

## FILE NAMING FOR PARADISEC ARCHIVING

It is understandable that you would want to use a clear, meaningful file-naming regime for your files while in the field; however, once you return and begin to organise your items to be archived, it is inevitable that these file names will change to reflect the shape of your corpus and to adhere to PARADISEC naming conventions. Here are some naming guidelines for files to be archived with PARADISEC.

Say you have, for example, a file that you named **20150908-wordlist-01.wav** while in the field. This in-field naming convention includes the date in ISO format (YYYYMMDD). The (-wordlist) marks this as being from a session involving wordlist elicitation. The (-01) indicates that this is the first track of the session. You will have your own conventions.

### ANATOMY OF A PARADISEC FILE NAME

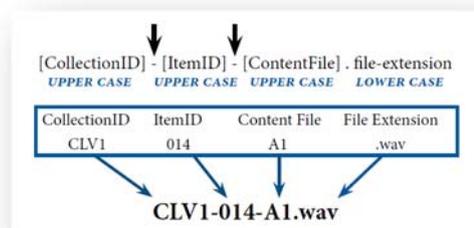
The name **20150908-wordlist-01.wav** does *not* fit with the PARADISEC naming conventions. First of all, there needs to be a collection ID followed by a hyphen, and then only two more parts separated by hyphens: CollectionID-ItemID-ContentFile.

- 1.) **CollectionID** is your PARADISEC collection name and it typically has the shape of your initials in capital letters and a number. This ID can be determined by you, but must be confirmed as available for use by the CoEDL Data Manager or a PARADISEC administrator. This ID is created when you set up a collection in PARADISEC (see [GETTING STARTED WITH PARADISEC](#)). Mine could be JCM1
- 2.) **ItemID** is also made up of alpha-numeric characters. You may use an underscore ( `_` ) if you need a separator; hyphens ( `-` ) are reserved PARADISEC operators, so cannot be used within a file name, excepting the two that separate the three parts of the file name. This ItemID can be used to differentiate recording sessions or events. You may have speaker initials in your item name (JB), abbreviation of the task (WRDLST), field site (BIMA).
- 3.) **ContentFile** is also made up of alpha-numeric characters. This part of the name allows you to enumerate files of the same format, i.e. photos with 001, 002, etc, or multiple tracks of a single audio or video session.

If you want to retain some of the original name, try something like this: **JCM01-20150908\_WORDLIST-01.wav**

CollectionID	ItemID	Content File	File Extension
JCM01	20150808_WORDLIST	01	.wav

The file name should be in all caps if using letters, the extension should be lowercase. A file name cannot exceed 30 characters, excluding the file extension. The image below diagrams the anatomy of a PARADISEC file name that is 11 characters, including the hyphens, excluding the file extension:



Below is an example of a set of recordings collected to address multilingualism. The collection includes wordlist data, sociolinguistic interviews, and natural speech. I have chosen to use very basic item names: 001, 002, 003, knowing that the specific information of the content language and speaker details will be contained in the metadata. All information contained in the **Description** column will appear in the Item description field in the PARADISEC catalog.

CollectionID	ItemID	Content File	File Extension	Description
JCM1	001	F45	.wav	All items "001" refer to recordings and annotations of a comparative Southern New Guinea wordlist recorded with speakers F45, M71, etc. I am planning to record 50 individuals, so will have a minimum of 100 content files under item 001.
JCM1	001	F45	.eaf	
JCM1	001	M71	.wav	
JCM1	001	M71	.eaf	
JCM1	001	SNG_WORDLIST	.pdf	
JCM1	002	F45	.wav	All items "002" refer to recordings and annotations of sociolinguistic interviews of speakers F45, F01, etc., as well as the questionnaire I used for the interviews. For this task there may be 150 interviews, and up to 150 annotation files, adding up to a possible 300 content files under item 002
JCM1	002	F45	.eaf	
JCM1	002	F03	.wav	
JCM1	002	F03	.eaf	
JCM1	002	SOC_QUEST	.pdf	
JCM1	003	MAR_YAMS	.wav	All items "003" refer to <i>one</i> of many recording events that took place in the garden hamlet of Mār, including associated video, audio, annotations, photos, etc. The part of the file name that refers to the different content files must be distinctive if they have the same file extension (i.e. a set of photos); however, an audio/video/annotation set all have different extensions, so can have the same content name.
JCM1	003	MAR_YAMS	.eaf	
JCM1	003	MAR_YAMS	.mxf	
JCM1	003	MAR_A	.jpg	
JCM1	003	MAR_B	.jpg	
JCM1	003	MAR_C	.jpg	
JCM1	003	MAR_D	.jpg	
JCM1	003	MAR_MAP	.pdf	

Example file name: JCM1-001-F45.wav

### WHY CORRECT FILE NAMES ARE IMPORTANT FOR PARADISEC

Files are automatically sent to specific locations within our archive structure. The first part of the file name (JCM1) tells our system that these files are to be sent to the collection JCM1. Similarly, the second part of the name (001) will direct files to item 001 in our archive structure. Items will have already been created in the catalog by the depositor before sending any files to PARADISEC. The third part of the file name (F45) distinguishes it from other files under that item. Distinctive names avoid conflicts and errors for our automated system.

For more information, please take a look at the PARADISEC webpage (<http://www.paradisec.org.au/naming.html>), or send me an email with your specific questions: [CoEDL Data Manager](mailto:julia.miller@anu.edu.au)

After reading through this guide, if you still have questions, or you wish to request a service, feel free to email me ([julia.miller@anu.edu.au](mailto:julia.miller@anu.edu.au)), or better, visit the **CoEDL Service Request Form**. CoEDL members use the Member login at the bottom of the CoEDL webpage. Then click the General Members tab, the link to the request form is in the left-hand panel. (<http://members.dynamicsoflanguage.edu.au/general-members/job-request/>).