Access and accessibility at ELAR, a social networking archive for endangered languages documentation

David Nathan

Introduction

This paper considers the issue of access in relation to archives holding documentation of endangered languages. Access cannot be reduced to a single parameter, and includes at least the following: the means of finding a resource, the availability of the resource, the delivery of the resource to the user, the relevance and accessibility of resource content to the user, and the user's perceptions of the their experience interacting with the archive and its resources.

The discussion is centered on the Endangered Languages Archive (ELAR) in London and its online catalog.[1] ELAR is one program of the Hans Rausing Endangered Languages Project, based at the school of Oriental and African Studies.[2] It has two sister programs: ELDP (Endangered Languages Documentation Program) provides grants for documentation work, and ELAP (Endangered Languages Academic Program) runs postgraduate courses in language documentation and field linguistics. ELAR opened in 2005 and launched its new catalog system in June 2010. The system uses features pioneered in Web 2.0 or 'social networking' applications (Nathan 2010), so it is innovative only in the sense of applying those features to language archiving. However, we believe that the results represent a true departure from conventional archives in our field.

To simplify things a little, until now access has been thought of as online resource discovery through querying standardized metadata. Where access control has been applied it has typically been based on a formal membership criterion, such as an account on a university's network. ELAR's goal was to provide an archive that is closer tied to the needs of those working with endangered languages, and, of course, the needs of members of those speaker communities. This turned out to be a rich area of exploration, and, coupled with the rise of social networking applications and conventions over the last five years, yielded a system which highlights the nuances and dynamics of access.

Discovering Language Documentation

Language documentation, also known as documentary linguistics, is a subfield of linguistics that emerged in the 1990s as a response to predictions that the majority of human languages will disappear within a century (Krauss 1992). The discipline aims to develop "methods, tools, and theoretical underpinnings for

---

compiling a representative and lasting multipurpose record of a natural language" (Gippert et al 2006:v). It weaves its focus on endangered languages together with traditional descriptive linguistics and a strong emphasis on the use of media and information technologies. It also encourages ethical practices such as involving language speakers as participants and beneficiaries (Grinevald 2003). Its central features are:[3]

- *focus on primary data* – documentation is based around collecting and analyzing a range of primary language data

- *interdisciplinarity* – documentation requires expertise from a range of disciplines, not just linguists. Its data should be available and accessible to a wide range of users

- *involvement of the speech community* – collaboration with community members not only as consultants but also as co-researchers

- *archiving* – materials should be preserved and made available to a range of potential users into the distant future

We can identify the participants and stakeholders in documentation as a prelude to considering what should be provided in terms of access. Firstly, there are the documenters themselves, typically linguists (and, occasionally, academics from other fields) who have received grants to do various kinds of documentation projects, together with the others in their teams who perform the various activities associated with running a project. Crucially, there are the language speakers and consultants, their families and communities. But not to be forgotten are the more peripheral stakeholders such as various institutions who host projects (typically universities) or are interested in evaluating the work or reputation of particular documenters, and governmental authorities interested in language planning. Finally – but most importantly when considering access issues – there are many categories of users: linguists and other researchers, teachers and applied linguists who are interested in resources for language revitalization, heritage users (community members generally interested in resources related to their culture), journalists (who always want poignant stories about last speakers), and, finally, curious people who are interested in all kinds of 'exotica'.[4]

Typically, however, archives in our field have provided a narrow, one-way access strategy, enabling academic documenters to provide materials, and linguistic researchers to access them, as depicted in Figure 1 (Nathan and Fang 2009).

---

[3] For further information, see Austin and Grenoble 2006, Himmelmann 2006:15.
[4] See also Woodbury 2011:162, 177.

Figure 1: A narrow channel between documenters (providers) and linguists (users)
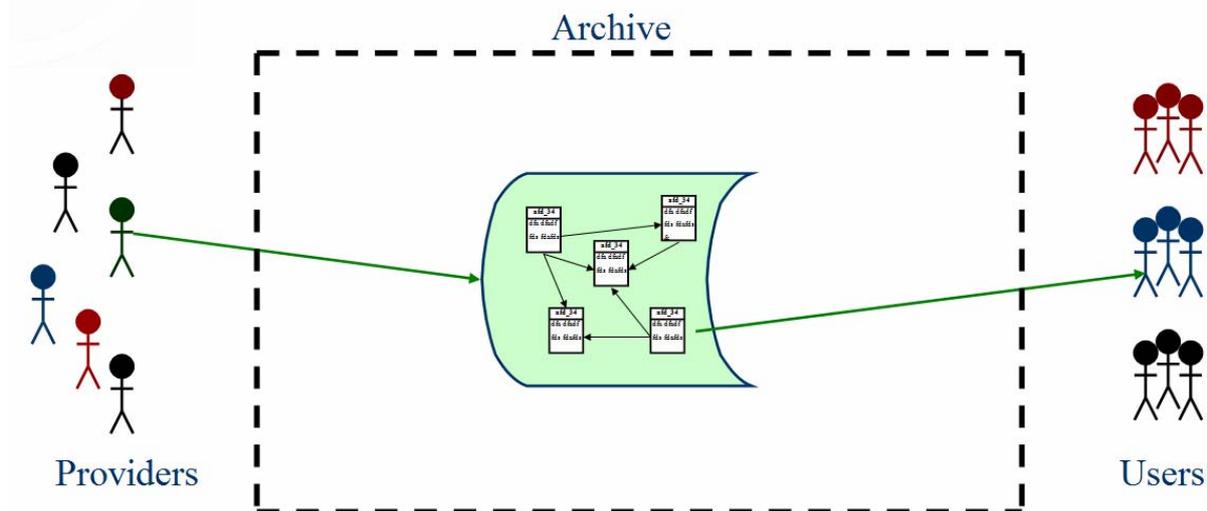
From Documentation to Archiving

At the time we began to develop policies and plans for ELAR around 2004, documentary linguistics was not yet a mature discipline and its archiving needs unclear. Even today many of its basic parameters remain open to discovery rather than being fact or convention:

> documentary linguistics is new enough [so] … that its scope, its scientific and humanistic goals, its stakeholders, participants and practices are still being explored and debated both inside and outside academic contexts … (Woodbury 2011:171)

We asked which aspects of documentation were both central to its practices and relevant to archiving and access. We were able to distil two such characteristics: diversity and protocol.

Himmelmann's seminal description of a language documentation as "a *multipurpose* … record of the linguistic practices characteristic of a speech community" (1998:166; emphasis DN) depicts its methods and outputs as inherently heterogeneous. Such records cannot then conform to a single template. Diversity is most clearly represented in the wide range of projects: ELDP's funded projects range from recording the 'whistled language' of a tiny Amazonian community[5] to a documentation of a language in China with thousands of speakers yet expected to decline quickly.[6] Layered on project contexts are their specific goals; whether, for example, they aim to describe particular linguistic phenomena, focus on annotated recordings, apply ethnomusicological understandings to songs, or create pedagogical resources for language revitalization. Within each project, the cultures, communities and

---

[5] See Julien Meyer's project on the Gaviao and Surui languages, described at www.hrelp.org/grants/projects/index.php?projid=148.

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

individuals with whom the documenter works all bring their unique skills, verbal styles, outlook, and motivations for participation. Documenters themselves are typically lone fieldworkers in remote locations (Austin 2005), so their practices are relatively unharmonized. Finally, of course, languages and their usages vary in yet unknown ways: that is what our awareness of language endangerment and the urgency of documentation tell us, for in truth we know relatively little about most of the world's 7,000 human languages.

Turning to the *form* of documentations, there are few clear conventions for what actually counts as a language documentation (Himmelmann 2006:10, Woodbury 2011:171, 184). We find them containing a wide range of media, text types, and data formats, for which there are few agreed or settled standards; language data are not (yet) captured by an agreed framework of attributes. Compare this situation to that of libraries or businesses whose data is anchored in concepts such as title, author, page, quantity, cost, and item code all of which are well-established, stable, and correspond to real-world objects, rather than the contestable interpretations of linguistics. It is an open question as to whether a universal and stable set of concepts and categories will ever be formulated and agreed, although efforts are being made in that direction, e.g. GOLD Ontology,[7] Leipzig Glossing Rules,[8] and genre inventories (Johnson and Dwyer 2002).

The second key characteristic is *protocol*. ELAR uses this term as shorthand for the sum of processes involved in the formulation and implementation of language speakers' rights and sensitivities, and the consequent controlled access to materials. Protocol extends from the beginning of any documentation activity (e.g. when a documenter seeks informed consent from speakers, and then collects sensitivity and access metadata from them for each recording) through to the mechanisms for providing, restricting, or negotiating about archived materials. To understand its pervasive importance for the discipline, consider that endangered language communities and their speakers are typically under various pressures and deprivations that are also contributing causes to the decline of their languages. These difficulties are amplified by the methodology of documentary linguistics, which most highly values the recording of spontaneous, naturalistic speech. When languages are contracting in their domains of usage (which is what primarily drives endangerment), people tend to use them more and more to speak of private, local, sensitive and secret matters. So the primary data of documentary linguistics maximizes the likelihood of including content that can cause embarrassment or harm to the recorded speakers.[9]

---

[7] See http://linguistics-ontology.org/.

[8] http://www.eva.mpg.de/lingua/resources/glossing-rules.php.

[9] In addition, documenters, unless they are community members, are likely to know less about sources of sensitivities – and are therefore less able to avoid them – than they might in other research contexts.

A Documentation Archive

Archiving is an integral part of language documentation, for it would be pointless to document endangered languages without securing the safety and sustainability of the recorded data (Bird and Simons 2003). Today, several archives are devoted to endangered languages documentation.[10] Most of these are digital archives because documentation is inextricably linked with digital technologies in four ways: digital recording has made portable, high quality recording affordable; long term preservation of audio and video is possible only through lossless digital copying (IASA 2005); most researchers use computers to annotate media and create data and analysis in general; and the World Wide Web has become the ubiquitous platform for accessing documentation materials.

A digital documentation archive has to be more than a data repository. It has to find ways to preserve diverse materials and disseminate (or publish) them to a variety of stakeholders while safeguarding access where required. Most archives have collection policies (Conathan 2011:240), some have policies which describe the types of access offered[11] or classes of users who they exist to serve,[12] however few explicitly link the architecture of their access system with the characteristics of their users. ELAR has done the latter by designing an archive with 'Web 2.0' (also known as 'social networking') features:

> archive access management can be effectively served and enhanced by the new [Web 2.0] technologies and the conventions that have quickly grown up around them. In Facebook … 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 … (Nathan 2010:122).

In this design, the archive is reconceived as a platform for building, maintaining and conducting relationships between information providers and their users, just as many libraries see their mission as supporting learning rather than tending books.[13]

ELAR aims to 'level the playing field' by offering more equitable access to various types of users rather than privileging the single-channel provision to researchers. We can cater better for language-speaker community members in several ways. The first is through our implementation of a nuanced protocol system to manage access and provide security and accountability. In Figure 1, green arrows show the workflow through a traditional archive; providers lodge their materials with the archive and users can (if permissions allow) find and access them. The archive functions as a searchable container for those materials.

---

[10] For a list, see http://www.delaman.org/participants.html.

[11] See, e.g. http://www.mpi.nl/DOBES/ethical_legal_aspects/DOBES-access-v2.pdf.

[12] http://www.ailla.utexas.org/site/welcome.html.

[13] See, e.g. http://www.lib.uchicago.edu/e/about/mvv.html.

ELAR uses Web 2.0 interactivity to provide a dynamic access process. Depositors can edit metadata for their collection at any time, including the metadata that governs access. More importantly, the archive 'plays out' protocol throughout its interface (see Figures 3-6), always letting users know which resources they can and cannot access, and offering a method for individual access to otherwise restricted resources through direct application to the depositor (via 'subscription': see below). A simplified representation of ELAR's subscription process is shown in Figure 2.

Figure 2: Dynamic access via 'subscription' at ELAR
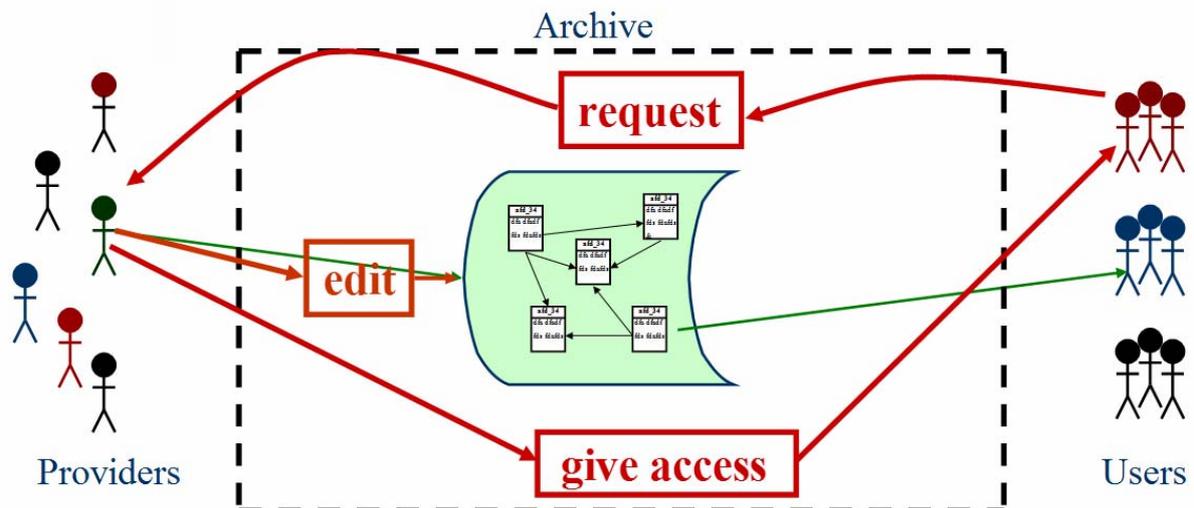
URCS protocol roles

Before further outlining how ELAR's system works, I describe its set of protocol roles. The protocol system is based around four roles (U, R, C and S) that have been defined as a result of research into depositors' preferences and through consultation with groups of depositors and archivists (Nathan 2010).

- U = ordinary User (must have an ELAR account)
- R = Researcher role
- C = Community member role (for a particular deposit only)
- S = Subscriber role (for a particular deposit or resource only)

Users are those people who have created an ELAR account. ELAR staff check account applications for bogus or scam attempts, but applications are in general automatically approved. Researcher role is available to relevant practitioners, for example linguists or teachers; applications for Researcher role are evaluated by ELAR staff and if approved apply across all collections in the archive.

Community member and Subscriber roles, however, are granted in relation to particular collections, and these applications are evaluated by the relevant depositor (or the depositor's delegate). A Community member is, as the

name implies, someone recognized as a member of the language-speaker community. This category can also be used to set up other community-oriented categories such as a family, a set of individuals, or any other group that a depositor and his/her language consultants permit to access their data.[14]

A <u>S</u>ubscriber is anyone who has identified a resource, requested permission to access it, and had their request approved by the depositor (see Figures 5 and 6). When a user submits a subscription request, the request is queued in the depositor's collection management panel. The depositor can see which item is being requested, together with information about the user (information about that the user entered when they first registered for an account, including the user's identity, affiliation, and a statement about involvement with endangered languages). For further information, see *ELAR's access protocol* in ELAR's help system.[15]

Depositors can also use the subscription system as a managed sharing mechanism (e.g. for limiting access to a project team).

The subscription system is a significant breakthrough in terms of broadening access to sensitive materials that in other archives would be under closed access. <u>S</u>ubscription applications are channels for communication between owners and potential users of resources: in other words, users and depositors *gain access to each other*.

How Protocol Works

As users navigate the ELAR website, its management system matches the URCS values of the resources in focus with the URCS rights of the logged-in user. Anyone can view a collection home page (see Figure 3), and see a resource's metadata, but only logged in account holders can access ELAR resources. Although requiring accounts limits wider access to ELAR's open (U) resources, we think this is a cost worth bearing. As described above, the subscription process supplies depositors with reliable information about requesters, including validated identities and archive usage history.[16] We do not support user anonymity; rather, we provide depositors with information about access of their collections. These components of a protocol system help to build and maintain a high level of trust and confidence on the part of depositors and their language consultants.

---

[14] Currently, eligibility for access under <u>C</u>ommunity member is decided by the depositor or depositor's delegate. We hope to develop a more flexible approach to managing this role in the future. The AILLA archive, for example, has a system using special passwords as answers to questions that only eligible community members would know (http://www.ailla.utexas.org/).

[15] http://elar.soas.ac.uk/help.

[16] We are in the process of implementing some of these features.

Figure 3: ELAR Home page for *Documentation of ... Galesh*, protocol labels and controls circled


As can be seen from Figure 3, we made a bold commitment to make protocol a prominent feature of the archive interface. It inverts the navigational design of other archives where one searches and navigates to a resource of interest, only to be faced by a 'not available message or a pop-up demanding a log in to an unknown service; users do not discover that a given resource is closed until having *completed* a possibly complex search. In such archives it can even be difficult for depositors themselves to know what access conditions currently hold for their own materials.

How does a user make use of ELAR's protocol information? Information at the top right of the collection's Home page (Figure 3) provides an overview, showing the default access protocol for the collection, together with the default access rights for the presently logged-in user. For performing search/navigation, controls are provided in the navigation panel. These also give more information. Figure 4 shows the user that 37 resources are available (because 'U' is outlined in solid green), while three Subscriber-only resources are unavailable (indicated by the 'S' in dotted red outline).

Figure 4: Access protocol controls in the navigational panel

Access protocol
U R C S (37)
U R C S (3)
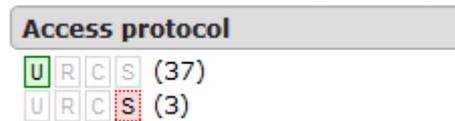
Users who only want to be shown resources for which they have access rights can thus search or browse by clicking on the appropriate protocol category. On the other hand, if a user browses all resources and reaches one which is Subscriber-only, he/she is offered an option to *Apply for access rights*, which, if clicked, triggers the subscription application process described above.

Figure 5: Display of a Subscriber-only resource

▼ **Documentation of the language and lifestyle of the Galesh**

Film from Ziarat.wmv   Access protocol: U R C **S**

You do not currently have access to this file. Apply for access rights
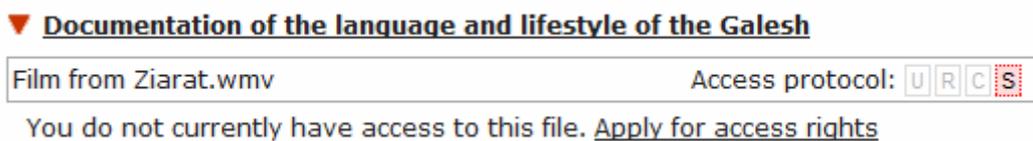
After a subscription application is approved by the depositor, the user will see the 'S' icon outlined in green, as shown in Figure 6, where a (different) resource is available, in this case an audio file which can be either played or downloaded.

Figure 6: This user has subscription rights to this resource

▼ **Cornelio 2**

co_cornelio2.wav   Access protocol: U R C **S**

◀ ▶
download

Users of this system are always aware of their access protocol context. They can choose to only search for accessible items, or they can request access to items where necessary. And at any point users know *why they can or cannot access* particular resources.

Searching, Browsing and Metadata

So far I have described the role of protocol in navigating ELAR's resources. ELAR also provides search and browse functions. Its search is fairly standard, offering a stemmed search over all archive metadata.[17] ELAR places higher priority on enabling users to browse. Browsing reflects the diversity of documentation; with its wide array of resources, formats, and metadata, users need a way to find out what is available. Browsing provides a user friendly 'road

---

[17] Stemmed search means that terms are searched according to their uninflected forms; for example, searching for "cats" will find all resources containing "cat" in their metadata and vice versa.

map' rather than potential responses to specific queries. It is implemented using a dynamic 'faceted browse' system, visible in the left hand panel in Figure 3; a detail for another collection appears in Figure 7.

Figure 7: User-friendly discovery using faceted browsing

Topic *less* ▲

Ais Island (1)
Akalao Bird and Daughter (1)
Akalao and Mother (1)
Aore Island (1)
Before Going to War (1)
Bird Story (7)
Cardinal (2)
Chatting (9)
Circumcision (1)
Coconut Oil (3)
Conch and Sea Snail (1)
Devilish Pig (2)
Directions (1)
Dying (1)
Engagement (1)
First Coconut (1)
Five Fingers (1)
Flying Fox and Parrot (2)
Laplap (2)
Linguo-labials (2)
Numbers (1)
Pig Attack (1)
Pig-killing Ceremony (1)
Piria (1)
Pledge (1)
Plover and Red-head Bird (1)
Prawn (2)
Rat, Short-leg and Octopus (1)
Sickness (9)
Six Sisters (1)
Surae (1)
Swadesh (2)
Talk (62)
Troll (1)
Turtle and Old Man (1)
Turtle and Shark (2)
Tutuba Wild Man (1)
Two Wild Men (1)
Wedding (1)
Where Wild Things Are (5)
White Heron (1)
Wild Apple (1)
less ...

There are, of course, good arguments for providing search over standardized metadata – for example ISO 639 codes enable users to accurately find all resources for a certain language, despite the variety of names it might have. [18] Such strategies have been the backbone of traditional library and indexing practice. But it is important to remember that while they serve certain classes of users and purposes very well, they also diminish access to other users and purposes. Researchers, for example, are likely to know – or know how to find – standard codes for languages. Searches via such codes yield high recall (returning most of the relevant resources, not missing many) and high precision (returning relevant resources, with few irrelevant ones). However, for many of the users and purposes we wish to serve, query interfaces provide low recall due to their "ontological flatness" (Christie 2005: 13).

A non-researcher language community member, for example, is likely to get better results when looking for a story about a particular animal or place if they can see the names of the animal or place displayed, and even better results if the colloquial or language term for that animal or place is shown (rather than, say, the scientific or official name). Depending on the level of literacy in community, even the colloquial or language terms may not normally be written, or may have variant spellings, so users' are better supported by being able to browse and select rather than being forced to type in search strings.

---

[18] For example, Ethnologue (http://www.ethnologue.com/) lists 7413 "primary" language names, but these have over 43,000 alternate (and dialect) names. However, many of theses names are exonyms (i.e. not the community's own term for their language).

Metadata underlies these searching and browsing functions. ELAR takes a permissive approach to metadata, encouraging each depositor to supply as rich and descriptive set as possible (Nathan 2011). ELAR also attempts to expose as much as possible of this metadata. Examples can be seen in Figures 3 and 7, where topics include butter, cheese, and pigs. In other cases, terms in local languages, such as Kastom,[19] or phonetic terms and symbols[20] appear.

Here is an example showing how ELAR's approach works to 'level the playing field'. If depositors provide names of the speakers/performers of recordings, these can be displayed for browsing on the collection's home page (see under 'Participants' in Figure 3). Speakers now appear right 'up front' in the interface; their status is represented similarly to that of the depositor. Community members – or others with no connection to the documenter or linguistic goals – can find and browse performances by those speakers, without having to remember the name of a fieldworker who once visited, the linguist's name for the project, or the ISO code for their language.

Access and Accessibility

ELAR's approach to protocol, search, and browse aims to enhance access, but we have not yet asked the question: what *counts* as access? Searching and browsing, and file display or download, are not ends in themselves. Ultimately, access has to take into account *accessibility to the content* of interest to users. Different people want different things. Depending on users' goals, and the content they desire, access could mean viewing metadata, playing an audio or video in the browser, or downloading the file to play or manipulate it later (see Figure 8). Formal linguists might want to download interlinearized marked-up material; community members might want to 'click and play' recordings of songs, stories, and events; language planners or educationalists might want to assess the range and quality of the available resources.

Some people mistakenly look to in-browser delivery as a strategy to *prevent* users receiving a digital copy of a file. This confuses access to content with the apparatus that delivers that content.[21] Instead, we have to shift our focus from access to accessibility. Take, for example, someone with little technical interest in their internet-connected computer who wants to learn a song. A simple 'play' button will maximize the accessibility of the song. But someone who wants to acoustically analyze speech or transcribe it in specialized software like ELAN[22] will not be able to do so without downloading.

---

[19] See http://elar.soas.ac.uk/deposit/guerin2007mavea.

[20] See http://elar.soas.ac.uk/deposit/connell2010somyev.

[21] In these days of 'always on' broadband, cloud computing, and a myriad of software for capturing YouTube media, in-browser media players are not a bulwark against file download or copying.

[22] http://www.lat-mpi.eu/tools/elan/

Figure 8: Playing a video in the browser – is *this* access?[23]

Providing accessibility goes beyond allowing a choice between playing and downloading; suitable renditions of content might need to be made for different audiences (Nathan 2006, Holton 2011). Not all users want, or can use, audio or video with time aligned morphological annotation. Eli Timan's ELAR collection [24] includes time aligned morphological annotation, but it is accompanied by a community resource that forgoes most of the 'linguistic' content, and provides what Eli, as a community member himself, knows that they might use: transliteration in Arabic and translation into English, together with pictures drawn by the story teller.[25] Another alternative we are working on is an in-browser video player (see Figure 9) that uses speech bubbles, a very conventional (and therefore accessible) method to present the written content of a conversation.[26]

---

[23] From the collection http://elar.soas.ac.uk/deposit/caballero2009raramuri.

[24] http://elar.soas.ac.uk/deposit/timan2008jewishiraqi.

[25] See these materials at http://jewsofiraq.com/.

[26] The speakers, conversing in Pite Saami, are Henning Rankvist (left) and Elsy Rankvist (right). From an ELAR collection deposited by Joshua Wilbur (http://elar.soas.ac.uk/deposit/wilbur2009pitesaami). The speech bubble player was created by Edward Garrett.

Figure 9: Experimental speech bubble player
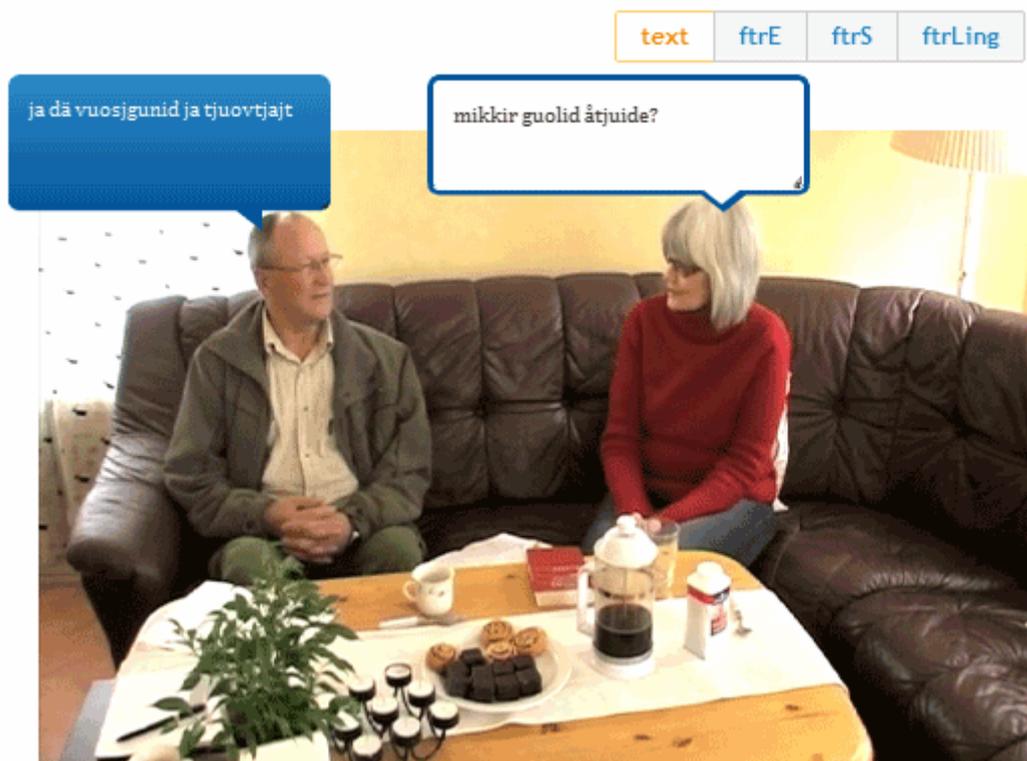
Perceptions and the Interface

Accessibility also depends on users' perceptions. Much of this paper has been about the nature of an archive's user interface; its design, layout, interactivity, controls and navigation. While many of these factors are based on underlying functional decisions, the overall effect – often called 'the user experience' – is greater than the sum of such decisions. Interface design plays a significant role in achieving goals. ELAR chose a contemporary look, echoing features of Facebook and blogs, because these genres reduce the perception of distance and power disparity, and encourage productive interaction (Bozarth 2010:55). Another example was given earlier; ELAR's decision to prominently signpost protocol throughout the website. We did this not only to embody our commitment to depositors' protocol choices, but also because users need guidance to take advantage of ELAR's new approach to access.[27]

Sometimes things play out in unpredictable but serendipitous ways. Recently a researcher described a West African community's responses to some archive websites. The community has only recently been connected to the Internet, and they mainly use sites such as Facebook, for social purposes. So for them, a prototypical website looks and works like Facebook, and after being shown a few online archives, they judged that ELAR was the only 'real' one.

---

[27] New to language archives, although we borrowed heavily from the social networking genre (Nathan 2010).

Interfaces can be misleading. For example, archives may give a false perception of access control. Some linguists believe that a particular language archive does not allow downloading of files, although investigation revealed that it is quite possible to download from that archive. The opacity of the archive's interface makes it so difficult to accomplish a download that they had concluded that it was impossible. This situation disadvantages those who legitimately want to access materials and gives a false sense of security to depositors who imagine a level of control which does not exist. In this case, perceptions have conflated difficulty of access with control of access.

Interfaces can also be subtle and unpredictable. Nariyo Kono's documentation of Kiksht[28] (Warm Springs, Oregon USA) contained sensitive materials, so they were deposited at ELAR under Subscriber-only access, available only to the depositor and the small community team she worked with. However, after the collection was accessioned and online, and the community members saw themselves displayed, they felt uncomfortable and wrote urgently to ask us to 'turn off' access. I replied, explaining the benefits of them being able to see and check the site before allowing others to access it (or indeed to decide against access). However, I had misunderstood; the fact that they could see themselves appearing in the browser, on the screen – in the place where normally only 'others' appear – was disturbing. We negotiated time to allow further discussion back in Warm Springs, and after a month the go-ahead was given to re-open the collection to community members only.[29]

Conclusion

The issue of access to archive resources is multifaceted, and goes far beyond designating resources as open or closed. I have illustrated some of the advantages of custom solutions for a specific field – here, endangered languages documentation. The central concept is a nuanced set of protocol values 'URCS', of which two describe a relation between an individual user and a particular resource which is negotiated between the user and depositor. We have not yet encountered a case where these roles and their associated mechanisms did not provide an appropriate solution for the protocol needs of a depositor or community. In fact we have been surprised at the number of apparently complicated cases that can be handled by the flexibility of the Subscriber role.

The response from depositors to ELAR's access system has been unanimously positive. Some have elected to deposit materials with ELAR that they would not deposit elsewhere, because our attention to protocol has inspired their trust. Others have approached ELAR for archiving as a result of searching for an archive with such a model for protocol and accountability. Some depositors who are preparing collections for deposit, on realizing that ELAR can

---

[28] http://elar.soas.ac.uk/deposit/kono2009kiksht

[29] I am grateful to Nariyo Kono, Valerie Switzer, Radine Johnson, and Pam Cardenas for sharing their views of this experience with me, and I apologize for any errors or remaining misunderstandings.

directly provide resources to the communities they work with, have reshaped their collections and revised their metadata to take advantage of the systems described here.

There is still much work to do. Depositors can edit the content of their collection Home page (Figure 3) to add translations in the documented language or a lingua franca,[30] but we would also like to be able to present the whole navigational interface in a variety of languages.[31] With our small team we do not have the resources to accomplish that, but some depositors have already offered to help. It would be great to complete the social networking dynamic by allowing users to contribute comments, links and materials, and to collaborate with depositors, but all of these moves will require careful consideration of moderation and protection of moral rights and intellectual property.

Until now, access has more or less meant providing 'insiders' with the means to locate specialist materials by using constrained ontologies. ELAR has sought to help 'outsiders' to access content they hope to find or perhaps never imagined finding. In doing so we are replacing a 'stork and baby' approach to archiving – deposit and abandon – with a platform for ongoing relationships and activities around the data. This does require an increased commitment on the part of depositors, but it may result in an enrichment of documentary linguistics and greater support for speakers of endangered languages.

David Nathan
Endangered Languages Archive
School of Oriental and African Studies
University of London

References

Austin 2005 Peter K. Austin. "Training in language documentation: The SOAS experience." Unpub. paper presented at the Linguistics Society of America Conference on Language Documentation: Theory, Practice, and Values. July 9-11. Harvard University.

Austin and Sallabank 2011 Peter K. Austin and Julia Sallabank eds. *The Cambridge Handbook of Endangered Languages*. Cambridge: Cambridge University Press.

Bird and Simons 2003 Steven Bird and Gary Simons. "Seven dimensions of portability for language documentation and description." *Language*, 79(3):557-582.

Bozarth 2010 Jane Bozarth. *Social Media for Trainers: Techniques for Enhancing and Extending Learning*. San Francisco: Wiley 2010.

---

[30] For example, Shenkai Zhang has provided summary information in Chinese for her ELAR deposit; see http://elar.soas.ac.uk/deposit/zhang2010pingjiang.

[31] AILLA, for example, enables users to toggle the entire interface between English and Spanish (see Footnotes 12 and 14).

Christie 2005 Michael Christie. "Aboriginal Knowledge Traditions in Digital Environments." Unpub. ms. Charles Darwin University. http://www.cdu.edu.au/centres/ik/pdf/CHRISTIE_AJIEpaper.pdf.

Gippert et al. 2006 Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel eds. *Essentials of language documentation* (Trends in Linguistics. Studies and Monographs, 178). Berlin: Mouton de Gruyter, 2006.

Himmelmann 2006 Nikolaus P. Himmelmann. "Language documentation: What is it and what is it good for?" In Gippert et al. 2006:1-30.

Holton 2011 Gary Holton. "The role of information technology in supporting minority and endangered languages." In Austin and Sallabank 2011:371-399.

IASA 2005 IASA Technical Committee, *The safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*, Dietrich Schüller (ed.). Version 3, 2005 (= Standards, Recommended Practices and Strategies, IASA-TC 03). International Association of Sound and Audiovisual Archives. http://www.iasa-web.org/tc03/ethics-principles-preservation-strategy.

Johnson and Dwyer 2002 Heidi Johnson and Arienne Dwyer. "Customizing the IMDI Metadata Schema for Endangered Languages." In *Proceedings of The International Conference on Language Resources and Evaluation* 2002. http://www.mpi.nl/lrec/2002/papers/lrec-pap-05-JohnsonDwyer.pdf.

Krauss 1992 Michael Krauss. "The world's languages in crisis." *Language*, 68:6-10.

Nathan 2006 David Nathan. "Thick interfaces: mobilising language documentation." In Gippert et al. 2006:363-379.

Nathan 2010 David Nathan. "Archives 2.0 for Endangered Languages: from Disk Space to MySpace." In *International Journal of Humanities and Arts Computing*, 4.1-2:111–124.

Nathan 2011 David Nathan. "Better data about metadata: A survey of depositor metadata submitted to the Endangered Languages Archive." Unpub. paper presented at the Annual Meeting of the Linguistic Society of America, Organized Session on Metadata, Pittsburgh, 7 January 2011.

Nathan and Fang 2009 David Nathan and Meili Fang. "Language documentation and pedagogy for endangered languages: a mutual revitalisation." In *Language Documentation and Description, Vol 6*. Ed. by Peter Austin London: SOAS. pp. 132-160.

Woodbury 2011 Anthony Woodbury. "Language Documentation." In Austin and Sallabank 2011:159-211.