**Pacific and Regional Archive for Digital Sources in Endangered Cultures**



**Sydney Conservatorium of Music, C41 University of Sydney NSW 2006**

**Tel: +61 2 9351 1279 Fax: +61 2 9351 1287**

**PARADISEC archiving workflow and deposit procedures**

**Contents**

[Becoming a PARADISEC user and depositor 2](#_bookmark0)

[Creating a collection 3](#_bookmark1)

[Collection ID 3](#_bookmark1)

[What constitutes a collection? 3](#_bookmark1)

[PARADISEC catalog at the collection level 4](#_bookmark2)

[What is an item? 5](#_bookmark3)

[File naming conventions 6](#_bookmark4)

[Additional examples of item organization and file naming 6](#_bookmark4)

[Why careful file naming is important 7](#_bookmark5)

[Metadata 7](#_bookmark5)

[Metadata Method 1: .XLS upload of minimal metadata spreadsheet 7](#_bookmark5)

[Metadata Method 2: ExSite9 7](#_bookmark5)

[ExSite9: Starting a new project 8](#_bookmark6)

[ExSite9: Metadata fields 8](#_bookmark6)

[Uploading metadata files using the minimal metadata spreadsheet or ExSite9 methods 9](#_bookmark7)

[Metadata Method 3: Adding/editing metadata directly via the catalog 9](#_bookmark7)

[Editing Item details 10](#_bookmark8)

[Editing Archive Information at the item level 11](#_bookmark10)

[Editing Admin Information at the item level 12](#_bookmark11)

[Item level access conditions 13](#_bookmark12)

[Inheriting details from Collection to populate Item metadata fields 14](#_bookmark13)

[OLAC (Online Language Archives Community) 15](#_bookmark14)

[Preferred deposit formats 16](#_bookmark15)

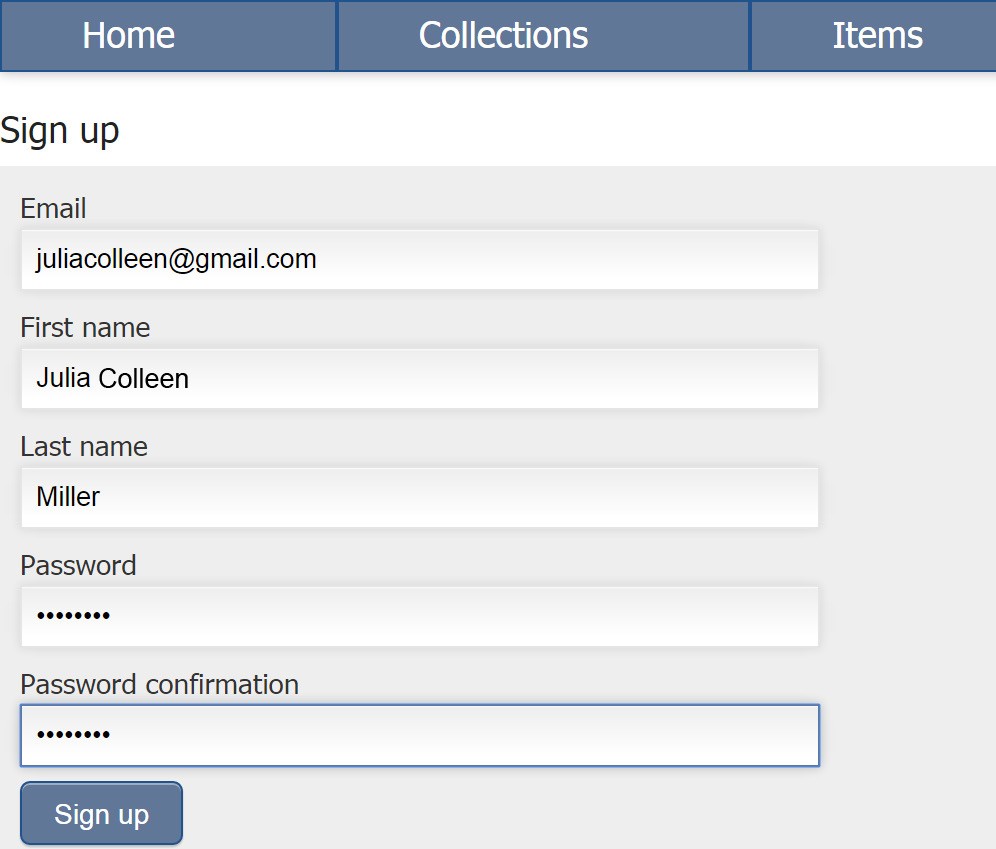
[Deposit form 16](#_bookmark15)

# Becoming a PARADISEC user and depositor

Becoming a user is necessary for viewing files within a collection and for depositing materials with PARADISEC.

Simply go to the catalog page to sign up

#### <http://catalog.paradisec.org.au/>



If you have already signed up as a user, this is also where you **sign in**. To

**sign up** as a user, you need to provide your name, email. and a password.

Once you select sign up, the following message pops up:

“A message with a confirmation link has been sent to your email address. Please open the link to activate your account.”

“Welcome Juliacolleen!

Content of automated email:

You have just signed up for an account with the PARADISEC catalog.

Before you can log into the system, we require you to confirm your account email through the link below:

**Confirm my account**

If you are unable to click on the link, copy and paste this into your browser, to confirm your account: [**http://catalog.paradisec.org.au/users/confirmation?confirmation\_token=fgFRLMu1jfMok9qCop9f**](http://catalog.paradisec.org.au/users/confirmation?confirmation_token=fgFRLMu1jfMok9qCop9f)Once you have confirmed your account you can log in here: [**http://catalog.paradisec.org.au/**](http://catalog.paradisec.org.au/)

The Conditions of Access are attached below for your reference. Please take the time to familiarise yourself with them.”

At this point you should read over the Conditions of Access. As a PARADISEC user and depositor, you need to familiarise yourself with what is expected of you when you want to access someone else’s material and what you, as a depositor, can expect from people who want to access material you have deposited.

Once you have clicked on the “**Confirm my account**” link, you are directed back to the catalog, where you can see that you are logged in and now have a collector’s dashboard.



Your account has been successfully confirmed. You are now signed in.

# Creating a collection

Only PARADISEC administrators can create a collection in the catalog. When you are ready for us to create a collection, contact the PARADISEC administration at [**admin@paradisec.org.au.**](mailto:admin@paradisec.org.au) You will need the following information:

* Collection ID (see below)
* A brief title of your collection
* A more detailed description of your collection
* Funding body (if applicable) (e.g. ARC, ELDP, etc.)
* Grant identifier (if applicable) (e.g. CoEDL = CE140100041 or Wellsprings = FL130100111)

You can edit the collection information after it has been created, adding details such as the originating university, country, language(s), etc. You also can set the map, define access, and assign editor privileges at the collection level.

## Collection ID

Once you are added as a user, we can help you start a collection. First you need to create a name for your collection; it should be unique and include your initials (all caps and no spaces)and a digit:

##### Example collection IDs:

#### JM1, JCM4, NT6, PK1, PK2, etc.

The digit is required and it distinguishes you from someone with the same initials (John Mansfield, Julia Miller, etc), but also allows for multiple collections, should you require, e.g. JM1, JM2... JM85.

## What constitutes a collection?

A collection is a personal corpus you have created that contains recordings of a language/performance/group of people you have researched, recordings that address a particular research question, or includes all the recordings collected during the life of a researcher. Each individual depositor will have one or more collections, distinguished by the collection ID. Below are some examples of collections held by PARADISEC:

* **LB1** - Theatre performances recorded in different locations and years, but all within the Philippine province of Ilocos Sur.
* **GB1-GB50** - 50 collections, each containing fieldnotes of a different Australian language.
* **TH1-TH4** - Each collection based upon different field trips, but the same language.
* **NT** collections - NT2 contains .pdf files of PhD thesis; NT8 contains dictionary files and texts of a language; NT4 is specifically for images, mainly plant identification photos.
* **SOCCOG** - Collection for all participants of the Social Cognition project using the “Family Problems” picture task to collect data from different languages around the world.

## PARADISEC catalog at the collection level

Come up with a brief descriptive

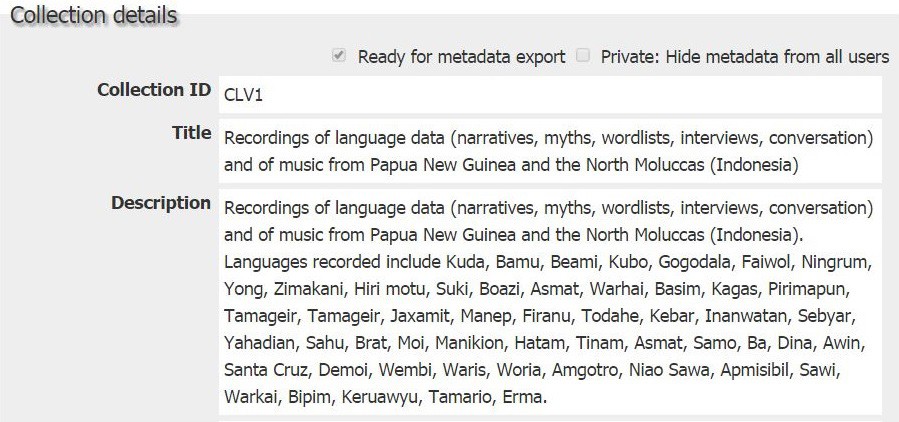
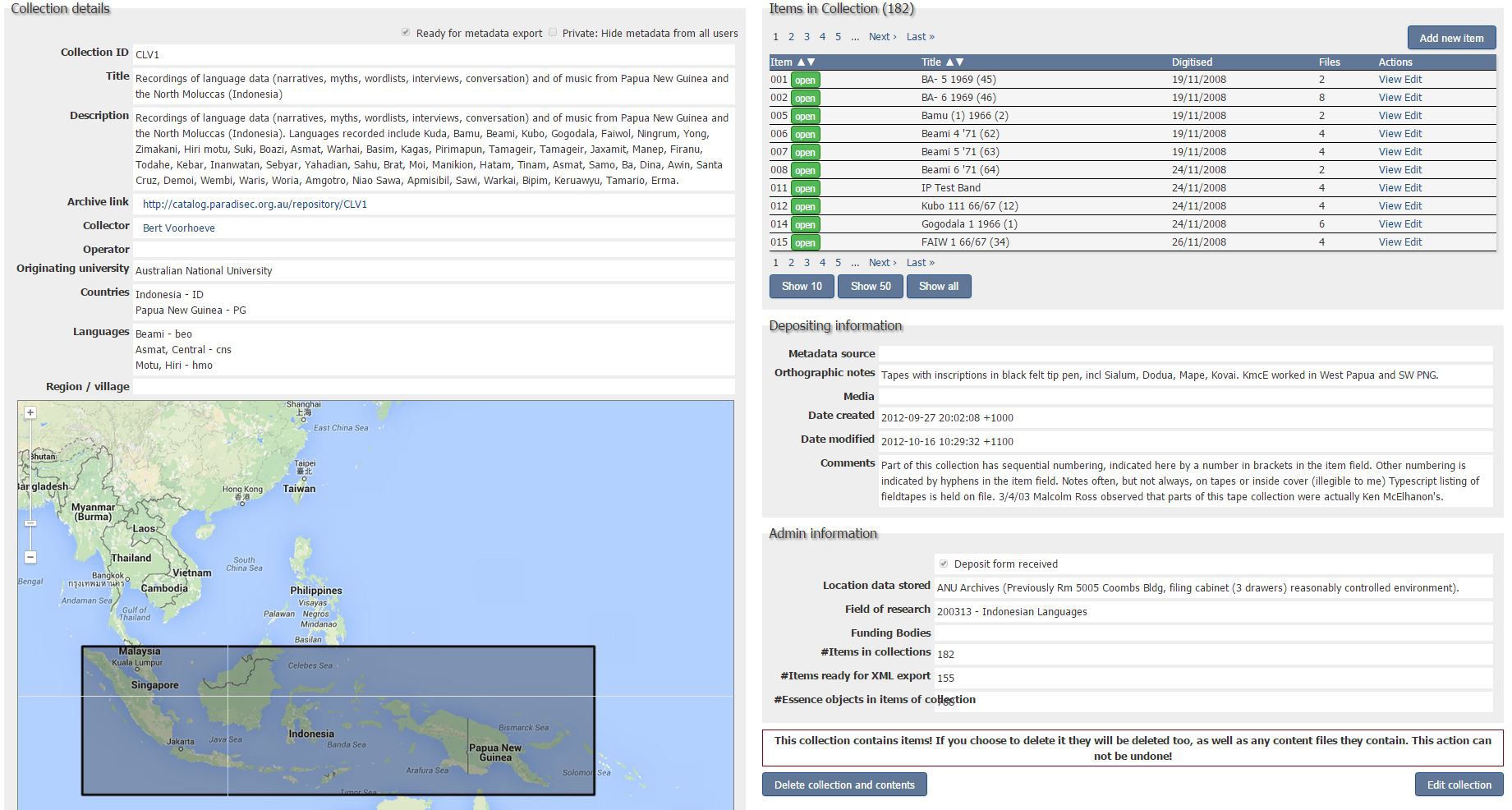
**title** for your collection.

Then provide a more detailed

**description** of what is in your

collection. Details could include

discussion of elicitation tools used,



particular research questions,

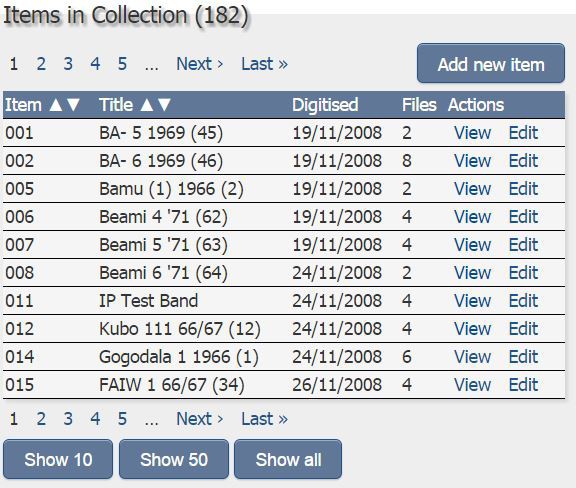
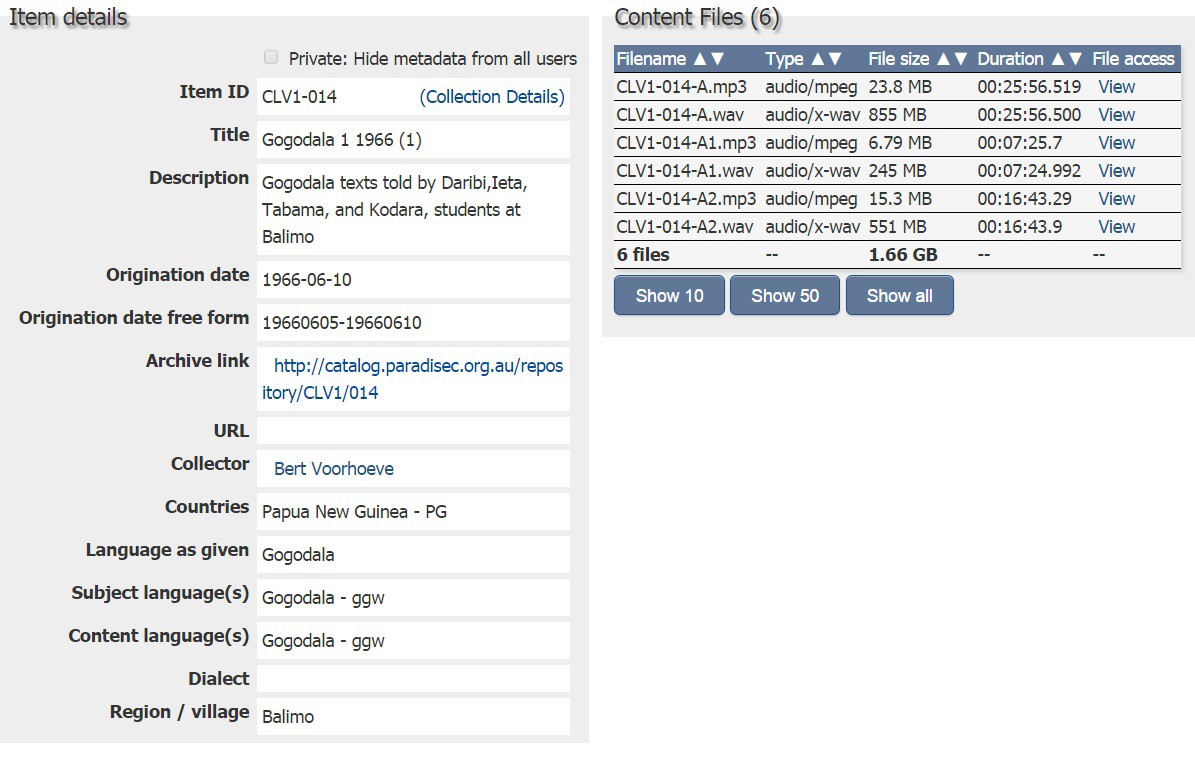
or motivations for creating your

corpus.

## What is an item?

Items are the next level of organisation below the collection. An item once referred to a physical tape, such as a reel-to-reel or a cassette tape. Now that much of the new data will be collected already in digital format, with multi-media, concurrent recording devices, it may be better to refer to an item as a recording session or event.

The item names within a collection do not have to be meaningful; ALL important information about each item will be contained in the metadata.



Within a **collection** you can have multiple items. In this collection, there are 182 items. The collector has chosen to name them “001”, “002”, “003”, etc. Do not use more than 30 characters for an item name.

Each **item** can have multiple content files. Under item 014, there are 6 content files, differentiated by the third part of the file name (“A”, “A1”, and “A2”). Three of these files were archived by the collector (the .wav files) and three were

generated by PARADISEC during the ingest process as online display versions (.mp3 files).

# File naming conventions

#### <http://www.paradisec.org.au/naming.html>

Per PARADISEC file naming rules, file names should be uppercase. File extensions should be lowercase. Hyphens are

special operators, used to separate the parts of the file name so should not be used. The two hyphens contained in the

following examples illustrate their use as special operators for PARADISEC. If you must use a separator, an underscore is

acceptable ( \_ ). Do not use spaces. More information on file-naming conventions can be found following the above link to

the PARADISEC website.

### [CollectionID] - [ItemID] - [ContentFile] . file-extension

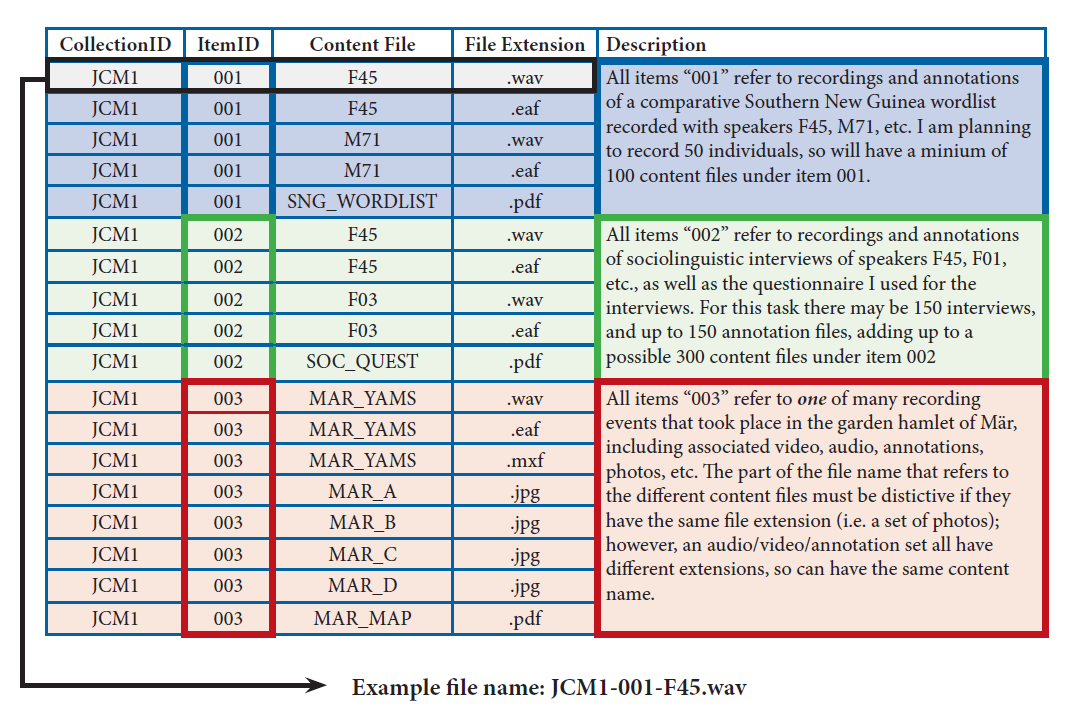
##### upper case upper case upper case lower case

|  |  |  |  |
| --- | --- | --- | --- |
| CollectionID | ItemID | Content File | File Extension |
| CLV1 | 014 | A1 | .wav |

**CLV1-014-A1.wav**

## Additional examples of item organization and file naming

Below is an example of a set of recordings collected to address multilingualism in the village of Bimadbn, PNG. The collection will include wordlist data for lexical and phonetic comparison, sociolinguistic interviews, and natural speech in the form of narratives, personal histories, procedurals, etc., in 4 different languages spoken in the village. Information on the subject and content languages will appear in the relevent metadata **item** description fields; it is unnecessary to have that information in the file name. I have chosen to use very basic **item** names: 001, 002, 003, knowing that the specific information of the content will be contained in the metadata.



## Why careful file naming is important

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 avoids conflicts and errors for our automated system.

# Meta data

At this time, there are three ways in which you can add metadata to the PARADISEC catalogue. This section offers a brief overview of each, with links to more information.

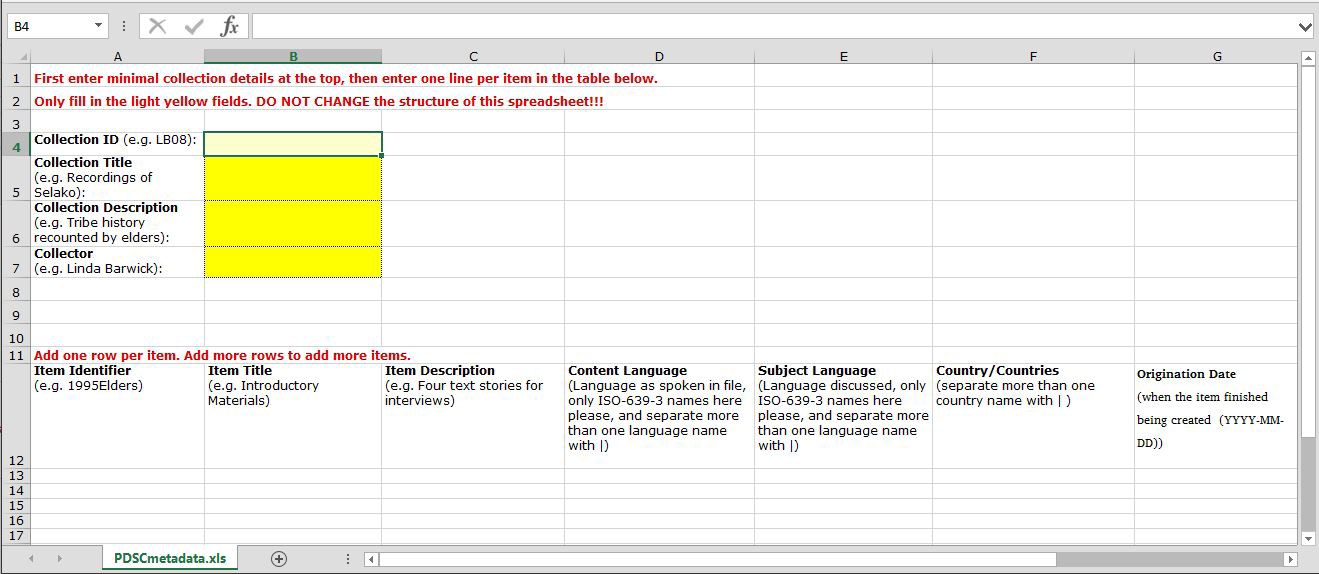
Metadata records contain descriptions or attributes about a given resource. Essentially, metadata answers questions such as ***WHO***, ***WHAT***, ***WHEN***, ***WHERE***, ***HOW*** and ***WHY*** about your files. This information will help you and other researchers find your materials. The more details you provide, the easier that process will be. PARADISEC offers 3 methods for you to add metadata to the catalog.

## Metadata Method 1: .XLS upload of minimal metadata spreadsheet

This method allows you to create a bulk set of items for your collection, requiring very few mandatory fields. Once you upload the file, you can then enrich the metadata of each item directly in the catalog (as described on page 10). As you can see in the image of the spreadsheet, there is a small number of fields to fill in. In order for the batch upload of metadata to work, you must not change the structure of the spreadsheet.

Download the minimal metadata spreadsheet from the PARADISEC site:

#### [http://www.paradisec.org.au/](http://www.paradisec.org.au/naming.html)PDSC\_minimal\_metadata.xls



Collection metadata: 4 required fields

Item metadata: 7 required fields per item

## Metadata Method 2: ExSite9

ExSite9 is a metadata editor created for PARADISEC and can be downloaded from our GitHub site:

#### <https://github.com/IntersectAustralia/exsite9/wiki/Install-packages>

For complete instructions on how to use this program, consult the manual available from PARADISEC:

* Linda Barwick’s PARADISEC version: [**http://paradisec.org.au/PARADISEC-ExSite9guidance.pdf**](http://paradisec.org.au/PARADISEC-ExSite9guidance.pdf)

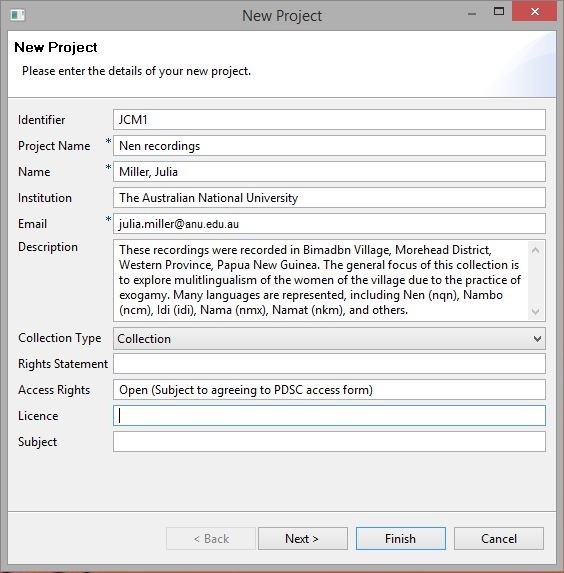
An additional manual fom the developer can be found here:

* Developer’s version: [**https://github.com/IntersectAustralia/exsite9/wiki/User-Manual**](https://github.com/IntersectAustralia/exsite9/wiki/User-Manual)

Below is a brief overview of the required fields for ExSite9

*ExSite9: Starting a new project ExSite9: Metadata fields*

### Begin a new project in ExSite9, filling in the necessary fields for your project. A project equates to a PARADISEC collection.



\* Indicates a required metadata field

**\*Project Name:** this should be a descriptive title for your collection.

**\*Name:** Insert your name, in the order last name, first name.

**Institution:** Insert the standard name of your institution as it appears in PARADISEC’s catalog.

**\*Email:** Insert your email address. This will be used by PARADISEC’s administrators to contact you should there be any questions about your collection. Note that this will not be publicly available through our catalog.

**Description:** Insert a summary of the context, purpose and contents of your collection.

**Collection type:** Select “collection” (PARADISEC does not use the alternative collection type “dataset”).

**Rights statement:** [Enter any access condition details here - see the “Access rights” field]

**Access rights:** Enter one of the following five options, using the exact wording below:

* 1. “As yet unspecified”

**2. “Closed (subject to the access condition details)”**

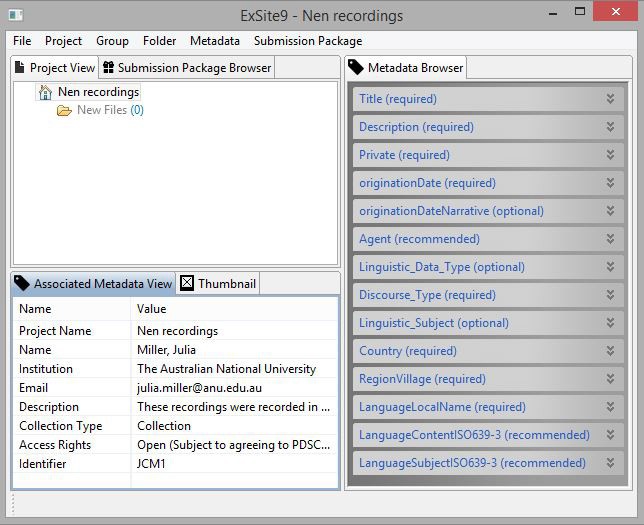
1. “Mixed (check individual items)”
2. “Open (subject to agreeing to PDSC access conditions)”

*If you have chosen the second option here, enter the access condition details in the “rights statement” field above.*

**Licence:** Leave blank

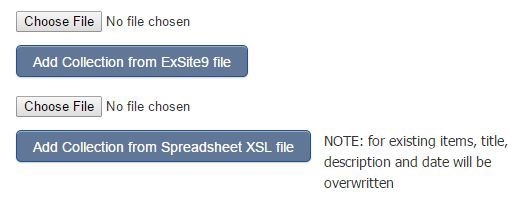
**Subject:** Leave blank

### Once you have created a project in ExSite9, you can then start adding metadata for each item you plan to archive in your PARADISEC collection. Remember, an item is not necessarily a ***file***; perhaps it can be better thought of as a ***recording event*** which may include many associated files.



*Uploading metadata files using the minimal metadata spreadsheet or ExSite9 methods*

Once you log in to the catalog, you will see your collection dashboard. Click the “Upload metadata file” button.



You are then presented with the option to upload an ExSite9 file or .xls spreadsheet.

Once you have used one of these methods to automatically create items in your collection, please follow the instructions on [**page 10**](#_bookmark9) to further enrich your metadata by directly editing items in the catalog, as what you have just uploaded is still missing important metadata.

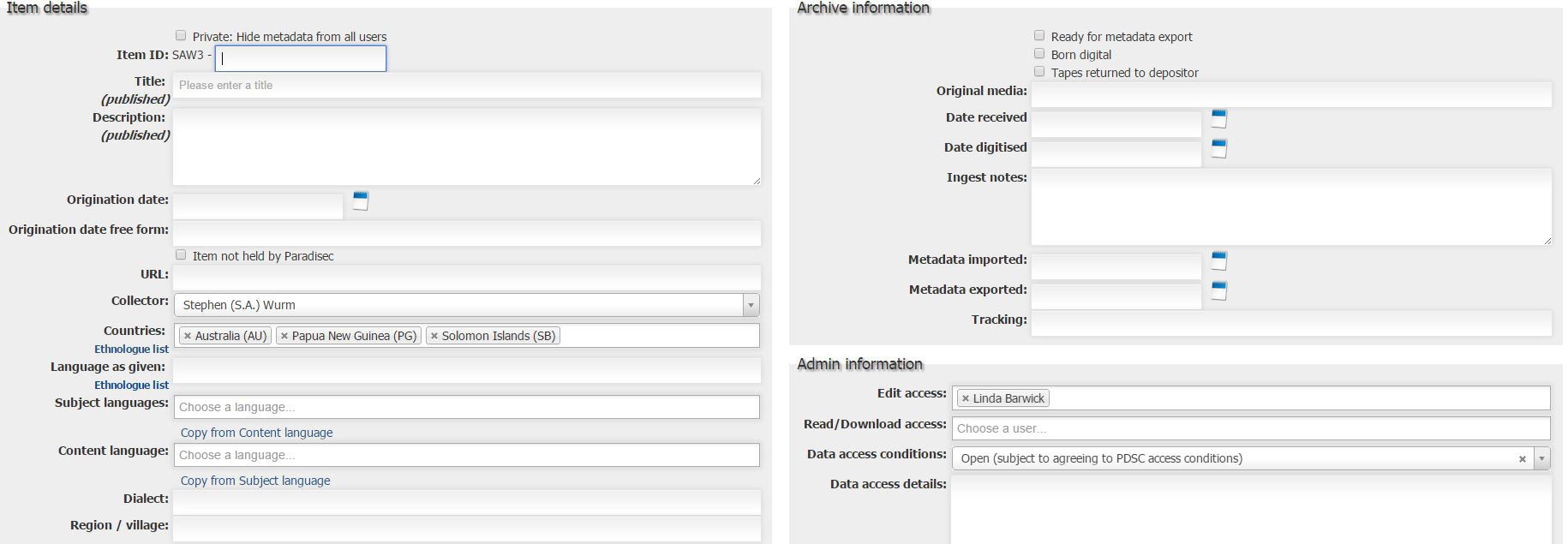
## Metadata Method 3: Adding/editing metadata directly via the catalog

The third method of adding metadata to your collection is by entering information directly into the catalog. Go to



your collector’s dashboard, then click on the “Add new item” button.

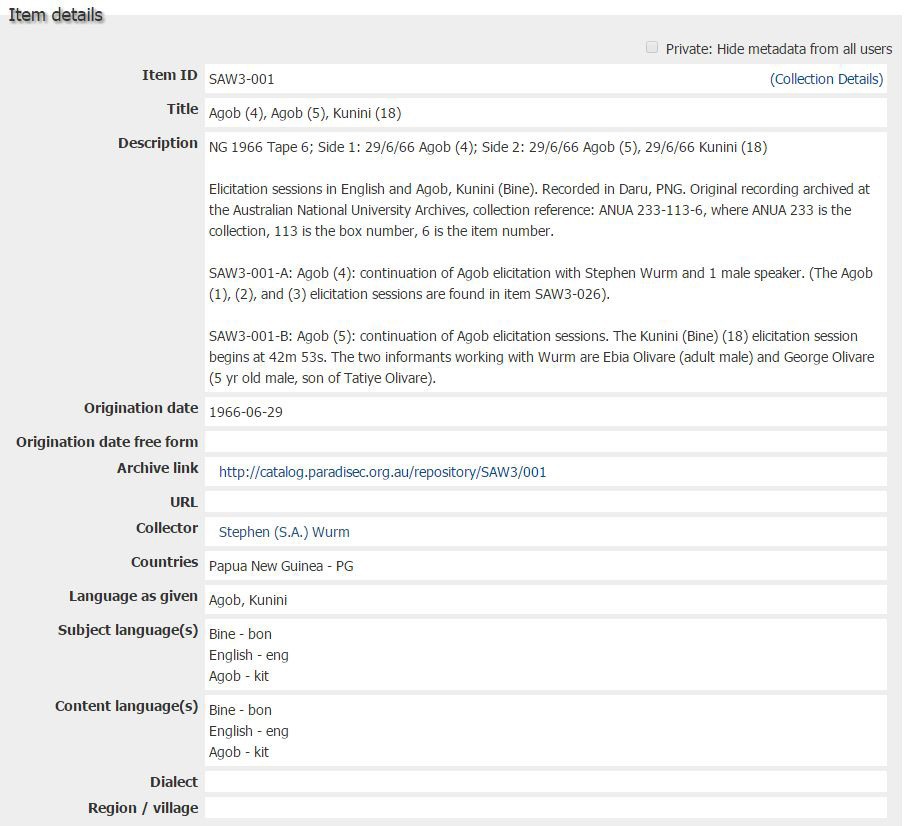
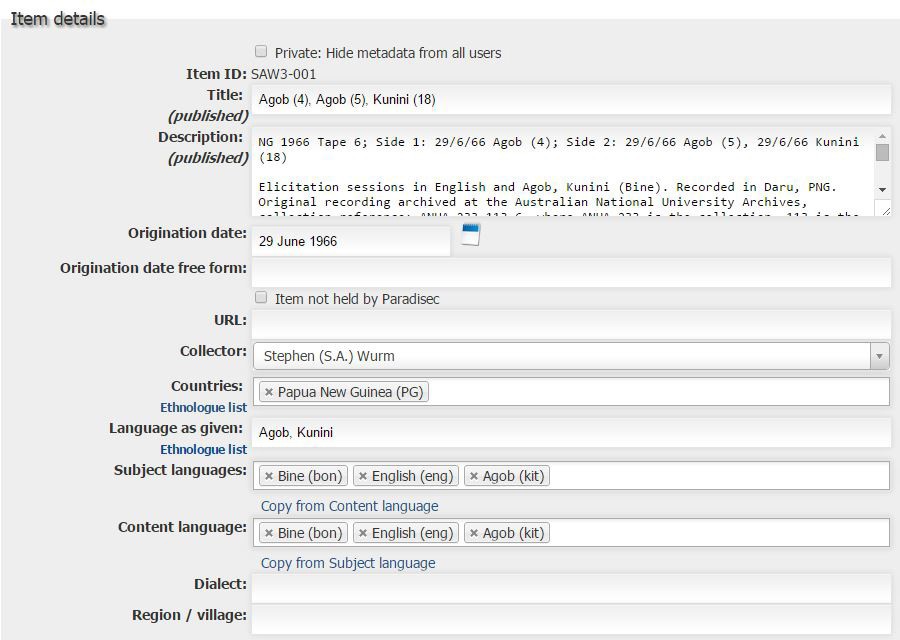
Once you select to add a new item, you are presented with an editable form, with three main sections: Item details, Archive information, Admin information. Just start typing metadata directly into the catalog.



|  |  |
| --- | --- |
|  |  |
|  |
|  |

When you have finished editing your metadata, just click **Add Item**

# Editing Item details



Item details (Edit mode):

Tick this box if you do not want your metadata searchable via PARADISEC, OLAC (further discussed on page 14), or other online search engines. Unless you are still in the process of editing this information, we would suggest you do not hide your metadata.

If your date is less specific, use this field, e.g. June1966

Item details (after saving):

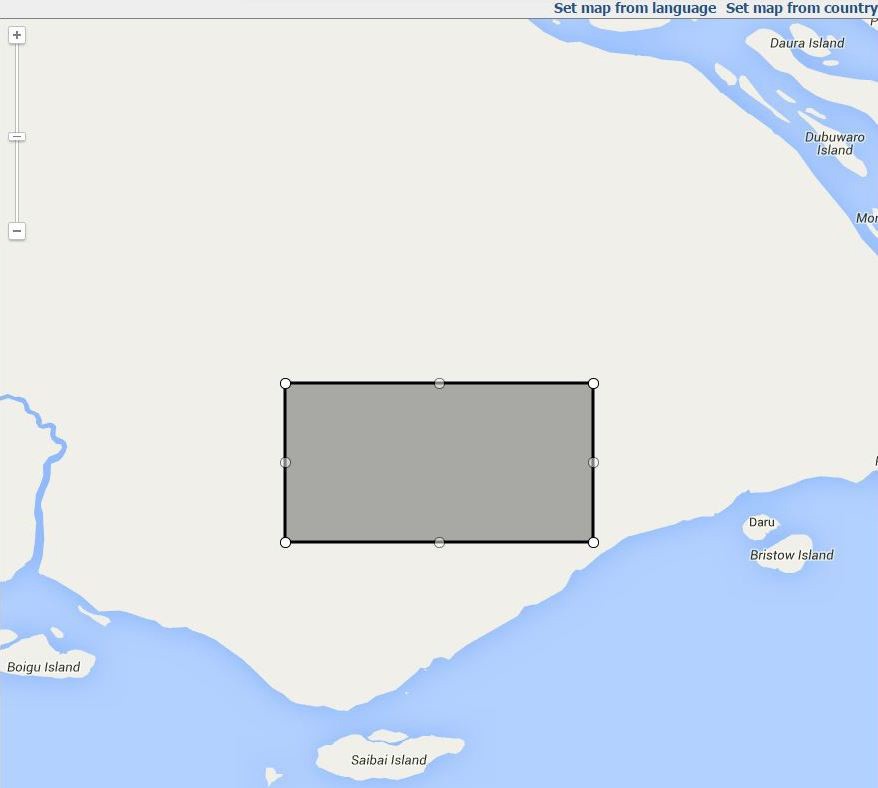
If dialect or village is relevant to your research, please add info here

**Subject language** is the target language for your research.

**Content language** will include all subject languages, plus any lingua franca used (e.g. Kriol, Tok Pisin, English)

Item details (Edit mode):

To add geographic information as to where each recording is from, you have three options:



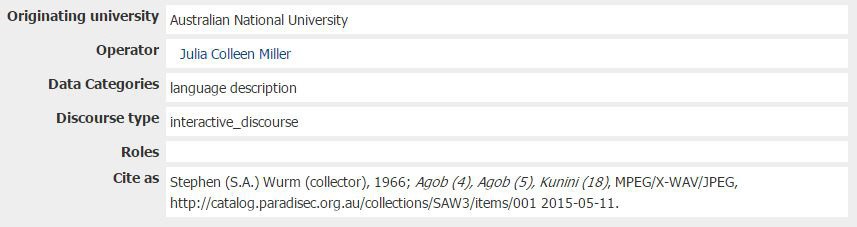
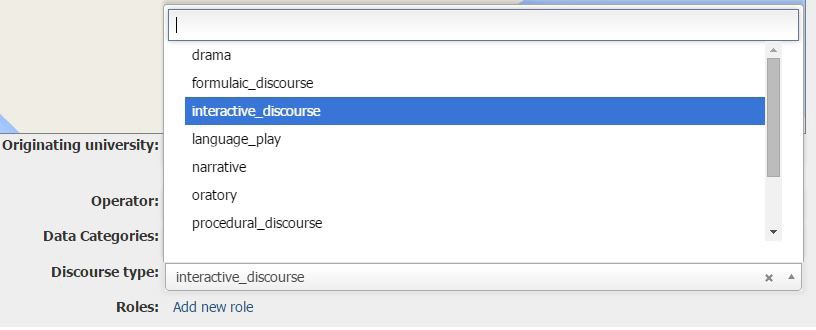
* 1. Set map from country
  2. Set map from language
  3. Drag bounding box to surround chosen area

*Editing Item details (cont’d)*

Item details (Edit mode):

Item details (after saving):

Note the automatically generated “cite as” information. This is created from the metadata you add to each item.



Drop down menus allow you to choose your originating university/institute and further describe your recordings.

**Data categories** include: *histroical reconstruction, historical text, instrumental music, lexicon, moving image, photo, primary text, song, sound.*

**Discourse types** include: *drama, formulaic discourse, interactive discourse, language play, narrative, oratory, procedural discourse, report, singing, unintelligble speech.*

## Editing Archive Information at the item level

Archive information (Edit mode):

When you have ***completed*** adding metadata to your item, tick this box



Archive information (after saving):

It is likely that most of your recordings will be created in a digital format. If so, tick the “Born digital” box. If your recordings originated from cassette or reel-to-reel (r2r) tapes, do not tick that box, put media information in the “Origi- nal media” field. Then, fill in any other relevent fields.

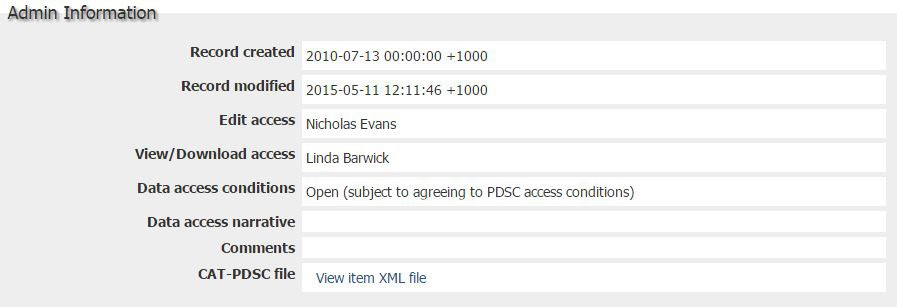
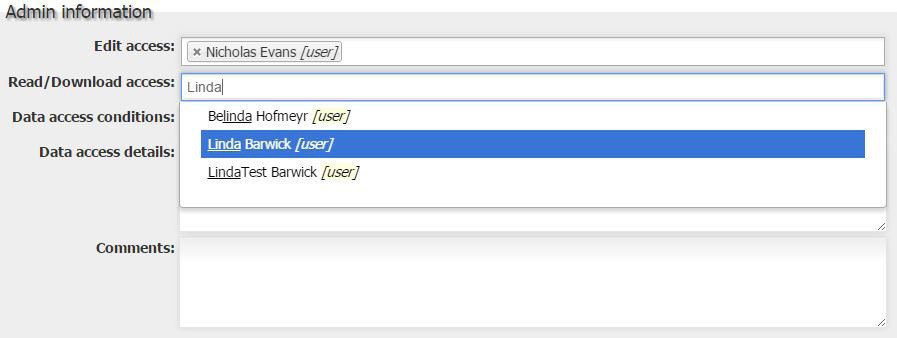
**}**

Leave these fields blank

## Editing Admin Information at the item level

You, as the collector, are the administrator of your collection and the items held within. In this section of the catalog, you may assign certain rights to selected individuals. *Note: Before you can add these individuals, they will need to sign up as PARADISEC users first.*

Admin information (Edit mode):



Admin information (after saving):

**Edit access**

**View/download access**

If you wish to allow someone to view or download a restricted access item file, you would add their name here; they will have to be added in as a PARADISEC user. Open access files can be downloaded by any registered user.

**Data access conditions**

You choose what access setting you would like to set per item.

1. “As yet unspecified”
2. “Closed (subject to the access condition details)”
3. “Mixed (check individual items)”
4. “Open (subject to agreeing to PDSC access conditions)”

If you have chosen the second option here, enter the access condition details in the “Data access narrative” field.

If you want to assign metadata editing rights to someone, such as a project collaborator, community linguist, or a research assistant, you would add their names here. If they are not already in the

PARADISEC system as a user, they will have to be added before they can be assigned editing rights.

**Data access narrative**

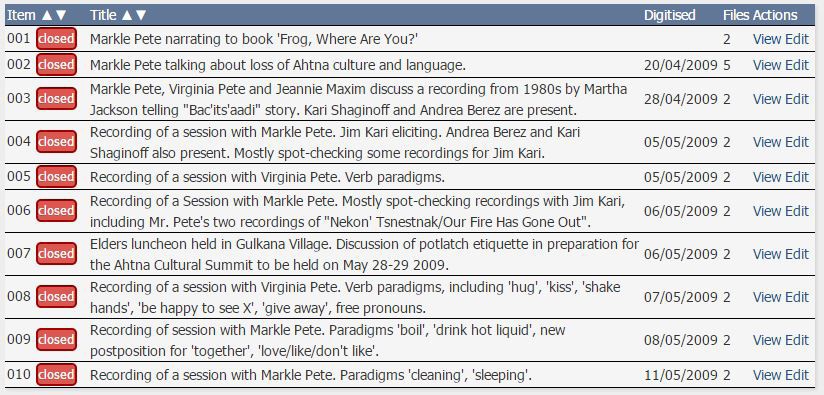
Here is where you offer instructions to gain access, if you would like to set

any conditions of access other than the PARADISEC “open” conditions. This could include instructions as to how to contact the person(s) responsible for making access decisions, or you could designate a date after which

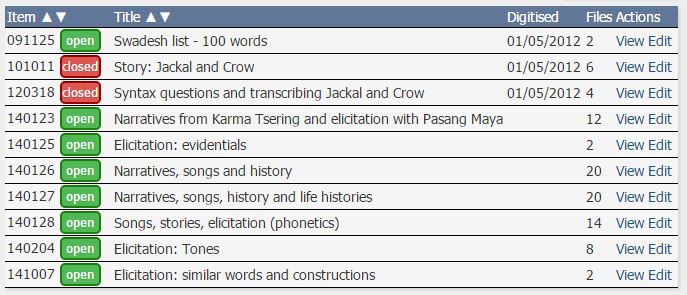
the material will become open, as per PARADISEC conditions, etc.

## Item level access conditions

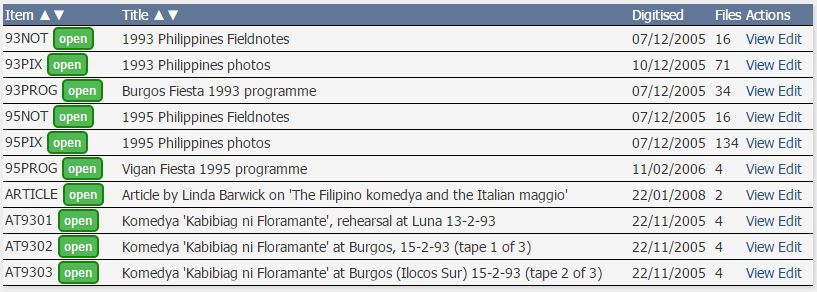
Collection with **closed access** - Access to all items requires interested parties to follow instructions in the Data Access Narrative field in order to gain access.



Collection with **mixed access** - Access to some items requires interested parties to follow instructions in the Data Access Narrative field in order to gain access. Open items can be accessed by agreeing to the PARADISEC Conditions of Access.

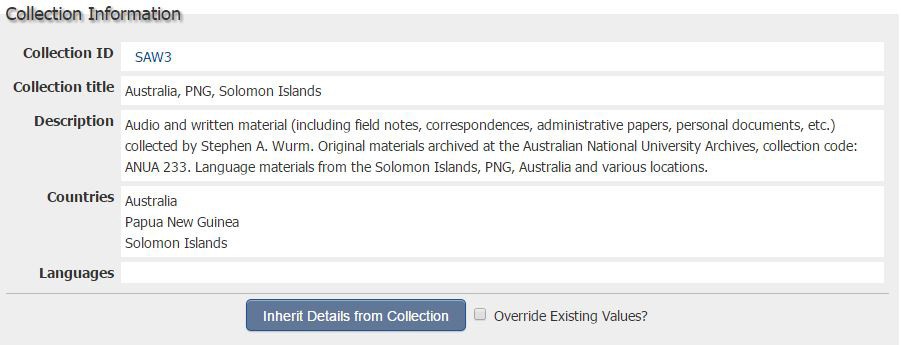


Collection with **open access** - Access to all items can be granted by agreeing to the PARADISEC Conditions of Access.

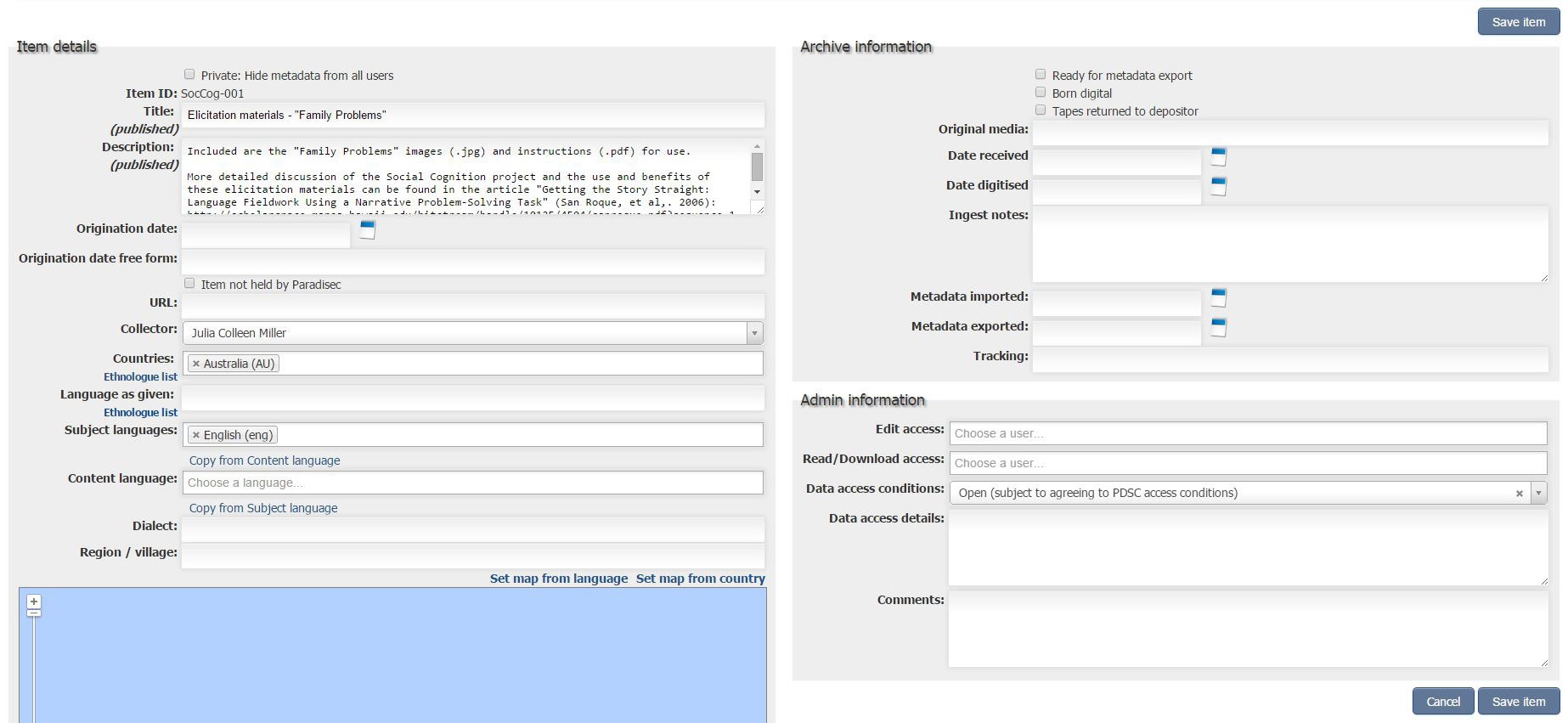


## Inheriting details from Collection to populate Item metadata fields

If an item shares information with your overall collection (such as same country, language, operator, access conditions, and region/village), you can automatically populate these fields by pressing “Inherit Details from Collection”.

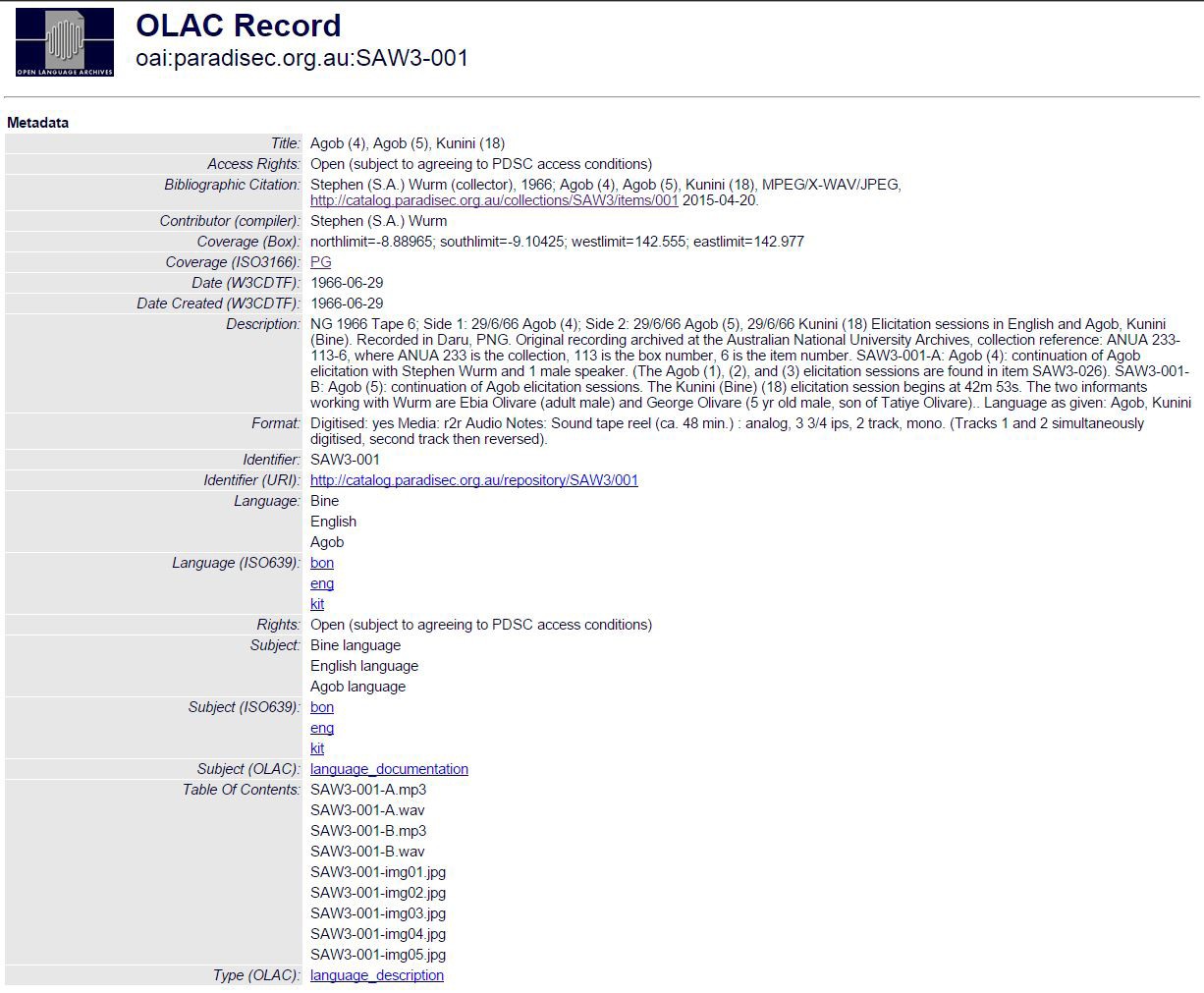


Once you fill in all the item metadata, click on the **Save Item** button, and the new item will be listed in your collection.



# OLAC (Online Language Archives Community)

Having rich metadata records not only facilitates searches directly via the PARADISEC catalog, it allows archiving search engines, such as OLAC, to find your material and make it more widely publicised. Below is an example of an OLAC search on the language Agob, and an example OLAC record. All information is extracted directly from the PARADISEC item record.



# Preferred deposit formats

We will generate access versions of all audio files (.mp3) and video (.mp4) files, allowing for easy online streaming. Below is a list of what files we would prefer to receive from you.

\* If you have files such as .mp3 (audio) or

|  |  |
| --- | --- |
| **Media** | **Files from you** |
| Audio | .wav, .aiff, .mp3\* (16bit 48khz, or as close to the archival 24bit 96khz as possible) |
| Video | .mts (AVCHD), .avi., .mov, .mpg\* |
| Text | .txt, .xml, .pdf, .rtf |
| Annotations | .eaf, .xml |
| Lexicons | .xml |
| Images | .tif, .jpg |

.mpg (video), we will certainly accept them;

however, if you are collecting new recordings,

please avoid these formats, as they are lossy,

compressed formats.

Regarding Toolbox and FLEx files, while .rtf or .pdf is acceptable, it is preferable that you deposit full-content versions of your projects. As an example of Toolbox files in PARADISEC, take a look at the NT8 collection:

#### ([http://catalog.paradisec.org.au/collections/NT8/items/TEXT?files\_per\_page=14](http://catalog.paradisec.org.au/collections/NT8/items/TEXT%3Ffiles_per_page%3D14))

“Texts are in interlinear format. Toolbox requires settings files (TYP) that accompany the main data. Thus the file NT8- TEXT-DIC\_[DATE].txt has the accompanying file NT8-TEXT-DICTYP\_[DATE].txt. The collection of texts in NT8- TEXT-TX\_[DATE].txt has the TYP file in NT8-TEXT-TXTYP\_[DATE].txt.”

# Deposit form

When you are ready to deposit your materials with PARADISEC, you must also fill in a **Deposit of Material** form. You can download this from the following sites:

* PARADISEC: [**http://www.paradisec.org.au/PDSCdeposit.rtf**](http://www.paradisec.org.au/PDSCdeposit.rtf)
* CoEDL Data Archives page: [**http://www.dynamicsoflanguage.edu.au/research/data-archives/**](http://www.dynamicsoflanguage.edu.au/research/data-archives/)

The purpose of the deposit form is to safeguard the interests and sensitivities of relevant indigenous people. There is helpful information located at the end of the form, such as explanations of ownership and intellectual property and an outline of PARADISEC’s conditions of access. You need to supply information such as your contact details, details of the material you are depositing, and an explanation of your chosen acces conditions. Sign your filled-in form and send a scanned copy of it along with your material. It will be archived in your collection (as a hidden admin file).

If you would like more information on depositing with PARADISEC, please visit our site:

#### <http://www.paradisec.org.au/deposit.html>