Graded Access to Sensitive Materials at the Archive of the Indigenous Languages of Latin America

Heidi Johnson

The University of Texas at Austin
Dept. of Anthropology
Austin, Texas 78712
+1 512-495-4604

ailla@ailla.org

Abstract

The Archive of the Indigenous Languages of Latin America (AILLA) is a web-accessible repository of multi-media resources in and about the indigenous languages of Latin America. In this paper, I describe the Graded Access System developed at AILLA to protect sensitive materials by allowing resource producers - academics and indigenous people - finely-grained control over the resources they house in the archive.

1. Introduction

The Archive of the Indigenous Languages of Latin America (AILLA) was founded at the University of Texas at Austin in 2000. It is a collaborative effort among Anthropology (Dr. Joel Sherzer), Linguistics (Dr. Anthony C. Woodbury) and the Digital Library Services Division of the UT General Libraries (Mark McFarland). Its mission is the preservation and dissemination of resources in and about the indigenous languages of Latin America, which is one of the most linguistically diverse regions of the world [4]. These resources are eclectic in both format and content: audio and video recordings, texts ranging from out-of-print books to word processor documents to handwritten field notes, databases, photographs, drawings, etc. All materials are digitized, archived in standard formats and made available on the Web through parallel interfaces in English and Spanish.

Resources are catalogued using the metadata schema developed for the International Standards for Language Engineering Metadata Initiative (IMDI) by the Max-Planck Institute for Psycholinguistics, Nijmegen, Netherlands [3]. AILLA's resources are also locatable through the

Electronic Metastructures for Endangered Language Data (EMELD; [1]) portal using a standard mapping to the metadata elements employed by the Open Language Archive Community (OLAC; [5]). The IMDI schema forms the basis for the relational database (MySQL) that houses information about AILLA's resources.

2. The Collection

AILLA's resources come from its community of users: the linguists, anthropologists, and indigenous people who have been investigating the indigenous languages of Latin America for more than half a century. Language documentation requires the collection of vast amounts of data, especially recordings of naturally-occurring discourse in a wide range of genres, including narratives, oratory, ceremonial chants, songs, and conversations. These recordings are usually accompanied by textual annotations such as transcriptions and translations, and by analytical materials such as datasets, word lists, elicited forms, ethnographic and grammatical sketches.

Researchers and their indigenous consultants have been creating such multi-media corpora since at least the 1950's. Until the advent of digital archives, there was no place to publish primary data, so these valuable materials have remained in the care of the individual researchers. AILLA is now actively acquiring and archiving these privately-held corpora, in order to make them available to the international community of researchers and more importantly, to the indigenous people themselves.

3. The Graded Access System

As long as these corpora remained in the hands of their original producers, intellectual property rights were not an issue. Now, however, the ability to disseminate recordings of indigenous people's cultural and personal expressions to

a global audience is provoking a major re-evaluation of the ways in which these works should be treated.

AILLA has the responsibility of ensuring that the intellectual property rights of the creators of resources in the archive are preserved and that the wishes of the indigenous people whose languages are recorded in the archive are respected. The legal and ethical situation is extremely complex due to the diversity of the region that we serve. There are over 500 indigenous languages spoken in Latin America, in communities which are far from homogenous. The laws of the countries in which the recordings were made vary and in many cases are now undergoing revision in response to indigenous concerns about protecting their cultural heritages.

Some materials are inherently sensitive and must be narrowly restricted. Religious or other cultural practices may require that only certain people be allowed to hear certain works, that have been recorded for fear of losing them [2], but which must be closely guarded. Identification of the narrators of accounts of historical events could in some cases lead to reprisals against the speakers. But AILLA also serves as a medium for collaboration between researchers, educators, and artists, both indigenous and non-indigenous, across the region. Depositors may want to restrict access to some materials to a group of collaborators for some period of time to develop a work in progress.

3.1. The Solution

Given the diversity of the resources and their producers, we determined that no single access restriction mechanism could adequately serve the needs of AILLA's community. We have devised a graded protocol that offers depositors a variety of means for controlling access to their sensitive materials.

All users are required to register with the archive in order to access any resource. Registration is free, but users must agree to abide by the conditions of responsible use in order to establish their accounts, which keep track of which resources they have deposited and which restricted resources they have access to. The Graded Access System provides four levels of access:

1. **Free public access.**
2. **Automatic control.** Three automatic protocols:
   a. <u>Passwords</u> offer fine or coarse control. For example, a secret word disseminated individually finely distinguishes authorized users. For coarse control, a password such as the word for 'friend' in an indigenous language, which only speakers and a few researchers would be likely to know, would let in all appropriate users, with little risk of an inappropriate user gaining access.

   b. <u>Time limits</u> would typically be the lifetime of a speaker, or some generally agreed upon limit for a researcher's exclusive use of data (e.g., 3 years).
   c. <u>Conditions</u> that the user must agree to, such as always crediting speakers by name, can be specified by depositors This control would not prevent a dishonest person from accessing a resource, but would provide electronic evidence that the user had agreed to the conditions, should a dispute arise.
3. **Depositor control.** Users contact the depositor directly to ask for permission to access the resource. The depositor contacts AILLA, and archive staff adjust the user's access list. This allows depositors to know who is interested in using their materials. Most depositors will be reachable by email, so that the communication cycle can be automatically facilitated by archive software.
4. **Indigenous control.** Users contact AILLA and we request permission on their behalf from the indigenous person or group that controls the resource. This allows indigenous people to know who is interested in their materials. Note that this cycle could take months, since communication with the indigenous community would most likely take place by surface mail.

4. Conclusions

The Graded Access System has only recently been implemented, so it is too soon to know how well it will fulfill its intended functions. The technological implementation is simple, but the sociological implications are complex and will take time to develop. We have plans to conduct evaluation workshops at sites in the U.S. and Latin America over the next 5 years, and fully expect it to take that long to determine whether our system is both sensitive enough and complete enough to serve our users' needs. We will adjust the system in response to user feedback as the evaluation period progresses.

The two most problematical areas will certainly be levels 3 and 4, which require communication with external parties. How well will we be able to keep track of depositors over time? What will we do for indigenous people who don't receive mail, like the hunter-gatherer tribes of Amazonia? It may also be possible to devise additional automatic (level 2) protocols, which will probably be the easiest for everyone to use. We hope, however, that we have now at least laid a foundation for supporting finely-grained control over access to sensitive archive resources, and thereby protecting the intellectual property rights of the indigenous people who have created them.

Acknowledgements

References

[1] Aristar, Anthony and Helen Dry. (2001). "The EMELD Project." *Proceedings of the IRCS Workshop on Linguistic Databases*, Philadelphia, 11-13 December 2001. pp. 11-16.

[2] Gnerre, Maurizio. (1986). *The decline of dialogue: Ceremonial and mythological discourse among the Shuar and Achuar of Eastern Ecuador.* In Joel Sherzer and Greg Urban Eds., <u>Native South American discourse.</u> Berlin: Mouton de Gruyter. pp. 307-341.

[3] ISLE Group. (2001)."Metadata Elements for Session Descriptions." *ISLE Metadata Initiative*, Draft proposal version 2.4 (7), May, 2001. http://www.mpi.nl/ISLE/documents/docs_frame.html

[4] Nichols, Johanna. (1990). "Linguistic diversity and the first settlement of the New World." *Language* 66(3).

[5] Simons, Gary and Steven Bird. (2001). *OLAC Metadata Set.* Draft, Open Language Archive Consortium.