

User manual

APE Data Preparation Tool (DPT v1.2)

Document name: ManualDPT-v1.2

Publish date: April, 2013

Authors: Yoann Moranville, Lucile Grand, Kerstin Arnold

TABLE OF CONTENT

SUMMARY / INTRODUCTION	2
1. INSTALLATION AND LAUNCH OF THE TOOL	2
2. OVERVIEW OF THE MAIN PAGE AND ITS OPTIONS	3
2.1. The menus	5
2.1.1 File menu.....	5
2.1.2 Options menu	6
2.1.3 Actions menu.....	8
2.1.4 Windows menu.....	8
2.1.5 Help menu	8
2.2 The list.....	8
2.3 The tabs.....	11
2.3.1 Summary tab	11
2.3.2 Validation tab.....	11
2.3.3 Conversion tab.....	12
2.3.4 Europeana conversion tab.....	12
2.3.5 Edition tab.....	12
3 EAD TO ESE/EDM CONVERSION	13
3.1 ESE conversion	13
3.2 EDM conversion	16
4 UPDATE OF THE TOOL	17
5 TROUBLESHOOTING	17

Summary / introduction

The Data Preparation Tool is meant to help the partners to prepare their data locally, off-line if they wish to. It offers the same range of functionalities as the Dashboard, with some differences. Basically, it offers more possibilities than the Dashboard.

This tool is altogether a validation tool, a conversion tool and an edition tool. These different actions can be combined, processed in batch etc.

- Validation tool:

Against existing schemas; this allows the partners to check whether their data are compliant to the EAD schema for instance.

After a conversion; this is a mandatory step to obtain correct apeEAD, EAG 2012 or – in future steps – apeEAC-CPF files that can afterwards be uploaded and published in the Portal.

- Conversion tool:

You can convert from a lot of formats to a lot of others. You can even import your own style sheet to convert your data: from EAD 2002 to apeEAD, from EAG 0.2 to EAG 2012 etc. You can also create your own date conversion table to get normalised dates (for instance from 20th century to 1901/2000).

- Edition tool:

Once the conversion is done, you can edit your file and check that everything went smoothly; on top of that, the tool provides you with a report on the points that can be improved (data quality check): normalise the dates, add titles and the type of digital objects.

1. Installation and launch of the tool

The tool is distributed as a ZIP archive file, whose name is *APE_data_preparation_tool_X.Y.Z.zip*, X.Y.Z being the version number. The last version is downloadable at this address:
<http://www.apex-project.eu/index.php/outcomes/tools-and-manuals>

Requirements in order to use the standalone tool:

- Have Java 6 installed on your computer, or a more recent Java version. You can verify which version is installed on your computer by launching the console (Windows) or terminal (UNIX) and typing *java -version*.

Please do the following once you have the ZIP archive:

- Uncompress the ZIP archive using *7-zip*, *winzip* or any program able to uncompress ZIP archives.
- You will have a folder containing 1 folder (named *xsl*) and 4 files.
- Depending on your Operating System (OS):
 - Using Windows: Double-click on the *launch.bat* file,

- Using UNIX: Using your terminal launch the *launch.sh* file (i.e. Type *shlaunch.sh* – make sure that the file has at least the execution rights, *chmod +x launch.sh*),
- Using OS X: Double-click on the *APE_data_preparation_tool-jar-with-dependencies.jar* file¹.

- On the first launch of the tool, a dialogue box will ask for an identifier for your archives: please enter your registered ISIL identifier (if any, else you will still need to enter a valid iso15511² code). It will also ask for your country code (2-letter country code according to iso3166). Note that you can change those anytime you want in the *options*. Be aware that the dialogue box might stay in the background so that you could not see it at first.
- The program should be now running (see figure 1), if that's not the case, please refer to the troubleshooting section of this manual.
- The folder now contains 4 folders (apeoptions, output, temp and xsl) and 5 files.

2. Overview of the main page and its options

¹ Double-clicking the *APE_data_preparation_tool-jar-with-dependencies.jar* file can also be used as an alternative to the *launch.bat* file with Windows.

² The ISO standard 15511 consists of up to 16 characters, usually starting with a two-letter country code (or a one-, three- or four-letter organisational code), followed by a hyphen, followed by an alphanumeric code for the single institution. The latter can contain the digits 0 to 9, the characters A to Z and the characters -, / or \.

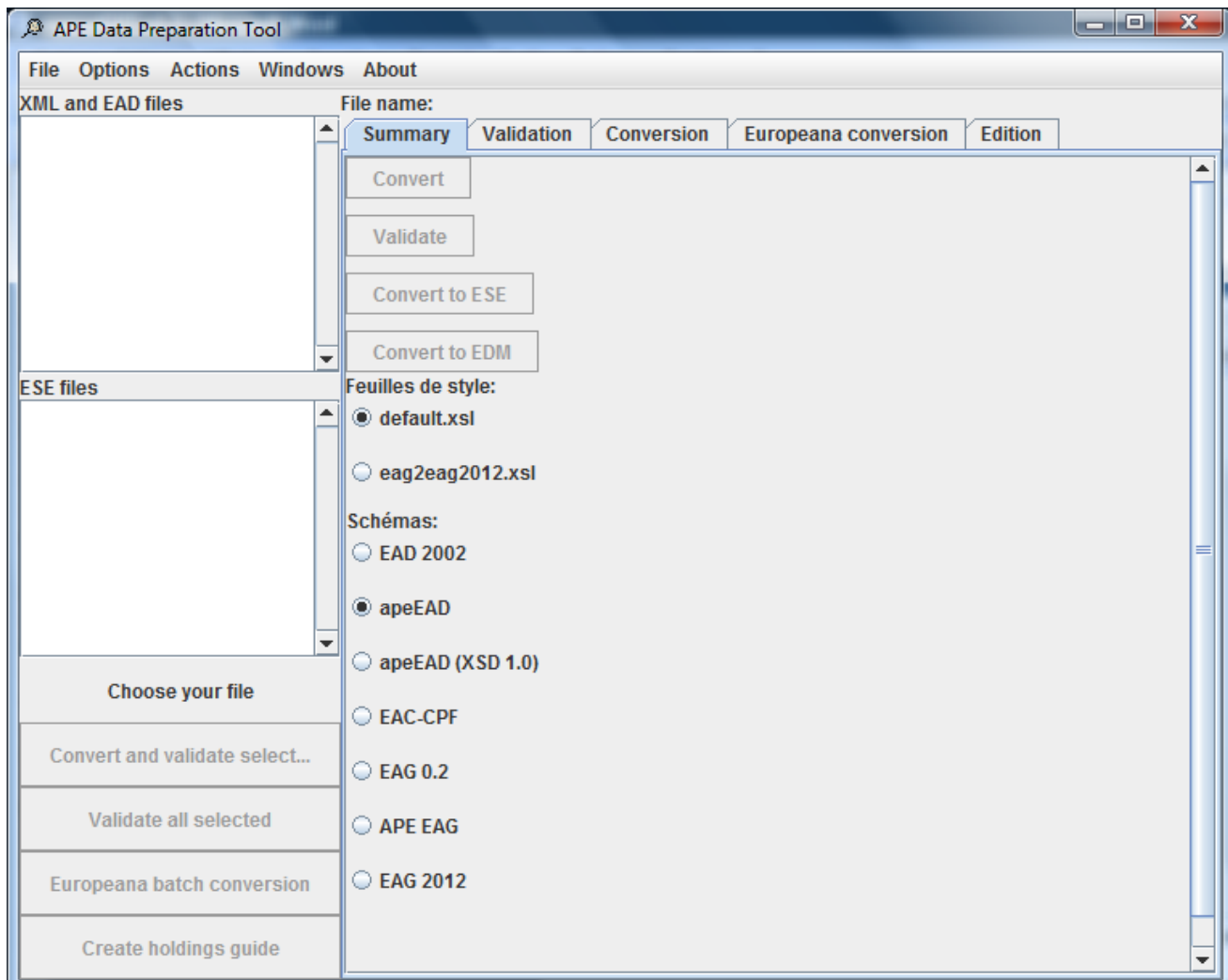


Figure 1: “home page” of the DPT, summary tab activated by default

In this main page, you see 5 menus at the top: *File*, *Options*, *Actions*, *Windows* and *About* (see next chapter 2.1).

The rest of the screen below is divided in two parts:

- On the left side two empty lists named “XML and EAD files” and “ESE files” with 4 action buttons at the bottom: *Convert and validate selected*, *Validate all selected*, *Europeana batch conversion*, *Create holdings guide*. These 4 buttons are placed just below the indication “Choose your file” (see chapter 2.2).
- The right side contains 5 tabs at the top: *Summary*, *Validation*, *Conversion*, *Europeana conversion* and *Edition* (see chapter 2.3).

The summary tab has 4 action buttons (*Convert*, *Validate*, *Convert to ESE*, *Convert to EDM*) and 2 radio button lists. The first list is for choosing the style sheet to be used (*default.xsl*, *eag2eag2012.xsl*); the second list is for choosing the schemas and lists the different schemas/standards used (*EAD 2002*, *apeEAD*, *apeEAD (XSD 1.0)*, *EAC-CPF*, *EAG 0.2*, *EAG 2012*). The other tabs stay empty until a file is processed. They then display the according information.

2.1. The menus

2.1.1 File menu

You have 4 items in the File menu:

- **Select file or directory** (shortcut: Ctrl+O or Cmd+O):

This is where you can open your files or directory to be able to process them in the tool. You have the possibility to select one or several files, or one directory. If you select a directory, it will import all the files located in this directory; however, subdirectories will not be included. Once you have done your selection, the files appear in the “XML and EAD” list on the left side of the window and are ready to be processed.

Please note that you can process hundreds or even thousands of files together.

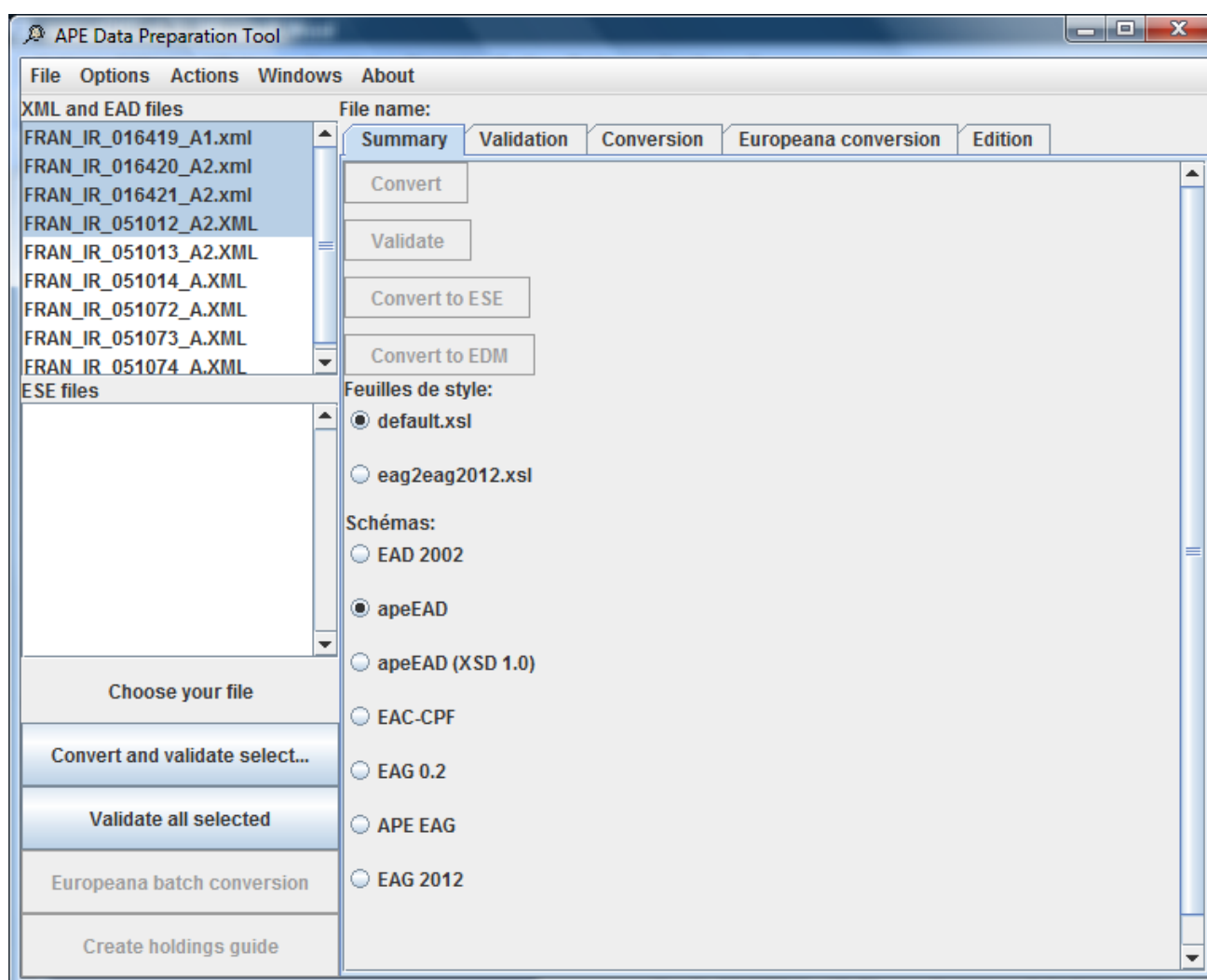


Figure 2: Activating the batch-buttons by selecting more than one file

- **Create EAG 2012** with three sub-menus: *Select an existing EAG (0.2) file*, *Select an existing EAG (2012) file*, *Create a new EAG 2012 file*.

- Select an EAG file (0.2 or 2012) allows you to open the file that you want to process in the tool.
 - Create a new EAG 2012 file opens a form in a pop-up window (figure 3). The form is subdivided in 7 tabs that have to be completed following the logical order of the standard: *Institution*, *Identity*,

Contact, Access and Services, Description, Control, Relations. You move from one tab to the other by clicking on the next/previous buttons.

The screenshot shows a web-based form titled 'eag2012.yourInstitutionTab'. It contains several sections for data entry:

- Your institution:** Fields for 'Person/institution responsible for the description', 'Country code*', 'Identifier of the institution*', 'Name of the institution*', and 'Parallel name of the institution'. There is an 'Add further IDs' button.
- Visitors address:** Fields for 'Street*', 'City / Town (with Postalcode)*', 'Country*', 'Coordinates (latitude)', and 'Coordinates (longitude)'. There is an 'Add postal address, if different' button.
- Other fields:** 'Continent*' (set to Europe), 'Telephone', 'E-mail address', 'Webpage*', 'Opening times*', 'Accessible to the public*' (set to yes), 'Facilities for disabled people available*' (set to yes), 'Reference to your institution's holdings guide', and 'Work places*'. There are also three 'Select a language' dropdown menus.
- Link information:** Fields for 'Link title' and 'Link title' (with a pre-filled value 'Go to our homepage'), and 'Closing times'.

At the bottom, there are 'Exit' and 'Next tab' buttons.

Figure 3: form to create an EAG 2012 file

- **Save selected APE file(s)** (shortcut: Ctrl+S or Cmd+S):

Once a file has been converted and validated, it has to be saved. The newly created files will be saved in the output directory by default but the options allow you to indicate another place. The name of the file changes when it is saved. If the file has been validated, its saved name will be *apeEAD_{original_name}.xml*, if not it becomes *NOT_apeEAD_{original_name}.xml*. Of course, this does not affect your original files.

- **Quit:**

Closes the program.

2.1.2 Options menu

You have 11 items in the Options menu:

- **Country code:**

Here you can change your 2-letter country code, the one you had to enter at the first launch of the tool.

- **Repository code:**

Here you can change the identifier for your archives, the one you had to enter at the first launch of the tool.

- **Checks when loading files:**

This option allows the tool to check if the files that you are uploading are valid XML files. This is only an option because it might take time if you have to upload numerous files.

- **Digital object type:**

This option allows you to indicate whether the digitised archives are Text, Image, Sound or Video. If you don't use this option, the default value will be Unspecified. This option is used within the Archives Portal Europe for refining the search results but it is mandatory in Europeana. Those values are reported inside your apeEAD XML files after you convert them. The chosen type will then be available inside your apeEAD XML file within each DAO element as attribute @xlink:role. In case, your local data already contain this information, you also have to specify that this shall be used if found.

Note: Converting the file and then choosing a type will *NOT* change the type inside the newly created file. A new conversion is needed.

- **Default folder for saved files:**

This option allows you to change the default folder where the tool saves the files, which is the output folder of the tool (one of the four folders mentioned above).

- **Use own style sheet:**

This option allows you to use your own style sheet (XSLT) in order to do your conversions alone. This option opens a dialogue box where you can select a style sheet and copy it in the xsl folder of the tool.

- **Use own schema:**

This option allows you to use your own schema (XSD). This option opens a dialogue box where you can select a schema and copy it in the xsl directory of the tool.

- **Choose default style sheet:** either default.xsl, or eag2eag2012.xsl;

By default the style sheet selected for newly imported files is the default style sheet, you can use this option if you wish to use another default style sheet.

- **Choose default schema:** lists the different possibilities offered by the tool to validate the files (*EAD 2002, apeEAD, apeEAD (XSD 1.0), EAC-CPF, EAG 0.2, EAG 2012*)

By default the system sets the data type to apeEAD files, you can change this by using this option.

- **Languages:**

The tool is available in six languages: Dutch, English, French, German, Greek and Hungarian. The tool will open in the language of your operating system if it is Dutch, French, German, Greek or Hungarian and if not found will rollback to English.

- **List of date conversion rules:**

This option lists in a dialogue box the rules that will be applied during the conversion on the date elements of your files in order to transform text dates into normalised dates.

In the left column, you indicate the values read (to be found in the original files), in the right column, the values converted (to be found in the converted files). For instance you can indicate that 18th Century will be converted in 1701-1800, circa 1877 in either 1877, or 1876-1878 etc.

Note that these conversion rules can be saved to be re-used later on. You can modify them as often as needed.

2.1.3 Actions menu

You have 2 items in the Actions menu:

- Validate (shortcut: Ctrl+N or Cmd+N)
- Convert (shortcut: Ctrl+M or Cmd+M)

Those items do the same as the buttons in the main summary tab that we will see in one of the following chapters (see *Chapter 2.3 Tabs*).

2.1.4 Windows menu

You have 5 items in the Windows menu:

- Summary (shortcut: Ctrl+1 or Cmd+1)
- Validation (shortcut: Ctrl+2 or Cmd+2)
- Conversion (shortcut: Ctrl+3 or Cmd+3)
- Europeana conversion (shortcut: Ctrl+4 or Cmd+4)
- Edition (shortcut: Ctrl+5 or Cmd+5)

In order to change to a selected tab, you can either click on it or select it in this menu.

2.1.5 About menu

You have 2 items in the About menu:

- **Visit project website:**

This option opens a page in your browser (you need an internet connection) and you will access the APE_x project website (www.apex-project.eu).

- **APE DPT v1.2.2:**

This indicates the version of the DPT you are working with. We will see in a following chapter (see *Chapter 3. Update of the tool*) that an automatic check is done at the opening of the program to see if a new version is available.

2.2 The list

On the left side of the window there are two empty lists that you populate either by opening files or directories (see 2.1.1) or with the files created within the DPT (for instance holdings guide, EAG or ESE files).

By clicking on the name of one or more files, you select it/them, and then can convert and/or validate it/them by using the radio buttons in the summary tab on the right part of the window and clicking on the action buttons for processing the files. The radio buttons indicate which schema and style sheet you want to apply to your file(s). By default, the buttons are selected to convert or validate EAD files in apeEAD by using the default style sheet.

The action buttons on the right part are activated when only one file is selected and depending on the status of your file (e.g. Convert to ESE will only be available for validated apeEAD files).

The actions buttons on the bottom of the left side are activated when at least two files are selected. These are batch buttons. If you click on the *Convert and validate selected* button while having 5 files selected in the list, the tool will convert and validate all 5 files.

When a file has been successfully converted and validated, its colour changes in the list display to green. If an error occurred, its colour changes to red. Same for the colour of the tab corresponding to the process conducted.

You can use the right click of your mouse to remove a file from the list.

On the main part of the window, you will see the progression and status of the operations thanks to a progress bar. You can stop the program when you want by pressing the *Abort* button that is made available during the process, but please note that after pressing the *Abort* button, the tool will finish its last operation before stopping.

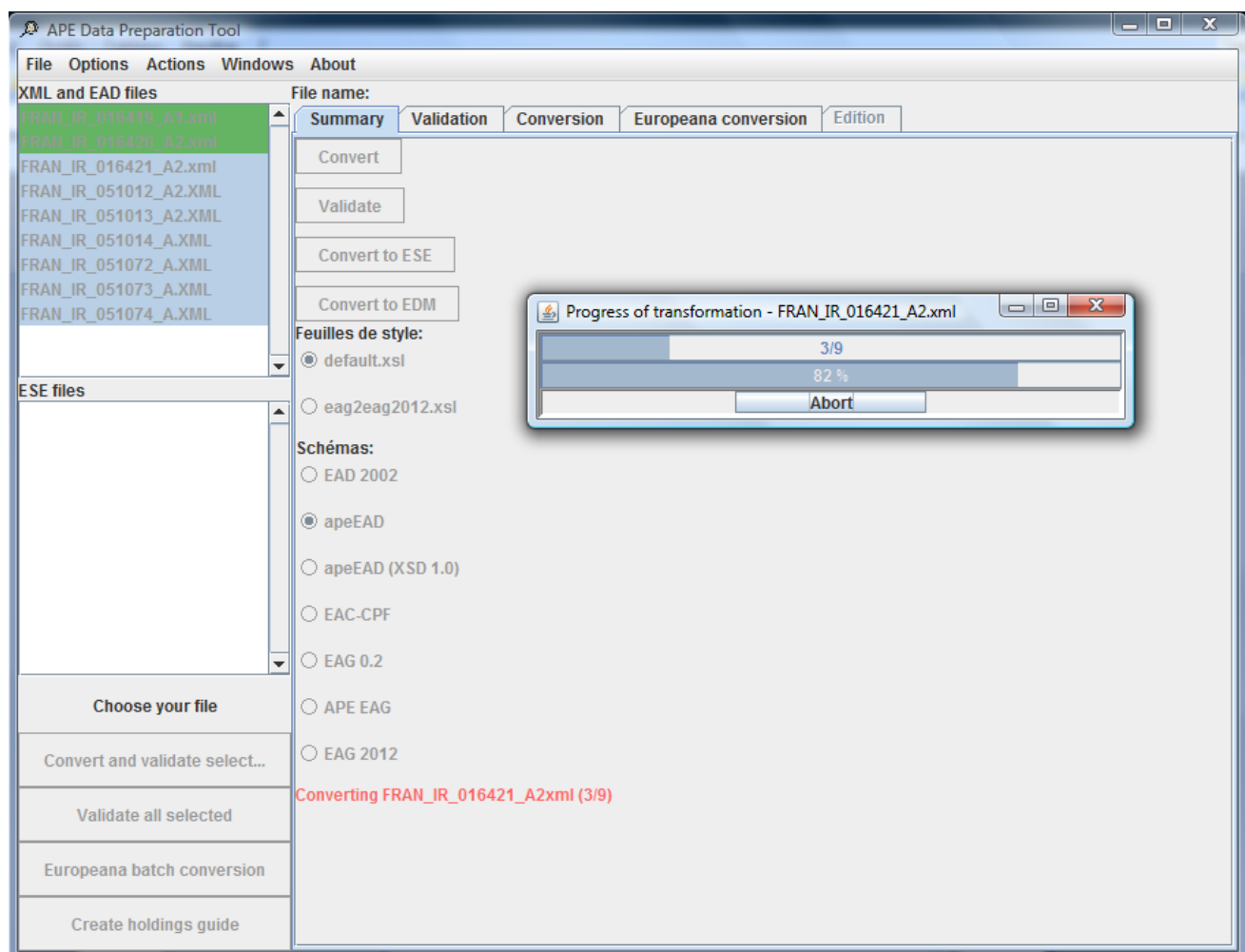


Figure 4: progress bar and information displayed during conversion and validation batch process. In this case, 9 files have been selected. 2 are already converted and validated; the third one is being processed.

The third button on the bottom-left is used to convert the EAD finding aids into the Europeana formats ESE and EDM (see chapter 3).

The last button is used to create a holdings guide from a list of already converted and validated finding aids. It's preferable to save your finding aids before beginning (see chapter 2.1.1)

- Select in the list the files to be integrated in the holdings guide and click on the create holdings guide button.
- A dialogue box opens (see figure 5). On the right side, you have the list of all finding aids to be linked within the holdings guide. You have first to provide some information to describe your holdings guide: identifier (unitid), name (unittitle) and description (scopecontent). This information has to be repeated for the lower levels if you do create lower levels.
- You can create as many levels as needed to organise your holdings guide.
- Once you have created your hierarchy, you simply drag and drop the files in the corresponding levels.
- If you need to rearrange the hierarchy or the order of your files: the up and down buttons on the left enable you to move files or levels inside the same level from one place to the other. To move an entire level to another level, you have to use the drag and drop facility.
- Once all finding aids have been inserted in the holdings guide, click on save.
- Back to the summary tab, validate your file that is now in the list on the left.
- Save it before closing the tool.

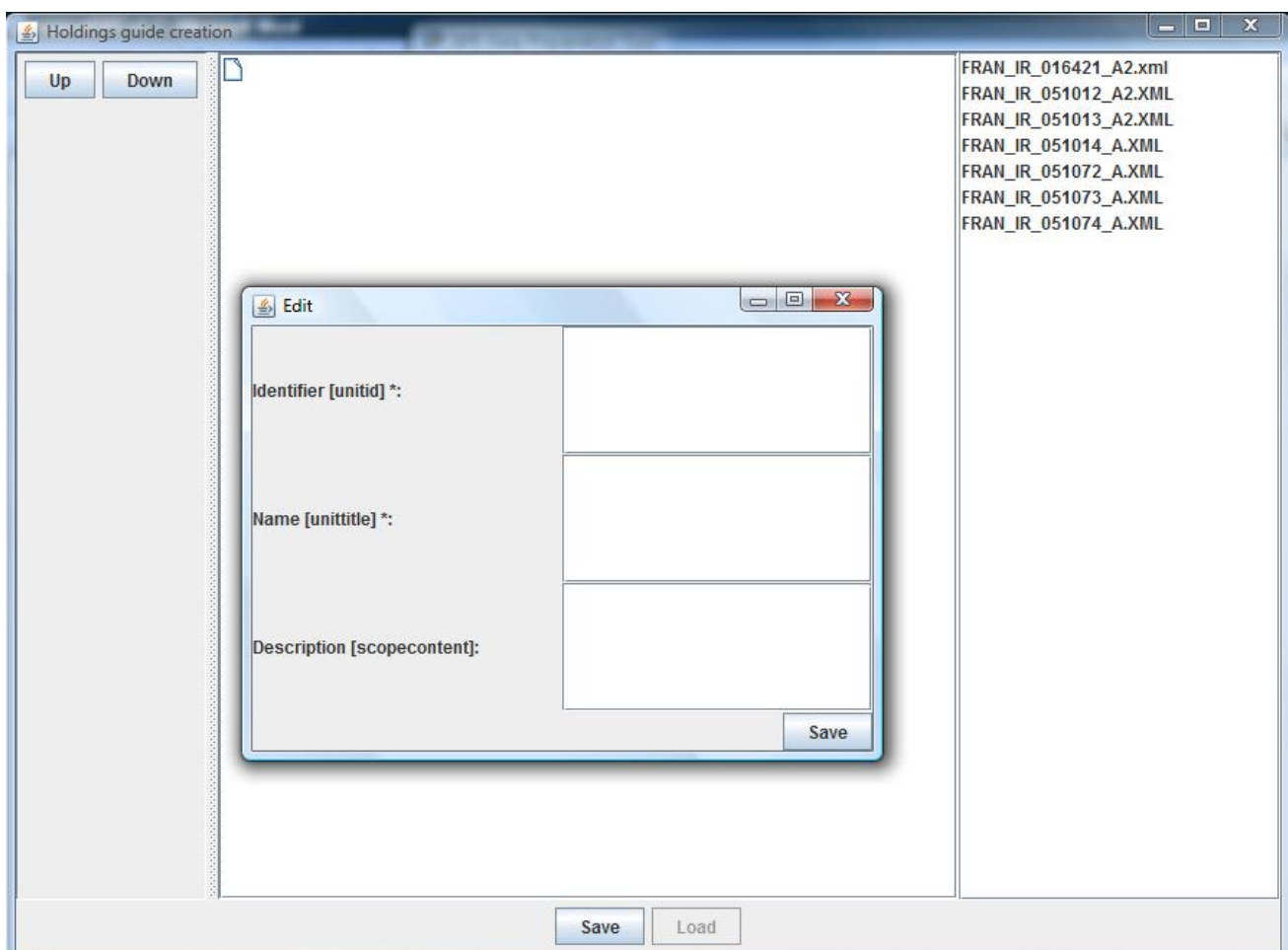


Figure 5: dialogue box to create a holdings guide

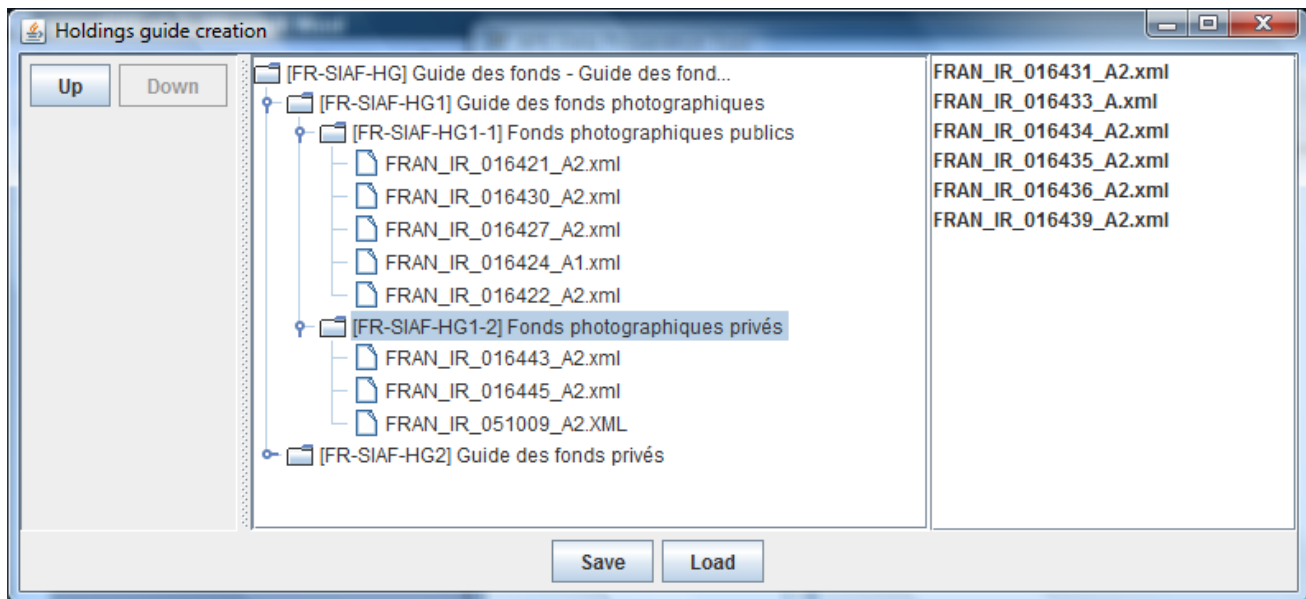


Figure 6: holdings guide being created: the levels have been created and some finding aids inserted in them. On the right, the remaining finding aids to be dragged and dropped in the proper place. On the left, the up button is activated, indicating that the selected level can be placed above the previous level.

2.3 The tabs

You have five tabs: *Summary*, *Validation*, *Conversion*, *Europeana conversion*, *Edition*.

Depending on the process, some of them turn green or red to indicate that a process is completed and that you might check them.

2.3.1 Summary tab

The first and main one is the **Summary**, and is opened by default when you enter the tool. This is where you can apply operations to your XML files. When a file is selected in the list, you can, from this tab, convert the file to the APE format, pass it to the validator or convert it to Europeana format. You have the possibility to choose the style sheet to be used for the conversion – of course only if you installed your own style sheet (see chapter 2.1.2 Options - Use own style sheet). The same goes for the schemas.

The status of the current transformation or validation is displayed beneath those buttons and radio buttons.

2.3.2 Validation tab

The second tab is **Validation** and here you will find the results of a validation for the selected file. It can either contain a sentence saying the validation succeeded or a list of all the errors found during the validation process.

Please note that if you validate a file before the conversion, the lines indicated in the error report will be the ones of your original file, but if you validate a file after the conversion, the lines indicated will be the ones of the new converted file, therefore will not exactly correspond to your original file.

Besides, a report on the data quality or completeness of the file will be displayed:

----- Data quality -----

Missing content for title information (unittitle): 0

Missing normalised date information (unitdate@normal): 0

Missing information on digital object type (dao@xlink:role): 0

This information also can be downloaded: it is saved in the report file to be found in the output folder of the tool.

You can improve the data quality of your files by editing them (Edition tab see chapter 2.3.5), or by adding rules for date conversion and by indicating the digital object type (Options, see chapter 2.1.2)

2.3.3 Conversion tab

The third tab is **Conversion** and here is displayed information on the transformation operated. For example, it will tell you all the elements of your original file that have not been transformed in the resulting apeEAD file.

2.3.4 Europeana conversion tab

The fourth tab is **Europeana conversion** and here is displayed information on the ESE/EDM conversions. Files that are transformed into ESE/EDM formats are automatically saved in the output folder (or an alternatively chosen default folder for saved files, see chapter 2.1.2 Options). Their names stay the same with an additional "-ese" or "-edm" before the file extension. ESE files will also be displayed in the second list of files for further processing to EDM.

2.3.5 Edition tab

The last tab is **Edition** and here you can see the XML tree of your selected document. Some elements only can be edited:

- eadid @mainagencycode
- eadid @countrycode
- eadid / text()
- language / language
- language / language / text()
- langmaterial / language
- langmaterial / language / text()
- unitdate @normal
- titleproper / text()
- c @level

The eadid is mandatory to validate your file.

The elements and attributes related to the language are useful for the display and refinement of the search results.

Same goes for the date: a date is searchable only if normalised.

An empty titleproper will be displayed as "no title".

Please note that after editing your file, you should validate it again and save it.

3 EAD to ESE/EDM conversion

The EAD to ESE/EDM conversion can only happen after a file has been successfully validated against an apeEAD schema. Once the file is valid, the button "Convert to ESE" is available in the summary tab to convert your apeEAD file into ESE.

3.1 ESE conversion

By clicking this button, a new window will open up, where you will have to provide some additional information:

- **Data provider (text field, mandatory):**

This field will either be prefilled with according information found in your apeEAD file for the archival unit containing the digital object (<c> with <dao> / did / repository) or you will be provided with the option to inherit this information from one of the higher hierarchical levels. The default value for inheriting this information is taken from the level describing the whole fonds or collection (<archdesc>). Whatever the situation in your files, you always can change the proposed text manually.

- **Provider (text field, mandatory):**

In addition to the data provider, i.e. your own institution, you can also indicate whether you intend to deliver your data via an aggregator, e.g. the Archives Portal Europe, to Europeana. The name of this aggregator should then be given in the field "Provider". Should you intend to deliver your data yourself, you can also repeat your institution's name here.

- **Type (radio buttons, mandatory):**

This refers to the type of digital objects included in the apeEAD file you want to convert. When you have used the option "Digital object type" already during the conversion of your local data to apeEAD or when your own EAD data already includes the corresponding information (in the attribute @xlink:role for <dao>), you will find the type pre-selected in the form. In case, your apeEAD file does not yet contain this information or the type "UNSPECIFIED" has been included during the conversion to apeEAD, you will be asked to select one of the four possible types (TEXT, IMAGE, VIDEO, SOUND) here.

- **Hierarchy prefix (text field, optional):**

As the Europeana formats, especially ESE, but also the current EDM implementation at the Europeana portal, follow an object-oriented approach, the Europeana conversion as offered by the DPT includes different aspects of inheriting information from higher hierarchical levels to the archival units containing the digital objects. In this case, the titles of all higher levels leading to a specific archival unit are concatenated in one alternative title element. In order to explain this alternative title, a hierarchy prefix can be included (default value: "Context:" or an according term in one of the other languages available in the DPT).

- **Inherit parent / Inherit origination (radio buttons, mandatory):**

Similar to the titles, other information can be inherited from higher levels, e.g. the information about the records creator (name). In case, you would like to use this option, you can switch the radio button from its default "No" to a "Yes".

- **Inherit language (radio buttons with optional drop-down selection, mandatory only for the object type TEXT):**

Although the label for this field states its mandatory application, this only refers to digital objects of the type TEXT. In case, you have chosen one of the other three object types, you could also skip this option and would nonetheless be able to complete the conversion. When the object type TEXT is selected, the conversion will first check whether according information already is available for the archival unit containing the digital object itself. Only if that isn't the case, you will be asked to either inherit the language information from a higher level of description or to select a language manually via the drop-down menu provided.

- **Specify license (radio buttons with additional drop-down selection, mandatory):**

This field is used to specify the license applying to the digital objects themselves. Please note, that in order to provide data to Europeana your institution will have to have signed the Data Exchange Agreement with Europeana. The therein-specified CC0 license however only refers to the metadata, i.e. the apeEAD finding aids, not to the digital objects. Therefore you are asked to provide this additional information here. You can choose between several Creative Commons licenses (including CC0 and the Public Domain Mark) or a Europeana rights statement with the values "Free access", "Paid access", "Restricted access" or "Unknown".

- **Specify additional rights information (text field, optional):**

If wanted, you can also provide some additional rights information, which especially could be of interest in combination with the Europeana rights statement (see above).

Konvertierung von APE-EAD nach ESE

All fields marked with an asterisk (*) must be filled

Data provider:*	Data provider	
Provider:*	Provider	
Type:*	<input checked="" type="radio"/> TEXT <input type="radio"/> IMAGE <input type="radio"/> VIDEO <input type="radio"/> SOUND	
Hierarchy prefix:	Context:	
Inherit parent:*	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Inherit origination:*	<input type="radio"/> Yes <input checked="" type="radio"/> No	
Inherit language:*	<input type="radio"/> Yes <input checked="" type="radio"/> No <input type="radio"/> Provided:	
Specify license:*	<input type="radio"/> Creative Commons <input type="radio"/> Creative Commons CC0 <input type="radio"/> Creative Commons Public Domain Mark <input checked="" type="radio"/> Europeana rights statements	Abkhazian Free access
Specify additional rights information:		
<div>Cancel</div> <div>Create ESE</div>		

Figure 7: Europeana conversion window for some additional information to be provided.

When clicking the button "Create ESE", the conversion process is started. Similar to the actions described above, you will also see a progress bar while the conversion is done. You will automatically be prompted back to the main view of the DPT. Successful conversion to ESE will be signalled by the message "Finished" written in red in the "Summary" tab, by the converted ESE file now being shown in the "ESE files" list on the left and by the colour of the "Europeana conversion" tab changed to green. In the latter you will be told that conversion has succeeded and that your file has been saved automatically.

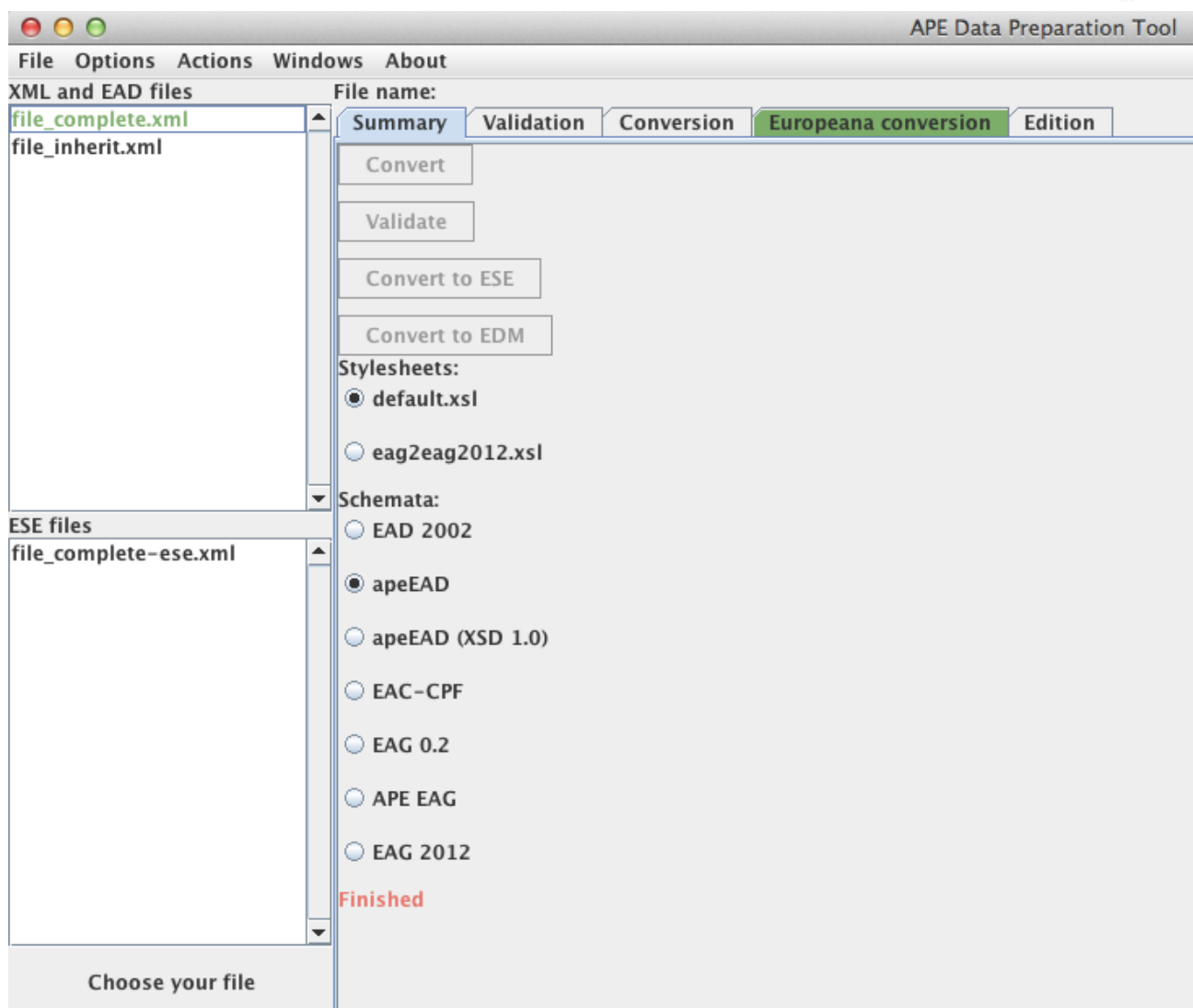


Figure 8: The main page of the DPT after successful conversion to ESE.

3.2 EDM conversion

Selecting a newly converted ESE file in the list will make the "Convert to EDM" button in the "Summary" tab available for further processing. The conversion to EDM will run in the background, i.e. without having to provide additional information and you will be notified about the process being finished by the "Europeana conversion" tab once more changing its colour to green. The EDM file will then be automatically saved.

As for conversion to apeEAD and validation, both steps of the Europeana conversion also can be done in batch processing.

For details on the mapping, please see the annexes of the "Accompanying report for D2.1 – First version apeEAD to EDM conversion tool (central / local)".

4 Update of the tool

A semi-automatic update is available in the tool. It is only semi-automatic because once detected, it requires the user to download the new version and install it. Please note that you need to be connected to the Internet for this.

At the launch of the tool, the tool will check on the Internet if a new version is available. This check is done at a minimal interval of 2 weeks. If it does find a new version, then the tool proposes to go to the Internet page in order to download the new version.

Please note that downloading the tool will also get the last version of the default style sheet where additions might have been made.

5 Troubleshooting

If problems were to be found, please send us an email via

<http://www.archivesportaleurope.net/contact> and select “Technical issues” in the proposed subjects.

If the tool contains bugs or you suspect it does, please do the following before sending an email:

- Go to the output directory of the tool and retrieve the file named errors.log,
- Attach that file report to the email you will send with some explanations on the error that occurred.