Filenaming conventions

The present document comprises a description of the metadata files that should be archived and the filenaming system employed for Jonathan Amith projects.

1. Recordings (one or two speakers) that cover a range of genres

METADATA CONTROL.
This is a list of codification for the filenaming.

For most types of recordings the following system is used

5_5_6_50_12
village_genre_narratorcode(s(_title/description_UID

This system is used when there is one person on the recording (e.g., in storytelling) or when the second speaker is an interviewer with insignificant or little intervention. In this case, however, the interviewer is listed in the metadata with role = interviewer. The interviewer is on the right track and the targeted/principal speaker on the left track.

Alternate: If there are two narrators (not including the interviewer) then both are listed with an intervening dash. The speaker of the LEFT channel is listed first: _LEFTxx-RIGHTxxx_
If the file is one channel (e.g., field recordings-botany) then the filename has the speaker codes in order that they speak in the recording. This is the system for botanical field recordings with a Zoom H4n.

5_5_6-6_50_12
village_genre_narratorcode-narratorcode_title/description_UID

Example: Tzina_Botan_DZD352-MSO325_koweeloot-Araceae_2012-07-17-n.wav
          Tzina   This is the code for San Miguel Tzinacapan
          Botan   This is the code for the genre of 'Botany'
          DZD352   This is the code for the speaker on the left channel (see database/file on personnel)
          MSO325   This is the code for the speaker on the right channel (see database/file on personnel)
          koweeloot-Araceae This is the Indigenous (Nahuat) name followed by the relevant botanical family
          2017-07-17-n   This is the unique identifier (UID). In this case it is the date of the recording + letter

All file names will have the same basic format:

a. Village Five characters (followed by a dash)
These are village codes as noted below. The codes represent where narrators were from. The codes are given below.
NOTE: in the event that the recording is a conversation between two individuals the village code will be that of the speaker in the left channel or, in the event that the speaker in the right channel is clearly dominant (i.e., speaks considerably more often), that of the right speaker. There are never two village codes even though there might be two narrator codes and the two narrators might come from a different village). The different origins of the speakers are, however, always recorded in the metadata.

b. Genre   Five characters (followed by a dash)
These are codes for the basic genre of the recording. The codes are given below. If there are two individuals from different villages only the village of the left channel speaker is given in the filename.

c. Speaker   Six characters (followed by a dash)
This is the code for the main narrator. The six-character codes comprises the initials of the narrator followed by a unique 3-digit code. The codes starting at 500 are for speakers of Yoloxóchitl Mixtec, for example. Note that archived is also a database of all project participants with their codes and demographic data (sex, birth year, village)

d. Description of content ('title')   Varying number of characters always < 50
This is a brief descriptive "text" of the recording, e.g., the title of a story or the nature of a life history testimonial.

e. UID   Final 12 (rarely 13) characters.

This is the unique identifier code of the item. The identified code comprises a year-month-day (yyyy-mm-dd) followed by a dash and letter. The year-month-day is a date that may, but does not necessarily, reflect the actual day of recording (the metadata entry for the item has the information of date of recording). It may be the date the material was entered into the metadatabase. Or it may be somewhat arbitrary. The key is that it is a unique identifier.

2. Botanical field recordings. E.g., Yolox_BotFl_CTB501-FEF537-EGS503_40196-Solanaceae-Datura_2017-01-05-a

In general Bota
Semántica de nombres de archivo de grabaciones en el campo.

Yolox_	Siempre para las grabaciones de Yoloxóchitl el nombre del archivo empieza con Yolox_ Nótese que siempre se separan los campos por guión bajo
BotFl_	Significa "Botany field" esto es, grabaciones de campo sobre botánica
CTB501-FEF537-EGS503_	Aqui van los códigos de cada hablante/entrevistado Debe seguir el orden en que hablan, asi que en este habla primero Constantino, después Filomeno y finalmente Esteban
40196	Esta es el número de colecta
-Solanaceae-Datura_	seguido (nota guión en medio, no bajo) de la familia y, si se sabe, género, si no se deja en familia
2017-01-05-a	Primero la fecha de la colecta y segundo a, b, c, d etc. para la primera, segunda, tercera, etc. grabación

3. Miscellaneous   (elicitation)
Occasionally derivative files may follow the filename as described above. For example, a sound file may have been converted from stereo to mono. This occurs with some elicitation files where the native speaker linguist, or another "prompter" is on the right channel and the consultant whose recording is targeted is on the left channel. The left channel only may be copied into a mono file and saved.

Original file: Yolox_Elict_EGS505_Lista-001-tonos-completos_2010-12-10-d.wav
Left channel only: Yolox_Elict_EGS505_Lista-001-tonos-completos_2010-12-10-d-mono.wav

4. Recordings of botanical and zoological names (code = Botnm and Zoonm, respectively)
Many recordings have been made of plant names that relate to a particular collection. The most important part of the filename is the collection number of the plant that is being named. Given the length of time it may take to get an accurate identification to species the best way to associate this recording to a scientific species is to search for the collection number in the most up-to-date list of plant collections and scientific identifications. The structure of the filename follows one or two forms:

Sample
a.   72004_Ecatl_Botnm_RBC405.wav

72004	Collection number
Ecatl	Ecatlán, community name
Botnm	Botanical naming (genre of recording: Plant names)
RBC405	ID of speaker

With taxon including in the filename
b.   72015_Ecatl_Botnm_AGB420_Anacardiacea-Spondias.wav

72015	Collection number
Ecatl	Ecatlán, community name
Botnm	Botanical naming (genre of recording: Plant names)
RBC405	ID of speaker
Anacardiaceae-Spondias	Family-genus of the referent for the Indigenous name

The pattern is the same with the exception that a botanical family and genus is added at the end. A species name is not given (and sometimes a genus name is not given) because these are provisional names often given before a taxonomic expert has reviewed the specimen. Note that the most up-to-date species identification will be available in a csv file (exported from Excel) that will be deposited every year. Even in the event that the genus, or even family, changes, the filename will not be changed.

5. In situ plant photographs
Example filename 31397o_IMG_4570_ed-300nr_mjc.jpg
                  82016o_IMG_1631_ed-cr-300_jda.jpg

5 digits (first 5 characters)	The first 5 digits of the filename reference the collection number, in this case 31397. To get the species name it is necessary to look up the most recent IDs in the collection database. In this case 31397 is *Philodendron inaequilaterum* Liebm. (Araceae). The database of plant collections gives all relevant data (coordinates of the collection; collectors; description of the plant; and who provided the identification to species [in this case Michael Grayum of the Missouri Botanical Garden upon examination of the specimen])
Letter following first 5 characters: "o" or "x".	The "o" (for "original") means that the photo was taken in situ of the actual specimen that was collected. An "x" indicates that the photo was taken by a plant that was judged to be the same as the species of the collection number but that the plant was not collected.
	Thus 82006x_IMG_1504_ed-300nr_jda.jpg indicates that this is a photo of what was considered, in the field, to be the same species as #82006, *Solanum aligerum* Schltdl. But the plant was not collected so there is no physical specimen reference. In general photos are taken of plants not collected only when there is little or no doubt as to their species identification.
_IMG_4570_	Whatever occurs between the 5-digit with letter ("o" or "x") beginning of the filename and the "ed" part of the name. This section of the filename is the original filename that came out of the camera. It is maintained in the filename to ensure a unique identity.
_ed	This indicates that the photo was reviewed and edited in Photoshop. All photos are at the least processed with an "unsharpen mask", however, minor. Others might be lightened, have sections edited out, or have the color adjusted through a warm or cool filter.
_cr	If the original photo was cropped, this appears in the filename.
300nr	This is simply an "arbitrary" character that reflects the fact that the file has 300 or greater dpi and it was **n**ot **r**esampled in resizing from the original file from the camera.
_jda	The final three characters before the extension represents the initials of the photographer (see table 3 below).

6. Editions of transcriptions.
For transcriptions only, an edition date is added to the end of the filename, before the extension. The date is YYYY-MM-DD and preceded by _ed-
        _ed-2010-07-27
Thus one may find the pair
        Yolox_Cuent_CRS513_El-flojo-y-el-rey_2008-11-01-d.wav
        Yolox_Cuent_CRS513_El-flojo-y-el-rey_2008-11-01-d_ed-2010-07-27.trs

Table 1   First five characters: Village code

Village code: Code for village where recording was conducted

Oapan	San Agustín Oapan, Guerrero
Tecui	San Miguel Tecuiciapan, Guerrero
Ahuel	Ahuelicán, Guerrero
Ozoma	San Francisco Ozomatlán, Guerrero
Ameya	Ameyaltepec, Guerrero
Acatl	Acatlán, Guerrero
Zitla	Zitlala, Guerrero

Village code: Mixteco

Bvsta	Buenavista
Cuana	Cuanacaxtitlan
Cumia	Arroyo Cumiapa
Yolox	Yoloxóchitl

Mpio. Cuetzalan del Progreso

Ayotz	Ayotzinapa, Puebla
Atemo	Atemolón, Puebla
Chilc	Chilcuauhta, dependencia San Miguel Tzinacapan, Puebla
Ctama	Cuauhtamazaco, Puebla
Ctapa	Cuauhtapanaloyan, Puebla
Ecatl	Ecatlan, Puebla [totonac]
Huitz	Huitziquitepec, mpio. Tuzamapan de Galeana (official name Huitziltepec)
Ixtla	Ixtlahuata, Puebla
Limon	Limonco, Puebla
Tecol	Tecoltepec, Puebla
Tclap	Tecolapan, Puebla
Teptn	Tepetzintla, Puebla (Nahuat)
Teptt	Tepetzintla, Puebla (Totonaco)
Tlaqm	Tlaquimpa, Puebla
Tzina	San Miguel Tzinacapan, Puebla
Xalcu	Xalcuauhta, Puebla
Xaltp	Xaltepec, Puebla
Xaltn	Xaltipan, Puebla
Xalpn	Xalpantzingo, Puebla
Yancu	Yancuitlapan (Saniago)
Yohua	Yohualichan, Puebla
Zacat	Zacatipan, Puebla
Zilte	Ziltepec, Puebla
Zoqui	Zoquiapan
Zuapl	Zuapilaco, Puebla

Totonaco

Atleq	Atlequizayán
Chico	Chicontla, Puebla
Coahu	Coahuitlán, Veracruz
Ecatl	Ecatlán, Puebla
Mecat	Mecatlán, Veracruz
Nanac	Nanacatlán, Puebla (mpio. Zapotitlán de Méndez)
Patla	Patla, Puebla
Teptt	Tepetzintla, Puebla (Totonaco)
Tonlx	Tonalixco, Puebla
Zihua	Zihuateutla
Zongo	Zongozotla

Nahuatl

Tlaqm	Tlaquimpa
Tepez	Tepetzintla (Nahuatl)

Table 2: Basic genre (5 characters)

Adivi	Adivinanzas (Riddles)
Agric	Agricultura (Agriculture)
Botan	Botánica (Botany)
BotFl	Grabaciones en el campo sobre etnobotánica Field recordings (most often Zoom H4n or H6n) of botanical collections
Bionm	Nombres de plantas, grabaciones cortas de nombres (Botanical naming: These are short recordings of plant names)
Cazas	Pesca y trampas, como hacer, incluyendo arcos, resorteras, etc. (Fishing and hunting, how to make traps, including slingshots, etc.)
Cielo	Cielo, estrellas (The heavens and stars)
Comid	Comida, preparación de (Food and food preparation)
Creer	Creencias, espíritus, "sobrenatural" (Beliefs, including about 'spirits' and non-human beings)
Cuent	Cuentos ficcionales (Fictional stories)
Elict	Elicitación lingüística (Linguistic elicitation)
Medic	Enfermedades y remedios tradicionales, prácticas medicinales
MatCl	Cultura material, construcción de casas (Material culture, house construction)
Music	Música, Música y narración (Music or music and narration)
Narra	Narración: historia de vida, historia del pueblo, experiencias personales
Relig	Religioso: vida de santos, fiestas (Religion, including saints' lives, fiestas)
Ritua	Textos rituales: pedir novia, pedir lluvia, levantamiento de sombra (Ritual texts: Bride-asking speeches, rain ceremonies, 'raising a spirit or shadow')
Tradi	Tradiciones: bautismos, bodas, etc. (Traditions: Baptisims, weddings, etc.)
Zoolo	Zoología (Zoology)
Zoonm	Nombres de animales, grabaciones cortas de nombres (Zoological naming: These are short recordings of animal names)

Table 3: Photographer

jda	Jonathan D. Amith
clc	Canek Ledesma Corral
mgs	Mariano Gorostiza Salazar
olf	Osbel López Francisco
csc	Ceferino Salgado Castañeda
mjc	Miriam Jiménez Chimil
ash	Anastacio Sotero Hernández