elpublishing

Language Documentation and Description

ISSN 1740-6234

This article appears in: *Language Documentation and Description, vol 12: Special Issue on Language Documentation and Archiving*. Editors: David Nathan & Peter K. Austin

# Access and accessibility at ELAR, an archive for endangered languages documentation

DAVID NATHAN

Cite this article: David Nathan (2014). Access and accessibility at ELAR, an archive for endangered languages documentation. In David Nathan & Peter K. Austin (eds) *Language Documentation and Description, vol 12: Special Issue on Language Documentation and Archiving*. London: SOAS. pp. 187-208

Link to this article: http://www.elpublishing.org/PID/144

This electronic version first published: July 2014

# Access and accessibility at ELAR, an archive for endangered languages documentation

David Nathan

Department of Zoology, Oxford University

## 1. Discovering language documentation[1]

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 new 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:[2]

- *focus on primary data* – it is based around collecting and analysing a range of primary language data
- *interdisciplinarity* – it requires expertise from a range of disciplines, not just linguists. Its data should be available to and useful for a wide range of researchers and other users
- *involvement of the speech community* – it encourages collaboration with community members not only as consultants but also as co-researchers and end-users

[1] This article is a revised and updated version of Nathan 2013. It focuses on access to language resources – the discovery, identification, downloading, viewing, negotiation of restrictions, and engaging with the content of resources – rather than end-usage of those resources. It draws on the designs and experiences of the Endangered Languages Archive at SOAS University of London; see elar-archive.org. Several important and complementary issues such as intellectual property, copyright, privacy, and usage licenses are beyond the scope of this paper. See Conathan (2011: 250) for an overview of these from an archive perspective; a good web source on some of these issues is Michael Brown's website Who owns native culture? (2012) web.williams.edu/go/native/ [accessed 16 August 2012]

[2] For further information, see Austin and Grenoble (2007), Himmelmann (2006: 15).

David Nathan 2013. Access and accessibility at ELAR, an archive for endangered languages documentation. In David Nathan and Peter K. Austin (eds.) *Language Documentation and Description, Vol 12,* 187-208. London: SOAS.

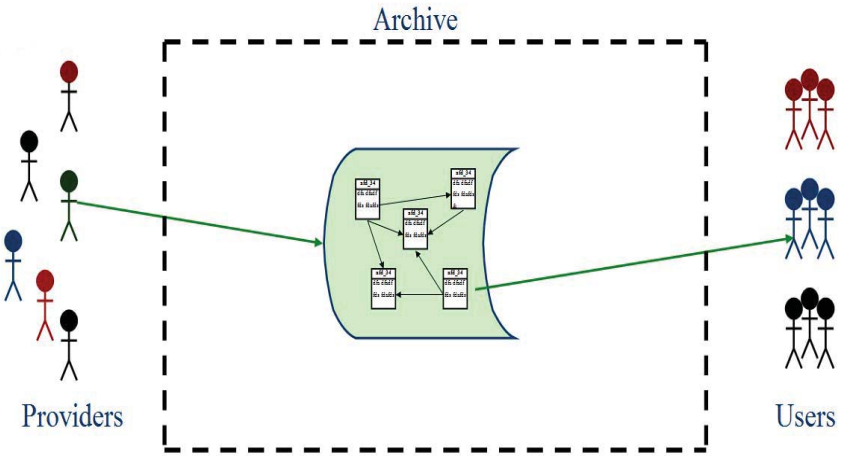
- *archiving* – its outcomes should be preserved and made available to a range of potential users into the distant future.

We can identify language documentation's participants and stakeholders 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 who host, support and provide language knowledge and performances to these documenters (in some cases community members are also the documenters). 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 importantly when considering access issues – there are many categories of users (which can overlap with the above-mentioned): linguists and other researchers, teachers and applied linguists who are interested in resources for language revitalisation, heritage users (community members broadly 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'.[3]

Typically, language archives have provided a narrow, one-way access strategy, with academic documenters providing materials, and linguistic researchers accessing them, as depicted in Figure 1 (Nathan and Fang 2009).

[3] See also Woodbury (2011: 162, 177).

*Figure 1: Traditional digital archiving: a narrow and one-way channel between documenters (providers) and linguists (users)*




## 2. From documentation to archiving

When policies and plans for The Endangered Languages Archive (ELAR)[4] were in initial development around 2004-5[5] documentary linguistics was not yet a mature discipline and its archiving needs were unclear. Still today many of its basic parameters remain open to discovery rather than being fact or convention; Woodbury (2011: 171) advises that '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'. ELAR staff asked which aspects of documentation were both central to its practices and relevant to archiving. We distilled two characteristics: *diversity* and *access protocol*.

[4] The ELAR in London, and its online catalogue is one programme of the Hans Rausing Endangered Languages Project, based at the School of Oriental and African Studies. For ELAR see www.elar-archive.org [accessed 16 August 2012]. ELAR's archiving activities are complemented by training, depositor support, outreach, and publishing. For the Hans Rausing Project, see www.hrelp.org [accessed 16 August 2012]

[5] ELAR opened in 2005 and launched its current catalogue system in June 2010.

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. It follows that such records cannot conform to a single template. Diversity is most clearly represented in the wide range of projects which form the main source of ELAR's deposits; the grantees of the Endangered Languages Documentation Programme (ELDP). ELDP's funded projects range from recording the 'whistled language' of a tiny Amazonian community,[6] to a documentation of a language in China with thousands of speakers yet expected to decline quickly.[7] 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 revitalisation. Within each project, the communities and 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 2008, Crippen and Robinson 2013), so their practices are relatively unharmonised. 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 the majority 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; and 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

[6] See Julien Meyer's project on the Gaviao and Surui languages, *Documentation of Gaviao and Surui Languages in whistled and instrumental speech* www.hrelp.org/grants/projects/index.php?projid=148 [accessed 16 August 2012]

[7] See Ross Perlin's project on the Dulong language, *Documentation and description of Dulong* www.hrelp.org/grants/projects/index.php?projid=123 [accessed 16 August 2012]

direction, e.g. the ISOCat Data Category registry,[8] GOLD Ontology,[9] Leipzig Glossing Rules,[10] and genre inventories (Johnson and Dwyer 2002).

The second key characteristic is *access protocol*. ELAR uses this term as shorthand for the sum of activities involved in the formulation and implementation of language speakers' rights and sensitivities, and the consequent methods and processes for controlled access to materials. The relevance of protocol extends from the beginning of any documentation activity (e.g., when a documenter seeks informed consent from speakers, and then collects metadata on sensitivity and access from them for each recording) through to the mechanisms used by the archive for providing, restricting, or negotiating about archived materials. To understand the pervasive importance of protocol for language documentation, 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. The importance of addressing access protocol is amplified by the nature of documentary linguistics, whose methodology most highly values the recording of spontaneous, natural speech. As languages cease to be spoken in a broad range of contexts (which is a primary marker of endangerment), people tend to use their language more and more to speak of private, local, sensitive and secret matters. So the primary data of documentary linguistics maximises the likelihood of including content that can cause discomfort or harm to the recorded speakers.[11]

## 3. 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 materials (Bird and Simons 2003). Today, several archives are devoted to endangered languages documentation.[12] Most

[8] See www.isocat.org/index.html.

[9] See linguistics-ontology.org [accessed 16 August 2012].

[10] See *Leipzig Glossing Rules*, Max Planck Institute for Evolutionary Anthropology Department of Linguistics www.eva.mpg.de/lingua/resources/glossing-rules.php [accessed 16 August 2012]

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

[12] For a list see *Digital endangered languages and musics archive network* www.delaman.org/participants.html [accessed 16 August 2012]

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; and the World Wide Web has become the ubiquitous platform for disseminating and 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[13] or classes of users who they exist to serve,[14] however few explicitly link the architecture of their access system with the characteristics of their holdings and their users. ELAR has done so by designing an archive with Web 2.0 (or social networking) features, as Nathan (2010: 122) points out:[15]

> 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.

---

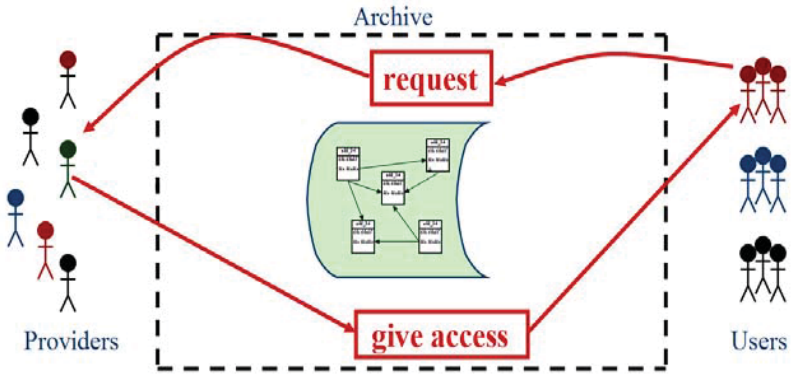
[13] For example Peter Wittenburg (2005) *Data Access and Protection Rules DAPR-V2* www.mpi.nl/DOBES/ethical_legal_aspects/DOBES-access-v2.pdf [accessed 16 August 2012]

[14] See The Archive of the Indigenous Language of Latin America www.ailla.utexas.org/site/welcome.html [accessed 16 August 2012]

[15] Although Facebook (www.facebook.com) is used here to exemplify social networking, Google+ (plus.google.com), released in June 2012, better resembles ELAR's model because it distributes a user's friends into 'circles' just as an ELAR deposit provides various user roles; Facebook generally treats all a user's friends as a single group, although it is possible to set up and edit a list of 'close friends' [both accessed 17 August 2012]

In this way, the archive is reconceived as a platform for building and conducting relationships between information providers and information users, just as many libraries see their mission as supporting learning rather than merely tending books.[16] A simplified representation of ELAR's subscription process is shown in Figure 2.

*Figure 2: Dynamic access via subscription at ELAR*




ELAR also 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 that was illustrated in Figure 1. ELAR caters for language speakers and community members in several ways. The first is through implementation of a nuanced protocol system to manage access and provide security and accountability. Figure 1 shows the workflow through a traditional archive; providers lodge their materials with the archive and users can (if permissions allow) find and access them. That kind of traditional archive (whether physical or digital) functions as a searchable container for those materials. ELAR uses Web 2.0 interactivity to provide a dynamic access process. Depositors can edit their collection's metadata 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 subscription access to otherwise restricted resources through direct application to the depositor.

[16] See, e.g. *Library Mission, Vision and Values*, The University of Chicago Library (2004), www.lib.uchicago.edu/e/about/mvv.html [accessed 17 August 2012]

## 4. URCS protocol roles

Before further outlining how ELAR's protocol implementation works, I describe its set of roles. The system is based around four roles (U, R, C and S) that were defined following 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
C = Community member (for a particular deposit or resource)
S = Subscriber (for a particular deposit or resource)

Users are those people who have created an ELAR account. Accounts are freely available by applying online and are automatically enabled. The integrity of accounts is protected by anti-spam measures in the application process, and in addition ELAR staff check all applications for spam/scam/bogus attempts. Researcher role is available to relevant practitioners, such as linguists or teachers; applications for Researcher role are individually evaluated by ELAR staff and if approved the user can access all R-marked resources in the archive.

Community member and Subscriber roles, however, are granted in relation to particular collections, and applications are evaluated by the relevant depositor (or the depositor's delegate). A Community member is, as the name implies, someone recognised as a member of the language-speaker community. This category can also be used by the depositor to set up other community-like 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 materials.[17]

A Subscriber is a user who has identified a resource in the ELAR catalogue, 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, they can write a message outlining the nature of the request or intended usage. The request is queued in the depositor's collection

[17] Currently, eligibility for access under Community 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. See www.ailla.utexas.org/ [accessed 17 August 2012]

management panel, where the depositor can see, for each such subscription request, the item being requested, together with the user's message and the user's profile (information that the user entered when they first registered for an account, including their name, affiliation, and a statement about association with endangered languages). The depositor uses this information to decide whether or not to permit access, and can input the decision directly into their collection management panel. In addition, the depositor can also write a message for the applying user, to let them know the reason for the decision, or to provide further information. This innovative message-passing function has proved to be extremely valuable: a survey of the messages being transacted showed that depositors and users were sharing useful information about how to use materials, further sources, research questions, and suggestions for meetings and collaborative work.[18] Depositors can also use the subscription role to establish managed sharing (e.g., for providing access only to a project team).

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

For further information about ELAR's access protocol, see ELAR's website.[19]

## 5. How protocol works

As users navigate the ELAR website, its management system matches the access rights of the logged-in user with the access protocol (URCS) values of the resource(s) that the user has in view. Anyone can view a collection home page (see Figure 3), and see a resource's metadata, but only registered account holders can view or download the actual resources: audio, video, text, images etc.. Although requiring users to register with ELAR does limit spontaneous access to ELAR's open (U) resources, this is a cost worth bearing. ELAR does not support user anonymity; rather, we provide depositors with information about access of their collections. User profiles supplied at registration enable

[18] The messages are sent only between the relevant user and depositor, but are also tracked by ELAR's catalogue system and accessible to ELAR staff for research purposes and to ensure that the system is used appropriately.

[19] See www.elar-archive.org/using-elar/access-protocol.php [accessed 1 November 2013]

depositors to be provided with reliable information about subscription requests (depositors are also provided with other information such as a user's archive usage history). These components of a protocol system help to build and maintain trust and confidence amongst depositors and their language consultants.

*Figure 3: ELAR Home page for Documentation of the language and lifestyle of the Galesh, Carina Jahani. Protocol labels and controls circled. (elar.soas.ac.uk/deposit/0094).*

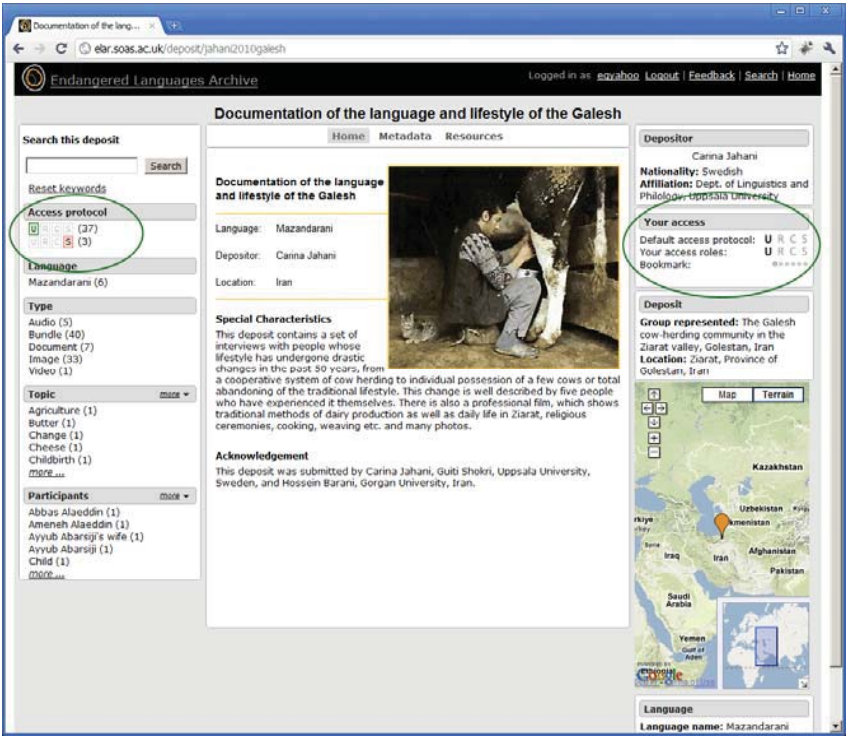


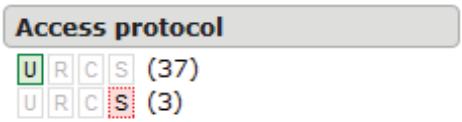

Figure 3 illustrates ELAR's bold commitment to making protocol a prominent feature of the archive interface.[20] The interface inverts the

[20] ELAR's systems are in continuous improvement and some details of the present interface will differ slightly from screenshots shown in this paper. Note that web projects by Kimberley Christen and her research partners have also focused on interfaces for cultural protocols 'that both limit and enhance the exchange, distribution

navigational design of archives where one searches and navigates to a resource of interest, only (if the resource is not open) to be ultimately faced by a 'not available' message or a pop-up demanding a log in to an unknown service. In such archives users do not discover that a given resource is closed until having *completed* a possibly complex search.[21]

How does a user make use of ELAR's protocol information? Information at the top right of the collection's Home page (see 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. Search/navigation controls are provided in the navigation panel. These also give more information about access. In Figure 4, the user is shown 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*



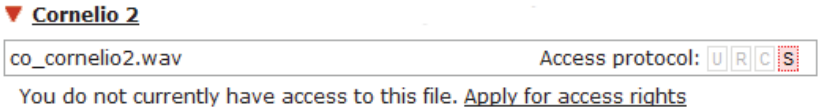

Using these controls, users who only want to be shown resources for which they have access rights can search or browse by clicking on the appropriate protocol category. On the other hand, if a user is browsing/searching all resources and reaches one which is Subscriber-only, he/she is offered an option to 'Apply for access rights'. Clicking on the link triggers the subscription application process described above.

---

and creation of knowledge'. See www.kimchristen.com/projects.html [accessed 17 August 2012]

[21] One prominent archive has recently improved its usability by adopting a protocol flagging system like ELAR's. Meanwhile, ELAR has removed the red colouring (warning of non-access) in response to its current funder's disdain for managed access.

*Figure 5: Display of a Subscriber-only resource*




After a subscription application has been approved by the depositor, the user sees the 'S' icon outlined in green, as illustrated in Figure 6, which shows the Subscriber-only resource now available to this user, in this case an audio file which can be played or downloaded.

*Figure 6: This user has subscription rights to this resource*

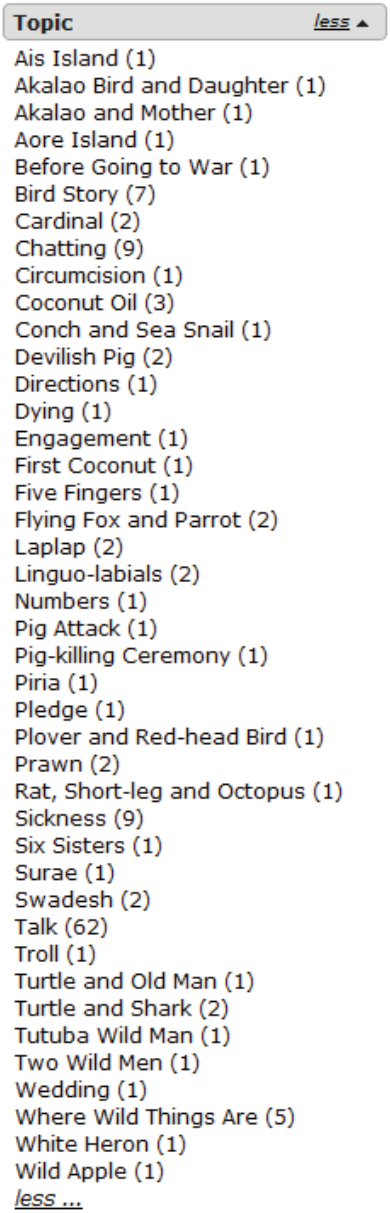

Users of this system are always aware of their access protocol context. They can choose to only search for accessible items, or they can request access to items where necessary. And users know *why they can or cannot access* particular resources.

## 6. 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.[22] ELAR places 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 map' rather than possible 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 as Figure 7.

[22] 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.

*Figure 7: User-friendly discovery using faceted browsing*



There are, of course, good arguments for providing search over standardised 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.[23] Such strategies have been the backbone of traditional library and indexing practice. While they serve certain classes of users and purposes very well, they also actually *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 ELAR wishes to serve, query interfaces only 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 a community, even the colloquial or language terms may not normally be written, or may have several variant spellings, so users are better supported by being able to browse and select rather than being forced to type in search strings hoping for an elusive match.

Metadata underlies these searching and browsing functions. ELAR takes a permissive approach to metadata, encouraging each depositor to supply as rich and descriptive a 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,[24] or phonetic terms and symbols, appear. [25]

ELAR's approach 'levels the playing field' in several ways. For example, 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 appear right 'up front' in the interface;

---

[23] For example, Ethnologue (www.ethnologue.com) lists 7,413 'primary' language names, but these have over 43,000 alternative (and dialect) names. Note that this does not generally include names in languages other than English. [accessed 17 August 2012]

[24] See *Documentation of Mavea*, Valérie Guérin, elar.soas.ac.uk/deposit/0015 [accessed 17 August 2012]

[25] See *Somyev (Sombə; KGT) Segmental and Tonal Contrasts*, Bruce Connell, elar.soas.ac.uk/deposit/0067 [accessed 17 August 2012]

they are a primary part of the site's navigation rather than merely being data points (see Garrett, this volume). Community members – or others with interests other than linguistics – can find and browse performances by those speakers, without having to remember the name of a fieldworker who once visited, a linguist's name for the originating project, or the ISO code for their language.

## 7. Access and accessibility

ELAR's approach to protocol, search, and browsing aims to enhance access by making it easier for our target users to find the resources they want, and by making any applicable access restrictions more transparent, accountable, and negotiable wherever possible. But another important question remains: what actually *counts* as access? Searching and browsing, and file display or download, are not ends in themselves. A broader account of access has to also consider *accessibility to the content* of interest to users (see also Holton, this volume). Different people want different things. Depending on users' goals, the content they desire, and the technical resources at their disposal, access could mean viewing metadata, playing an audio or video in the browser, or downloading a file to play or manipulate it later (see Figure 8). While formal linguists might want to download interlinearised marked-up material, community members might want to 'click and play' recordings of songs, stories, and events (as they can on YouTube, for example); language planners or teachers might want to assess the range and quality of the available resources.

Consider a mistake that some people make: looking to in-browser delivery as a strategy for preventing users from receiving their own digital copy of a file. This confuses access to content with the apparatus that delivers that content.[26] Instead, we have to shift focus from access to *accessibility*. Take, for example, a user with little technical interest in computing who wants to learn a song. A simple 'play' button will maximise the accessibility of the song (see Figures 6 and 8) . But someone who wants to acoustically analyse speech or transcribe it in specialised software like ELAN[27] will want to download and save the resource.

---

[26] In these days of 'always on' broadband, cloud computing, and a myriad of software for capturing YouTube media, in-browser media players are no defence at all against media file download or copying.

[27] See *The Language Archive: ELAN*, Max Planck Institute for Psycholinguistics www.lat-mpi.eu/tools/elan [accessed 17 August 2012]

*Figure 8: Playing a video in the browser – is this access?*[28]


Providing accessibility goes beyond allowing a choice between playing and downloading; suitable renderings of content might need to be made for different audiences (Nathan 2006; Holton 2011). Not all users want, or can use, audio and video with time-aligned morphological annotation. Eli Timan's ELAR collection includes time-aligned morphological annotation,[29] but it is accompanied by community-oriented resources that forgo most of the 'linguistic' content, and provide 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.[30] Another alternative we are working on is an in-browser video player (see Figure 9) that uses

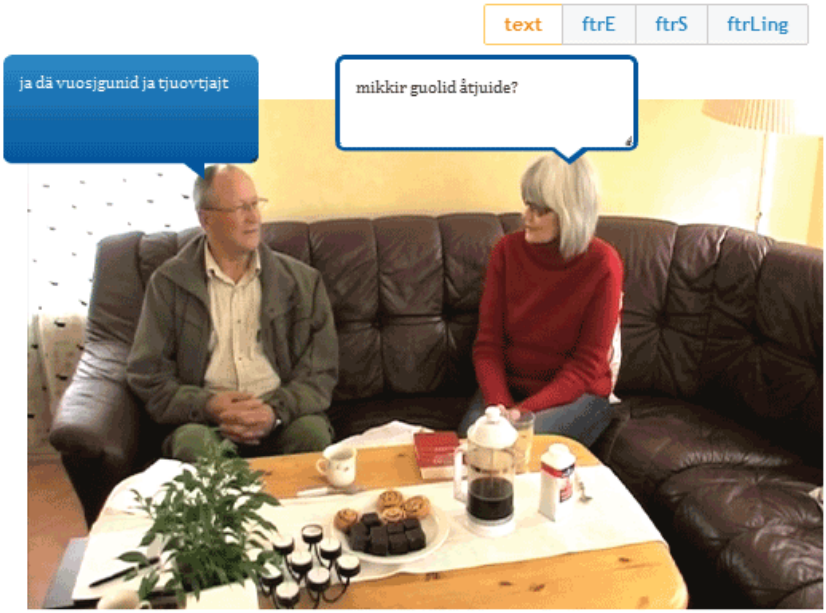
[28] From the collection *Choguita Rarámuri description and documentation*, Gabriela Caballero, but with the speaker's face pixilated. elar.soas.ac.uk/deposit/0056 [accessed 17 August 2012]

[29] See *Preservation of the Jewish Iraqi spoken language*, Eli Timan elar.soas.ac.uk/deposit/0026 [accessed 17 August 2012]

[30] See these materials at *Jews of Iraq* jewsofiraq.com [accessed 17 August 2012]

speech bubbles, a very conventional (and therefore in principle highly accessible) method to present a written representation of a conversation.[31]

*Figure 9: Experimental speech bubble player. Note that the preferred orthography for Pite Saami has changed and the bubbles would now be written as follows: Henning (left): ja dä vuosjgunijd ja tjuovtjajd; Elsy (right): mikkir gulijd åtjojde?*




## 8. Perceptions and the interface

Accessibility also depends on 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

[31] The speakers, conversing in Pite Saami, are Henning Rankvist (left) and Elsy Rankvist (right). From an ELAR collection deposited by Joshua Wilbur, *Pite Saami: documenting the language and culture* elar.soas.ac.uk/deposit/0053 [accessed 17 August 2012]. The speech bubble player was created by Edward Garrett .

role in achieving goals. ELAR chose a contemporary look, echoing features of Facebook and blogs, because these genres reduce the perception of disparities of distance and power, and they encourage productive interaction (Bozarth 2010: 55). ELAR prominently signposts protocol throughout the website not only to guide users through the interface,[32] but also to embody a commitment to depositors' protocol choices.

Sometimes things play out in unpredictable but serendipitous ways. Recently a researcher described a West African community's responses to browsing some archive websites. The community had only recently been connected to the Internet, and they mainly used social websites such as Facebook. So for them, a prototypical website looks and works like Facebook, and after their survey of online archives, they felt that ELAR was the only 'real' one.

Interfaces can be misleading. For example, archives may give a false perception of access control. Some linguists believe that a prominent language archive does not allow downloading of files, although in fact it is quite possible to download from that archive; the opacity of that archive's interface makes it so difficult to accomplish a download that it had been perceived as disallowed.[33] Such an interface disadvantages those users who legitimately want to access materials, and it also gives a false sense of security to depositors who imagine a level of access control that 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 (Warm Springs, Oregon USA) contains sensitive materials,[34] so it was 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 placed online, and the community members saw themselves displayed, they felt uncomfortable and wrote urgently to ask ELAR to 'turn off' access. We replied, explaining the benefits of them being able to see and check the display of their materials before considering allowing others to access them (or to decide against allowing access). However, we had misunderstood; it seems that the fact that the community members could see themselves framed in the browser, on the

[32] ELAR's social-networking style interface is new to language archives, although it is borrowed heavily from the existing social networking genre (Nathan 2010).

[33] That archive has recently slightly improved its access interface by adopting a protocol flagging system like ELAR's.

[34] See *Conversational Kiksht*, Nariyo Kono, elar.soas.ac.uk/deposit/0066 [accessed 17 August 2012]

computer screen – a place where normally only 'others' appear – disturbed them. We all agreed to await further discussion back in Warm Springs, and after a month they gave the go-ahead to re-open the collection, to community members only.[35]

## 9. 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 values (C and S) describe a relation between an individual user and a particular resource which is negotiated between the user and depositor. The system has proved robust and flexible during its six years of operation; we have not yet encountered a case where these roles and their associated mechanisms did not provide an appropriate solution for the access 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 combination of roles, judiciously applied.

ELAR's depositors have responded positively to the access system. Some have told us that they elected to deposit materials with ELAR that they would not deposit elsewhere, because our attention to access protocol has inspired their trust. Others have approached ELAR for archiving as a result of searching for an archive with just such a model for flexibly and accountably managing access. Some depositors who are preparing collections for deposit, on realising that ELAR can directly deliver 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 fully manage the access to their collections, and edit the content of their collection Home page (Figure 3) to add translations in the documented language or a lingua franca,[36] but we would also like to be able to present the whole navigational interface in a

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

[36] For example, Shenkai Zhang has provided summary information in Chinese for her ELAR deposit; see *Pingjiang traditional love songs* elar.soas.ac.uk/deposit/0079 [accessed 17 August 2012]

variety of languages.[37] ELAR's small team does not have the resources to accomplish that, although some depositors have already offered to help. It would be great to complete the social networking circle by allowing users to contribute comments, links and materials, and to collaborate with depositors, but any of these moves will require careful consideration of moderation and protection of moral rights and intellectual property. See also Mary Linn's contribution to this volume, where she argues that true accessibility lies in community involvement in the creation and curation of collections.

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 diverse audiences including '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 is likely to result in greater participation amongst and support of speakers of endangered languages, and an enrichment of the methods and outcomes of documentary linguistics.

## References

Austin, Peter K. 2007. Training for language documentation: Experiences at the School of Oriental and African Studies. In Margaret Florey and Victoria Rau (eds). *Documenting and Revitalising Austronesian Languages*, 25-41. Language Documentation and Conservation Special Publication No. 1. Hawaii: University of Hawaii Press.

Austin, Peter K. and Lenore Grenoble. 2007. Current trends in language documentation. In Peter K. Austin (ed.) *Language Documentation and Description, Vol 4*, 12-25. London: SOAS.

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

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

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

[37] AILLA, for example, enables users to toggle the entire interface between English and Spanish www.ailla.utexas.org/site/welcome.html [accessed 17 August 2012]

Christie, Michael. 2005. *Aboriginal Knowledge Traditions in Digital Environments*. Unpublished paper, Charles Darwin University. www.cdu.edu.au/centres/ik/pdf/CHRISTIE_AJIEpaper.pdf [accessed 17 August 2012]

Conathan, Lisa. 2011. Archiving and language documentation. In Peter K. Austin and Julia Sallabank (eds). *The Cambridge Handbook of Endangered Languages*, 235-254. Cambridge: Cambridge University Press.

Crippen, James A. and Laura C. Robinson. 2013. In defense of the lone wolf: collaboration in language documentation. *Language Documentation and Conservation* 7: 123-135.

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

Himmelmann, Nikolaus P. 2006. Language documentation: What is it and what is it good for? In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds). *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178)*, 1-30. Berlin: Mouton de Gruyter.

Holton, Gary. 2011. The role of information technology in supporting minority and endangered languages'. In Peter K. Austin and Julia Sallabank (eds). *The Cambridge Handbook of Endangered Languages.* Cambridge: Cambridge University Press, 371-399.

IASA Technical Committee. 2005. *The safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy* (Dietrich Schüller ed., Version 3) (= Standards, Recommended Practices and Strategies, IASA-TC 03), International Association of Sound and Audiovisual Archives, www.iasa-web.org/tc03/ethics-principles-preservation-strategy [accessed 17 August 2012]

Johnson, Heidi and Arienne Dwyer. 2002. Customizing the IMDI Metadata Schema for Endangered Languages. *Proceedings of The International Conference on Language Resources and Evaluation*. www.mpi.nl/lrec/2002/papers/lrec-pap-05-JohnsonDwyer.pdf [accessed 17 August 2012]

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

Nathan, David. 2006. Thick interfaces: mobilising language documentation. In Jost Gippert, Nikolaus P. Himmelmann and Ulrike Mosel (eds). *Essentials of language documentation (Trends in Linguistics. Studies and Monographs, 178)*, 363-379. Berlin: Mouton de Gruyter,.

Nathan, David. 2010. Archives 2.0 for Endangered Languages: from Disk Space to MySpace. *International Journal of Humanities and Arts Computing*, 4: 111–124.

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

Nathan, David. 2013. Access and accessibility at ELAR, a social networking archive for endangered languages documentation. In Mark Turin, Claire Wheeler and Eleanor Wilkinson (eds). *Oral Literature in the Digital Age: Archiving Orality and Connecting with Communities*, 21-40. Cambridge: Open Book Publishers.

Nathan, David and Meili Fang. 2009. Language documentation and pedagogy for endangered languages: a mutual revitalisation. In Peter K. Austin (ed.) *Language Documentation and Description, Vol 6*, 132-160. London: SOAS.

Woodbury, Anthony. 2011. Language Documentation. In Peter K. Austin and Julia Sallabank (eds) *The Cambridge Handbook of Endangered Languages*, 159-211, Cambridge: Cambridge University Press.