



Cómo Organizar tus Materiales y Datos para un Archivo de Lenguaje

Originalmente dado en CoLANG 2018, University of Florida, junio 18-21, 2017

Traducido al español y modificado por Ryan Sullivant y Susan Kung, octubre-noviembre 2018

Quienes somos

Susan Kung

- Administradora de AILLA
- Lingüista descriptiva
 - Tepehua de Huehuetla, Hidalgo
- Depositante temprana en AILLA (2003)

Ryan Sullivant

- Curador de AILLA
- Lingüista descriptivo
 - Chatino de Tataltepec de Valdés, Oaxaca
- Depositante en AILLA
- Usuario de archivos

Nuestros objetivos

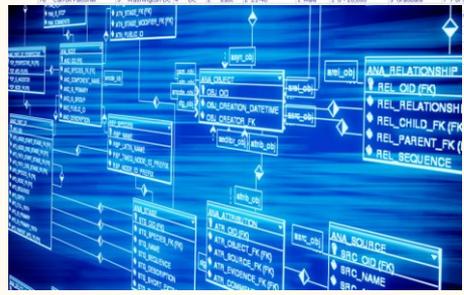
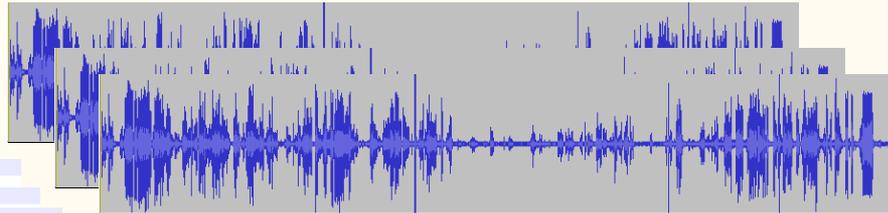
Darles a conocer ...

- qué es el archivado de datos
- el porqué de la preservación
- la importancia de metadatos
- ejemplos de buena descripción
- asuntos de materiales sensibles
- la propiedad intelectual
- los derechos de autor

...y aclarar las dudas que tengan

¿Tengo yo “datos”?

Customer	Customer	Customer	Customer	Customer	Customer	Customer	Household	Missing	Customer	Phone	Customer	Phone	Analytics	Lead	
ID	NAME	ADDRESS	CITY	STATE	ZIP	DOB	INCOME	AGE	STATUS	TYPE	NUMBER	NUMBER	FUNCTION	DATE	
1	Victor Pardo	San Francisco	CA	1	West	1 below 20	1	Owner	1	High	1	High	36	25000	
5	Helena Sawyer	New York	NY	2	East	1 45-60	2	Female	50,000 upward	High School	3	2	Owner	2	18000
7	Max Hernandez	New York	NY	2	East	1 15-20(00)	1	Female	15,000-20,000	Other	1	1	Owner	2	18000
8	Ernie Colburn	Washington DC	DC	2	East	2 31-40	1	Female	80,001-100,000	Other	3	1	Owner	1	130000
11	Eric Keller	Washington DC	DC	2	East	1 45-60	2	Owner	20,001-40,000	High School	1	1	Owner	1	15000
14	Carole Robinson	San Francisco	CA	1	West	1 61 and above	1	Female	50,001 upward	Other	7	7 or more	Owner	1	120000
15	Heather Palmer	New York	NY	2	East	2 21-40	1	Female	15,000-20,000	High School	1	1	Owner	1	15000
16	Shirley Schreiber	New York	NY	2	East	1 45-60	1	Female	40,001-60,000	Graduate	7	7 or more	Owner	1	130000
18	Constance Green	San Francisco	CA	1	West	1 15 and above	1	Female	40,001 upward	Graduate	4	4	Owner	1	150000
19	John Condit	San Francisco	CA	1	West	1 61 and above	2	Female	40,001-60,000	Graduate	4	4	Owner	1	150000
21	Isaac Vannucci	Washington DC	DC	2	East	2 21-40	1	Female	15,000-20,000	High School	1	1	Owner	1	100000
23	Paul Lindberg	Virginia	VA	2	East	1 below 20	1	Male	15,000-20,000	Other	5	5	Owner	1	100000
26	Frederic Buchner	Pittsburg	PA	2	East	1 45-60	1	Male	20,001-40,000	Graduate	3	1	Owner	1	100000
27	John J. Powell	New York	NY	2	East	1 20(00)	1	Male	20,001-40,000	Undergraduate	1	1	Owner	1	15000
30	James Lind	San Francisco	CA	1	West	1 61 and above	2	Female	40,001-60,000	Other	2	1	Owner	1	14000
31	Christie Padua	Los Angeles	CA	1	West	1 21-40	1	Female	50,001 upward	Graduate	2	2	Owner	1	18000
32	Theresean Condon	Washington DC	DC	2	East	1 45-60	1	Female	20,001-40,000	Other	1	1	Owner	1	10000
34	Keith Edmondson	Washington DC	DC	2	East	2 21-40	1	Male	50,001 upward	Graduate	7	7 or more	Owner	1	18000
35	Harvey Duff	New York	NY	2	East	1 45-60	1	Female	20,001-40,000	High School	4	2	Owner	1	10000
37	Eric Leblond	Pittsburg	PA	2	East	1 below 20	1	Male	50,001 upward	Graduate	5	2	Owner	1	130000
38	Heather Harris	New York	NY	2	East	1 below 20	1	Female	15,000-20,000	Other	4	4	Owner	1	10000
41	Lynwood Stahl	San Francisco	CA	1	West	1 31-40	1	Male	40,001-60,000	Graduate	4	2	Owner	1	130000
42	Constance Hoffman	San Francisco	CA	1	West	1 15 and above	1	Female	40,001-60,000	Undergraduate	1	1	Owner	1	10000
43	Joseph Hughes	San Francisco	CA	1	West	1 31-40	1	Male	50,001 upward	Graduate	6	1	Owner	1	200000
44	Chingmei Seayates	San Francisco	CA	1	West	1 15 and above	1	Female	40,001-60,000	Graduate	2	2	Owner	1	100000
45	Frederic Bayr	New York	NY	2	East	1 61 and above	1	Male	20,001-40,000	High School	7	7 or more	Owner	1	200000
46	Gregory Borchardt	New York	NY	2	East	1 45-60	1	Male	40,001 upward	High School	2	2	Owner	1	100000
47	Paul Gaudin	Los Angeles	CA	1	West	2 21-40	1	Male	50,001 upward	High School	2	2	Owner	1	200000
48	Barbara Butler	New York	NY	2	East	1 45-60	1	Female	20,001-40,000	Undergraduate	5	1	Owner	1	200000
49	John Gaudin	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	7	7 or more	Owner	1	200000
50	Simon Seayates	New York	NY	2	East	1 45-60	1	Female	50,001 upward	Graduate	1	1	Owner	1	250000
51	Ray Schreiber	San Francisco	CA	1	West	1 45-60	1	Male	20,001-40,000	Graduate	6	2	Owner	1	200000
52	Sam Baker	San Francisco	CA	1	West	1 below 20	1	Male	15,000-20,000	Undergraduate	1	1	Owner	1	20000
53	Frederic Schreiber	Los Angeles	CA	1	West	1 61 and above	1	Male	20,001-40,000	Undergraduate	1	1	Owner	1	20000
54	Dave Knicker	Virginia	VA	2	East	2 21-40	1	Male	20,001-40,000	Undergraduate	6	6	Owner	1	200000
55	Mark Sawyer	New York	NY	2	East	1 below 20	1	Male	15,000-20,000	Undergraduate	5	1	Owner	1	200000
56	Caroline Williams	San Francisco	CA	1	West	1 45-60	1	Female	50,001 upward	Undergraduate	1	2	Owner	1	200000
59	Leah Green	Washington DC	DC	2	East	1 45-60	1	Female	40,001-60,000	Undergraduate	1	1	Owner	1	100000
62	Clare Anderson	Seattle	WA	2	West	2 21-40	1	Female	20,001-40,000	Other	2	2	Owner	1	200000
63	Lucy Green	New York	NY	2	East	1 15 and above	1	Female	40,001-60,000	Other	8	8	Owner	1	100000
64	John Green	Pittsburg	PA	2	East	1 below 20	1	Male	15,000-20,000	High School	7	7 or more	Owner	1	200000
65	Lucy Green	Los Angeles	CA	1	West	1 45-60	1	Female	20,001-40,000	High School	2	2	Owner	1	100000
70	Patricia Knox	Washington DC	DC	2	East	1 45-60	1	Female	50,001 upward	Other	4	4	Owner	1	200000
71	Camille Spencer	Washington DC	DC	2	East	1 21-40	1	Female	15,000-20,000	Graduate	1	1	Owner	1	200000
72	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
73	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
74	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
75	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
76	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
77	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
78	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
79	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
80	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
81	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
82	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
83	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
84	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
85	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
86	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
87	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
88	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
89	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
90	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
91	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
92	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
93	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
94	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
95	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
96	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
97	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
98	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
99	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000
100	John Green	San Francisco	CA	1	West	1 45-60	1	Male	50,001 upward	Graduate	1	1	Owner	1	200000



```

SochChin-MapNav-2-1
M- SochChin-MapNav-2-1
M- SochChin-MapNav-2-1
M- M-C
TC M-M-Chin: Pél ré13
M-M-Spa: Pedro.
P-TC M-Eng: Pedro.
P- TC
P-P-Chin: he13
P-E P-Spa: eh?
M-TC P-Eng: What?
M- TC
M-M-C
TC M-M-Chin: ma3tsau13 juá32 cheih32 fi1 tash32 cá2mlau21 réi13
M-M-Spa: Vamos por en frente de la terminal.
M-Eng: Let's go by the front of the terminal.
TC M-Eng:
TC 00:02:47.480 - 00:02:50.490
    
```

Time	Speaker	Text
00:02:47.480 - 00:02:50.490	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:48.320 - 00:02:49.980	P-TC	TC
00:02:49.980 - 00:02:50.490	P-E	eh?
00:02:50.490 - 00:02:51.000	M-TC	TC
00:02:51.000 - 00:02:52.000	M-M-Chin	he13
00:02:52.000 - 00:02:53.000	M-M-Spa	Pedro.
00:02:53.000 - 00:02:54.000	M-Eng	What?
00:02:54.000 - 00:02:55.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:55.000 - 00:02:56.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:56.000 - 00:02:57.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:57.000 - 00:02:58.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:58.000 - 00:02:59.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:02:59.000 - 00:03:00.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1
00:03:00.000 - 00:03:01.000	M-Ch	laun2 m2tsau13 di2 juá32 cheih32 fi1 hna2 tsu2 m3 ré1

el desarrollo del archivo

—

Al principio

Se almacenaron materiales físicos (sobre todo textuales) en acervos, universidades y museos

Se difundieron en revistas y libros académicos o redes informales de colectores

Para ver materiales, hacía falta...

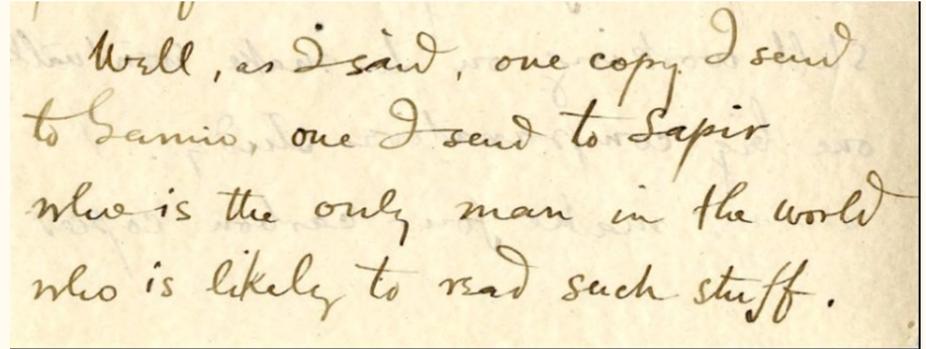
- suponer que existieran
- los fondos y oportunidad de viajar
- conocer a aquellos que los conservaban

Unquachogs. about 20, settled they an institute the Pusepa lock settlement in the town of Brookhaven S. side of Long island.	bread. ap. me. nye coco. cane homin. samp. meat. weewow. stone. sun. clay. pookwe squawb. sand yaac water. nup dirt. pookwe the world. waamapamakee. skiy. ke-ik cloud. pamayaxen rain. sukum snow. soachpo ice. coprat hail. mosécar sun. haquagna moon. nee pa. star. aragusac fire. ruht. yukt. oil. miltapagum. grease. pum. miltap. ikale. fish. operamac oyster. apoonahac. clam. poquackoe. a man. run. woman. squah. child. peewitotat boy. magnehac. lad. nungcum. girl. squasoc. last. yunksquas. head. theygunne. hair. we-usk eye. shasuc. nose. coohoy mouth. cutth teeth. kaput lips. kussis'it. chin. etumyuar cheek. uargnono ear. catawoc. neck. keesquis. shoulder. paguear arm. coputia elbow. keesquar. hand. coit. ho.	back. cu belly. i hip. cor thigh. cop knee. cu leg. cor foot. cu gr. toe. cur little toe. pee father. cu mother. cu brother. e sister. le child. n aunt. e uncle. n gr. father gr. mother husband. wife. gr. child milk. w peas. no beans. m black. shi white. wa red. squ yellow. w blue. se green. je rain. bon bow. atu arrow. n tomahawk a pot. a bed. a blanket axe. oc house. w door. s. chimney. h gourd. ap water. w margarum. moccasin sord. wo sh. ma sh.
---	--	--

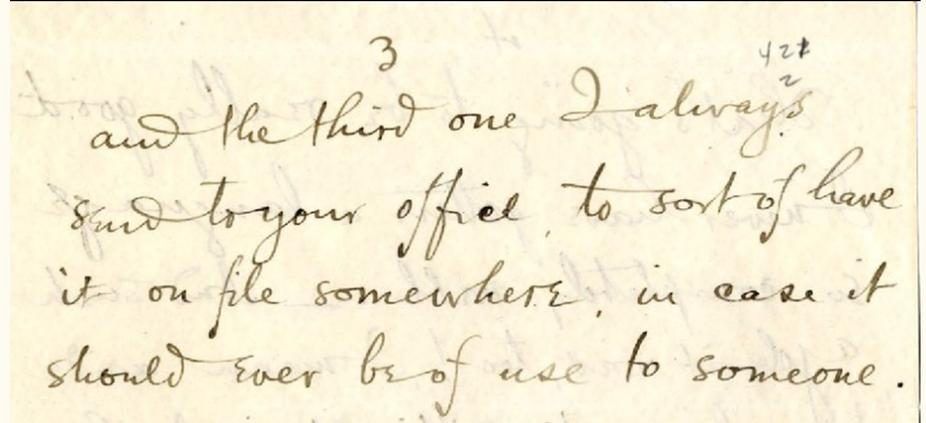
Circulación limitada

Jaime de Angulo escribe a Franz Boas, 1924:

Pues, como dije, una copia [de mis papeles] envió a Manuel Gamio, una envió a Edward Sapir quien es el único en el mundo quien probablemente lee tales cosas, y la tercera siempre la envió a tu oficina para tenerla registrada de cierta manera en alguna parte en caso de que sea útil para alguien.



Well, as I said, one copy I send to Gamio, one I send to Sapir who is the only man in the world who is likely to read such stuff.



3
and the third one I always ^{42t} send to your office, to sort of have it on file somewhere, in case it should ever be of use to someone.

Materiales analógicos

Más allá del papel, los recopiladores siempre estaban al día con la tecnología, y grabaron voces e imágenes en diversos materiales.

- Cilindros de cera
- Alambre magnético
- Discos de vinil
- Cinta magnética
- Película
- Video

...y todos requieren aparatos especiales para tocarlas

Arriba: Frances Densmore escucha un fonógrafo de cilindro de cera con Mountain Chief. <https://www.loc.gov/item/2004667752/>
Abajo: Profesor de antropología Melville Jacobs graba la voz de Annie Miner Peterson del tribu Coos con su nuevamente construido fonógrafo eléctrico portátil durante su visita a Charleston, Oregon, julio de 1934.
<http://digitalcollections.lib.washington.edu/cdm/ref/collection/portraits/id/684>



Desde entonces...

- Los 1990
 - Se prestó más atención a idiomas en peligro
 - Se aumentó el enfoque en la preservación de estos idiomas y sus registros
- Los 2000
 - El archivado se convierte en componente esencial de la documentación lingüística
- Hoy
 - Tendencias emergentes
 - Expandir los públicos de los acervos de datos lingüísticos
 - Eliminar las fronteras tradicionales entre depositantes, usuarios y archiveros

Archivos digitales de lenguaje

Los repositorios de hoy en día son

- De multimedia
- Llenos de materiales de origen digital
- Más complejos
- Más extensos
- Más accesibles
- Un deber de muchos recopiladores
- Un requisito para unos recopiladores



CALIFORNIA LANGUAGE ARCHIVE



Archivos digitales de lenguaje

Los acervos digitales difieren de acervos tradicionales

- Las colecciones se centran por idioma, comunidad o proyecto
 - no todos los productos de un recopilador particular
- Se recopilan los materiales con el propósito de crear un archivo
 - no constan de materiales que se crearon por otros fines
- Poco tiempo pasa entre la creación y el archivado
 - no se esperan décadas

Archivos digitales de lenguaje NO son para el almacenamiento temporal.
Son colecciones PERMANENTES de datos.

preservación y acceso

—

El propósito de un archivo

**Preservación y acceso a
largo plazo**

No es para almacenamiento
temporal

(no es DropBox)

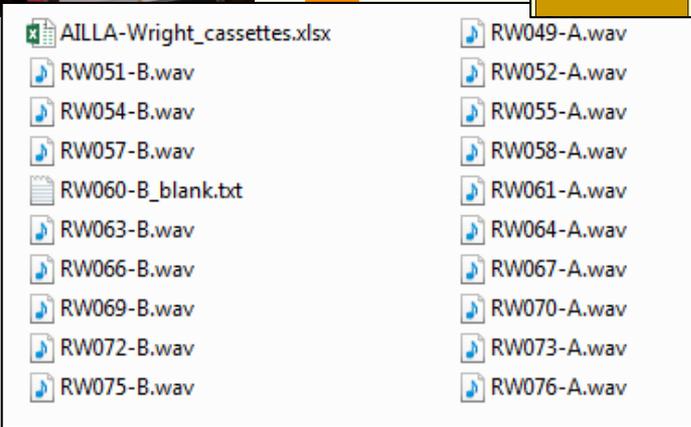
No es primariamente para
divulgación

(no es YouTube)



<http://jessalingel.tumblr.com/post/66108958850/this-is-not-an-archive>

Preservación incluye migración



Preservación para el futuro y el acceso

Los archivos preservan materiales para que los usuarios futuros puedan ocuparlos

¿Quiénes son estos usuarios futuros?

- Los recopiladores mismos
- Recuperadores del idioma/ hablantes de idiomas heredados
- Miembros de la comunidad
- Investigadores futuros (lingüistas, historiadores, genealogistas)
- Organizaciones gubernamentales y no gubernamentales
- El público general

Preservación para el futuro y el acceso

Los acervos lingüísticos son importantes en contextos de sustitución lingüística.

Proveen un registro de la ecología del idioma.

Proveen un registro de un pueblo y una cultura en un dado momento.

Los repositorios especializados para lenguaje pueden ofrecer útiles campos de metadatos contextuales que el usuario puede buscar o navegar.

la reutilización

—

Reutilización: Los Papeles de J.P. Harrington

- Cuadernos de apuntes y grabaciones de audio
- Se han redactado muchas tesis, gramáticas y diccionarios basados en sus datos.

- Ineseño Chumash (Applegate 1972)
- Barbareño Chumash (Wash 1995, 2001)
- Mutsun (Okrand 1977)
- Salinan (Turner 1987)
- Rumsen (Miller 1999)
- Chocheño (Miller & Callaghan 1999)
- Chimariko (Jany 2009)
- Cupeño (Hill 2005)
- ... y seguramente hay otros ejemplos!



J.P. Harrington y una muchacha tule graban el idioma y canciones gunas con un dictáfono en el Instituto Smithsonian. http://ids.si.edu/ids/deliveryService?id=NMNH-gn_04305b

Reutilización: Ejemplos de ELAR

- Nivel académico:
 - Tesis de Maestría externo basado en materiales de gorwaa recopilados por Andrew Harvey <https://elar.soas.ac.uk/Collection/MPI1014224>
- Nivel de la comunidad:
 - Desarrollo de una aplicación de finalización de texto para minderic para fomentar la comunicación digital en el idioma <http://www.poio.eu> basado en datos recopilados por Vera Ferreira
- Nivel de divulgación:
 - Instalación de video por cineastas profesionales para presentar idiomas y sus colecciones a un público más grande <https://elar.soas.ac.uk/Collection/MPI156291>
 - Last Whispers [Últimos Susurros] - Exhibición en el Museo Británico por Lena Herzog con materiales archivado de diferentes coecciones de ELAR: <http://www.lastwhispers.org>

Reutilización: Ejemplos de AILLA

Una gramática de selk'nam

(Rojas Berscia 2014)

Análisis acústicos de idiomas mayas
2015)

(Burnett-Deas 2009, Bennett 2010, Baird 2011, Noguchi

Presentación de fenómenos prosódicos de diversos idiomas
(Clemens 2016)

Análisis sociolingüístico de géneros de canción amazónicos
2015, 2017)

(Hosemann 2013,

Extracción de información léxica de grabaciones mixtecas
(Sullivant 2012)

Reutilización: Revitalización y Recuperación

Materiales archivados han sido esenciales para varios esfuerzos de revitalización y reclamación lingüística

EEUU: **Myaamia** (Baldwin 2003 i.a.)

Wampanoag (Baird 2013)

Mutsun (Warner et al. 2007)

Eyak (Holton 2012)

Tunica (Anderson 2017)

Chitimacha (Hieber 2015) **Tongva**

(Munro 2014)

y otros como **Hupa, Natchez, Osage, Otoe-Missouria**

Canadá: **Huron-Wendat** (Dorais et al. 2011; Lukaniec 2017)

y muchos más **Kaska, Ktunaxa, Kwak'wala, Nisga'a, Northern St'at'imcets, Nuučaan'ut, Secwepemc, Tlingit, Tsilhqot'in, Xa'isla, Xaayda kil, XaiXais, ʔayajuθəm**

Australia: **Kaurna** (Amery 2000)

Palawa Kani [Tasmanian] (Berk 2017)

Reutilización: Enseñanza

Creación de materiales pedagógicos de colecciones archivadas

WordByWord:

Enseñanza de vocabulario mindérica

<http://media.cidles.eu/labs/wordbyword/>



The screenshot displays a user interface for a word-based quiz. On the left side, there are four white buttons with black text: "Open", "Previous", "Next", and "Solution". The main content area on the right has a grey background. At the top, the word "universidade" is written in bold blue text. Below it, there are three radio button options: "classe da da borra", "classe ancha do touquim" (which is selected with a blue dot), and "classe do choane / cartuxa do choane". Below the options, the text "classe ancha do touquim" is displayed in bold blue, followed by the message "This was the correct answer!". At the bottom right, there is a white button with black text that says "Read Solution".

Reutilización: *Afirmación de Identidad*

El movimiento de doble nombramiento en palawa kani (tasmaniano) busca nombres originales en fuentes archiviales



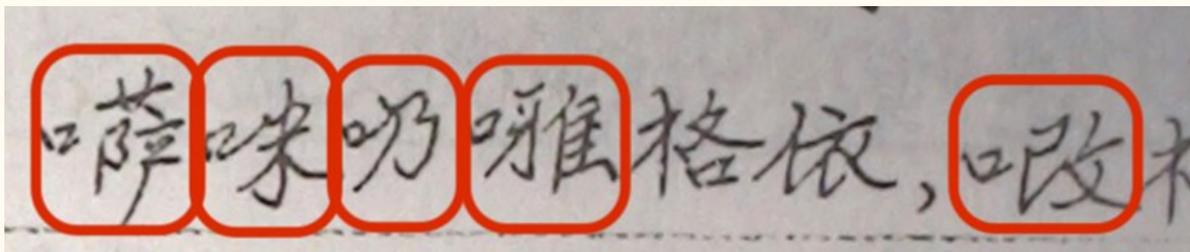
nipaluna/Hobart

Reutilización: Mejorar la documentación

Hoy duoxu se habla por 9 ancianos.

Hablante Wu Wancai recopiló datos de duoxu desde 1955, inventando caracteres para poder escribirlo.

Un proyecto del ELDP de 2013-2017 reutilizó sus datos para ayudar a los hablantes a recordar el idioma



Reutilización: Arte & Música

Grabaciones han sido reutilizadas para fines artísticas ([artículo del New York Times](#))

- *Tree of Codes* (Lim) incluye fragmentos del turco silbado
 - ([ponencia de Lim](#); [entrevista con Lim](#); [fragmento de la obra](#))
- *Ishi's Song* (Bresnick) “un arreglo de un canto por el último miembro de los yahi
 - ([espectáculo por Moore](#); [grabación archivado en CLA](#))
- *Last Whispers*
 - ([sitio web](#); [fragmentos](#))
- *Counting in Quileute* (James)
 - ([soundcloud](#))

Ishi's Song
for Piano Solo

MARTIN BRESNICK

Gently, with swing ♩ = 66

mf (Ye - ni - ni - wo - ki - ma wo - ki - ma ye - ni - ni - wo - ki - ma)*

Reutilización: Personal

“Especialmente con la colección del idioma osage, hay personas que vienen y nos dicen más acerca de las cosas. Algo como ‘Ésa fue mi abuela,’”

- Mary Linn, antigua Curadora de la Native American Languages Collection en el Sam Noble Oklahoma Museum of Natural History

“Lo que distingue los pedidos de estos clientes es que reflejan un interés específicamente en el contenido no lingüístico.”

- Gary Holton, antiguo Director del Alaska Native Language Archive ([Holton 2015:106](#))

la gestión de datos

—

Buenas Prácticas de Manejo de Datos

- Gestionar los datos a medida que los recopila ahorra tiempo a largo plazo.
- El tiempo necesario para la gestión de datos es mínimo en comparación con otras tareas.
- ¡No te quedas atascado bajo la avalancha de datos atrasados! = el diluvio digital

Consideraciones Prácticas

- Formatos de archivos
- Encriptación
- Sistemas de nombramiento de archivos

Formatos y encriptación de archivos

Los archivos deben de estar en formatos *abiertos y duraderos*.

Los formatos *propietarios* (que tienen un dueño) están cerrados en software particular que

- puede no estar disponible en el futuro, o sólo disponible en otra versión
- puede tener un costo.

Nunca metas archivos encriptados en un acervo! En cambio:

- utiliza las restricciones del acceso del repositorio
- modifica archivos para eliminar información sensible
- no entregues tales archivos al repositorio

Tipos de Archivos (VIDEO)



Sistemas de nombramiento de archivos

- ¿Qué puede haber en un nombre?
 - Letras (a-z)
 - Números (0-9)
 - Guión (-)
 - Guión bajo (_)
- ¿Qué no debe haber en un nombre?
 - Diacríticos
 - Espacios

Sistemas de nombramiento de archivos

Nombres secuenciales

Nombres semánticos

Nombres semánticos-secuenciales

Sistemas de nombramiento: Nombres secuenciales

File naming

Press `Esc` to exit full screen

Name	Date Modified	Size	Kind
50000001.JPG	Jun 18, 2013, 9:43 AM	4 MB	JPEG image
50000002.JPG	Jun 18, 2013, 9:45 AM	1.9 MB	JPEG image
50000003.JPG	Jun 18, 2013, 9:32 AM	4.1 MB	JPEG image
50000004.JPG	Jun 18, 2013, 9:34 AM	4.3 MB	JPEG image
50000005.JPG	Jun 18, 2013, 9:48 AM	5.3 MB	JPEG image
50000006.JPG	Jun 18, 2013, 9:50 AM	5.7 MB	JPEG image
50000007.JPG	Jun 18, 2013, 9:52 AM	3.9 MB	JPEG image
50000008.JPG	Jun 18, 2013, 9:53 AM	3.6 MB	JPEG image
50000009.JPG	Jun 18, 2013, 9:54 AM	4.1 MB	JPEG image
50000010.JPG	Jun 18, 2013, 9:56 AM	4 MB	JPEG image
50000011.JPG	Jun 18, 2013, 9:58 AM	5.1 MB	JPEG image
50000012.JPG	Jun 18, 2013, 10:01 AM	4.7 MB	JPEG image
50000013.JPG	Jun 18, 2013, 10:03 AM	3.8 MB	JPEG image
50000014.JPG	Jun 18, 2013, 10:05 AM	5 MB	JPEG image
50000015.JPG	Jun 18, 2013, 10:06 AM	4.6 MB	JPEG image



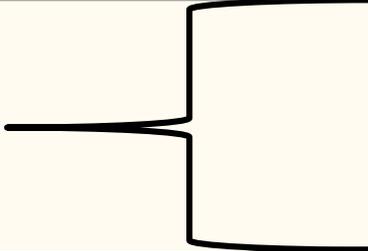
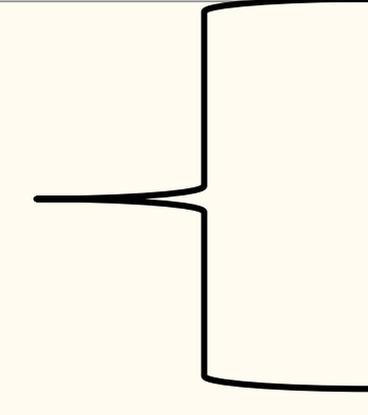
0:30 / 4:29

CC HD

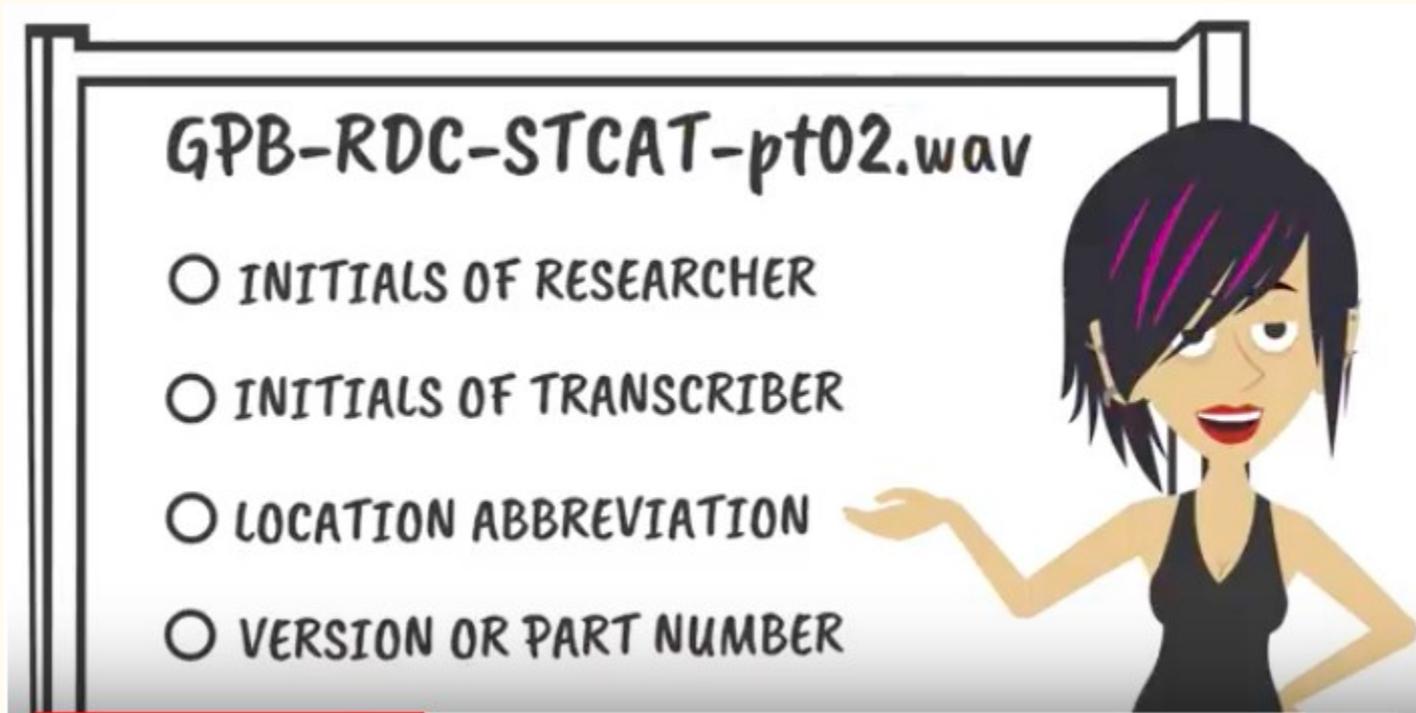
Sistemas de nombramiento: Nombres semánticos

110718-TXT-BG-2.wav	110718	18 de julio de 2011
	TXT	texto narrativo
	BG	hablante Benita Gregorio
	2	la segunda grabación de la sesión
	wav	extensión de formato de audio
Cacra_RW_Herranza_C.mpeg	Cacra	pueblo del hablante
	RW	hablante Regina Huamán
	Herranza	fiesta ganadera andina
	C	comentario sobre un video silencioso
	mpeg	extensión de formato de video

Sistemas de nombramiento: Semántico-Secuencial

minderico223.wav		minderico	palabra clave (nombre del idioma)
		223	número secuencial
		wav	extensión de formato de audio
minderico0001-001.jpg		minderico	palabra clave (nombre del idioma)
		0001	número secuencial
		-	indica archivo es parte de una serie
		001	primero en la serie
		jpg	extensión de formato de imagen

Video de nombramiento de archivos



<https://www.youtube.com/watch?v=c-Mcp5ozgx0>

la organización

—

¿Quién debe organizar tus datos?

¡Tú!

El recopilador o depositante es la mejor persona para organizar los datos.

Los recopiladores tiene la **capacidad, oportunidad e información contextual** para administrar y organizar sus datos mejor.

Tienen los conocimientos contextuales y areales que el archivero no tendrá.

Tienen una familiaridad con los materiales que el archivero no tendrá.

Pueden encontrar, recordar, o reconstruir metadatos a los cuales el archivero no tendrá acceso.

Diferentes clases de juegos

Conjunto

todos los materiales

- archivos en proceso
- fuentes tempranas
- versiones múltiples
- borradores y copias de respaldo
- efímero

Colección

material archivado

- subgrupo del conjunto
- curado
- con metadatos que ayudan al usuario a encontrar y entender materiales

Corpus

un juego de análisis

- uniforme
- creado a propósito

Diferentes clases de repositorios

- Archivos Digitales de Lenguaje
- Acervos Comunitarios
- Repositorios de Datos
- Repositorios Institucionales

Escoger un repositorio

<https://www.linguisticsociety.org/content/finding-archive-your-endangered-language-research-data>

Es posible que tu institución o patrocinador exige cierto repositorio.

Escoger: Archivos Digitales de Lenguaje

Archivos digitales de idiomas son repositorios que tienen un enfoque en datos de idiomas

Campos de metadatos están diseñados para destacar información relevante para investigación lingüística

DELAMAN: Digital Endangered Languages and Musics Archives Network delaman.org

Escoger: Archivos Digitales de Lenguaje

Algunos archivos digitales de lenguaje tienen un **ámbito regional** y probablemente no aceptarán materiales de otras regiones

Pacific and Regional Archive for Digital Sources in Endangered Cultures ([enlace](#))

Archive of Languages and Oral Resources of **Africa**

[\(enlace\)](#)

Alaska Native Language Archive

[\(enlace\)](#)

Archive of the Indigenous Languages of **Latin America**

[\(enlace\)](#)

...pero...

Escoger: Acervo Comunitario (del Idioma)

Acervo, biblioteca, o museo de una universidad tribal

Miami University <https://www.lib.miamioh.edu/digital>

Centro, acervo, biblioteca, o museo de una comunidad

Owens Valley Career Development Center <http://www.ovcdc.com/blog/>

Mukurto u otro Sistema de Gestión de Contenidos

Plateau Peoples' Web Portal <https://plateauportal.libraries.wsu.edu/>

Bridge to the Future Web Portal: <http://bridge-to-the-future.com/>

Escoger: Repositorios de Datos

Repositorios digitales de datos son buenísimos para unas colecciones

- Corpora entera que no ocupan metadatos a nivel del archivo o la sesión
- Colecciones de grabaciones cortas hechas para experimentos acústicos.

Si tus datos están designados a ser utilizados en total, debes considerar un repositorio de datos

Avisos

- Los campos de metadatos puede ser no óptimos para datos lingüísticos
- Puede haber límites en tamaño o cantidad de archivos y/o restricciones de formatos

Escoger: Repositorios Institucionales

Repositorios institucionales puede ser bueno si haya uno en tu universidad

Es posible que te haría falta de la ayuda de un miembro de la facultad

RIs son diseñados principalmente para documentos, pero es posible que funcione por otros materiales también

Avisos

- Puede haber dificultades con audios, videos, o archivos relacionados
- Asegúrate que usuarios fuera de tu institución pueden acceder a los materiales (especialmente si tienes un Plan de Administración de Datos)

Plan de Administración de Datos (DMP en inglés)

Berez-Kroeker, Andrea, Lauren Collister & Susan Smythe Kung. 2017. Workshop on Data Management Plans for Linguistic Research, LSA Summer Linguistic Institute, University of Kentucky.

Grabaciones de audio y diapositivas [en inglés] archivadas en AILLA:

<https://www.ailla.utexas.org/islandora/object/ailla:254604>

(Son disponibles incluso para los que no tengan una cuenta de AILLA.)

Organización de AILLA

- Colección
 - Carpeta
 - Archivos de medios
-

Unidades de organización

Unidades comunes de organización archival:

Colección: Conjunto de materiales de un solo recopilador o proyecto.

Carpeta: Una carpeta de archivos multimedia relacionados de alguna manera (Otros nombres incluyen *recurso* (en AILLA), *sesión*, *resource*, *session*, *conjunto de datos* o *dataset*)

Archivo Multimedia: El archivo digital que contiene los datos o análisis. La cosa que puede verse o descargarse

Anidación Profunda

Rukiga	Colección
Datos de Rukiga	Subcarpetas anidadas
2016	
Verano	
Archivos de audio	
A. Historia y Orígenes	
Sesión 1	
20160711_CGG_APN_Ba chwezi.m4a	Archivo de media



Estructura plana

Al contrario de una estructura anidada, una colección consta de una lista plana de carpetas que solo contienen archivos

Content Files (2)

View file contents

Filename ▲▼	Type ▲▼	File size ▲▼
GH2-aol_ak_061219-elic.mp3	audio/mpeg	99.8 MB
GH2-aol_ak_061219-elic.wav	audio/x-wav	1.08 GB
2 files	--	1.17 GB

Show 10

Show 50

Show all 2

Items in Collection (59)

Item ▲▼	Title ▲▼
aol_ak_061219 open	Alor elicitation session
aol_aw_060324 open	Dolu
bi_am_061021 open	Publia war song (Indonesia),
lev_061023 open	Wedding Discussion
lev_061113 open	Soli
lev_061114 open	Soli
lev_061115 open	Soli
lev_070418 open	Rice Harvest Lunch (Oikos):
lev_am_040806 open	Bu Manema - Kalangburi
lev_am_061018_BuManema open	Bu Manema (abridged, with
lev_am_061021 open	Ang Tubbe's descendents, M Pubila war song
lev_am_070404 open	Air Mama, Baranusa riot 198

Organización: Personal vs. Archival (VIDEO)



Organización de Archivos (VIDEO)



Organización de carpetas

Hay distintas maneras para la organización de carpetas

- por sesión de grabación / evento de lenguaje
- por hablante / localidad / variedad lingüística
- por protocolo
- por tema / forma o función lingüística
- por publicación

Organización de carpetas

- por sesión de grabación / evento de lenguaje

Muchos recopiladores optan por carpetas de medios mixtos organizados por sesión de grabación

- La grabación de algún evento
- Las anotaciones, traducciones, transcripciones y medidas de ella

Puede dar un enfoque en el contenido del recurso

Es mejor para usuarios que están preocupados en el evento de lenguaje mismo.

Una carpeta de sesión puede tener múltiples grabaciones.

Content Files (28)

View file contents

Filename ▲▼	Type ▲▼	File size ▲▼	Duration ▲▼	File access
DGB1-2011_005-waiwaik_1_collecting_plant.mxf	application/mxf	949 MB		View
DGB1-2011_005-waiwaik_1_collecting_plant_video_narration_flextext.txt	application/xml	56.7 KB		View
DGB1-2011_005-waiwaik_1_collecting_plant_video_narration.mp3	audio/mpeg	2.05 MB	00:02:14.609	View
DGB1-2011_005-waiwaik_1_collecting_plant_video_narration.pdf	application/pdf	23.3 KB		View
DGB1-2011_005-waiwaik_1_collecting_plant_video_narration.wav	audio/x-wav	74.2 MB	00:02:13.473	View
DGB1-2011_005-waiwaik_1_collecting_plant.mp4	video/mp4	61.4 MB	00:01:40.400	View
DGB1-2011_005-waiwaik_1_collecting_plant.webm	video/webm	61.8 MB	00:01:40.400	View
DGB1-2011_005-waiwaik_2_cutting_root_video_narration_flextext.txt	application/xml	92.6 KB		View
DGB1-2011_005-waiwaik_2_cutting_root_video_narration.mp3	audio/mpeg	1.84 MB	00:02:00.737	View
DGB1-2011_005-waiwaik_2_cutting_root_video_narration.pdf	application/pdf	30.6 KB		View
DGB1-2011_005-waiwaik_2_cutting_root_video_narration.wav	audio/x-wav	66.6 MB	00:02:00.721	View
DGB1-2011_005-waiwaik_2_cutting_root.mp4	video/mp4	67 MB	00:01:49.200	View
DGB1-2011_005-waiwaik_2_cutting_root.mxf	application/mxf	906 MB		View
DGB1-2011_005-waiwaik_2_cutting_root.webm	video/webm	67.1 MB	00:01:49.200	View
DGB1-2011_005-waiwaik_3_making_drink_video_narration_flextext.txt	application/xml	163 KB		View
DGB1-2011_005-waiwaik_3_making_drink_video_narration.mp3	audio/mpeg	4.68 MB	00:05:07.16	View
DGB1-2011_005-waiwaik_3_making_drink_video_narration.pdf	application/pdf	44.5 KB		View
DGB1-2011_005-waiwaik_3_making_drink_video_narration.wav	audio/x-wav	169 MB	00:05:06.990	View
DGB1-2011_005-waiwaik_3_making_drink.mp4	video/mp4	183 MB	00:04:59.759	View
DGB1-2011_005-waiwaik_3_making_drink.mxf	application/mxf	2.27 GB		View
DGB1-2011_005-waiwaik_3_making_drink.webm	video/webm	184 MB	00:04:59.759	View
DGB1-2011_005-waiwaik_4_drinking_drink_video_narration_flextext.txt	application/xml	100 KB		View
DGB1-2011_005-waiwaik_4_drinking_drink_video_narration.mp3	audio/mpeg	1.71 MB	00:01:51.960	View
DGB1-2011_005-waiwaik_4_drinking_drink_video_narration.pdf	application/pdf	32 KB		View
DGB1-2011_005-waiwaik_4_drinking_drink_video_narration.wav	audio/x-wav	61.7 MB	00:01:51.942	View
DGB1-2011_005-waiwaik_4_drinking_drink.mp4	video/mp4	24.5 MB	00:00:39.880	View
DGB1-2011_005-waiwaik_4_drinking_drink.mxf	application/mxf	321 MB		View
DGB1-2011_005-waiwaik_4_drinking_drink.webm	video/webm	24.5 MB	00:00:39.880	View
28 files	--	5.42 GB	--	--

A veces una carpeta puede contener materiales de diversas fuentes

Por ejemplo, el recopilador puede disponer elicitaciones léxicas por campo semántico, p.ej. agrupando grabaciones de términos de parentesco grabados con diversas personas en distintas ocasiones en una carpeta.

Media Files

There are 18 objects in this resource

Object
120322-KIN-CC-abuela.wav
120322-KIN-CC-abuelo.wav
120322-KIN-CC-ahijado.wav
120322-KIN-FM-madrina_de_vela.wav
120703-KIN-FM-abuelos.wav
120703-KIN-FM-abuelos_voc.wav
120703-KIN-FM-amigo_voc_entre_hombres.wav
120703-KIN-FM-amigo_voc_entre_mujeres.wav
120703-KIN-FM-concuño.wav
120703-KIN-FM-cuñada.wav
120703-KIN-FM-cuñado.wav
120703-KIN-FM-hija_voc.wav
120703-KIN-FM-hijo_voc.wav
120703-KIN-FM-nombre_de_cariño.wav
120703-KIN-FM-padrastro.wav
120703-KIN-FM-saludo_de_compadres.wav
120703-KIN-FM-tía.wav
120703-KIN-FM-tío.wav

Los materiales pueden organizarse por publicaciones.

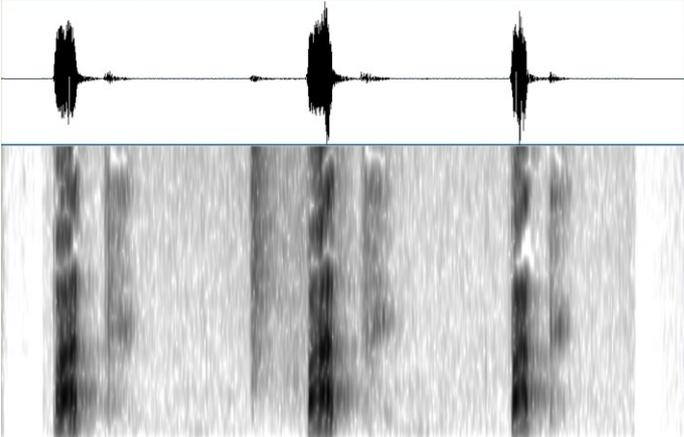
Una carpeta puede contener:

- Grabaciones de las palabras o los textos que aparecen en una publicación

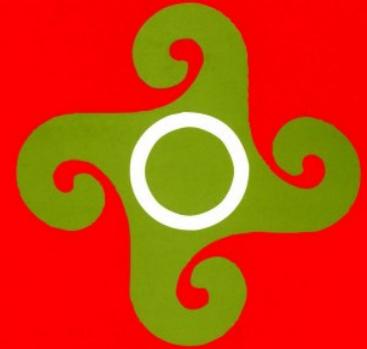
FONEMAS¹

1. Posición inicial de palabra. 2. Posición interna de palabra. 3. final de palabra.

<i>Fonema</i>	<i>Ejemplos</i>	<i>Traducción</i>
/p/	1. pat	casa



The image shows a waveform and spectrogram of the phoneme /p/. The waveform is a black line on a white background, showing three distinct pulses corresponding to the three positions of the phoneme. The spectrogram is a grayscale image below the waveform, showing the frequency spectrum of the phoneme. The spectrogram is divided into three vertical sections, each corresponding to one of the three positions of the phoneme. The first section shows a strong vertical line at a low frequency, indicating a strong pulse. The second section shows a strong vertical line at a higher frequency, indicating a strong pulse. The third section shows a strong vertical line at a low frequency, indicating a strong pulse.



cristina buenrostro

**chuj de san mateo
ixtatán**

archivo de lenguas
indígenas de méxico

Los materiales pueden organizarse por publicaciones.

Una carpeta puede contener:

- Datos y código usados en el análisis de una publicación (*datos de replicación*)

The screenshot shows a data repository page for a dataset titled "Local environment interview data for the dissertation, Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino Sign Language". The page includes a description, subject, and a list of six files for download or exploration.

Local environment interview data for the dissertation, Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino Sign Language Version 1.0

Mesh, Kate. 2017, "Local environment interview data for the dissertation, Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino Sign Language", doi:10.18738/T8/PJXZJI, Texas Data Repository Dataverse, V1, UNF:6:/LRkUBtzvXxkLQ+WovbKXA== [Cite Dataset](#) [Learn about Data Citation Standards.](#)

Description Metalinguistic interview data for Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino Sign Language

Subject Social Sciences

Files Metadata Terms Versions

Search this dataset...

6 Files

<input type="checkbox"/>	<input type="checkbox"/>	LEI_analysis_aggregated_hearing_participants.R R Syntax - 2.8 KB - Sep 1, 2017 - 2 Downloads MD5: 7288984518e8d9e3513df2a104de382	<input type="button" value="Download"/>
<input type="checkbox"/>	<input type="checkbox"/>	LEI_analysis_Koyu.R R Syntax - 1.3 KB - Sep 1, 2017 - 0 Downloads MD5: ca85d5db22ac95c12c894084be0688a	<input type="button" value="Download"/>
<input type="checkbox"/>	<input type="checkbox"/>	LEI_analysis_Sendo.R R Syntax - 1012 bytes - Sep 1, 2017 - 0 Downloads MD5: 802a748561edf8e491f7cde82f4ead2	<input type="button" value="Download"/>
<input type="checkbox"/>	<input type="checkbox"/>	LEI_data_aggregated_hearing_participants.csv text/comma-separated-values - 178.9 KB - Sep 1, 2017 - 3 Downloads MD5: ffe40598938dd98c877c89653f22d507	<input type="button" value="Download"/>
<input type="checkbox"/>	<input type="checkbox"/>	LEI_data_Koyu.tab Tabular Data - 41.8 KB - Sep 1, 2017 - 1 Download 30 Variables, 184 Observations - UNF:6:/W/ftrGPoh0HLZSP7Mw==	<input type="button" value="Explore"/> <input type="button" value="Download"/>
<input type="checkbox"/>	<input type="checkbox"/>	LEI_data_Sendo.tab Tabular Data - 15.7 KB - Sep 1, 2017 - 1 Download 27 Variables, 88 Observations - UNF:6:/TINWoyi0XrhV9Qdtf3ww==	<input type="button" value="Explore"/> <input type="button" value="Download"/>

Mesh, Kate, 2017, "Local environment interview data for the dissertation, Points of Comparison: What Indicating Gestures tell us About the Origins of Signs in San Juan Quiahije Chatino Sign Language", doi:10.18738/T8/PJXZJI, Texas Data Repository Dataverse, V1, UNF:6:/LRkUBtzvXxkLQ+WovbKXA==

el descubrimiento y
los metadatos

—

Descubrimiento

La capacidad de descubrir tu colección y los materiales dentro de ella, está apoyado por metadatos en todos los niveles de la estructura de un repositorio

- Ayuda la localización y recuperación de materiales
- Hace visible los enlaces entre archivos

Descubrimiento

Trabajos con buenos títulos se encuentran y se citan más

Why do academics and PhDers carefully choose useless titles for articles and chapters?

Six ways to get it wrong, and four steps to get it right



C

[Clinics \(Sao Paulo\)](#). 2012 May; 67(5): 509–513.

doi: [10.6061/clinics/2012\(05\)17](https://doi.org/10.6061/clinics/2012(05)17)

PMCID: [PMC3351256](#)

PMID: [22666797](#)

Articles with short titles describing the results are cited more often

[Carlos Eduardo Paiva](#)^{I,II}, [João Paulo da Silveira Nogueira Lima](#)^I and [Bianca Sakamoto Ribeiro Paiva](#)^{II}

[Author information](#) ► [Article notes](#) ► [Copyright and License information](#) ► [Disclaimer](#)

Descubrimiento - Relaciones entre archivos

Recopiladores de lenguaje suelen indicar relaciones entre archivo por nombres que varían solamente en extensión.

Algunos archivos como EAF tienen vínculos a otros archivos

- Los archivos vinculados deben ser nombrados en los metadatos
- Especialmente si fueron renombrados
- Incluso si esta información puede recuperarse dentro de los archivos mismos

CAM20548_01.mpg	grabación de video original
CAM20548_01.wav	audio del video
CAM20548_01.eaf	transcripción alineado con el tiempo
CAM20548_01.txt	traducción libre

```
<HEADER MEDIA_FILE="" TIME_UNITS="milliseconds">
  <MEDIA_DESCRIPTOR EXTRACTED_FROM="file:///Volumes/Sicoli-1.5TB-Seagate/Chinantec/CAM20548_01.mpg"
    MEDIA_URL="
file:///Volumes/Sicoli-1.5TB-Seagate/Chinantec/FieldSochiapamNov2011/SochNov2011Audio/FogWhistlers/FromVideo/CAM20548_01.wav" MIME_TYPE="audio/x-wav" RELATIVE_MEDIA_URL="./CAM20548_01.wav"/>
  <MEDIA_DESCRIPTOR MEDIA_URL="file:///Volumes/Sicoli-1.5TB-Seagate/Chinantec/CAM20548_01.mpg"
    MIME_TYPE="video/mpeg" RELATIVE_MEDIA_URL="../../../../CAM20548_01.mpg"/>
  <PROPERTY NAME="lastUsedAnnotationId">105</PROPERTY>
</HEADER>
```

Descubrimiento

Hay dos maneras de descubrimiento: navegación y búsqueda

En **navegación**, el usuario abre una colección y revisa las carpetas para encontrar lo que busca

En **búsqueda**, el usuario busca palabras por todos los metadatos de AILLA y revisa los resultados por lo que busca

Recientemente, AILLA estrenó una **búsqueda avanzada** que permite que el usuario filtre los resultados

Navegación

Puede ser complicado si los títulos no son suficientemente descriptivos

En una sola colección (con enfoque en muchos idiomas mayas), hay 10 carpetas con el mismo título.

Si estás interesado en materiales en kaqchikel, ¿cuál es la carpeta que debes abrir?



The Compadres



Comparative Vocabulary



Comparative Vocabulary



Comparative vocabulary



Comparative Vocabulary



Comparative Vocabulary



Comparative vocabulary



Comparative Vocabulary



Comparative vocabulary



Comparative vocabulary



Comparative Vocabulary



Complement

Búsqueda

Y una búsqueda únicamente funciona si hay palabras claves en los metadatos.

Es dudoso que esta grabación sea descubierta.

Tape 7-1b

Cinta 7-1b

Detalles del Objeto

Idioma	shuar
Idioma PID(s)	ailla:119715
Título [Indígena]	
Idioma del Título Indígena	
Título	Cinta 7-1b
Comunidad del Idioma	
País	Ecuador
Lugar de Creación	
Fecha de Creación	0000-00-00
Descripción [Indígena]	
Idioma de la descripción indígena	
Descripción	Este lado del casete no lleva apuntes. El lado opuesto es Cinta 7-1a.
Géneros	Conversación
Nota Sobre la Fuente	Tape_7_1_B
Referencia	
Contribuidor(es) Individual / Rol de colaborador	Gnerre, Maurizio (Investigador)
Contribuidor(es) Corporativo / Rol de colaborador	

Descubrimiento

Buenos títulos con suficiente metadatos descriptivos ayuda al descubrimiento por navegación y por búsqueda

Descripción	Para distinguir los tonos Receptivos de los Fuertes entre las palabras de tono Relajado o Bajo, grabamos unos verbos en el aspecto potencial y completivo.
Descripción	Florencia Mejía habla de su viaje a Acapulco, un paseo en lancha para ver unas estatuas de santos costa afuera, como un clavista recogió una moneda tirado en el agua por un turista, el calor de Acapulco.
Descripción	Todas las lenguas chatinas tienen dos estrategias para la flexión personal de sustantivos y sustantivos relacionales/preposiciones/marcadores de caso. Los que son poseídos alienablemente indican su poseedor por poner el poseedor (procedido por el marcador jiʔin) después del sustantivo poseído. Los que son poseídos inalienablemente y los que indican relaciones espaciales (la mayoría derivada de sustantivos poseídos) demuestran su poseedor por la yuxtaposición sencilla o, en el caso de un poseedor de la primera o segunda persona singular, por un cambio de tono y/o en la vocal tónica. Estos archivos son grabaciones de muchas palabras así tanto sin flexión que con flexión por un poseedor de la primera persona singular y luego de la segunda persona singular.

el archivado progresivo

—

El modelo del archivado progresivo

Una recomendación para proyectos en curso

- Deposita las partes inmutables de la colección (“datos brutos”) primero
- Agrega archivos adicionales al finalizarlos

Las tareas del archivado

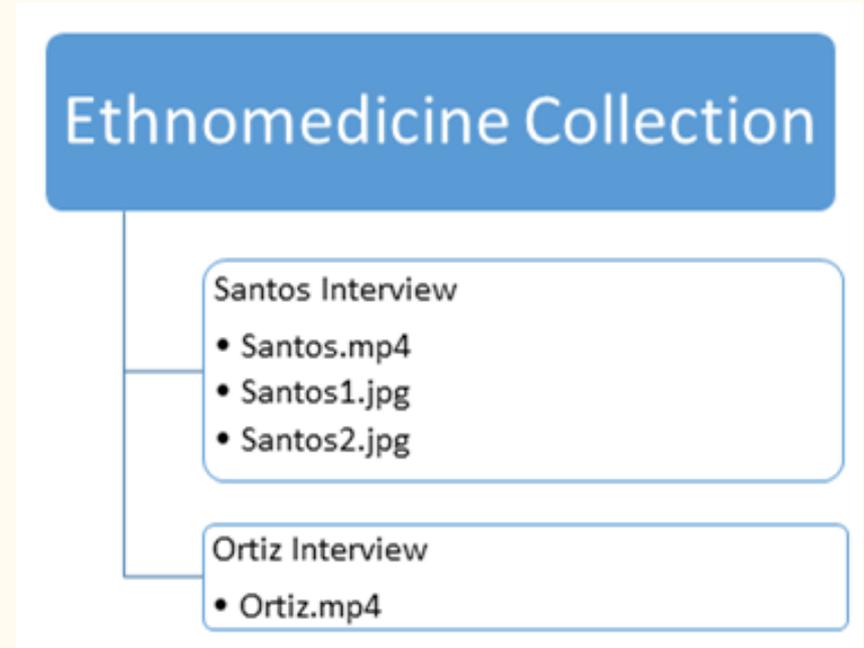
- están distribuidos tras el tiempo
- están realizados más próximamente a la colección o producción de los archivos
- requieren metadatos administrativos para saber cuales archivos ya se archivaron

El archivado progresivo (VIDEO)



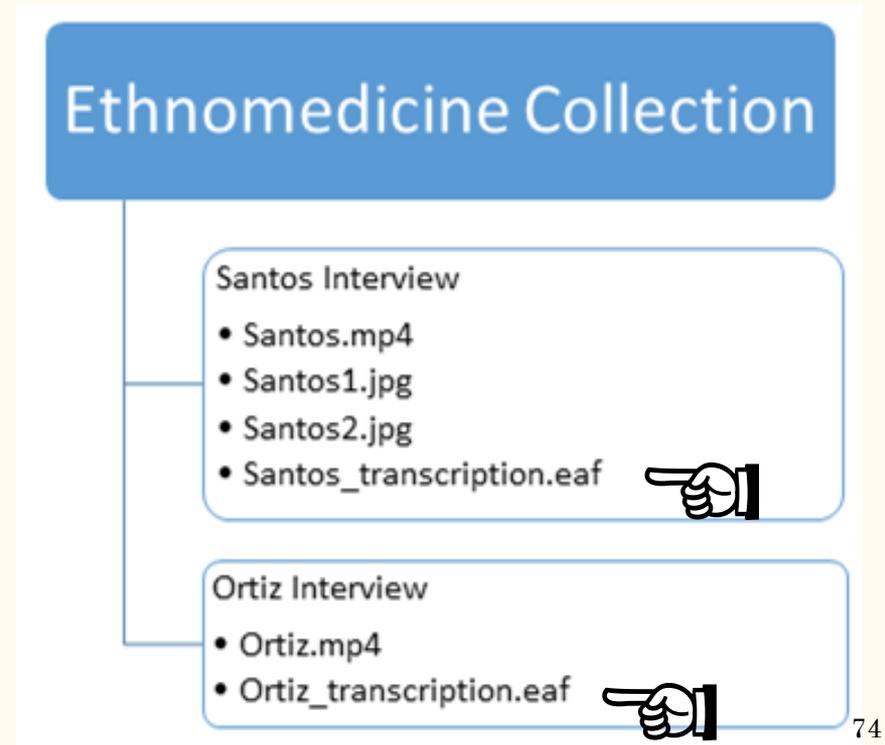
Ejemplo del archivado progresivo

Deposita videos y fotos de entrevistas después de la primera temporada de investigación



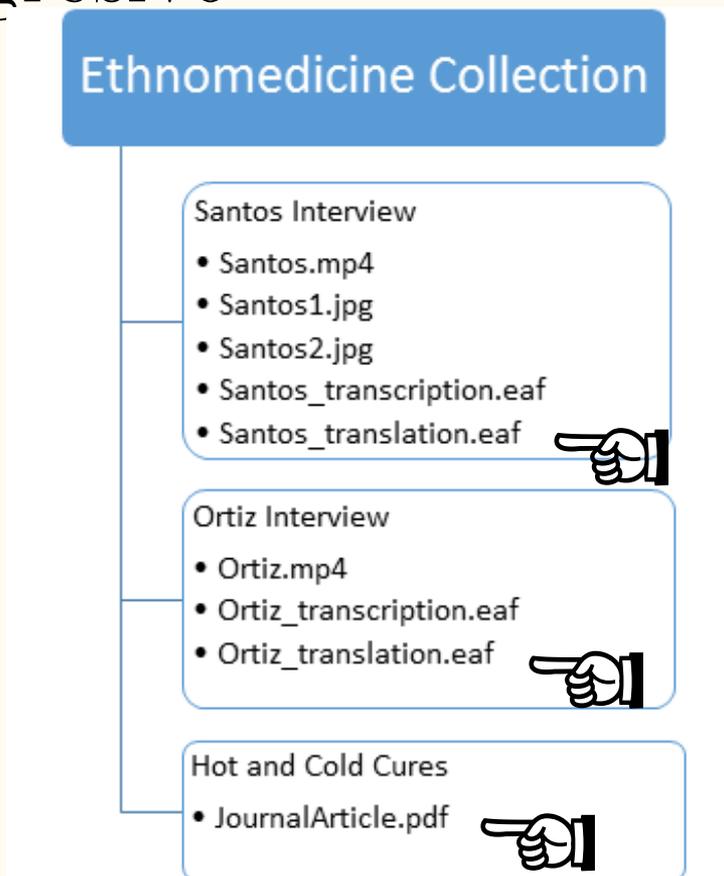
Ejemplo del archivado progresivo

Agrega las transcripciones acabadas durante la segunda temporada de investigación



Ejemplo del archivado progresivo

Agrega otros objetos como traducciones y documentos analíticos al acabarlos.



Avisos

Puede haber restricciones estructurales al archivado progresivo.

Por ejemplo, en AILLA no se puede agregar archivos grandes a carpetas existentes.

descripción archivística

—

Descripción archivística

Como los metadatos están distribuidos por toda la colección, comprender qué hay en una colección y localizarlo puede ser retador por muchos usuarios.

Aun con búsquedas avanzadas, puede ser útil tener un solo documento que explica información contextual y descriptiva de una colección.

Vemos cuatro posibles formas de descripciones...

Descripciones, Guías, Ayudas, Reseñas

Documentos que pueden avisar los contenidos de una colección al usuario.

Tienen públicos y propósitos diferentes y así tienen diferentes contenidos.

	Info. básica	Info. detallada	Lista de materiales	Evaluación crítica
Descripción	Requerido	Opcional	Opcional	No
Guía	Requerido	<u>Requerido</u>	Opcional	No
Ayuda de Localización	Requerido	Opcional	<u>Requerido</u>	No
Reseña	Requerido	Recomendado	Opcional	<u>Sí</u>

Descripciones de Colecciones

Proveen información básica de la colección:

- Título
- Identificadores & Repositorio
- Idioma(s)
- Ubicación geográfica
- Comentarios breves del contenido
- Comentarios breves de la comunidad

Si, Aung. Documentation of Danau, an endangered language of Myanmar (Burma). ELAR. <https://elar.soas.ac.uk/Collection/MP1172411>

Documentation of Danau, an endangered language of Myanmar (Burma)

Deposit

Bundles and resources

Documentation of Danau, an endangered language of Myanmar (Burma)

Language:	Danau (ISO639-3:dnu)
Depositor:	Aung Si
Location:	Myanmar
Deposit Id:	0321
ELDP Id:	SG0205
Level:	Deposit

Summary of deposit

This collection represents a preliminary documentation of the Austroasiatic language Danau, which is spoken primarily in 3 or 4 villages near the town of Aungban in Shan State, Myanmar (formerly Burma). These villages include: Taungbohla Village (lat: 20.575; long: 96.790), Htinyugon Village (lat: 20.603; long: 96.736), Chaungya Village (lat: 20.614; long: 96.767), Aung Chan Tha Monastery in Aungban (lat: 20.657; long: 96.631).

Audio recordings representing a range of speech genres have been deposited here: these include lexical and grammatical elicitation, conversation, description and rituals. Recordings containing names of local, culturally important plants are accompanied by colour photographs of these plants. Around 90 minutes of audio (not including the wordlist recordings) have been transcribed and translated so far.

The collection also includes a preliminary analysis of the phonetics, phonology and grammar of Danau has been achieved to date, and the transcriptions presented here should be regarded as provisional and subject to revision.

Group represented

Speakers of Danau language in Myanmar

Language information

The Danau language of the Shan State of Myanmar (Burma) was reported to be "dying" as long back as 1965. Anecdotal evidence from the field suggests that since the language has no official protection in the country, and is not formally taught, it is only spoken by a handful of scattered communities. Although around 10,000 people identify themselves as 'Danau', it has never been properly documented, and its future remains highly uncertain. 80

Evaluando Descripciones en AILLA

- Colección del Proyecto de Documentación de la Lengua Iquito
- Colección Chatino de Leslie y Kitty Pride
- La Colección Livingsteña del Idioma Garífuna
- Colección de Lenguas Amazónicas de Arthur P. Sorensen, Jr.
- El Proyecto de la Documentación del Chatino de Zacatepec

¿Qué opinan de estas descripciones?

¿Cuáles datos les faltan?

Evaluando Descripciones en AILLA

Visita la [Colección Oxlajuuj Keej Maya' Ajtz'iib' de Lenguas Mayas](#)

Lee la description de ella en “[Unrestricted Mayan Language Resources In the Archive of the Indigenous Languages of Latin America](#)”

¿Te ayudó a comprender la colección esa descripción?

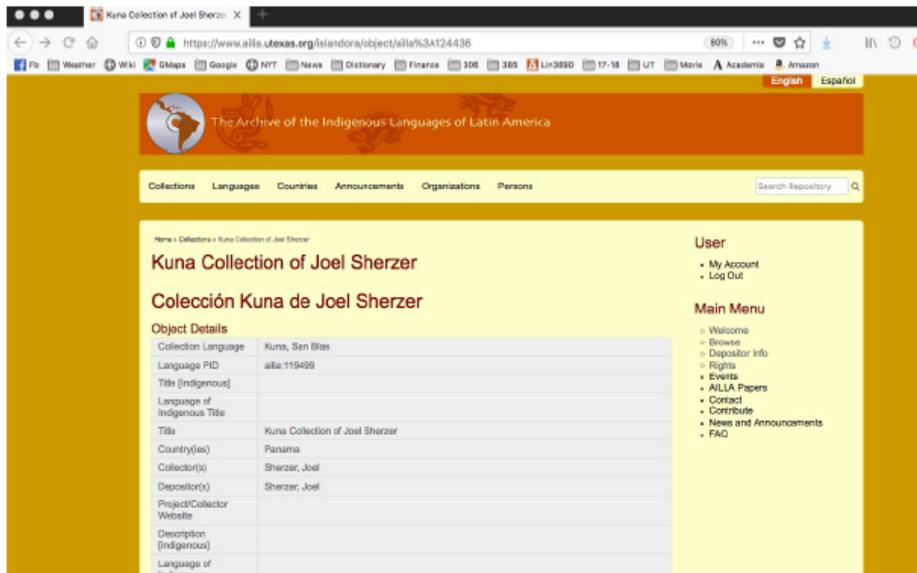
¿Qué más información te ayudaría a encontrar materiales en esta colección?

Ayudas de Localización

Información básica junto con una lista completa de carpetas y/o archivos

Appendix B: Navigating the Kuna Collection of Joel Sherzer

I. MAIN PAGE: <https://www.aila.utexas.org/islandora/object/aila%3A124436>



The screenshot shows a web browser window displaying the AILLA website. The page title is "Kuna Collection of Joel Sherzer". The main content area is titled "Colección Kuna de Joel Sherzer" and "Object Details". The details table includes the following information:

Collection Language	Kuna, San Blas
Language PID	aila:119499
Title [Indigenous]	
Language of indigenous Title	
Title	Kuna Collection of Joel Sherzer
Country(ies)	Panama
Collector(s)	Sherzer, Joel
Depositor(s)	Sherzer, Joel
Project/Collector Website	
Description [Indigenous]	
Language of indigenous	

On the right side of the page, there is a "User" section with links for "My Account" and "Log Out", and a "Main Menu" section with links for "Welcome", "Browse", "Depositor info", "Rights", "Events", "AILLA Papers", "Contact", "Contribute", "News and Announcements", and "FAQ".

II. FOLDER TITLES

- 1 Agouti Story (chanted)
- 2 Agouti Story (spoken)
- 3 AILLA Administrative Documents - Sherzer
- 4 Anselmo texts
- 5 Apsoket Ikar
- 6 Arkan Kae
- 7 Basic word list
- 8 Campos counsels child birth medicine
- 9 Ceremonial dialogic greetings
- 10 Chanting
- 11 Chicha song
- 12 Chief Dionisio chants
- 13 Chief Gonzalo of Niatuppu counsels
- 14 Chief Mantiwekinya Chants
- 15 Chief Mastaletat chants, José Cristiano interprets
- 16 Chief Mastayans chants
- 17 Chief Muristo Chants
- 18 Chief Ninabkinya chants, Pedro Arias interprets

Ayudas de Localización

Pueden contener descripciones y otros metadatos también

p.ej. si una grabación está transcrita o traducida

Contents:

A - Iskonawa songs

A collection of songs and chants in Iskonawa totaling 00:12:14 of audio and video files, many of which are accompanied by EAF annotations made by the ELAN application that contain transcriptions and/or translations of utterances into Spanish.

	<u>Date</u>	<u>Description</u>	<u>Length</u>	<u>Transcribed?</u>	<u>Translated?</u>
1	2015-06-15	Juanita sings the song of the bihao plant, narrating how the wind moves its leaves	0:00:42	no	no
2	2015-06-15	Juanita sings the song of the coto monkey, who lives in the hills	0:00:32	yes	yes
3	2015-06-15	Juanita sings the song of the hualo toad, describing its legs and body	0:00:29	yes	yes
4	2015-06-15	Juanita sings the song for the communal drinking	0:01:48	yes	yes
19	2015-12-15	Elías Rodríguez performs a song in Iskonawa.	0:00:32	yes	no
23	2014-12-13	Juanita sings the song for when a son dies	0:00:18	yes	no
24	2014-12-13	Juanita sings the song for when a father dies	0:00:23	yes	no
29	2014-12-13	José sings the song for the collective drunkenness; he interrupts it repeatedly.	0:03:51	yes	no

Evaluación de una ayuda de localización

Visita la [Colección del Idioma Iskonawa](#), abre una carpeta y lee los metadatos de unos archivos

Revisa el [Guía para la colección](#)

¿Te ayudó a entender qué hay en la colección?

Con la ayuda de la guía de localización, ¿piensas que puedes encontrar materiales mejor en la colección iskonawa?

Guías de Colecciones



*Farming, food and yam:
language and cultural
practices among Ikaan
speakers*

(Salffner 2015)



*Choguita Rarámuri
description and
documentation*

(Caballero 2017)

Arriba: Grace Adeola y Sophie Salffner caminan por una finca (ikaan317_14.jpg, fotógrafo: Fred Adekanye)
Abajo: Bertha Fuentes Loya describe términos deicticos y nombres de lugares (“deictics2”)

Esbozo común

1. Introducción
2. Antecedentes
 - a. El idioma y sus hablantes
 - b. Documentación existente
 - c. Proyecto de investigación
 - d. Convenciones de la colección
 - i. Formatos
 - ii. Metadatos
 - iii. Anotaciones
3. El contenido de la colección
 - a. Descripciones de series
4. (Restricciones al acceso)
5. Sigüientes pasos
 - a. Trabajo en proceso
 - b. Trabajo previsto
 - c. (Citando y usando los datos)
6. Conclusión/Resumen

Guías de Colecciones - Salfner (2015) contenido

1. Enfoque lingüístico

- a. Paseos por la granja
- b. Historias del ñame y la comida
- c. Cuentos tradicionales y adivinanzas
- d. Historias fotográficas
- e. Talleres de campos de palabras
- f. Mundos de bloque
- g. Sesiones de elicitación

2. Enfoque sociolingüístico

- a. Entrevistas contextuales
- b. Parientes

3. Enfoque antropológico

- a. El ñame y la fiesta del ñame
- b. El funeral tradicional
- c. El trabajo de las mujeres
- d. Filmación observacional y apuntes de campo visuales
- e. Los apuntes de campo escritos como 'descripciones densas'

4. Materiales documentando la documentación

- a. Materiales administrativos
- b. Álbums de videos

5. Otros actividades y materiales

- a. La escuela de fotografía Ikakumo
- b. Productos académicos
- c. Divulgación y comunicación de hallazgos

Guías de Colecciones - Caballero (2017) contenido

1. **Materiales solo con enfoque lingüístico**
 - a. Sesiones de elicitación
 - b. Pláticas
 - c. Sesiones de enseñanza del idioma
2. **Materiales con enfoque comunitario y lingüístico**
 - a. Textos
 - b. Entrevistas
3. **Materiales con enfoque comunitario**
 - a. Música
 - b. Grabación en video de eventos de la comunidad

Reseñas de Colecciones

Pocos o ningunos ejemplos en el campo de la documentación y descripción de idiomas

Puede convertirse en un género paralelo a las reseñas de libros que aparecen en revistas académicas

¿Por qué no escribes una reseña de una colección de un idioma que estudias?

Descripción archivística

La descripción archivística es **iterativa** igual que el archivado progresivo.

Puedes empezar con una **descripción** básica

Crear una **ayuda de localización** de tu registro de metadatos

Actualizarla después de depositar nuevos conjuntos de archivos

Al terminar el proyecto/tesis/subvención, puedes redactar un **guía** que resume todos tus materiales con mejor comprensión que tienes del contexto, importancia, y relevancia de la colección que has creado

El archivado lingüístico y AILLA

Conferencia a distancia para CIESAS Sureste

Susan Kung

3 diciembre de 2018

Temas distintos pero muy relacionados

- La ética
- El consentimiento informado
- La privacidad
- La propiedad intelectual y los derechos de autor
- El Conocimiento Tradicional

ética



La ética: Reglas y normas para establecer el comportamiento

Códigos de ética

Linguistics Society of America:

[Declaración de Ética](#) y [página de Ética](#)



American Anthropological Association:

[Principios de Responsabilidad Profesional](#) y [blog de Ética](#)



Society of American Archivists:

[Declaración de Valores Principales y Código de Ética](#)

consentimiento
informado

—

Consentimiento informado

Los participantes-hablantes en un proyecto o estudio lingüístico deben dar su consentimiento

- para participar en la investigación o recopilación de lenguaje tanto como
- para archivar los materiales resultantes del proyecto.

Obtén permiso para archivar los materiales CUANDO obtengas el consentimiento informado para la investigación, en el mismo documento/pedido/grabación si es posible.

[Ejemplo de formulario de consentimiento](#) (Programa de Documentación Lingüística del Depto de Lingüística de UT-Austin) que incluye el permiso para archivar datos.

Consentimiento informado

El consentimiento puede tener diversas formas, y su documentación también.

El Proyecto para la Documentación de la Lengua Chatina solicitó el permiso del Comité de la Ética de su universidad para usar una declaración hablada en lugar de un formulario con firma.

Nótese que no hay una línea para firmar aquí:

Declaración de aceptación:

He leído o he oído la información arriba y hay suficiente información para decidir sobre la participación en este estudio. Acepto participar en este estudio.

Consentimiento informado

Algunos depositantes han preservado la documentación del consentimiento (típicamente como objetos restringidos)

- Grabaciones audiovisuales del consentimiento verbal
- Documentos firmados
- Fotos de personas firmando documentos

Consentimiento audiograbado de los líderes en Collakamani, Bolivia para participación en la investigación.

Rene Flores Huagua (Profesor/Director de escuela Collakamani), Teodoro/Tiburcio Ilafaya (Dirigente comunal), Donato Roque (Junta escolar)

Consulte I002 para mayor información.

Permisos

Genre

Interview

Observational Filming

Staged Communication

Consent

Elicitation (words, Phrases)

Narrative

Elicitation (phrases)

Consentimiento informado

Pero no te olvides,

“El consentimiento no debe ser un asunto de obtener una firma en una hoja de papel, sino el establecimiento de una colaboración informada.”

(Eckert, 2013:14, traducción por Ryan Sullivant)

privacidad

—

Privacidad

Derecho de un individuo a la privacidad vs. la invasión de privacidad: 4 partes (EEUU)

1. **Intrusión de soledad:** Intrusión (física o electrónica) en los asuntos privados ajenos - incluye la grabación sin el consentimiento informado
2. **Divulgación pública de hechos privados:** compartir la información privada de alguien que no querrían compartir.
3. **Falso aspecto ("False light" en inglés):** Publicación de hechos que presentan un falso aspecto de una persona; semejante a difamación (calumnia y libelo); en EEUU se aplica únicamente a las personas vivas.
4. **Apropiación:** el uso no autorizado de la imagen o nombre de alguien por el beneficio de otro.

propiedad intelectual y
derechos de autor

—

La propiedad intelectual

La propiedad intelectual es cualquier invención de la mente humana que tenga valor comercial y según las leyes incluye:

- Marcas (registradas)
- Patentes
- Secretos comerciales (Trade secrets)
- Derechos de autor

Organización Mundial de la Propiedad Intelectual www.wipo.int

Los derechos de autor

Un monopolio legal que les otorga a los creadores un conjunto de 6 derechos exclusivos para sus trabajos creativos por un plazo limitado. Los derechos de:

1. Reproducir la obra
2. Crear obras derivadas
3. Distribuir copias de la obra
4. Realizar interpretaciones públicas de la obra
5. Mostrarla públicamente
6. Transmitir digitalmente una grabación sonora



Todos los derechos reservados

Los derechos de autor

Los 6 derechos de autor son **transferibles** -- todos juntos o individualmente.

La **transferencia** de derechos de autor puede ocurrir por medio de **contrato, licencia y herencia**.

La posesión de una cosa NO ES IGUAL a la posesión de la propiedad intelectual. (Si compras una película, no estás comprando los derechos de autor.)

Si un acervo tiene un archivo de media, (probablemente) no tiene los derechos de autor del contenido del archivo.

Los derechos de autor

Requisitos para que una obra esté protegida por derechos de autor:

1. Tiene por lo menos un grado mínimo de originalidad o creatividad y
2. Se fija en un medio tangible.

La aplicación de derechos de autor es automático siempre que cumplan los 2 requisitos.

La registración de la obra no es requerido pero es aconsejable hacerlo en México y los EEUU si quieres ejecutar sus derechos de autor.

Duración de derechos de autor

México: 100 años después de la muerte del autor (o de la muerte del último autor si hay co-autores)

EEUU: 70 años después de la muerte del autor (o de la muerte del último autor si hay co-autores)

Ejemplos de obras protegidas

- Obras literarias como libros, artículos, poemas, cuentos
- Obras musicales como letras de canciones y la misma música
- Obras dramáticas incluso su música
- Obras coreográficas (pero no los pasos básicos, como de vals o huapango)
- Obras de arte como pintura, escultura, arte digital o gráfica
- Obras audiovisuales, como películas y programas de televisión
- Fotografías
- Arquitectura
- Programas de computadora
- Bases de datos que contienen obras protegidas
- Compilaciones, obras colectivas, obras derivadas ...

Ejemplos de obras NO protegidas

- Ideas NO fijadas en un medio tangible
- Obras en el dominio público
- Hechos
- Títulos, nombres, números, símbolos familiares
- Datos científicos o de la naturaleza
- Listas de contenidos o ingredientes (libros telefónicos, recetas)
- Pasos básicos de un tipo de baile (vals, huapango)
- Procesos, procedimientos, métodos (de la cultura material)
- Palabras de un idioma
- Folclore (mitos, leyendas, costumbres, cuentos tradicionales, refranes)*
- Bases de datos que contienen información no protegida

Donar vs. Depositatar

Escritura o contrato de donación - la opción de renunciar a todos los derechos de autor al repositorio, [ejemplo de Dartmouth](#)

Licencia (de depositante) - una licencia no exclusiva para que el archivo pueda copiar, mostrar y distribuir los archivos - los titulares *mantienen* sus derechos p.ej. [Licencia de Depositante](#)

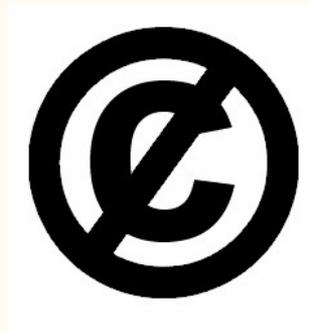


Derechos de autor y materiales archivados

Obras huérfanas

Dominio público

Trabajo por contrato o encargo



Licencias de [Creative Commons](#) y asuntos por datos de lenguaje

- Autoría y derechos de autor
- Licencias CC no se pueden revocar



conocimiento
tradicional

—

El(los) Conocimiento(s) Tradicional(es)

El **Conocimiento Tradicional** (TK por su sigla en inglés) es "un cuerpo vivo de conocimientos que se transmite de una generación a otra en una misma comunidad. A menudo forma parte de la identidad cultural y espiritual de la comunidad" (de <https://www.wipo.int/tk/es/>)).

Muchos conocimientos tradicionales no son protegidos (automáticamente) por leyes de propiedad intelectual.

The image shows the letters 'TK' in a large, bold, black serif font. The letters are centered on a white rectangular background. The 'T' is on the left and the 'K' is on the right, both with a classic, slightly ornate design.

Conocimientos Tradicionales

OMPI divide los conocimientos tradicionales en tres subcategorías:

1. **Conocimiento tradicional** (*sensu stricto*): capacidades y conocimientos técnicos, especialmente los relacionados con la biodiversidad, agricultura y salud.
2. **Recursos genéticos**: conocimientos y prácticas con relación al material genético de plantas, animales, etc.
3. **Expresiones culturales tradicionales** (folclore): expresión de cultura tradicional, p.ej. la danza, la música, las artesanías, diseños, símbolos, etc.

World Intellectual Property Organization (WIPO). Sin fecha. Traditional Knowledge and Intellectual Property – Background Brief. Accedido el 19 de mayo 2016 de http://www.wipo.int/pressroom/en/briefs/tk_ip.html

Conocimiento Tradicional: también se conoce como:

Conocimiento Ecológico Tradicional (TEK por su siglas inglesas)

Patrimonio Cultural

Propiedad Intangible Aborigen / Propiedad Intelectual Indígena

Panamá - Ley N° 20 del 26 de junio de 2000 sobre el Régimen Especial de Propiedad Intelectual sobre los Derechos Colectivos de los Pueblos Indígenas para la Protección y Defensa de sus Identidad Cultural y de sus Conocimientos Tradicionales

protocolos especiales y
materiales sensibles

—

Protocolos

Protocolos o Derechos consuetudinarios: Normas o reglas que “dictan derechos, responsabilidades y obligaciones culturales específicos” en comunidades indígenas (IPinCH 2015, p. 5).

Varían de una comunidad a otra.

"Esto no es material 'ordinario'. Es material que sigue siendo formado por y entendido dentro de los contextos locales y culturales donde la gente mantiene el vínculo crítico con estos materiales" (Anderson & Christen 2013: 113).

"Es bueno ver a viejas amigas [refiriéndose a las cestas karok en el Hearst Museum], pero es difícil dejarlas allí." (Susan Gehr, Plenaria de CoLang, 19 de junio de 2018)

Tabúes sobre nombres en Australia

En algunas culturas indígenas australianas, el **nombre**, la **imagen** y/o la **voz** de un fallecido no se debe usar durante el periodo de luto.

Para responder, algunas emisoras australianas tocan o reproducen una descarga de responsabilidad antes de programación::

*[Espectadores, escuchadores, lectores]
aborígenes e Isleños del Estrecho de Torres
están avisados que el programa siguiente puede
contener imágenes y voces de personas que han
fallecido.*



Etiquetas y Licencias de TK

Licencia - Una autorización por parte del titular para que el usuario utilice el material por un propósito específico.

Licencias TK - contrato legal; otorgan a las comunidades el poder de controlar la propiedad intelectual inherente en sus prácticas culturales, materiales, artefactos, e idiomas

Etiquetas TK - no son jurídicamente vinculante; "ayudan a prevenir o limitar el uso indebido y trato despectivo [o incorrecto] de materiales culturales indígenas, tradicionales y locales.

La reutilización legal y ética

Los repositorios tienen muchas audiencias, y los materiales archivados pueden ser reutilizados en maneras inesperadas, idealmente respetando la propiedad intelectual de los titulares (recopiladores y hablantes o sus herederos):

- Usar fragmentos de materiales archivados en obras artísticas
- Componer música inspirada por materiales archivados
- Hacer pinturas de fotos archivados

OJO: La creación de una obra derivada requiere el permiso o una licencia de los titulares y/o el uso cuidadoso de "Uso Leal" (Fair Use).

Resumen

Piénsalo bien si los materiales que recopilas deben depositarse en un repositorio.

- ¿Quién es el dueño de la propiedad intelectual o los derechos de autor?
- ¿Tienes el permiso de aquella persona para ponerlos en un acervo? ¿Has obtenido el consentimiento informado para archivar materiales junto con el consentimiento de participación?
- ¿Está en conflicto el conocimiento tradicional de la comunidad con la ley de la propiedad intelectual?
- ¿Hay protocolos comunitarios que debes de considerar?
- ¿Estará bien que estos materiales sean reutilizados en el futuro?

Agradecimientos de NSF

Este material se basa en trabajo apoyado por la National Science Foundation bajo subvención BCS-1653380. Cualquier opinión, hallazgo, conclusión o recomendación expresado en este material son de los autores y no necesariamente reflejan los de la National Science Foundation.