ to broader descriptions of languages in south-western China that may have hundreds or even thousands of speakers yet are expected to decline quickly under the pressure of rapidly mounting cultural and educational influences from that country's metropolitan centre.⁸ Layered upon those contexts are the particular goals of the documentation project (whether for example, describing particular linguistic phenomena, focusing on annotated recordings or on dictionaries and grammars, creating pedagogical resources for community-based language revitalisation, or an ethnomusicological approach to

⁶ For example, a transcription might be changed as the linguist better understands a language's structures. Chomsky aimed to lay foundations for linguistics that would ground out this problem but his work has not been influential in language documentation.

⁷ See Julien Meyer's project on the Gaviao and Surui languages, described at <http://www.hrelp.org/grants/projects/index.php?projid=148>.

⁸ See Ross Perlin's project on the Dulong language described at <http://www.hrelp.org/grants/projects/index.php?projid=123>

song poetics), and ultimately the nature of the language and its usages. Then within each project, the cultures, communities and individuals with whom the documenter works all bear their unique influences.

Deposited materials come mainly from individual language documenters, whose goals, projects, methods, skills, competence, and work locations are as different as the range of environments we can find across the entire planet. Documenters are typically lone fieldworkers in remote and often dangerous locations, and their practices are not easily harmonised. And their documentations can contain a wide variety of media, materials and formats, for which there are few agreed standards.

The audience of *users* served by documenters' data includes language community members, linguists, ethnographers, historians, language planners, journalists, and even members of the general public, all of whose interests in and interactions with documentary materials may be different.

Another source of diversity is change over time. Many depositors update their materials. We allow - and even encourage - this, because many materials are produced by single researchers who are transcribing and annotating audio and video of languages that they are only slowly gaining an understanding of. This work can take up to 250 hours to process a single hour of recording. Should the archive express a strong preference for receiving completed materials only, many resources will remain at risk without robust preservation, and ultimately less will be received since the process of "adding value" to primary language documentation is rarely comprehensively complete. Archiving such materials represents a tension between providing timely data security for stable resources such as media recordings, while encouraging ongoing intake of new materials by making it easy for documenters to update their deposits.

Of course, these sources of diversity also crosscut protocol and access. As we saw in Section 2, depositors might wish to apply various formulations of access conditions. To make things more complicated, different conditions can apply to various materials within a deposit. Many materials have access conditions that associate resources with users, rather as if particular books in a library are not only borrowed under different terms by staff and students, but may be only borrowable by particular named individuals - and, in many cases, only after securing permission from the author. In the archive, for example, only (certain) females might be permitted access to some items, while other items might be available only to members of the relevant speech community, and others might be unrestricted. Finally, protocol (unlike most other categories of metadata) can change over time, as conditions and attitudes change in the speaker community. A salient example is the taboo on encountering the name or voice of someone who has recently passed away in some Australian Indigenous communities: here, access needs to be differentially restricted, depending on the social proximity of the user and the amount of time that has elapsed since the event.

Thus, the protocol that governs access to materials is subject to a similar degree of diversity as the materials themselves. Collecting, representing, transmitting and implementing protocol amplifies the workload for communities, researchers, depositors, archives, and end-users; any solution which is a 'natural fit' with its

complexity and dynamics is likely to not only ease such workloads but also make the efforts of all these participants more effective.

4. Protocol information in ELAR's deposit form

ELAR's deposit form has five parts, the third of which is "Part C: Access Protocol", in which depositors are asked to "[d]efine the permissions for users to access materials, to observe sensitivities and protect against risk." The main part of this form is shown in Figure 1, where depositors select access options.⁹ This grid addresses the issues discussed above to cater for the most common scenarios, expressed with sufficient granularity, in order to capture the majority of expected choices, while at the same time being explicit enough to be implemented by a computer system. The latter is sidestepped in the case of [P3], where depositors are to be personally consulted on each request.

The protocol collection grid shown in Figure 1 offers, in summary, the options of [P1] open access, [P2] access to particular people by name or membership of nominated categories,¹⁰ [P3] the depositor is asked to decide each request, or [P4] no access at all.

Figure 1: Main part of ELAR depositors' form, protocol (access conditions) section

P1. Anyone	<input type="checkbox"/>
Any person may view/listen to or receive a digital copy of any part of the deposit	
P2. Certain people or groups	
Choose any combination of P2A, P2B, and P2C:	
<i>P2A</i>	<i>Research community members</i>
What level of access (choose one only)?	
P2A1.	They can receive a digital copy of requested material <input type="checkbox"/>
P2A2.	They can view/listen but cannot receive a digital copy <input type="checkbox"/>
<i>P2B.</i>	<i>Language community members</i>
See below regarding identifying members	
What level of access (choose one only)?	
P2B1.	They can receive a digital copy of requested material <input type="checkbox"/>
P2B2.	They can view/listen but cannot receive a digital copy <input type="checkbox"/>
<i>P2C.</i>	<i>Particular named people or bodies</i> <input type="checkbox"/>
See below regarding identifying people/bodies	
P3. Depositor is asked permission for each request	
You will be contacted and asked for permission on each request.	
How do you want to be contacted?	
P3A.	Requester is given address to contact you directly <input type="checkbox"/>
P3B.	ELAR will relay requests to you <input type="checkbox"/>
P4. Only the depositor has access	<input type="checkbox"/>
Persons other than the depositor will not be able to request access.	

⁹ There is another component of the schema, not shown in Figure 1, which asks depositors to tell the archive how to identify members of particular named groups or categories. The form is online at <http://www.hrelp.org/archive/depositors/depositform/>.

¹⁰ Another section of the form asks depositors to tell us how to determine membership of the categories they nominate.

Although we did not explicitly set out with the goal of treating what depositors entered in Part C as research data, we have long held a belief that language documentation is an incipient field for which methods must be discovered, rather than declared by fiat or by extension from traditional linguistic and archiving methods. In fact, the evidence of the trends in depositors' preferences, from about seventy forms over three years, turned out to have clear and interesting implications.

When the development of ELAR's archive information systems moved towards providing access to data during 2009, we revisited the protocol scheme and its responses to date. (Up to this point, we had not operationalised the protocol, nor were we systematically disseminating any data.) A simple analysis of depositors' choices showed that [P3] - where the depositor wishes to be asked permission for access, either directly or indirectly - was the most frequent choice. As a result of this, and further discussions with depositors, we believe that:

- many depositors feel uncomfortable with access that is unregulated (by themselves); they want to know who is accessing their data and why (perhaps in order to gain something from an interaction with the requester);
- several depositors want to close access for the sake of their own exclusive use of materials for academic purposes, but feel that perhaps they might also usefully or safely share the data with certain others (who they do not wish to, or cannot, identify in advance);
- some depositors feel "guilty" about denying open access, so [P3] is an attractive option that stops short of closing access entirely
- some depositors may be considering the use of controlled access in the future as a means to set up a small network of colleagues who could work together on the materials.

All these reasons for preferring individual handling of requests point to:

- a willingness to share data (only?) as a result of negotiation with particular individuals
- a preference for allowing access on the basis of a person-to-person transaction
- the depositors' sense that the potency of the materials requires direct negotiation to establish the requester's credentials

Perhaps they also point to depositors' belief that they control access on behalf of the language speakers who provided the data; and such depositors look for a selection that implements an ongoing custodial role, best realised by [P3].

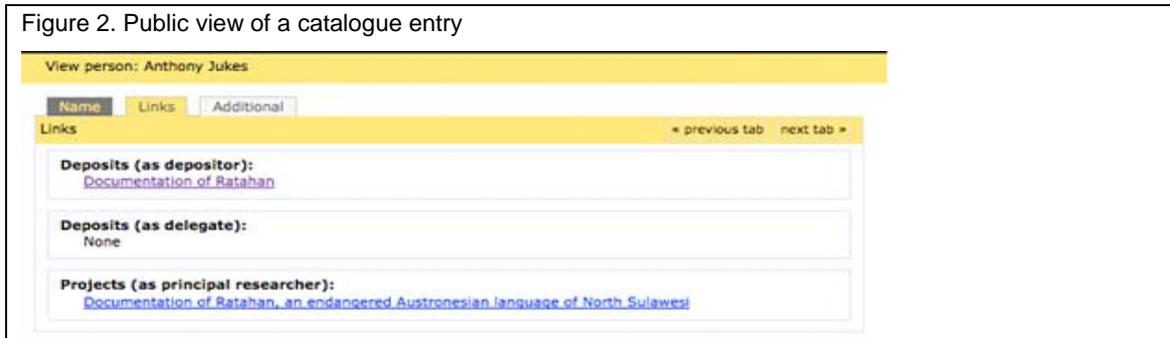
5. Implementation

We are currently extending our catalogue system to provide online access to deposited data, using the protocol grid as a "roadmap" for the implementation of access control.¹¹ The catalogue devolves much of the information management to depositors.

¹¹ The catalogue is based on a customised Drupal content management system. Drupal offers inbuilt support for controlling access among users and groups, but has had to be extensively customised to meet the requirements of our scheme.

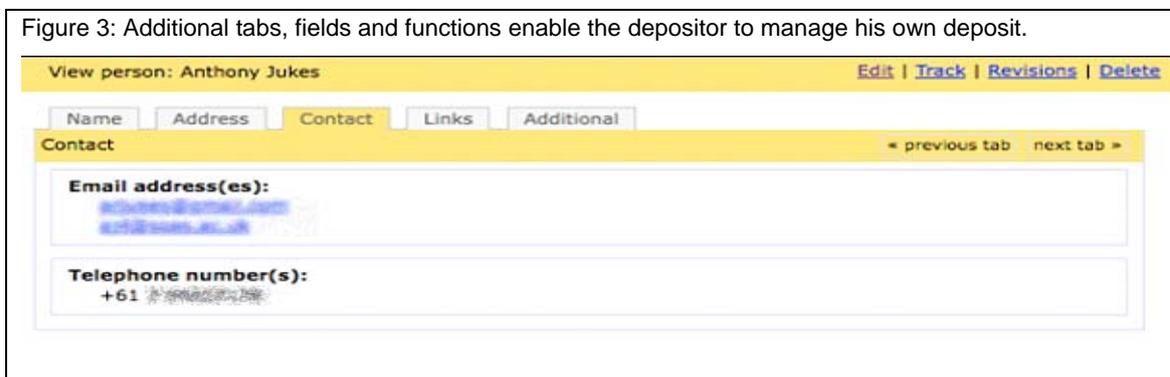
For example, a public view of the catalogue record for Anthony Jukes' deposit is shown in Figure 2.

Figure 2. Public view of a catalogue entry



However, the depositor's own view of this screen has additional tabs, fields and functions, allowing him to maintain and update the information himself (Figure 3).

Figure 3: Additional tabs, fields and functions enable the depositor to manage his own deposit.



We turn now to our draft implementation for negotiating access. Figure 4 shows a user - let's call him "Everyday Ed" - checking access to the Chaquita Rarámuri deposit, for which Everyday Ed presently has no access rights (he could get an outline of the present default access restrictions by viewing the "Protocol" tab). This could be because the depositor has applied any of the following:

- P2A, but Everyday Ed has not yet established his credentials with ELAR as a researcher¹²
- P2B or P2C, but Everyday Ed has not yet established his credentials with the depositor as a member of the deposit's language community (or other specific identity)
- P3 or P4

Thus, Everyday Ed is presented with a button inviting him to "Apply for access rights".

¹² Researcher is a global role across all deposits. Once the role is established for a given user (by any depositor entitled to confer it), that user can access any deposit which allows Researcher access.

Figure 4: User “Everyday Ed” presently has no access rights but he can apply for them.

The screenshot shows a web interface with a navigation bar containing tabs: 'Basic info', 'Depositor', 'Deposit', 'Protocol', and 'Access'. The 'Access' tab is selected and highlighted in yellow. Below the navigation bar, the 'Access' section displays the following information:

- File count:** 1028
- Access:** [Browse](#) the files which you have access to.
- Individual access rights:** For access to more files, you can apply for individual access rights to this deposit. Below this text is a button labeled 'Apply for access rights'.

Once Everyday Ed has applied for access, a permission request is queued in the depositor’s management page. The next time the depositor logs in, she receives a notice about any pending requests. Information already collected from Everyday Ed when he registered as an ELAR user is offered to the depositor to help her to decide.¹³ The depositor is able to confer two types of role on Everyday Ed: as a “subscriber “ to her deposit, which enables him to access its files; or as a member of the relevant language community, which provides cascading rights for other deposits for the same language, thus streamlining the process of accrediting users.

Figure 5: The depositor logs in and is presented with a panel for dealing with the access application.

The screenshot shows a web interface with a navigation bar containing tabs: 'Basic info', 'Depositor', 'Deposit', 'Protocol', and 'Access'. The 'Access' tab is selected and highlighted in yellow. Below the navigation bar, the 'Manage access rights applications' panel displays the following information:

- Manage access rights applications** [Role applications](#) | [User applications](#)
- Manage access rights (for all deposits)**
- Deposit:** [Choquita Rarámuri \(Tarahumara\) description and documentation](#)
- Name:** Everyday Ed
- Information provided at registration:**
 - Institutional affiliation: **SOAS**
 - Community affiliation: I am a native speaker of American English, and a member of the American community in London.
- Access roles:**
 - subscriber
 - community member
- Access rights:** [Approve](#) [Reject](#)

The next time Everyday Ed logs into the ELAR catalogue, his personal page will show him that Chaquita Rarámuri has now been added to the list of deposits for which he has individual rights, and he is offered a link to access files.

Figure 6: The depositor logs in and is presented with a panel for dealing with the access application.

The screenshot shows a web interface with a navigation bar containing tabs: 'Basic info', 'Depositor', 'Deposit', 'Protocol', and 'Access'. The 'Access' tab is selected and highlighted in yellow. Below the navigation bar, the 'Access' section displays the following information:

- File count:** 1028
- Access:** [Browse](#) the files which you have access to.
- Individual access rights:** You have individual access rights to this deposit ([details](#)).

¹³ All users who wish to access any data, whether on open access or otherwise, are required to first register their details with ELAR.

The system maps neatly onto the protocol scheme described in Section 4. For example, the difference between [P3A] and [P3B] maps onto the alternative between the “arm’s length” process illustrated in Figures 4 to 6 (corresponding to [P3B]), and the addition of a messaging facility whereby Everyday Ed and the depositor can communicate and exchange information directly (corresponding to [P3A]). However, the system is currently in initial testing mode only, and we expect it to evolve and improve as we receive feedback from depositors and users and accruing data about the use of the system, such as depositors’ patterns of decisions and the time taken for the parties to respond. We intend to extend the system in various ways including: developing further ways for depositors and users to communicate; allowing users to contribute moderated content; and providing detailed reports to depositors detailing accesses of their materials. We expect the sum of all these developments to represent a shift such that an archive deposit is no longer seen primarily a set of files, but as a dynamic resource at the centre of sharing and discussion.

6. Discussion

Web 2.0 has been described as the shift to the Internet and its users as the interaction platform, rather than software. Its hallmarks are social networking, “network effects” (what happens when participation and interactivity scale up to critical masses enabling new services and businesses to grow), and a preference for open, shared applications and data.¹⁴ The mantra of Web 2.0, according to Tim O’Reilly, the term’s originator, is: “Don’t build applications: build contexts for interaction” (Shuen 2008:9, 101). These three factors are exemplified in the social networking sites Facebook and MySpace, the volunteer-authored Wikipedia, and the “marketplaces” of eBay and Amazon.

Tim Berners Lee, inventor of the World Wide Web, has doubted that Web 2.0 represents an innovation because connecting “people to people ... was what the Web was supposed to be all along”.¹⁵ Keen (2007:2, 50), on the other hand, paints a thoroughly dystopian view of Web 2.0 as a place where “the words of a wise man count for no more than the mutterings of a fool” because it does not distinguish between “audience and author, creator and consumer, expert and amateur”.

However, in applying the description “Web 2.0” to ELAR’s systems, this paper is not trying to simply fly a populist phrase. Rather, I have attempted to show that social networking is indeed a good match to the needs of digital archives and in particular those working in endangered languages, by illustrating how archive access management can be effectively served and enhanced by the new technologies and the conventions that have quickly grown up around them. In Facebook and MySpace, account holders build and participate in virtual communities by choosing who are to be their “friends” - who are in effect the people who are permitted to see and interact with their presence on the site. In the same way, ELAR provides a channel for users to find and approach depositors to request access to materials, and for depositors to decide who will be their “subscribers”. Distinct roles of audience/subscriber and author/depositor are at the heart of ELAR’s design and operation.

¹⁴ Kelly, in Wired 07.09: 122, called this utopian aspect of Web 2.0 “digital socialism”.

¹⁵ See <http://www.ibm.com/developerworks/podcast/dwi/cm-int082206txt.html> [accessed 4 December 2009]

Even if, as Tim Berners Lee believes, the web was people-centred from its beginning, it is extremely unlikely that, without the creation and popularisation of sites like MySpace and Facebook, ELAR would have been able to convince depositors that it was reasonable to expect them to manage archive access themselves - despite the fact that their responses to the protocol scheme (cf Section 4) indicate that this was what they actually wanted.

Recently Oxford University Press USA nominated *unfriend* as “word of the year” for 2009.¹⁶ Its meaning is instantly obvious to anyone with a passing familiarity with social networking; there was no need for an acronym, technical term or neologism, because the conduct of social activity on web platforms is not a metaphor but reflects and complements real life. And note that *unfriend* is not about unconstrained file sharing and tagging, but about drawing boundaries and exerting control over interactions and resources, just as people do to mark their territories of friendship and trust in the physical world.

The vision we have for ELAR could be seen as an amalgam of Web 2.0 archetypes. Firstly, just as in the YouTube/Wikipedia model where members spend much of their own time creating resources, ELAR stores language documentations that are the outcomes of many months or years of work. Secondly, ELAR’s access component implements a Facebook-like model where the “product” is the set of site members and their relationships (Shuen 2008:101). And thirdly, we hope eventually to develop ELAR’s catalogue to support a variety of exchanges between depositors and others, thereby reflecting aspects of the eBay/Amazon marketplace of independent “shopfronts”. Today, language documenters are data managers, but only within the confines of their own (typically individual) projects. Soon, they will be able to extend their reach to manage their resources in a more widely accessible public sphere, a shift that will be welcome amongst the champions of language diversity and language community empowerment.

Language documentation is a young discipline whose methodologies are still being debated. A small number of archives (see Appendix) are the principal repositories for its materials. However, for various reasons, including the sensitivities attached to data, the sum of materials generally available from those archives remains limited. At some point in the future, when those materials become easily accessible, and when linguists begin to use language archives as an academic platform in the same way they exchange ideas at conferences and in journals, we can expect documentary linguistics to take on the appearance of a real discipline. Perhaps this point will be reached suddenly, as a result of a confluence of developments, and it is highly likely that the adoption of Web 2.0 into the operations of archives will be a key component.

References

- DAM-LR. 2006. ‘Live Archives: a checklist of principles and tasks’. Pamphlet of the DAM-LR partners. <http://www.mpi.nl/DAM-LR/>
- Dobrin, Lise M., Peter K. Austin and David Nathan. 2008. Dying to be counted: the commodification of endangered languages in documentary linguistics, *Language Documentation and Description* Vol 6. London: SOAS. [online at http://www.hrhelp.org/publications/ldlt/papers/ldlt_08.pdf]

¹⁶ See <http://blog.oup.com/2009/11/unfriend/> (accessed 22 Nov 09).

- Edward Garrett and David Nathan. 'Digital Language Archiving'. Course presented at 3L Summer School, SOAS, July 2009.
[<http://www.hrelp.org/events/3L/programme.html>]
- Himmelmann, Nikolaus P. 1998. 'Documentary and descriptive linguistics'.
Linguistics 36:161–195.
- Keen, Andrew. 2007. *The Cult of the Amateur: How Today's Internet is Killing our Culture and Assaulting Our Economy*. London: Nicholas Brealey.
- Kelly, Kevin. 2009. "The new socialism". *Wired UK* Edition. 07.09, pp 121-125.
- Nakamura, Lisa. 2008. *Digitizing Race: Visual Cultures of the Internet*. Minneapolis: University of Minnesota Press.
- Nathan, David 2006. "Thick interfaces: mobilising language documentation". In Joost Gippert, Nikolaus Himmelmann and Ulrike Mosel (eds.), *Essentials of language documentation*. Berlin: Mouton de Gruyter. 363-379.
- Shuen, Amy. 2008. *Web 2.0: A Strategy Guide*. Sebastopol CA: O'Reilly.

Appendix: Listing of some endangered languages archives

- Aboriginal Studies Electronic Data Archive (ASEDA). Australian Institute of Aboriginal and Torres Strait Islander Studies.
<http://www1.aiatsis.gov.au/ASEDA/>
- Alaskan Native Language Center Archives (ANLC). University of Alaska.
<http://www.alaska.edu/uaf/anlc/>
- Archive of the Indigenous Languages of Latin America (AILLA). University of Texas.
<http://www.ailla.utexas.org/site/welcome.html>
- Digital Endangered Languages and Musics Archives Network (DELAMAN).
<http://www.delaman.org/>
- Dokumentation Bedrohter Sprachen (DoBeS). Max Planck Institute Nijmegen.
<http://www.mpi.nl/DOBES>
- Endangered Languages Archive (ELAR). School of Oriental and African Studies.
<http://www.hrelp.org/archive>
- Langues et Civilisation et Traditions Orale (LACITO). Centre National de la Recherche Scientifique. <http://lacito.vjf.cnrs.fr/archivage/index.htm>
- Leipzig Endangered Languages Archive (LELA). Max Planck Institute Leipzig.
<http://www.eva.mpg.de/lingua/resources/lela.php>
- Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). University of Melbourne/University of Sydney/Australian National University. <http://paradisec.org.au/>
- Rosetta Project. Long Now Foundation. <http://www.rosettaproject.org/>
- Survey of California and Other Indian Languages. University of California, Berkeley,
<http://linguistics.berkeley.edu/Survey/>