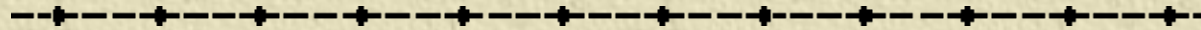
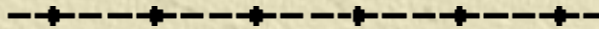


Archivos de recursos para la documentación de lenguas

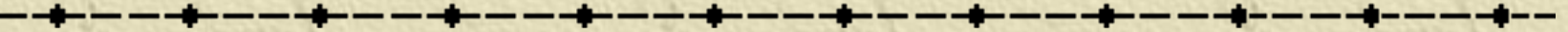


Heidi Johnson
Archivo de Lenguas Indígenas de
Latinoamérica
Universidad de Texas en Austin



INALI, México D. F., 2 junio 2005

Preámbulo



Los archivos son la mejor esperanza para preservar y mantener acceso a recursos irremplazables de lenguas en peligro de extinción.

Los archivos digitales pueden preservar recursos en varios medios indefinidamente.

Un poco de historia

- ✦ La tradición de Boas: gramática, diccionario, colección de textos.
- ✦ Los lingüistas dejaron materiales en museos y bibliotecas, v.g. Museo de Antropología. Buscaron un lugar permanente para esos recursos.
- ✦ Los museos y las bibliotecas tradicionales realmente no pueden conservar grabaciones más allá de guardarlas en un lugar fresco y oscuro.

Historia, cont.

- ✦ Lo que usualmente se publica es solo una destilación – el producto del análisis. Recursos secundarios y terceros.
- ✦ Hasta aquí no hay modo factible de preservar ni publicar materiales fundamentales.
- ✦ El nuevo milenio: archivos digitales pueden preservar y publicar cualquiera cosa.

¿ Qué es un archivo?

Un archivo es un repositorio fidedigno creado y mantenido por una institución con un **compromiso de permanencia demostrado** y un compromiso para la preservación a largo plazo de los recursos archivados.

¿ Por qué necesitamos archivos?

- ✦ Para preservar grabaciones de lenguas en peligro para las generaciones que vienen.
- ✦ Para facilitar el re-uso de materiales básicos para:
 - programas de enseñanza de lenguas;
 - estudios tipológicos, históricos, comparativos;
 - investigaciones en antropología, psicología, botánica.
- ✦ Para fomentar el desarrollo de literaturas orales y escritas de lenguas en peligro.
- ✦ Para facilitar colaboraciones entre investigadores y educadores.

Ventajas de los archivos digitales

Medios digitales en formatos estándares:

- ✦ Se pueden actualizar a largo plazo para guardarlos para siempre
- ✦ Se pueden copiar, ofrecer por la red, y transferir a medios analógicos como cintas de casete
- ✦ Pueden ser re-usados: una lista de palabras se pone en un diccionario de multimedios
- ✦ Se pueden transferir de una institución a otra en discos duros portátiles.

¿ Qué hay en un archivo digital?

- ✦ Grabaciones: audio y video
- ✦ Textos: manuscritos escaneados (cuadernos, fichas, artículos y libros fuera de publicación)
- ✦ Textos: digitales (v.g. Word), convertidos a formatos estándares (v.g. ascii, rtf, html, xml)
- ✦ Bases de datos: Excel, Shoebox, etc. convertidos a formatos estándares (v.g. xml)
- ✦ Fotografías: escaneadas o digitales

Desventajas de los archivos digitales

- ✦ El equipo es caro y tiene que ser actualizado cada 3 a 5 años
- ✦ El ambiente computacional tiene que ser mantenido por profesionales capaces
- ✦ Si un archivo digital es marcado incorrectamente o no tiene metadatos, dentro de 5 a 10 años podría ser imposible saber de qué consiste el archivo.

Establecimiento de un archivo: el contexto institucional

**Buscar la institución más duradera que se pueda
para ser la anfitriona del archivo.**

Esta institución puede proveer:

- ✦ Apoyo e infraestructura técnicos
- ✦ Personal de tiempo completo permanente
- ✦ Credibilidad: una reputación establecida
- ✦ Conexiones a otros trabajando en proyectos similares para compartir recursos tecnológicos.

Establecimiento de un archivo: el contexto tecnológico

Que esté más allá del ámbito de especialistas en un área de estudio, como lingüistas. Se necesita:

- ✦ Servidores que puedan manejar terabitos de recursos en multimedios.
- ✦ Capacidad de red para servir tanto a la comunidad nacional como la internacional.
- ✦ Expertos que estén al tanto de los cambios tecnológicos.

Establecimiento de un archivo, cont.

El problema tecnológico es la mayor razón para aliarse con un proyecto más grande, como:

- ✦ La red académica

- ✦ Internet 2:

 - ◆ <http://www.noc-internet2.unam.mx/>

- ✦ Biblioteca Digital Universitaria de la DGSCA:

 - ◆ <http://www.bibliodgsc.unam.mx/>

Declaración de la misión del archivo

Defina claramente:

- ✦ el ámbito y escala de la colección (e.g. lenguas mexicanas / mesoamericanas / latinoamericanas);
- ✦ de dónde van a venir los recursos (e.g. materiales viejos de investigadores, materiales nuevos de proyectos de comunidades, etc.)
- ✦ quiénes van a ser los usuarios principales del archivo.

Manejo del archivo: Personal

- ✦ Gerente general
- ✦ Gerente técnico
- ✦ Especialista de metadatos
- ✦ Traductor
- ✦ Digitalizadores

Gerente general del archivo

Debe ser un especialista en el área de estudio: un lingüista con conexiones en el campo de documentación de lenguas.

- ✦ dirección general, supervisión de otro personal
- ✦ recaudación de fondos
- ✦ adquisición de materiales nuevos
- ✦ talleres en el manejo de corpora
- ✦ participación en DELAMAN (Digital Endangered Languages and Music Archive Network)

Gerente técnico del archivo

Más probable alguien de las ciencias informáticas.

- ✦ mantiene programas del archivo: base de datos y sitio de la red
- ✦ se mantiene al corriente de cambios en las tecnologías digitales y los estándares técnicos
- ✦ talleres en grabación y digitalización
- ✦ supervisa digitalizadores, en y fuera del archivo
- ✦ participa en OLAC, DELAMAN

Especialista de metadatos

Alguien de ciencias informáticas o un lingüista.

- ✦ escoje y prepara a su gusto un esquema de metadatos.
- ✦ participa en consorcios internacionales de metadatos (OLAC, IMDI, Dublin Core)
- ✦ ayuda a depositantes a desarrollar metadatos para sus recursos – más talleres.
- ✦ introduce y repasa metadatos para materiales archivados.

Traductor

- ✦ Hay información abundante sobre archivos digitales, medios y formatos digitales, metadatos, documentación de lenguas, etc. en inglés que necesita ser traducido al español.
- ✦ Los metadatos y otros documentos del archivo a veces necesitan traducción a/de inglés u otras lenguas.
- ✦ Supervisa traductores trabajando en lenguas indígenas, como para hacer interfases en la red y metadatos.

Personal de digitalización

- ✦ Cualquier persona que sea paciente y orientada a detalles puede ser instruido para hacer este trabajo.
- ✦ ¡Advertencia!: digitalizar es algo muy aburrido, y requiere atención cuidadosa a detalles en cada paso. Es un trabajo muy lento.
- ✦ Permita que otras instituciones y depositantes ayuden a digitalizar sus materiales según las especificaciones del archivo. ¡Comparta el trabajo!

Defina políticas de manejo

Defina procedimientos y políticas para:

- ✦ adquisición de materiales incluyendo una estrategia para priorizar el horario de digitalización;
- ✦ diseminación de materiales, incluyendo restricciones de acceso, lenguas de las interfases, etc.
- ✦ seguridad de calidad;
- ✦ seguir con atención los estándares para la digitalización y la migración tecnológica a formatos digitales nuevos;
- ✦ recuperación después de una desastre – copias de seguridad, sitios espejos, etc.

Manejo de archivo - Lecturas

- ✿ OAIS Reference Model for Digital Libraries

 - ✿ <http://www.ccsds.org/documents/650x0b1.pdf>

- ✿ EU-US Working Group on Spoken-Word Audio Collections

 - ✿ <http://www.dcs.shef.ac.uk/spandh/projects/swag/>

- ✿ OLAC documentos:

 - ✿ http://www.language_archive.org

- ✿ DELAMAN miembro archivos:

 - ✿ <http://www.delaman.org>

Manejo del archivo: resumen de componentes

-
- ✦ Defina el objeto de archivo: qué es lo que se va a preservar y cómo se va a clasificar.
 - ✦ Diseñe identificadores para esos objetos.
 - ✦ Escoja un sistema de catalogación – un esquema de metadatos.
 - ✦ Decida cuáles formatos se van a usar para preservar y cuáles para presentar los datos a los usuarios.
 - ✦ Determine políticas para los derechos de propiedad intelectual.
 - ✦ Desarrolle métodos para acceder el archivo, v.g. sitio de la red.

El objeto de archivo I

Los recursos de la documentación de lenguas vienen en *juegos* de ítemes relacionados. En AILLA llamamos a todo el juego un *recurso del archivo*:

- ✦ Una grabación con textos (transcripción, traducción, anotaciones) y fotografías;
- ✦ Una tesis enlazada a muchos ejemplos de audio;
- ✦ Un poema kuna con una versión en español.
- ✦ Variaciones de formatos para cada recurso.

El objeto de archivo II

- ✦ Las relaciones entre objetos tienen que ser preservadas por el método del archivo y documentados en los metadatos.
- ✦ Es de ayuda que los identificadores codifiquen las relaciones, para que sea más fácil para los usuarios reconstruir el juego.

Relaciones entre objetos

- ✦ derivación: e.g. una transcripción se deriva de una grabación
- ✦ serie: e.g. una grabación larga que ocupa varias cintas/discos
- ✦ parte-totalidad: e.g. grabaciones de video y audio que forman parte del mismo evento
- ✦ asociación: (relación difusa) e.g. fotografías del narrador de una grabación, comentarios

Los tipos de contenidos de AILLA

(Elementos de metadatos no-estándares.)

texto primario

transcripción

traducción

transcripción y traducción

interlinearización

ilustración

anotación

comentario

muestra

fotografías

guía

mesa de contenidos

Identificadores de objetos I

Defina qué constituye un objeto archival y sea consistente en emplear su definición.

1. objetos digitales que corresponden a un medio original: Suárez cinta 1 + cuaderno 1, parte 1 (es especialmente útil que también se archiven los originales)
2. objetos digitales que corresponden a eventos documentales o a algún concepto de contenido intelectual: "El nahual", grabación + texto

Identificadores de objetos II

Los identificadores deben apoyar:

- ❖ cobranza del medio original (analógico) si estos se preservan en el mismo archivo;
- ❖ hacer juego del objeto a los relacionados;
- ❖ cita correcta de recursos archivados;
- ❖ reubicación del objeto en al archivo si se coloca fuera de su lugar.

Identificadores de objetos en AILLA: I

La ordenación primaria es por **lengua** (usando los códigos de 3 letras del Ethnologue):

✦ ZOH001R040I001.mp3

- ✦ ZOH = código de la lengua
- ✦ 001 = número del depósito (el primero)
- ✦ R040 = 40 recursos en este depósito
- ✦ I001 = primer ítem en este recurso
- ✦ .mp3 = extensión de formato del archivo

Identificadores de objetos en AILLA: II

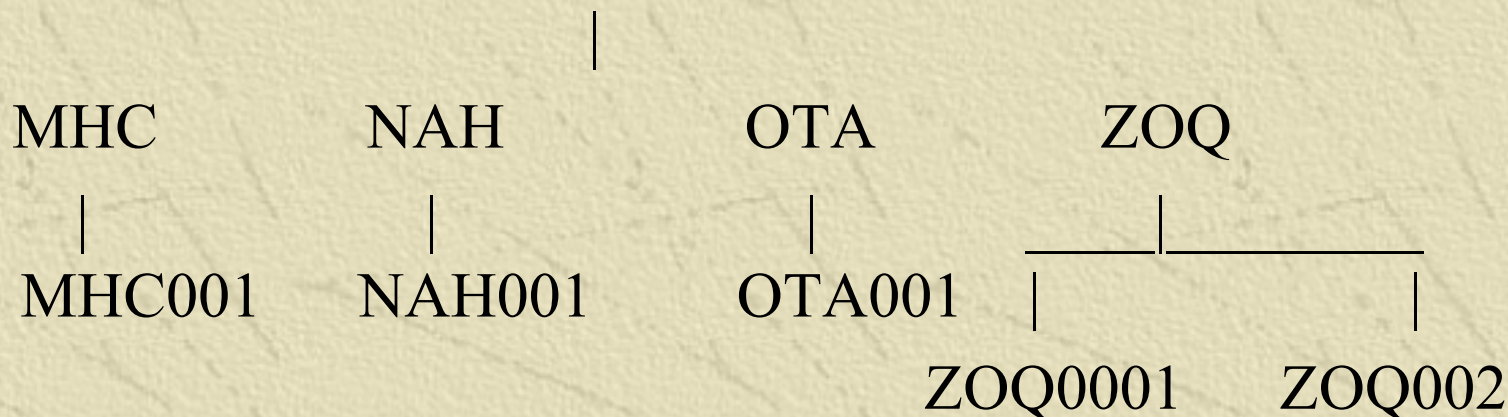
Archivos del recurso ZOH001R020:

- | | |
|-------------------------|-----------------------------|
| ✦ ZOH001R020I001.mp3 | audio zoque |
| ✦ ZOH001R020I001.wav | audio zoque |
| ✦ ZOH001R020I001-sa.mp3 | audio zoque 1-min muestra |
| ✦ ZOH001R020I002.mp3 | audio español |
| ✦ ZOH001R020I002.wav | audio español |
| ✦ ZOH001R020I001.txt | texto interlinear |
| ✦ ZOH001R020I001.pdf | formato de texto para bajar |

Los identificadores y la organización del archivo

Los identificadores de AILLA facilitan la organización del sistema de los archivos del archivo por lengua y por número de depósito:

/usr/local/media/ailla



Los identificadores y la organización del archivo

Otros archivos usan otros criterios de organización:

- ✦ DoBeS: organizado jerárquicamente por proyecto y por tipo de data
- ✦ PARADISEC: organizado por depositante / creador (identificadores empiezan con las iniciales)
- ✦ Archivos existentes no digitales podrían usar sistemas existentes para las versiones digitales.

Los identificadores de objetos: última palabra

- ✦ Cambiar los identificadores es un trabajo muy tedioso y arriesgado.
- ✦ Piense mucho sobre su sistema. Asegúrese de que refleje:
 - las clases de materiales que se van a archivar
 - el flujo de trabajo del archivo
 - los contextos institucional y tecnológico
- ✦ *Pruebe su sistema en un juego piloto de recursos.*

Metadatos: información de catálogo para recursos digitales

- ✦ Información para la administración, descripción del contenido, descripción del recurso, y derechos.
- ✦ Esquema de metadatos debe cumplir con los requisitos de OLAC.
- ✦ La mejor práctica es adoptar y preparar a su gusto un esquema (OLAC, IMDI) para llevar al máximo la interoperabilidad.
- ✦ Sea un participante activo en la comunidad internacional de archivos de lenguas.

Metadatos II

La información del catálogo apoya:

- ✦ manejo del archivo
- ✦ protección de materiales sensibles
- ✦ búsquedas
- ✦ uso de recursos por mucha gente
- ✦ cita correcta de recursos archivados

Metadatos III

Dos esquemas compatibles para recursos de lenguas:

1. OLAC: The Open Language Archives Community. (Univ. Pennsylvania)
<http://www.language-archives.org>
2. IMDI: International Standards for Language Engineering Metadata Initiative (MPI-Nijmegen)
<http://www.mpi.nl/IMDI>

Metadatos IV : información mínima

- ✦ Nombres completos de los hablantes.
- ✦ Lengua: Sea específico: zoque de San Miguel Chimalapa, Oaxaca, México.
- ✦ Fecha de creación: YYYY-MM-DD. Use la fecha primaria (de la grabación) para el juego.
- ✦ Lugar de creación. Sea específico: pueblo, estado, país, o valle de río, región, país...
- ✦ Restricciones de acceso e instrucciones, si son necesarios.
- ✦ Palabras claves para el género, e.g. narrativo

IMDI

Juego de una sesión de grabación = recurso

- ✦ Título, fecha, lugar, descripción general
- ✦ Depositante: información para contacto
- ✦ Proyecto: nombre, director, fundador, etc.
- ✦ Participantes: nombre, alias, papel, datos demográficos
- ✦ Recursos: origen, formatos, relaciones, etc.
- ✦ Contenido: contexto, género, descripción, etc.
- ✦ Referencias: publicaciones relevantes

OLAC

La definición del objeto archival se queda con Ud.

- ✦ Colaboradores / creadores
- ✦ Título, fecha, descripción
- ✦ Recursos: formatos
- ✦ Relación a otros objetos
- ✦ Materia, campo o sub-campo lingüístico, e.g. fonología
- ✦ Tipo lingüístico = género, e.g. discurso, léxico

Ejemplo de metadatos de AILLA (basado en IMDI)

ID: ZOH001R020

Título Saturnino 'i Soldao'øyti'

Fecha 1994-07-03

Lugar Fortín de las Flores, Veracruz, Mexico

Descripción: Saturnino y los Soldados Es una historia de un héroe local que fue contado por el mismo héroe quien lleva de nombre Saturnino. Germán Sánchez quien es mi asesor, lo aprendió de su padre Agripino Sánchez quien lo aprendió de Saturnino. La historia se lleva a cabo en San Miguel Chimalapa, Oaxaca, México, en los primeros años de vida adulta de Saturnino - posiblemente durante la revolución Mexicana, alrededor de 1918.

Referencias: [libro]

Participantes:

Germán Sánchez Morales: papel: Hablador, fecha-nació:1940; sexo: M; lengua-materna: ZOH; otros-idiomas: SPN, ZAP

Heidi Johnson: role:Researcher, Depositante, fecha-nació:1956; sexo:F; lengua-materna: ENG; otros-idiomas: SPN

Ejemplo, cont.

Detalles de medios de un archivo

ZOH001R018I001.wav

Idioma: ZOH

Tipo de contenido: texto_primario

Objecto-archival: No

Nivel de acceso: 1 (public)

Largo: 00:09:01

Tamaño de archivo: 46M

Formato: 44.1/24

Plataforma de digi: PC, Flying Cow, SoundForge

Medio original: cinta de cassette

Cualidad de grabación: 3 (1-5, bajo-alto)

Archivado por: haj

Fecha archivado: 2004-12-10

Descripción: <detalles sobre el medio original o problems con la grabación o la digitalización>

Metadatos - Enlaces

✱ OLAC

✱ <http://www.languagearchives.org>

✱ IMDI

✱ <http://www.mpi.nl/IMDI>

✱ Dublin Core

✱ <http://es.dublincore.org/>

✱ METS (Metadata Encoding and Transmission Standard)

✱ <http://www.loc.gov/standards/mets/>

✱ AILLA

✱ http://www.ailla.utexas.org/metadata_sp.html

Cita correcta de recursos archivados

Sánchez Morales, Germán. (1994). "Satornino y los soldados." [audio+interlinear text]
Heidi Johnson, (Researcher).

<http://www.ailla.utexas.org>: Archive of the Indigenous Languages of Latin America. Access=public. ZOH001R010.

Propiedad Intelectual

Desarrolle políticas que se apliquen a:

- ✦ problemas de sujeción por el archivo o la institución anfitriona
- ✦ derechos de creadores de recursos, tanto hablantes indígenas como investigadores
- ✦ cualquier condición particular que los creadores quieran ligar al recurso
- ✦ requisitos del uso y acceso para los usuarios

Propiedad Intelectual

- ✦ Provea guías para los productores de recursos para elicitación de consentimiento a archivar y publicar;
- ✦ Haga políticas con respecto a copiar o hacer sitio de espejo con otros archivos.

Todas las restricciones y metadatos sobre derechos tienen que viajar con cada recurso si las copian a otro archivo.

Propiedad Intelectual – Ejemplos de acuerdos de licencia

✿ AIATSIS:

✿ <http://coombs.anu.edu.au/SpecialProj/ASEDA/ASEDA.html>

✿ AILLA:

✿ http://www.ailla.utexas.org/site/use_conditions_sp.html

✿ OLAC:

✿ <http://www.language-archives.org/docs/license.html>

Propiedad Intelectual - Lecturas

- ✦ Lieberman artículo
[<http://www ldc.upenn.edu/exploration/expl2000/papers/liberman/liberman.html>]
- ✦ Información sobre propiedad literaria de la abogada de UT
[<http://www.utsystem.edu/OGC/intellectualproperty/index.htm>]
- ✦ World Intellectual Property Organization
[<http://www.wipo.int>]

Formatos I

El archivo tiene que considerar 3 tipos:

- ✦ Formato de trabajo: lo que usa el lingüista en el campo.
- ✦ Formato de presentación: lo que se usa para publicar.
- ✦ **Formato de archivo:** lo que preservamos.

Formatos II

	Texto	Audio	Video
	gramática	grabación	película
archivo	tiff / txt xml	PCM wav	mpeg2
presentación	pdf / html	mp3	qt, mpa
trabajo	Excel / MS Word	mp3	formato de camera

Formatos III

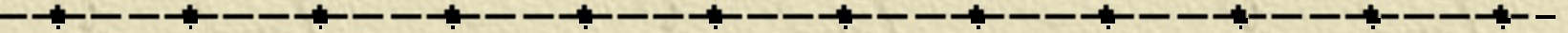
-
- ✦ Distinga claramente los formatos archiviales de los de presentación para que los usuarios entiendan que *materiales digitales en formatos de presentación no son materiales de calidad archival*.
 - ✦ Los archivos deben publicar sus estándares de digitalización como guías para los productores que quieran depositar materiales digitales.

Formatos IV

Requisitos generales para formatos de calidad
archival (copia principal):

- ✦ no-propietario; es decir, la codificación es en el dominio público;
- ✦ portátil, re-usable;
- ✦ la mejor reproducción posible del original.

Formatos V



- ✦ Los objetos archivales (archivos en formatos wav, tiff, xml, etc.) se actualizarán cuando sea necesario en el futuro.
- ✦ Se convertirán al formato nuevo.
- ✦ Todas las otras versiones se descartarán y se generarán nuevos formatos de trabajo y formatos para presentación de las nuevas copias archivales.

Formatos de audio I

- ✦ **PCM linear wav** es el formato estándar para la preservación de audio.
- ✦ Para habla, la calidad de CD es aceptable:
 - ✦ 44.1 kHz frecuencia de muestreo, 16 bits por muestra
- ✦ Un paso mejor: 44.1 / 24 bits por muestra
- ✦ Para música (y para habla, según algunos expertos): 96 kHz / 24

Formatos de audio II

- ✦ Frecuencia de muestreo (*sample rate*): indica cada cuánto se convierte el sonido analógico en valor digital y se mide en hercios (hz.), o en su múltiplo, el Kilohercio (Khz.) . Para alcanzar la calidad de sonido de un CD Audio necesitamos una frecuencia de 44100 hz.
- ✦ Tamaño de la muestra (*sample size=bit depth*): define el número de bits que se necesitan para grabar las amplitudes de los sonidos. Las tarjetas de sonido actuales tienen un tamaño de muestra de 16 o 24 bits. Con 16 bits puede conseguirse un sonido igual al de los CD Audio.

<http://www.educa.aragob.es/cursoryc/comunes/modulo3/unidad2.htm>

Formatos de audio III

- ✦ Método empírico: usar una frecuencia de muestreo dos veces más que la frecuencia más alta en la grabación. El habla no tiene frecuencias más altas que ~6000 hz y el oído humano solo puede oír hasta 20000 hz, entonces 44.1 es suficiente.
- ✦ Sin embargo, los valores más altos (96/24) darán reproducciones más fieles del medio ambiente y permitirán mejor reducción de ruido en el futuro.
- ✦ Los valores altos producen archivos más grandes.

Formatos de audio IV

Tamaños de archivos para una grabación de 10 minutos:

44.1/16 51M (megabitos)

44.1/24 76M

mp3 (128 kbps) 9.2M

cualidad-CD \approx 5M por minuto

** mp3 es un formato comprimido

Formatos de textos

-
- ✦ Meta: poder leer el archivo en 10+ años sin buscar programas obsoletos. (¿Quiere alguien WordStar?)
 - ✦ Expertos recomienden xml y Unicode (pero todavía no nos dicen cómo usarlos.)
 - ✦ xml y html son nada más archivos de texto sencillo con marcas estructuradas que describen cómo deben ser formateados.
 - ✦ Texto sencillo (ascii/rtf) y texto separado por paletas (salida de Word, Excel, Shoebox) son perfectos.

Formatos - Lecturas

✱ EMELD School of Best Practice:

✱ <http://emeld.org/school/index.html>

✱ Para imágenes:

✱ <http://www.library.cornell.edu/preservation/tutorial-spanish/contents.html>

✱ Colorado Digitization Program:

✱ http://www.cdpheritage.org/resource/audio/std_audio.htm

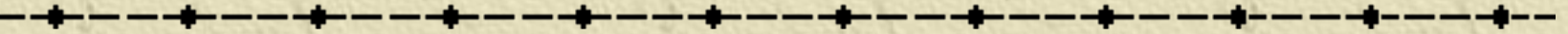
✱ Herramientas básicas de multimedios:

✱ <http://www.educa.aragob.es/cursoryc/comunes/modulo3/unidad2.htm>

Equipo para digitalización de audio

- ✦ PC/Mac con 1 Gb RAM, 40+ disco duro
- ✦ Programa: SoundForge (sólo PC), ProTools, CoolEdit
- ✦ Convertidor análogo-a-digital: Flying Cow, Lucid, otros (con los más caros se puede conectar más componentes a la vez)
- ✦ Media players: casete, rollo abierto, plato
- ✦ Cables, altavoces, audífono (2 por cada estación)

Equipo para digitalización de audio



Si tiene una institución anfitriona muy generosa (y un ingeniero de audio):

Quadriga (>\$12K)

<http://www.cube-tec.com/Quadriga.html>

Procedimientos para la digitalización de audio I

-
- ✦ Cintas de rollo abierto: tóquelas primero para determinar:
 - ◆ la velocidad, 1 o 2 lados, niveles de volumen;
 - ◆ limpie la cinta y repare empalmes viejos
 - ◆ rebobine regular la cinta.
 - ✦ Digitalice un lado entero y después corte la grabación digital en segmentos (e.g. cuentos) usando el programa de audio.
 - ✦ La normalización puede ser necesaria para aumentar el volumen.

Procedimientos para la digitalización de audio II

-
- ✦ Puede añadir un poco de metadatos a la grabación (Properties->summary; BWF)
 - ✦ Con SoundForge (PC), se puede usar el BatchConverter (convertidor de lota) para producir mp3, 1-min. muestras (por archivos > 10 mins.) y calidad-cd wav (44/16) versiones para sus usuarios. (Convierte ~20 a la vez.)
 - ✦ **Toma por lo menos 5 horas el tramitar 1 hora de grabación (para archivarla.)**

Flujo de trabajo de archivar: I

1. Ordene el depósito en recursos, asignando identificadores del archivo y poniendo etiquetas en **cada cosa** claramente.
2. Empiece un registro del proceso para llevar la cuenta de cada paso cuando sea terminado.
3. Digitalice los materiales análogos.
4. Haga los metadatos. Esto probablemente requerirá de múltiples consultas a los depositantes.

Flujo de trabajo de archivar: II

6. Produzca formatos de presentación (mp3, pdf).
7. Suba los objetos digitales al servidor y/o el disco duro externo y/o CD/DVDs archiviales.
8. Enlace los archivos a los metadatos en la base de datos.
9. Pruebe. Pruebe otra vez. Sería mejor que otra persona también haga pruebas.
10. Haga copias de seguridad diarias y semanales, y guárdelas en otro lugar.

Bases de datos para metadatos I

MySQL: <http://www.mysql.org>

✦ gratuito, documentación en varias lenguas, mucha ayuda por la red.

✦ <http://sqlzoo.net/es/>

✦ http://www.hospedajeydominios.com/mambo/documentacion-manual_mysql.html

✦ Se necesita un programador. Se puede empezar con la de AILLA y modificar esa.

Bases de datos para metadatos II

✦ FilemakerPro: \$200. Muchas plantillas, fácil de aprender (relativamente), no necesita programación (exactamente). PARADISEC.

● <http://www2.filemaker.fr/spain/>

✦ Oracle: \$10K-\$30K. Muy poderoso. Requiere un programador. Súper si su institución anfitriona tiene licencia. (EMELD)

Almacenamiento del archivo

-
- ✦ Lo mejor es una instalación completa con servidor en la red que otra gente (profesionales) maneje.
 - ✦ Lo siguiente es un servidor de escala media comercial. Hay que consultar a personas que conozcan la situation local para informaciones específicas.
 - ✦ A corto plazo (~5 años): computadora personal con un disco enorme (100+ Gb) y discos duros externos para copias de seguridad. En 5 años **tiene que transferir** todos los archivos a un disco nuevo a quedarse con la actualidad.

Interfases del archivo I

- ✦ Se pueden ofrecer sólo copias de presentación: cassetes, cds, textos en medios portátiles.
- ✦ Y/o ofrezca la colección por la red:
- ✦ PHP: lenguaje de scripting, gratuito, buena documentación, funciona con mySQL, muy popular mundialmente
(<http://www.php.net/manual/es/>)
- ✦ Programas comerciales para desarrollo de aplicaciones de la red: e.g. Dreamweaver

Interfases del archivo II

Con la red se puede:

- ✦ Unir su consorcio. Todos pueden enlazar sus propias páginas, contribuir al contenido, editar metadatos, subir/bajar recursos, etc.
- ✦ Ofrecer interfases en otras lenguas, como lenguas indígenas.
- ✦ Utilizar el archivo en, e.g., telesecundarias.
- ✦ Manejar su archivo a larga distancia.

Cuánto tiempo toma establecer todo eso?

-
- ✦ Desarrollo del sitio de la red y de la base de datos inicial: 1-2 años, dependiendo del personal, infraestructura y las selecciones de tecnología.
 - ✦ Después de 2 años se va a encontrar con 100 cosas que se quiere cambiar, por eso planea por lo menos algún tipo de redesarrollo en 2-3 años.

Detalles: espacio físico del archivo

- ✦ Espacio de oficina para cada posición de personal de tiempo completo.
- ✦ Laboratorio de digitalización: el lab de AILLA es ~ 5m x 7m., lo cual es suficiente para 3 estaciones para digitalización, una mesa de trabajo para ordenar depósitos y tener reuniones, y una cuarta computadora a usar mientras digitaliza en las otras.
- ✦ Muchos estante para guardar depósitos en curso.

Detalles: artículos misceláneos

- ✦ CDs y cajas para hacer copias para depositantes
- ✦ DVDs para copias de seguridad convenientes
- ✦ disco duro externo (pequeño) para hacer copias de seguridad de la computadora administrativa (del gerente;
- ✦ cada tipo de cable y conector que exista
- ✦ cajas blancas y ordenadas para ordenar y guardar transitoriamente depósitos en curs;
- ✦ artículos para limpiar y empalmar cintas

Detalles: una estación de digitalización

Computadora: Dell Dimension (PC) con 1 Gb RAM y 40 Gb disco duro. (~\$1500)

AD convertidor: M-Audio Flying Cow (44.1/24) (~\$300)

Programa: SoundForge 8.0 (~\$200)

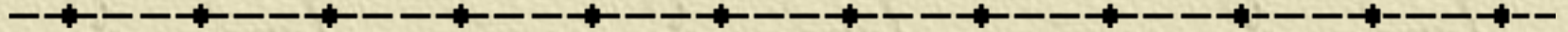
Tocadores de medios:

- rolo abierto: Revox, Aiva, Uhers (pedir de amigos)
- casete: Tascam (\$700)
- minidisco: Tascam (\$500)

Escaneador: HP (~\$400)

Totalidad: \$3450 (más cables, etc.)

¿ Realmente vale la pena?



¡ Absolutamente sí !

Un archivo bien establecido:

- ✦ Puede fomentar “archivos hijos” o “archivos jukebox” en comunidades locales: una computadora ordinaria con un juego de mp3s, refrescado del archivo materno de vez en cuando;
- ✦ Puede animar la documentación amplia de lenguas por proveer un lugar para preservar y publicar grabaciones;
- ✦ Puede distribuir materiales educativos por la red a maestros bilingües en todo el país.

Enlaces

✦ Integración tecnológica para educación:

✦ <http://distancia.dgsca.unam.mx/>

✦ “Experiencia de la red académica mexicana (CUDI)”:

✦ <http://lacnic.net/CUDI-MX-lacnicVI.pdf>

✦ Centro Universitario de Investigaciones Bibliotecológicas:

✦ <http://cuib.unam.mx/>

Más enlaces

✦ LINGUIST-LIST motor de búsqueda:

- ✦ <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/olac/olac-search-advanced.cfm>

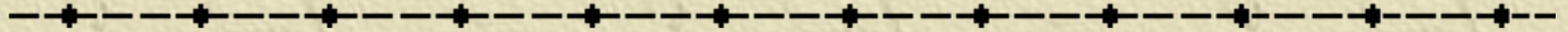
✦ Cyber-lexico comparativo (con mis gracias):

- ✦ <http://www.telefonica.es/fat/elex.html>

✦ Herramientas básicas de multimedia:

- ✦ <http://www.educa.aragob.es/cursoryc/comunes/modulo3/unidad2.htm>

Quejas y sugerencias a:



hjohnson@mail.utexas.edu

<http://www.ailla.utexas.org>