Creating Metadata

(CMDI Maker – Arbil Workflow)

Endangered Languages Archive, 27 April 2017

Overview

Sections 1 and 2 of this helpsheet are an overview of the workflow you will need to follow in order to create your metadata files for self-archiving. If you encounter any errors along the process, please refer to section 3.

You will need to open CMDI Maker (http://cmdi-maker.uni-koeln.de/) on your browser. You only need Internet connection the first time because the tool stays saved in your browser afterwards.

You also need to download Arbil (https://tla.mpi.nl/tools/tla-tools/ Arbil/download/) for the second part of this workflow.

Before starting it is important that you have an idea of how you are going to bundle your files and that this is reflected in your file naming system (refer to ELAR: instructions for depositors for further details on file naming).

Create a folder, name it for instance Metadata, and copy all the resources you want to archive to that folder.

Contents

Overview ......................................................................................................................................................... 1

1 Create your sessions/bundles with CMDI Maker ......................................................................................... 3
   Getting started on CMDI Maker .................................................................................................................. 3
   Backup your project ...................................................................................................................................... 7
   CMDI Maker Export .................................................................................................................................... 9

2 Final adjustments in Arbil .......................................................................................................................... 10
   Open IMDI files in Arbil .............................................................................................................................. 10
   Define subject language & working language ............................................................................................ 13
   Add keywords ............................................................................................................................................... 14
   Add topic ...................................................................................................................................................... 16
   Add access definitions ................................................................................................................................. 17
   Bulk editing access definitions .................................................................................................................... 19
   Bulk editing access definitions of specific file types .................................................................................... 22
   Arbil Export ................................................................................................................................................. 23
3 Troubleshooting........................................................................................................... 26
Linking missing resource files in Arbil ........................................................................... 26
Problems with unrecognized files in Arbil ...................................................................... 27
How to update Arbil file checker .................................................................................... 27
File encoding problems ................................................................................................... 29
1. Create your sessions/bundles with CMDI Maker

Getting started on CMDI Maker

1. Open CMDI Maker and select the IMDI profile. Please do not use the ELDP profile, because it creates metadata in a format (*.cmdi) that is not supported by our self-upload archiving system.

2. Fill in every tab in the CMDI maker with rich and comprehensive descriptions: corpus, resources, actors, and sessions. Click on the buttons in the upper green bar to move forward and keep adding information.
3. In the corpus tab, do not forget to set the content language. In Corpus, leave the fields ‘name’, ‘title’ and ‘description’ empty.

4. In the resources tab, you can drag and drop the resources you want to add to your sessions or browse them. Media files (audio, video and pictures) will appear in orange and text/annotation files (pdf, txt, ELAN files, etc.) will appear in green. Refer to ELAR: instructions for depositors for further details on accepted file formats.
5. In the Actors tab, click on ‘create one’ to add your first actor and then refer to the bottom menu to add more, delete or duplicate actors.

6. If the same actor undertakes different roles, you need to add him/her different times with different roles. Click on the actor that is supposed to have different roles and select ‘Duplicate this actor’ at the bottom menu. After that, you just need to change his/her role.
Creating Metadata

7. At the session level, in the project tab, please enter:
   I. Name of the project in the following format: name of the language-surname of depositor-deposit key (example: ikaan-salffner-0259)
   II. The title of the project is the title of your grant (example: Farming, food and yam: language and cultural practices among Ikaan speakers)
   III. ID is your ELDP grant ID (example: IPF0178). If you don’t have an ELDP grant ID, please leave the ID field empty.

8. Use the session description field to enter not only the description of the session but also information about its content, otherwise the content description will not show up in the online catalogue. Leave the description field in the ‘content’ and in the ‘project’ empty. Add the actors and resources to the session.
Creating Metadata

Backup your project

9. Back up your project after each metadata working session (i.e. after you have created your sessions or every time you make changes or update information). Click on the hamburger icon at the top right corner (when you hover with you mouse over it, the background changes from green to pink) and select ‘Save project to file’, which will create a *.cmp file (CMDI Maker Project). We recommend that you name these files with the date so you can keep track of your updates.

10. To open the *.cmp file click on the hamburger icon at the top right corner and select ‘Open Project File’.

11. A window pops up that asks you to choose the cmp file you want to open. Click on ‘Choose file’.
12. Browse for the *.cmp file you have saved before and select it. Then click on Open.

13. When you open a CMDI Maker Project File, the information in the file will overwrite whatever other information you may have entered into the CMDI Maker before. Once it is overwritten, you cannot undo the overwriting and the previous data is lost. That is why the CMDI Maker double-checks with you that you really want to load the new file and overwrite the old information. If you are sure you want to overwrite, click on ‘Yes, overwrite data’.

14. This opens the CMDI Maker Project in your browser and you can see all the information you have entered before.
CMDI Maker Export

15. Export the IMDI file to your Metadata folder when you finished entering all the information you need in your metadata by clicking on ‘XML output’.

16. You have the following export options:

   Option 1 (for exporting a few sessions only):
   i. In front of each session name you have a save button, click on it to save each file individually in the folder where you also have the resources (i.e. the metadata folder you created before).
   ii. Scroll down to see the other sessions.

   Option 2 (for exporting all sessions at once):
   i. Click on ‘Download Corpus including all sessions’ at the left bottom of the page and it will open the save option for each individual IMDI file.
   ii. Select your metadata folder as the place to save the metadata.

   Option 3 (for compressing and exporting all sessions at once):
   i. Click on ‘Download zip archive’ at the left bottom of the page and it will compress all your IMDIs in a zip file that you can save and then extract the individual IMDI files into your metadata folder.
2 Final adjustments in Arbil

Open IMDI files in Arbil

17. Open Arbil. Go to File > Import. Navigate to your metadata folder and select your newly created IMDI files (the files with the extension *.imdi).

18. In the Import window, tick the option ‘Import resource files (if available)’ and click on start. This might take a few minutes.

19. The Import Branch window stays open during the import process, and you can see the progress in the moving bar under the numbers that show how many bundles of the total number of bundles have already been imported.
20. At the end of the import process, Arbil informs you how many files it has processed. Click on Details to check if there have been any import errors, and if yes, what the errors may have been.
21. This takes you back to the Import Branch window, where you can now see the error report, here the red writing in the pink circle.
   a. In this case, there are no errors.
   b. If there were errors, click on the Validation Errors tab and the Resource Copy Errors tab (in the pink circle) and either trouble shoot the errors yourself or copy them and send them to ELAR to ask for help.
   c. At the end, click on Cancel to close the Import Branch window.
Define subject language & working language

22. Your deposit has been copied to the Local Corpus. You can open your sessions by clicking on the pin next to the session node (the ones with the green sign next to it).

23. Open the bundle by clicking on the blue pin in its left.
24. Right click on Content > Add > Content Language.
Creating Metadata

25. Fill in the fields ‘ID’, ‘Name’ and ‘Description’ with the ISO code of the language (in the format ISO639-3:code), its name and whether it is the subject or the working language.

![Image of a table with fields and values]

26. The new values appear in blue, which means they are valid but not saved yet. Hit Ctrl + S or go to File > Save Changes to save the information you have entered. The values you just entered and saved will turn black. If the values appear red, enter a new value that the system recognizes, otherwise you will not be able to export the bundle correctly later on.

27. You can add as many subject and working languages as you need for each bundle by following the steps 19-21.

Add keywords

28. Keywords help you not only to organise and sort your data according to different thematic or grammatical categories but also to make more specific searches when your data is online. Keywords are specific to bundles. To add keywords, open the bundle, right click on Content > Add > Content Key.

![Image of a menu with options for content manipulation]
Creating Metadata

29. On the main window, right click on the field ‘Keys.Key.Standard Template Key’ and choose ‘Open in long field editor’.

30. Click on ‘Change key name’ to change the name of the field to ‘Keyword’ (with capital K).

31. Then click on Apply > OK and close the window. You can now enter your first keyword. Hit Ctrl + S or go to File > Save Changes to save your changes.
Creating Metadata

32. Each keyword must be in an individual field, i.e. for each keyword you need to add a new content key and follow the steps 23-25. You can add as many keywords as you want.

Add topic

33. Topics, like keywords, help you organise and sort your data and make detailed searches later. They are more general than keywords and normally comprise several bundles (topic example: cooking / keyword example: traditional dish). To add them, open the bundle, right click on Content > Add > Content Key.

34. On the main window, right click on ‘Keys.Key.Standard Template Key’ and choose ‘Open in long field editor’.
35. Click on ‘Change key name’ to change the name of the field to ‘Topic’ (with capital T). Then click on Apply > OK and close the window.

36. Write your topic and hit Ctrl + S or go to File > Save Changes to save your changes. You can add as many topics as want but each topic should be in an individual field (follow the steps 28-30 for each new topic).

Add access definitions

37. This time you will need to add information directly to the resource files. Open the bundle and go to ‘Written Resources’ or ‘Media Files’. Open them and click on each individual file. Have a look at the table displayed in the main window. (If your resource files are not showing as expected, i.e. if the paths to the resource files are broken, refer to section 3 in this document to know how to link them properly.)
38. You should be able to see that the field ‘Access.Availability’ is blank. At ELAR, we work with the following access protocols:

- **O (Open access)**
  Anyone can see and download the files, even if they are not registered with ELAR and have not agreed to the End User Terms and Conditions that ELAR requests to be accepted by all ELAR users. It also means that you cannot track who downloaded your materials. Some depositors do indeed want this, for example to take away a barrier for community users.

- **U (User)**
  All registered users of ELAR who have accepted the ELAR End User Terms and Conditions can see your files, after they have logged in. With this license, you as a depositor can also track who downloaded which files when.

- **S (Subscriber)**
  A person has to (a) be a registered user of ELAR who has accepted the ELAR End User Terms and Conditions, and (b) has to specifically request access from you as the depositor and be granted access by you (or your delegate) to see your files, and only after they have logged in. With this license, you as a depositor can also track who downloaded which files when. If you do choose to use the S category, you will need to explain why you are closing your resources and discuss with ELAR the terms of the embargo period.

Please, consider making your data available at least to ELAR users. You can get in touch with us, if you have any questions regarding further access issues.

39. Enter the value O, U or S into the ‘Access.Availability’ field and save your changes. If you decide to close the access to a resource, you will have to include a description in the field ‘Access.Description’ explaining why (e.g. PhD embargo, sensitive material, etc.). Please discuss the embargo conditions with ELAR before.
Creating Metadata

Bulk editing access definitions

Setting access definitions by bulk editing is fairly straightforward in Arbil. As the first step, you have to select all the files that need editing, and as the second step you edit the access. If you have decided that all your resources should have the same access definitions (U, for instance), this is the way to proceed.

40. Click on the corpus node or select all the bundles you want to edit.
41. Right-click and in the menu that opens select ‘Search’.
42. In the window that opens, select All as type and type Access.Availability in the second box. Leave the third box blank. Then click on search to start the search.
43. All files are listed. Move the column boundaries around to find and see the column with the header Access.Availability.

44. Click into the first cell in the Access.Availability, type the access definition you want to set for the entire selection and hit Enter.
Creating Metadata

45. Right-click into the cell. In the menu that opens, select ‘Copy Cell to Whole Column’.

46. Say Yes when a window warns you that copying will overwrite all existing entries in the column.
47. The information has been copied down. Press CTRL+S to save your changes or go to File > Save Changes.

48. Please, remember that if you are applying S to any files or bundles, you need to specify why. If the same reason applies to all bundles, you can also edit this in bulk by repeating the previous steps (35-42) but in the column Access.Description.

**Bulk editing access definitions of specific file types**

If you want to have all your media files (pictures, audio and video files) set to U but the written resources (pdf, txt, ELAN, Praat and toolbox files) should be set to S, then follow the instructions below.

49. Click on the corpus node or select all the bundles you want to edit.
50. Right-click and in the menu that opens select ‘Search’.
51. Select MediaFile (pictures, audio and video files) or WrittenResource as type and don’t type anything into the boxes.
52. All files are listed, as below:
53. Move the column boundaries around to find, see the column with the header Access.Avlailability and follow the steps 43-45.

**Arbil Export**

54. When you are done with editing your metadata, select your bundles, right click on them and choose ‘Export’.

55. Navigate to your metadata folder, create a subfolder called ‘Arbil Export’ to save your sessions from Arbil. Click on ‘Export Branch Destination Directory’.
56. Tick the box ‘Export resource files if available’ and then click on start. This process might take a long time (anything between a blink of an eye or ten hours!) depending on the amount of data.

57. If a window pops up telling you that the branch directory destination is not empty, tick ‘Remember my choice’ and click on ‘Overwrite’.

58. When Arbil is done, it informs you that it has processed the files.
   a. If there are no export errors, click on Close.
   b. If there are export errors and you want to investigate further, click on Details.

59. When you click on Close, Windows Explorer opens the ‘Arbil Export’ folder and shows you the metadata file it has created and the subfolders with the resources in them. There should be one folder per IMDI file. These are the files/folders you use later to upload to LAMUS and update the metadata record in the catalogue (refer to the helpsheet on Lamus and self-upload process).
To save space on your computer, delete the resources that you have previously copied to the metadata folder. They are now properly sorted within different subfolders in the ‘Arbil Export’ folder.
3 Troubleshooting

Linking missing resource files in Arbil

61. If you have your IMDI files and the resources that belong to them in the same folder (metadata folder) and you have been consistent with their names, Arbil should be able to find the files and link them to the metadata automatically. If it is not the case and the path to the resources is broken, follow these steps to link them manually.

62. Open a bundle by clicking on the blue pin in its left. This is how resources with broken paths files look like in Arbil:

63. To relink them to your bundles, there are two options:

**Option 1**

a. In the bottom left corner of your screen, right click on Working Directories > Add Working directory.

b. Navigate to the folder that contains your resource files, select it and click on ‘Add working Directory’. You work with as many directories as you want.

c. Now the folder should appear beneath the computer icon. Open it and find the file you want to link. Select the file and drag it to the top of the bundle.
Creating Metadata

d. To delete the unlinked file from your bundle, right click on it and select Delete > Yes. It is always safer to link first and then delete.
e. Remember to add the access definitions to the resource and to save your changes.

Option 2:
a. Right click on the resource that you want to relink and select ‘Browse for resource file’
b. A window pops up that allows you to browse for the resource that is in your metadata folder.
c. Select the resource you want to link and click on ‘Open’.
d. Remember to add the access definitions to the resource and to save your changes.

Problems with unrecognized files in Arbil

First, be sure that you have updated Arbil file checker, so that Arbil can check if your resources conform to ELAR guidelines.

Note: The Arbil file checker is a small file that defines and checks which resource files pass the archive guidelines and which files still need to be converted to be of archive quality. The ELAR definitions of what we consider archivable files differ slightly from the standard settings in Arbil. Therefore the file checker file needs to be manually updated within Arbil.

How to update Arbil file checker

For the Arbil file checker to work, it needs to sit in the right folder. To put it there, follow the steps below.

64. Download the file filetypes.txt from ELAR guidance website (https://www.soas.ac.uk/elar/) and keep it safe for pasting into a folder later.
65. Open Arbil.
66. In the menu bar, click on ‘Options’, and then on ‘Local Corpus Storage Directory’. You will see in grey writing the path to where your local storage for Arbil is.

67. In the Arbil shown here, the full path to the local storage folder for Arbil is C:\Users\Local Settings\ApplicationData\arbil\ArbilWorkingFiles.
In your own Arbil, it will be slightly different, and there might not be a full stop in the .arbil folder either. That’s nothing to worry about, that’s all fine.
68. Open a Windows Explorer and follow the path shown in your Local Corpus Storage Directory until you get to the folder .arbil [or just arbil if you are using a Mac].

69. Open the .arbil folder and paste the file filetypes.txt right into the .arbil folder. The file should be a sister node to ArbilWorkingFiles folder, in the same level in the hierarchy.

70. Close Arbil.
71. Open Arbil again.
72. It is possible that some of your files look like this (with a F) in Arbil:
73. Right-click on the resources that show up with a big F and select ‘Force Type Checking’.
74. This should redo the type checking and show the MTS/MP4 file now with a neat little camera next to it or the WAV file with a loudspeaker.
75. If the F doesn’t disappear, then further checks need to be done:
   a. Check that the file name does not contain spaces, tildes or special characters.
   b. Check the format. Remember that ELAR does not accept .mp3, .mov, .avi, .doc or .trs
      (for a complete list of the accepted formats please refer to ELAR: instructions for depositors)

File encoding problems

76. Considering the previous checks, if everything seems in order and your files still show up with a big F, the problem is related to the encoding of the file (normally the audio or the video file).
77. Run it through a converter without changing the settings. We recommend Handbrake (https://handbrake.fr/) for video and Audacity (http://www.audacityteam.org/) or Switch (http://www.nch.com.au/switch/) for audio files (refer to specific helpsheets on audio and video conversion and re-encoding).
78. Badly encoded .pdf files need to be saved as PDF-A. You will find this option when saving the document in your word processor or using Adobe Acrobat Pro.
79. Once you have converted the files and replaced them, right click on the file and choose ‘Reload’. It should appear with its normal icon now: