

AILLA

Archivo de los Idiomas Indígenas de LatinoAmérica
Universidad de Texas en Austin

Informaciones sobre archivar y
tecnologías para la documentación de lenguas

Octubre 2007

Dr. Heidi Johnson
www.aila.utexas.org
aila@aila.utexas.edu

Metadato = Información de catálogo

Información de su proyecto

Su nombre

Info de contacto email, dirección, teléfono, etc.

Patrocinador(es)

Descripción un párrafo sobre el ámbito y duración del proyecto

Información para cada ítem (grabación, texto, base de datos, spreadsheet...)

Identificador una etiqueta única (ver página de Etiquetas)

Título un título descriptivo y útil para usuarios

Fecha de creación use formato internacional = YYYY-MM-DD

Lugar de creación

Lenguaje sea tan específico como sea posible

ISO código de <http://www.ethnologue.com/>

Nombres de hablantes nombres completos de todos los colaboradores

Palabra clave de género e.g. narrativa, lista de palabras, poema, oratoria...

Equipo todo lo que usaba para crear este recurso

Relaciones liste todos los ítemes relacionados, e.g. texto interlineado, fotos, comentarios...

Descripción un párrafo sobre el contenido del recurso

Duración hh:mm:ss para grabaciones de audio & video

Información sobre colaboradores

Nombre completo

Pseudónimo e.g. iniciales usadas en transcripciones

Guardar Anónimo 'Sí' si metadato público no debe revelar el nombre

Fecha de nacimiento YYYY-MM-DD (año estimado es suficiente)

Lengua materna

Otras lenguas otras lenguas que habla este colaborador

Lugar de origen

Descripción información pertinente, e.g. que la persona es un curandero. También para otros nombres, familiares...

Etiquetas

****** Nada puede ser más importante que poner una etiqueta en cada ítem que produce con una CONSISTENCIA SIN PIEDAD *****

(Si no sabemos que es, no lo podemos archivar.)

I. Criterios de organización

Cual criterio tiene prioridad depende de su proyecto.

lengua	lugar (e.g. pueblo)
fecha (YYYY-MM-DD)	hablante

II. Media análogo (cintas, cuadernos, cds, shoeboxes...)

Ordene la colección e.g. por lengua, después por fecha. Nombre cada objeto desde 1 a N. Relacione ítems relacionados por medio del número (cinta1 y cinta2 van con cuaderno1.) Apunte estas relaciones en su catálogo de metadatos. Escriba las etiquetas en los objetos con tinta negra indeleble. No use notas pegajosas o puntos colorados para etiquetas – se caen.

III. Archivos digitales

Cree una sintáxis extensible para su proyecto. Use los componentes para ayudar a identificar lugar/lengua, hablantes & investigadores (iniciales), tipo de contenido (Txt, Lex), fecha, serie, etc.

Folder: G_Cruz_historia

SJQ-2007_09_11-Txt_GC-acw-1.wav

SJQ-2007_09_11-Txt_GC-acw-2.wav

SJQ-2007_09_11-Txt_GC-acw.eaf

Folder: C_Flores_Cuentos

2002_08_06_Pedro-Tierra_CF1_Am.wav

IV. File system organization

Un "recurso" es un juego de archivos relacionados, e.g. grabación + texto de anotación text.

** Ordene recursos en sobres separados para ayudarle controlar ítems relacionados.

Guía de 2-minutos para los Derechos de Propiedad Intelectual

Cuales son los derechos del autor?

Los derechos indican quien puede utilizar una obra y de cual manera. El poseedor de los derechos del autor tiene el derecho a:

hacer copias;	distribuir copias;
publicar;	mostrar públicamente;
ejecutar públicamente;	hacer obras derivadas.

Usos potenciales de documentación de lenguas:

- Archivar con acceso público: más fácil acceder para los hablantes.
- Archivar acceso restringido:
 - contraseña: se puede compartir con un grupo conocido.
 - fecha límite: restringir acceso hasta una fecha especificada (e.g. la vida de un hablante)
 - control por el depositante: los usuarios tienen que contactar al depositante a pedir permiso.
- Publicar la obra entera: e.g. texto anotado, CD, DVD
- Publicar extractos: e.g. gramáticas, artículos, libros
- Publicar obras derivadas: e.g. enciclopedia de multimedia
- Presentar en público: e.g. emisión de radio
- Usar todo/extractos en clases (fotocopia, bajar por el profesor)
- Usos comerciales: libros y CDs para vender.

Cómo puede documentar el consentimiento:

- Acuerdos de licencia firmados (si la gente quiere)
- Acuerdos de licencia grabados (transcritos y traducidos)
- Relacionar el acuerdo con los recursos por metadato
- Documente los Códigos de Conducta para referencia futura.

Tenga una conversación con sus consultantes sobre usos y abusos posibles; sobre quien puede/no puede oír/ver los materiales; sobre riesgos posibles – cárcel, violencia, vergüenza, pérdida de trabajo. Grabar la conversación y incluirlo en su depósito.

Formatos

	Archivables	Presentación
Audio	wav, 44.1 / 16 min ¹	mp3
Video	mpeg-2 (mpg)	mov
Texto: editable²	Unicode + txt, xml, eaf, trs, html	cualquier
Texto: no-editable	PDF/A ³	PDF/A

Comparación de tamaños para una grabación de 10 minutos:

5 Mb	64 kbps mp3
50 Mb	44.1 / 16 wav
100Mb	mpg

1. especificaciones mínimas para habla: 44.1Khz rato de muestra / 16 bit depth
para música: 96/24

2. Texto editable que puede ser bajado de un archivo y editado por los usuarios. Ejemplo: una anotación para que varias personas contribuyan. Tales textos sólo pueden ser preservados indefinidamente si son de texto Unicode que puede ser leído por un programa no-proprietario. Si no tiene que permitir a otra gente editar sus textos, escoja la próxima opción.

3. PDF/A es un formato estándar adoptado por el ISO. Es básicamente PDF 5.0 sin enlaces, todas fuentes implantadas, sin animaciones, etc. Texto sencillo. PDF/A documentos archivables se pueden producir con Adobe Acrobat 8.0, o puede entregar sus textos en su formato original (Word, WordPerfect, Excel, etc.) al archivo y los convertimos para Ud.

Cómo hacer un texto que se puede leer para siempre

1. Usar una fuente Unicode
 - Lucida Sans Unicode, Lucida Grande, Doulos SIL
2. Preparar un diseño del teclado que haga fácil teclear sus caracteres particulares:
 - SIL's Ukelele para Mac OSX:
http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&id=ukelele
 - Microsoft Keyboard Layout Creator:
<http://www.microsoft.com/globaldev/tools/msklc.mspx>
3. Transcribir sus grabaciones con programas que apoyen alineación de tiempo, turnos de hablantes, y filas:
 - Transcriber: <http://trans.sourceforge.net/en/presentation.php>
 - Elan: <http://www.mpi.nl/tools/elan.html>
4. Interlinear con una herramienta que produce XML:
 - SIL's Toolbox: <http://www.sil.org/computing/toolbox/>

Pautas generales para crear textos:

1. Suponer que algún día otra persona va a querer ver sus datos y asegure que es legible y interpretable por otros.
2. Documentar su ortografía. (El documento podría citar simplemente el IPA.)
3. Cuando sea posible, use términos lingüísticos y abreviaturas comunes a su comunidad de práctica. Cuando no sea posible, documente sus términos.
4. Recordar que formatos propietarios como Microsoft Word o Excel cambian casi cada año. Ud. tiene que ser diligente en migrar sus archivos en adelante (abrirlos y ahorrarlos otra vez) cada vez que suban de grado sus programas y/o su plataforma.
5. Ahorrar copias de backup en formato de texto tabulador- o coma- separado frecuentemente.

Equipo para grabación de audio

(desde el School of Best Practice: <http://emeld.org/school/>)

1. Micrófonos

- Audio–Technica ATR35S (\$50) headset mic
- Shure SM58 cardioid dynamic mic (\$100)
- Shure Beta 53 (\$600) or Sennheiser HSP2 (\$450), purchased with an in–line phantom power adapter (all headsets)

2. Grabadoras

- M–Audio MicroTrack 24/96 (\$400)
- Marantz PMD660 (\$500)
- Marantz PMD 671 (\$1000)
- Sound Devices 722 (\$2375)

3. Otro equipo

- parabrisas para micrófonos (\$5–\$50)
- cables, convertidores/conectores (e.g. XLR -> RCA)
- tarjetas de memoria compactos (~ \$50 / 1 Gb)
- pilas & cargador
- soportes para micrófonos y luces

4. Programas para digitalizar/editar/convertir audio

- Audacity (gratis pero muy básico) <http://audacity.sourceforge.net/>
- Sound Forge Audio Studio (\$65) PC only
- Adobe Audition (\$200) Mac & PC
- Adobe Soundbooth (\$200) Mac & PC

5. Programas para digitalizar/editar/convertir video

- Adobe Premiere Elements (\$65) PC
- FinalCut Express (\$150) Mac

Equipo para grabación de video (desde el School of Best Practice)

- Escoja una cámara que pueda grabar audio de alta-calidad (PCM linear) en un mínimo de 44.1 kHz, y asegure que ponga la cámara a grabar en este rato de muestra.
- Si puede hacer el gasto, escoja una cámara de 3-CCD (que tiene un CCD separado para cada uno de los tres planes de color), en vez de 1-CCD, porque 3-CCD provee mejor resolución.
- Use un micrófono externo de buena calidad (los mics en la cámara son insuficientes). Asegure que ligue el micrófono a la solapa del hablante, no en la parte superior de la camera.
- Siempre grabar en SP, no LP.
- Nunca use la función de zoom digital. Está bien usar el zoom optical, pero tenga cuidado -- algunas cámaras cambian automáticamente desde optical a digital zoom.
- Cuando filmen en la noche, sería mejor cambiar el control de noche a Apagado (Off), porque este control emite luz suficiente que le hace pensar que agarra video bueno, pero realmente, no lo agarra.
- Cuando está encuadrando el plano, mantenga la cámara a una distancia suficiente para capturar todos los gestos del hablante. [Pero suficiente cerca pra que el hablante ocupe toda la pantalla.]
- Un grabador de audio distinto proveería sonido de mejor calidad; sin embargo, sincronizando el audio y el video puede ser difícil.
- Use la técnica vieja de Hollywood de empezar cada escena con un estampido (ayuda en sincronizar audio y video después) y un plano de un pedazo de papel mostrando metadatos básicos (fecha, lugar, nombre, etc.)
- Un cable malo o aparato incompatible puede ser difícil de reemplazar en el campo. Pruebe todos sus aparatos juntos *antes* de su viaje, también con cualquier procedimiento de digitalización que se propone hacer en el campo.

Pautas y consejos para grabación

1. Si es posible, prepare un "estudio de grabación" que puede usar para la mayor parte de sus sesiones de grabación. Protección del viento, tela para disminuir ecos, una mesa para los aparatos, sillas, luces para video...
2. Preocúpese por el viento: una brisa agradable se produce ruido fuerte y constante, y se oculta en su grabación.
3. Desconecte generadores, refrigeradores, y otras máquinas que producen un zumbido bajo.
4. Practique con sus aparatos para aprender la mejor colocación para micrófonos, soportes, sillas, etc. En particular, pruebe mics auriculares (headset mic) para asegurar que no están tan cerca de la boca del hablante que algunos consonantes causan clipping.
5. Escuche sus grabaciones mientras grabe. En un medio ambiente nuevo, grabar un poquito, escuchar, ajustar, y seguir. Y/o escuche lo que grabó el día antes y anote cualquier ruido que podría cambiar/evitar la próxima vez.
6. Haga copias lo más pronto posible:
 - subir datos de tarjetas de memoria a su computadora cada día;
 - queme 2–3 cds para cada grabación. Entregue una copia a su archivo, de una copia a su consultante, y use una para transcripción, etc..
7. Escriba los metadatos cuando haga la grabación.
8. Anote la información básica – lengua, fecha, lugar, nombres completos de hablantes – en el inicio de cada grabación. Hable claramente en una lengua en gran escala (e.g. Español). No olvide incluir su propio nombre y el nombre de la lengua en los metadatos.
9. Para video, trate de poner un poco de luz sobre la cabeza del hablante para que su cara pueda ser vista.
10. Recuerde que archivos de video son enormes y no filme la tierra, sillas vacías, etc. Si no hay gente hablando, cantando, o bailando, considere tomar una foto. O, borre las partes sin interés cuando edite el video.

Equipo para digitalización

I. convertidor analog–digital

Edirol UA–5 24 bits / 96 KHz

Conecta a la computadora por cable USB (no requiere tarjeta de audio especial.)

(\$240) <http://www.zzounds.com/item--EDIUA25>

II. Tarjeta de audio

Para transferir señales digitales a la computadora (del minidisco)

M–Audio Audiophile 24–Bit/96kHz

(\$ 150) <http://www.bhphotovideo.com/>

III. Software

1. Audacity (gratis y sencillo) <http://audacity.sourceforge.net/>

2. Adobe Soundbooth CS3 (PC & Mac) (\$200)

<http://www.adobe.com/products/audition/compare/>

3. Sound Forge Audio Studio (sólo PC) (\$60)

<http://www.sonycreativesoftware.com/products/product.asp?pid=454>

IV. Tocadores

1. Casset

Tascam 322 Dual Cassette Deck (\$500)

<http://www.fullcompass.com/product/301542.html>

2. Minidisc (transferir a la computadora por la tarjeta de audio M–Audio)

Tascam MD350 MiniDisc Recorder/Player (\$470)

<http://www.zzounds.com/item--TASMD350>

** se necesita un convertidor especial para conectar el minidisco:

Optical to Coaxial Digital Audio Converter (\$40)

<http://www.cablestogo.com/product.asp?cat%5Fid=504&sku=40019>

3. Rollo abierto

Hay que buscarlos en donde se pueda, incluyendo eBay.

V. Formatos para archivar

1. Audio

mínimo para habla: PCM WAV, 44.1KHz sample rate, 16 bit depth.

96/24 mejor para música.

Estereo solamente si el original es estereo; si no, mono.

2. Video

MPEG-2 (.mpg)

3. Textos

a. textos "muertos" (que los usuarios no pueden editar): PDF/A

[Nota: PDF/A es un estándar nuevo de ISO para el formato pdf. Esta versión es básicamente PDF 5.0, sin enlaces, animación, etc. Ya está en el dominio público, entonces podemos garantizar que se pueda leer documentos en este formato en el futuro.]

b. textos "vivos" (que los usuarios pueden editar en el futuro):

fuentes de Unicode (Lucida Sans Unicode, Lucida Grande)

formatos: txt, xml, eaf, trs, html

VI. Escanear textos

1. Canon® CanoScan® 8400F Flatbed Scanner (\$200) (Office Depot)

2. Adobe Acrobat 8.0

– Professional (\$130) <http://www.softwaresurplus.com/Guaranteed/Low.Prices?>

permite conversión de PDFs viejos a PDF/A.

– Standard (\$120)

crea PDF/A pero no convierte los viejos

VII. Equipo misceláneo

1. conectores, cables, etc.

<http://www.cablestogo.com/index.asp?>

2. 3.5 floppy drive (\$20) <http://www.tigerdirect.com/>

3. CDs, etiquetas, cajas

4. cajas para poner en juegos los depósitos (preferimos los blancos)

5. para cintas en rollo: leader tape, splicing tape, splicing block

6. software que lee toda clase de texto digital: Word, WordPerfect, Excel, etc.

7. disco duro externo: (500 GB = \$170 en Office Depot)

Donde ir para más información

I. Información general

- **EMELD School of Best Practice:** emeld.org/school/index.html
- **Hans Rausing Endangered Languages Project (HRELP):** www.hrelp.org/documentation/whatisit/
- **Documentation of Endangered Languages (DOBES):** www.mpi.nl/DOBES/INFOpages/applicants/dobes-ling-aspects-lang-doc.html
- **Resource Network for Linguistic Diversity:** www.linguistics.unimelb.edu.au/thieberger/RNLD.html
- **Language Archives Newsletter:** www.mpi.nl/LAN/
- **The Vermont Folklife Center:** www.vermontfolklifecenter.org/res_audioequip.htm
- **Matrix Oral History Project:** www.historicalvoices.org/oralhistory/improve-ad.html

II. Estándares y organizaciones

Nota: escriba a cualquier miembro de DELAMAN con preguntas

- **DELAMAN:** Digital Endangered Languages and Musics Archive Network: www.delaman.org
- **OLAC:** Open Language Archives Community: www.language-archives.org

III. Programas

- **Unicode:**
 - www.unicode.org/
- **Doulos SIL Unicode IPA:**
 - www.sil.org/computing/catalog/show_software.asp?id=91
- **ELAN,** multimedia annotator:
 - www.mpi.nl/tools/
- **IMDI Editor,** java tool for IMDI metadata
 - www.mpi.nl/tools/
- **Transcriber,** transcription tool:
 - trans.sourceforge.net/en/presentation.php
- **Praat,** speech analysis tool:
 - www.fon.hum.uva.nl/praat/
- **Audacity,** free audio editing tool (can be used for digitizing):
 - audacity.sourceforge.net/
- **AILLA metadata forms & Shoebox templates:**
 - www.ailla.utexas.org/site/download_md_forms.html
- **Toolbox,** database & interlinearization:
 - www.sil.org/computing/toolbox/