

IMDI editor

User guide

Version 3.2

Last updated: 2008-04-15

IMDI editor: User guide

User guide

N. Wood (nick.wood@mpi.nl [<mailto:nick.wood@mpi.nl>])

Version 3.2

Last updated: 2008-04-15

Table of Contents

Introduction	iv
1. Overview	5
2. Navigating and use of the IMDI Editor	6
3. Generating an IMDI file	9
3.1. Project	10
3.2. Content	11
3.3. Actors	13
3.4. Resources	17
3.5. References	21
4. Repository panel	22
Glossary	24

Introduction

This is a basic user guide for making IMDI files using the IMDI Editor, and a brief overview of creating metadata; it is not in any way to be used as a reference manual. For the detailed IMDI Editor Reference manual please download it from here: <http://www.mpi.nl/corpus/manuals/manual-imdi-editor.pdf>.

The IMDI Editor has somehow got the reputation for being very user-unfriendly. This reputation probably came from users struggling with the very detailed reference manual, as opposed to the application itself being difficult to use - hence this basic user guide.

The difference between the existing reference and this user guide is that this guide is to get you started and talk you through the more commonly used fields you are most likely going to want to populate with data. It is also designed to be read right through. Once you are more used to the IMDI editor, the IMDI reference manual might be of more use to you as by then you will have become a more advanced user.

The editor is a very commonsense application, so you will only be limited by your own imagination. Once you have played with it and made some IMDI templates and IMDI files you will realize its bad reputation is totally unfounded and you will immediately want to furnish your corpus with layers upon layers of rich metadata.

Chapter 1. Overview

The IMDI (ISLE Metadata Initiative) Editor is used to create metadata for resource files that are going to be uploaded into the archive. The IMDI file is the metadata file that the resource files will link to. The IMDI file with the joint resource files is referred to as the session resource bundle. The best way to imagine this is as a vessel that contains all the media, text, annotation and lexical data, plus the metadata describing these resource files. This is why in the IMDI browser the IMDI file is represented graphically as a bag.



Note

A brief note on the use of the term session: You will notice, when talking with a corpus manager, or the Digi-Team, conflicting use of the term session. So, to clarify: to the Digi-Team session is the section of media you want cut from your DMF that you want to describe with metadata and upload to the archive. To a corpus manager session can be one of two things: the IMDI file and its attached media and resource files, a session resource bundle to give it its full title, or just another term for an IMDI file that has been archived. The two parties have tried to come to some consistent non-conflicting language, but old habits take time to change, so please be patient.

Regarding metadata, this really is not an area where you want to think of cutting corners. If you approach making your IMDI files just with a view to getting your media viewable in the IMDI browser, you will end up with a potentially dull corpus that will not reflect the effort you made in creating this body of work in the first place. Yes, you will need to supply a minimal metadata description to actually get media digitized in the first place, but you will be required to embellish this minimal amount to archive the resource. Imagine listening to a record with no information other than the artist and the title of the track you are currently listening to. Now imagine listening to the same record, but having the album cover in your hands with all the information regarding the recording on it, plus literature relating to the artist, supporting players, technical roles, reviews of the recording, liner notes relating to the motivation behind the album and links to outside information. You have to admit that the latter would add to a much richer listening experience. Which way would you prefer the future public research community to see your corpora?

Lastly, regarding metadata in the archive, remember that all information in your IMDI file is open; this data can't be blocked from Internet searches. So, bear this in mind if you are going to use people's names, addresses/email addresses, etc. This is particularly relevant regarding the Actors' data.

Chapter 2. Navigating and use of the IMDI Editor

The editor cannot be run on-line, but has to be installed. As long as you have Java on your machine it will run. You initially download the IMDI Editor from the LAT Tools Page (<http://www.lat-mpi.eu/tools/>) and select IMDI version 3.2. This will now install the editor, and from this point out you open it using your local shortcut. At the moment it is best to use version 3.2 and not 3.3. Version 3.3 is still in its untested beta stage.

See Figure 2.1 for the main view of the IMDI Editor.

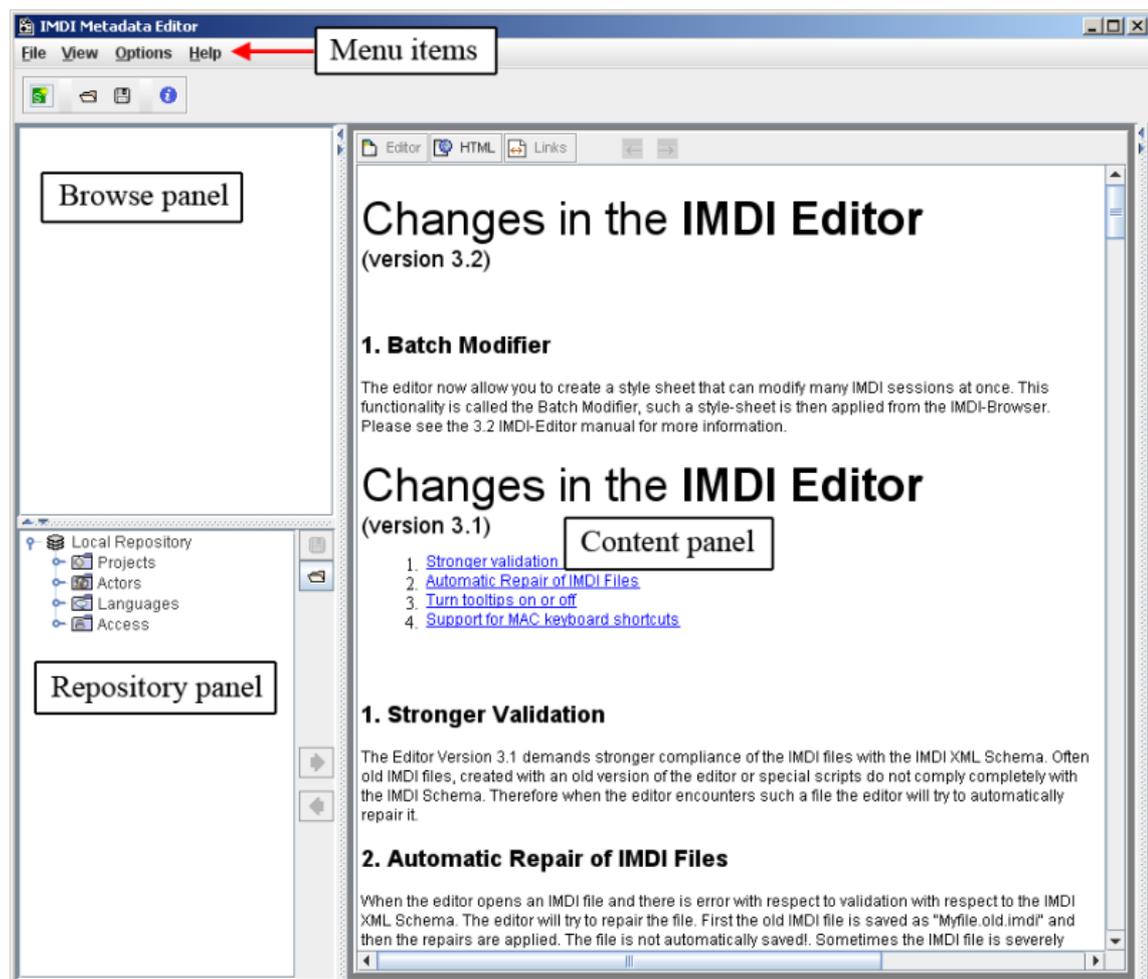


Figure 2.1. Main view

Menu items:	this is where you will find all the options of the application in drop-down lists
Browse panel:	this is where you get an overview of all the IMDI files you have open in a tree structure view. This is also where you can navigate through the various elements of the IMDI file
Repository panel:	this is where you keep stubs of recurring information
Content panel:	this is the graphical view of the IMDI file and the workspace where you populate the IMDI file

Nodes

In the Browse Panel the IMDI file is represented in a tree structure. Each node in the tree represents either an IMDI metadata element or a group of IMDI elements referring to the same issue.

	This is the tree top node that represents the IMDI file
	This is a node that represents a metadata element. When you select this type of node in the tree structure, in the Content panel, the editor will open the page which refers to the selected metadata element.
	This is a node that represents a group of metadata elements. Double click the node or click the blue switch to open it.

In Figure 2.2, in the Browse panel the Actors element group and in that group the Actor John Doe is selected (A). In the Content Panel this results in a view of all metadata elements referring to Actor John Doe (B). In the Content Panel you can browse through the metadata elements and sub elements by clicking the tabs (C). It is in the tabs that you can add values to the metadata fields.

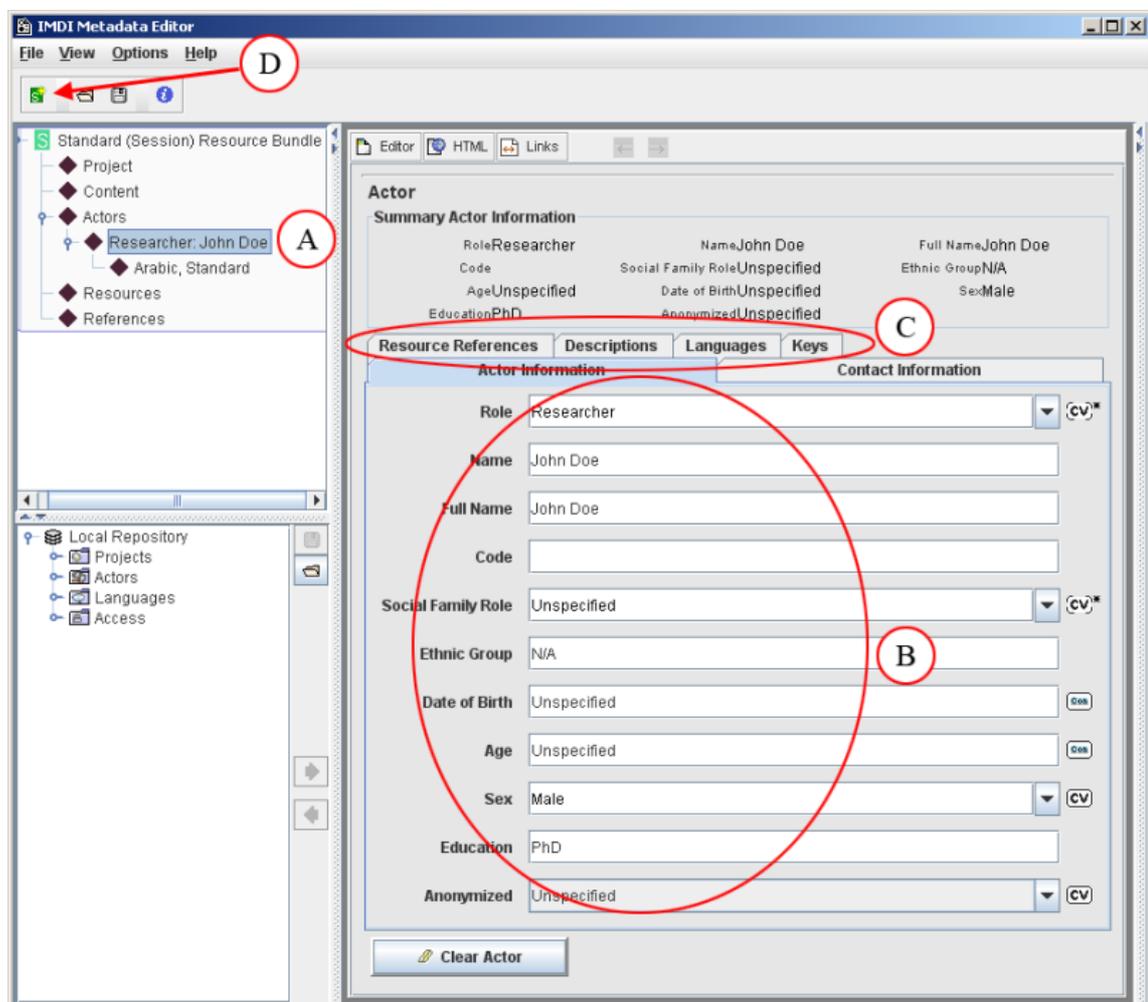


Figure 2.2. IMDI Session file

Tree view representation of the IMDI file in Browse Panel and Tabbed view of the selected IMDI element in the Content Panel

Some of the fields you need to populate have small icons to the right of them, indicating the use of controlled and semi-controlled vocabularies (see Table 2.1)

Table 2.1. Icon set indication restrictions to metadata values

	Controlled Vocabulary (solid outline): To populate this field you may only select from the drop-down options associated with the field
	Controlled Vocabulary (broken outline): To populate this field you can select from the drop-down options, or if there isn't a suitable choice you may add free-text, though if you can find a relevant choice from the drop-down options this is preferred
	Controlled Vocabulary (broken outline and asterisk): As with the controlled vocabulary with the broken line, but on a field with an asterisk you can add multiple values separated by a comma
	Constraint Description: This is normally associated with fields that contain either a date or a time code and need to be written in a certain format
Pop-ups:	If you are ever unsure of what data to populate a field with, or the format the data should take, then rest your mouse pointer on the bold text adjacent to the field and a pop-up will appear with the details of that field. This is the case for most bold titles and workspace areas throughout the IMDI file. These pop-ups can also be as annoying as they are useful though, so if at any point whilst making an IMDI file you are not using your mouse it is recommended to move the pointer off the IMDI file to prevent these pop-ups.



Note

A brief note about populating an IMDI file: The IMDI file is structured in groups of metadata elements and sub elements and in some cases sub-sub elements. Do not generalize data across metadata groups. For example, if you are populating the Descriptions field of the Language node for Actor John Doe in the earlier example of sub nodes, you would only be describing the language of that actor, and no other language related information at all.

Chapter 3. Generating an IMDI file

From the File drop-down menu select New Session (Figure 2.1), or click the New Session icon (Figure 2.2 D). This opens the main page of the IMDI file; in the Browse Panel the IMDI node is highlighted (Figure 3.1).

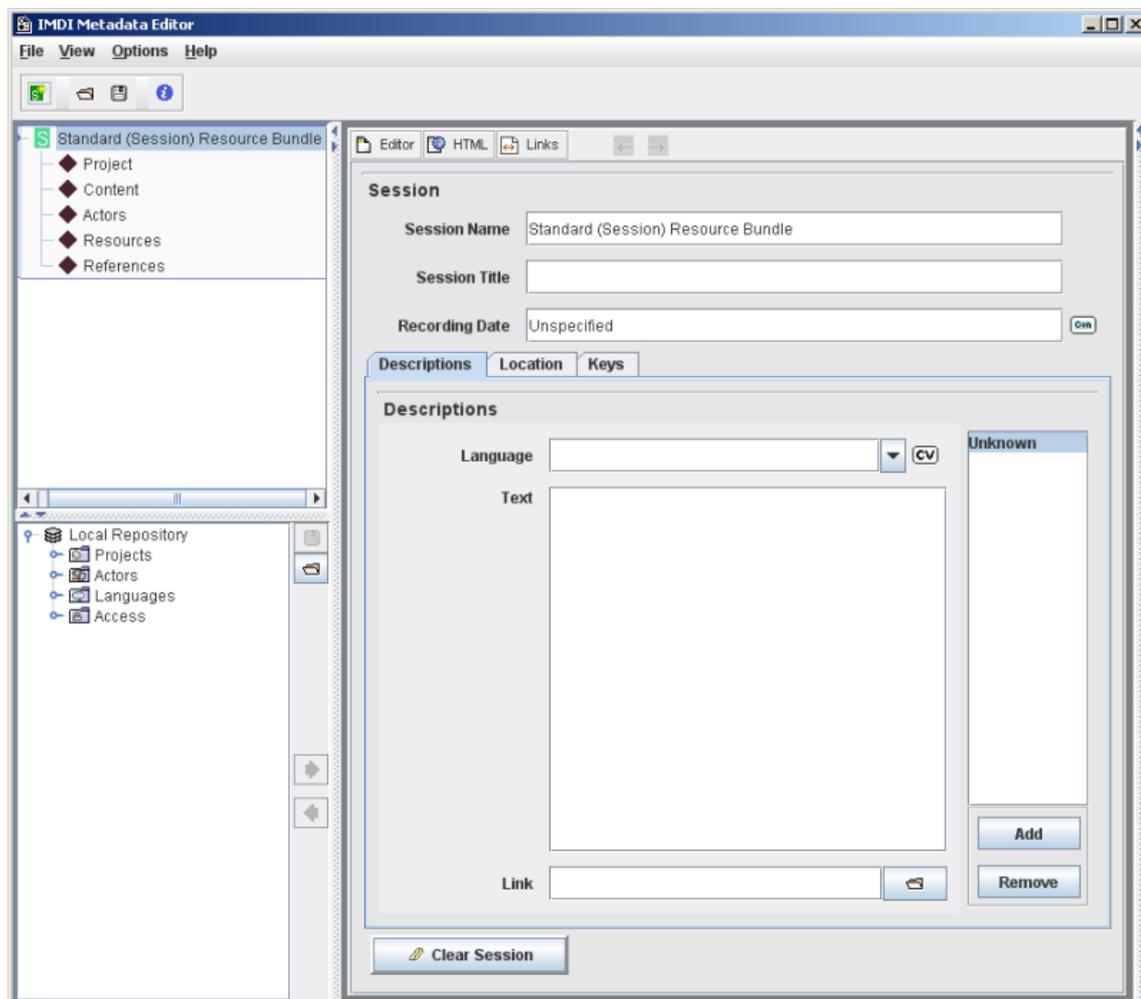


Figure 3.1. Blank IMDI, main page

The information in this main section is going to be about the data contained within this IMDI file, rather like an abstract of the file.

In the Control Panel there are two sections, Session in the top half and the tabbed sections in the lower half, Descriptions, Location and Keys.

Session contains three fields, Session Name, Session Title and Recording Date:

Session Name:	This is the name you give your session that will show in the IMDI browser and will also be the IMDI filename, unless you 'Save as' and specify otherwise. No spaces or symbols can be used in this field, just the alphanumeric characters and the underscore and the hyphen. Try to restrict the number of characters to twenty.
Session Title:	This is for a more detailed description of the Session Name and is not limited in filename size and what characters you can use.

Recording Date:	This is the date the recording associated with this IMDI file was made. If the recording(s) cover more than one calendar date then please use the first date and mention in the Descriptions section the true date duration of the recording(s).
-----------------	--

Descriptions

The Descriptions element has three fields, Language, Text and Link (see Figure 3.1).

Language:	This field is for you to list the language you are going to use in the Descriptions element. It is a controlled vocabulary field where you can only select from the drop-down list. Once you have selected the relevant language it appears in the column to the right, where initially 'Unknown' is indicated. If you want to add a Description in one language click the Add button in the lower right corner and you get 'unspecified' in the Language column, then select from the drop-down the desired language. If you have multiple languages here the editor only allows you to view one language's text at a time. Highlight the languages in the right column to switch between them.
Text:	This field is a free-text field to write a description of the metadata and an outline of the session.
Link:	Link fields are for you to put in urls to external information. As these are not commonly used fields they are not covered in this User Guide; please refer to the IMDI Reference Manual.



Note

A brief note on the Descriptions tab: Every node and sub-node will have a Descriptions tab; the process for completing them is identical for every one, so this instruction is relevant for all the other descriptions tabs throughout the IMDI file.

Location

The Location element is for you to signify the location of where you recorded the material for this IMDI file. There are two main sections in this tab: Location and Region.

Location:	Has three fields: Continent, Country and Address, which are all very self-explanatory.
Region:	This is a free-text field for you to specify in more detail the areas the data was gathered. Click the Add button and a text box opens with "Unspecified" highlighted in it, over-write this with your region name. Add as many regions as you need, but keep one region per line.
Keys:	The Keys fields run right through the IMDI file and are for you to add extra metadata elements, if you feel the need to describe your resource with more, or deviating, elements than are already available in the IMDI file.

3.1. Project

This element is for data regarding the project for which you are doing the data collection. There are three main sections: Project, Contact Information and Description (Figure 3.2).

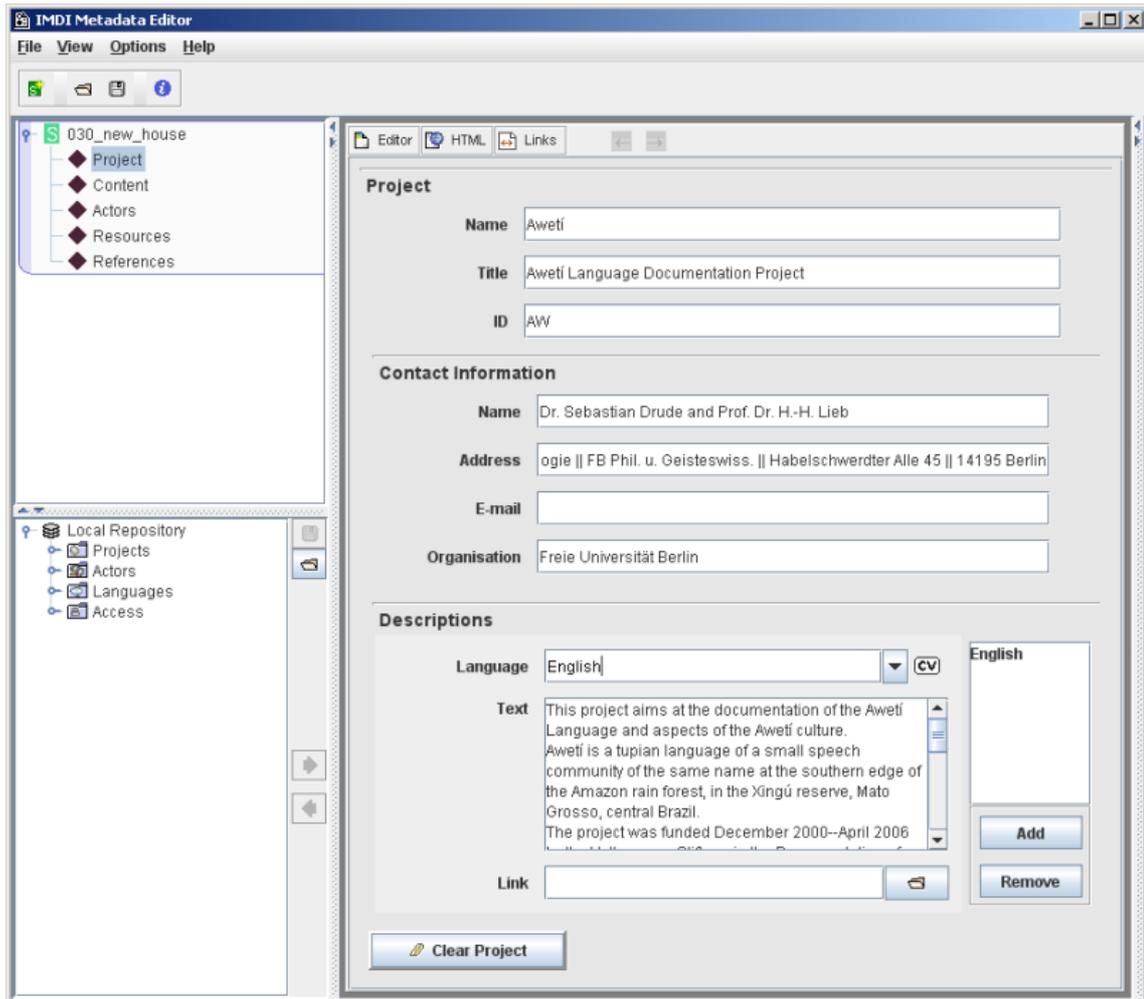


Figure 3.2. Project

Project:	Name, Title and ID. As with the IMDI node, Name is the name of the project and Title is a more descriptive title, for example, Name: Awetí, Title: Awetí Language Documentation Project. ID's for projects are often given by the financing organization.
Contact Information:	Here is where the details of the main contact for the project are entered. Even though this field is not compulsory, it is very important that you not only populate this field, but also update it should the listed contact be superseded. A project that doesn't have a valid contact is as good as redundant as far as the archive is concerned. Also, as was mentioned in the Overview, remember all information in the IMDI file is open, so check with the contact before you start putting in their address and email address.
Descriptions:	The Text field in the Descriptions element of Figure 3.2 gives a good example.

3.2. Content

This element is where you describe the content of the resource, or the experiment set-up (Figure 3.3). There are four sub elements in the Content group: Content Type, Descriptions, Languages and Keys. You will find them in the tabs in the Content Panel (A)

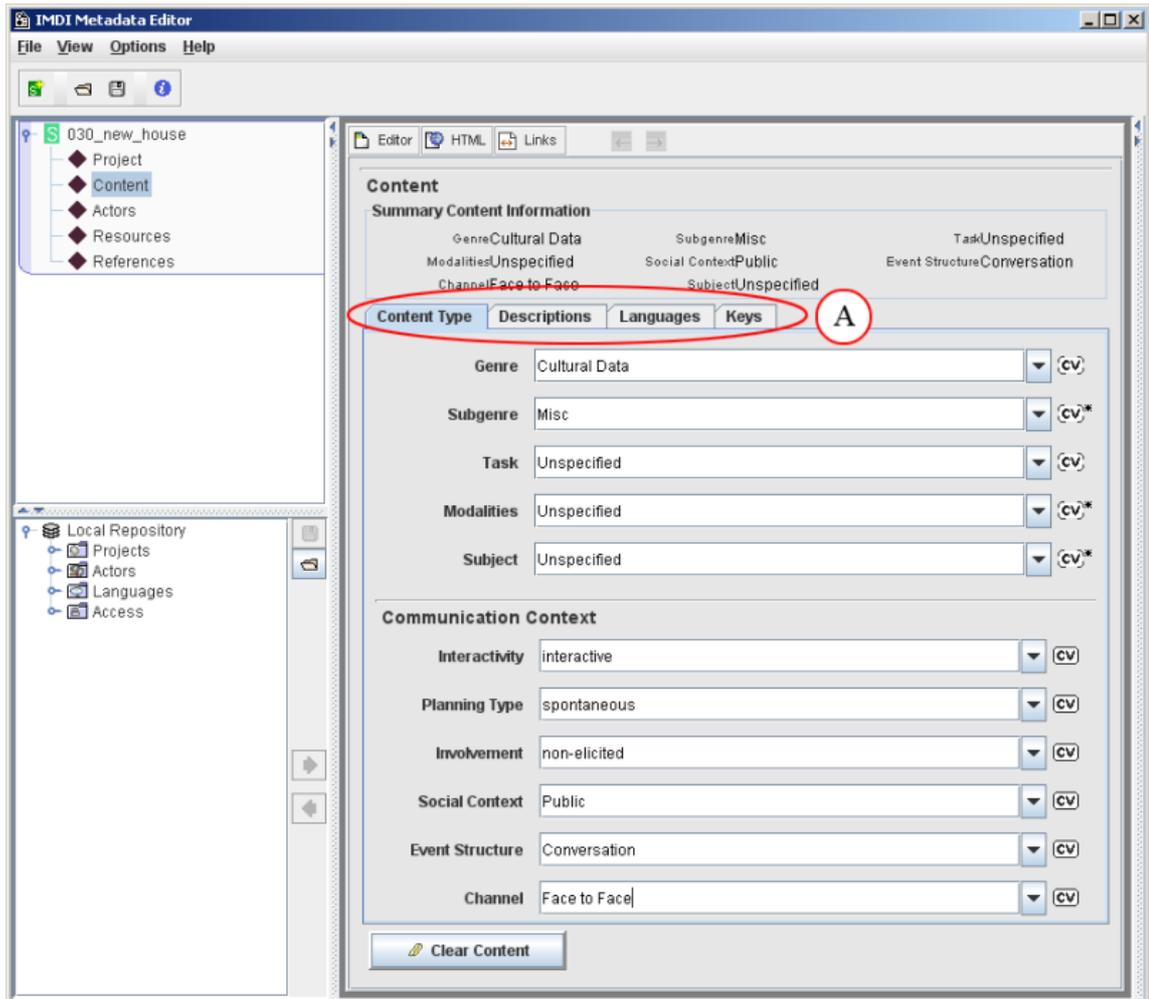


Figure 3.3. Content

Content Type:	There are two halves to this tab, the top half being where you describe the genre and tasks used, the lower being where you describe the communication context in which the data collection took place.
Languages:	This is where you list all the languages used in this session. When you access the Languages tab you will first need to click Add to get the Language element. This element screen consists of two areas: Language in the top half, at the tabbed lower half: Reference Resources and Descriptions (see Figure 3.4).

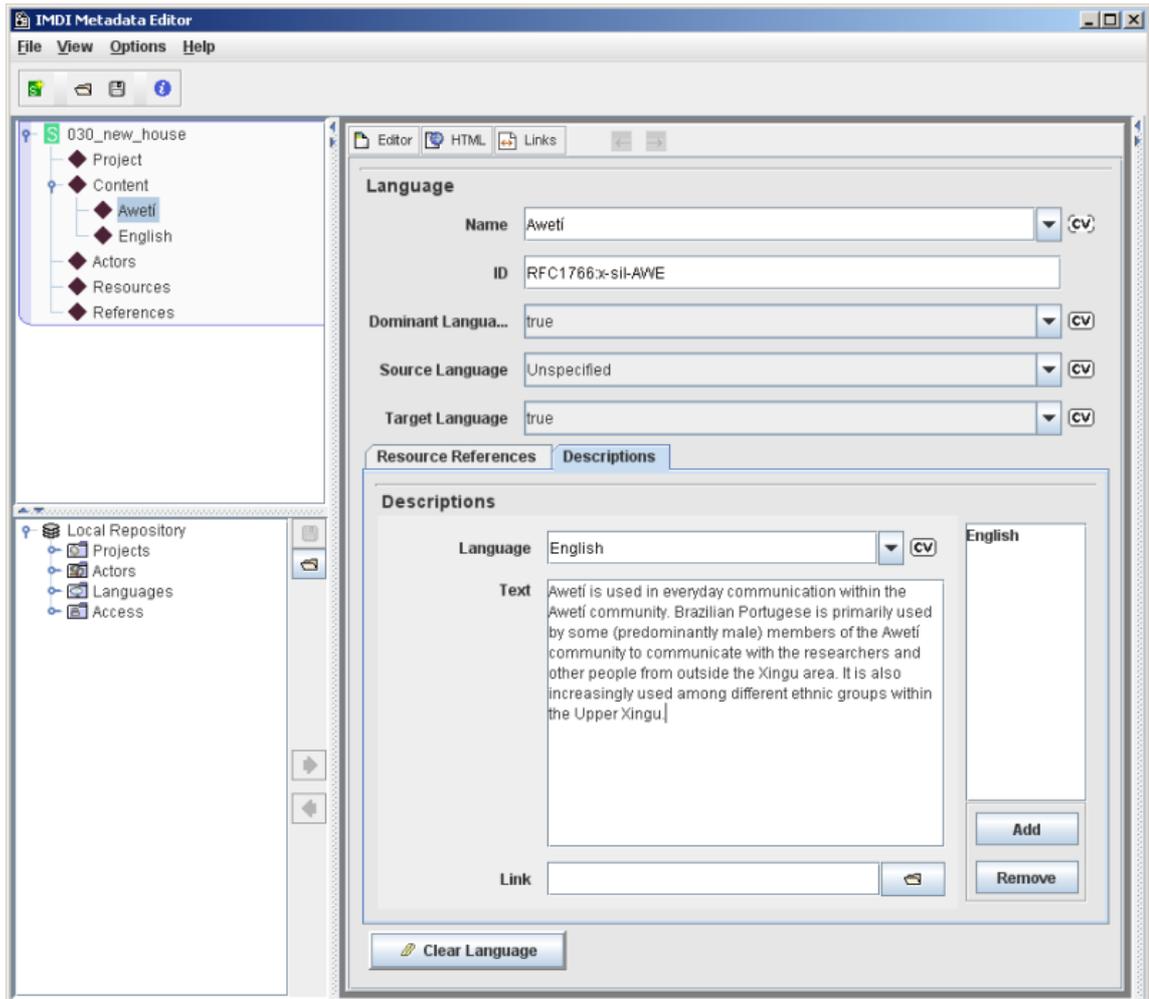


Figure 3.4. Content / Language / Description

Language

The following fields are simple true/false choices. See Figure 3.4 for this element populated.

Dominant Language:	This is for you to specify whether this language is the dominant, i.e. most frequent, one used in the session.
Source Language:	This is for you to specify whether this language is L1 for most actors in the session or not.
Target Language:	This is for you to specify whether this language is the language under study.

3.3. Actors

The Actors element is where you record information on the subjects that were part of this session. For each Actor the first thing you need to do is add the actor. Click the **Add** button in the upper right corner of the Control Panel. This will open the Actor Element (Figure 3.5). The Actor element is the most complex and detailed one you have to complete. It is also the one where you have to exercise the most caution about the information you include.

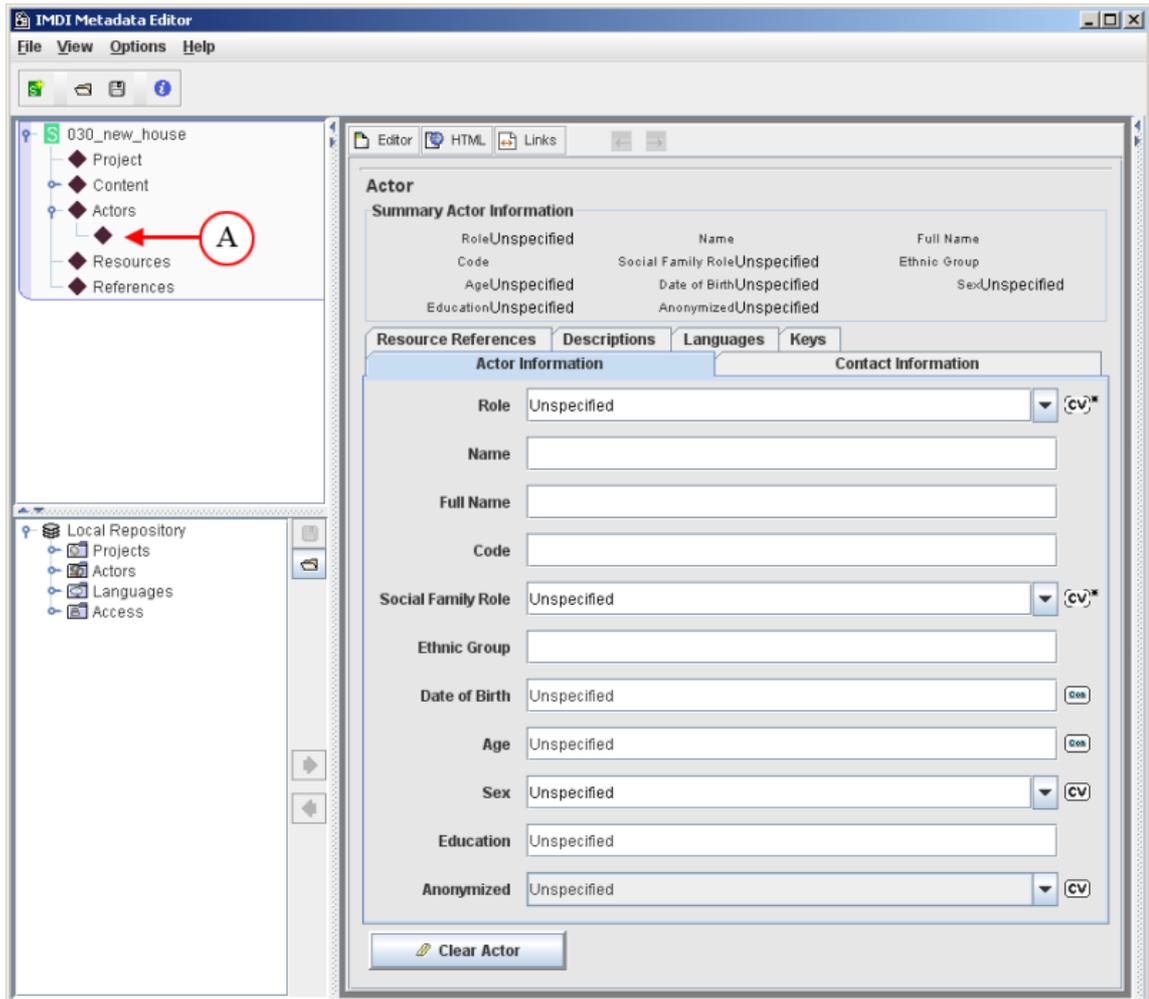


Figure 3.5. Actor information

After adding an Actor, you will notice that in the Browse Panel you have a new node under the Actors node (Figure 3.5 A). This is just for one actor, and the view in the Content Panel relates only to this individual actor.

In the Content Panel there are six sub elements: Actor Information, Contact Information and Descriptions, Keys, Languages and Resource References. Please refer to the Reference Manual for information regarding the Keys and Resource References.

Actor Information:	This is where you describe the role of the actor. Bearing in mind that actor is a generic term for anyone involved in the session, so it also includes researchers, recorders, teachers, students, etc.
Contact Information:	This is the contact information for the actor, not necessarily actor's details, but those of the contact.
Languages:	This is where you list all the languages the actor uses. As this information is regarding the actor generically, and not just unique to this session, then add all the languages, and not just the ones used in this recording.

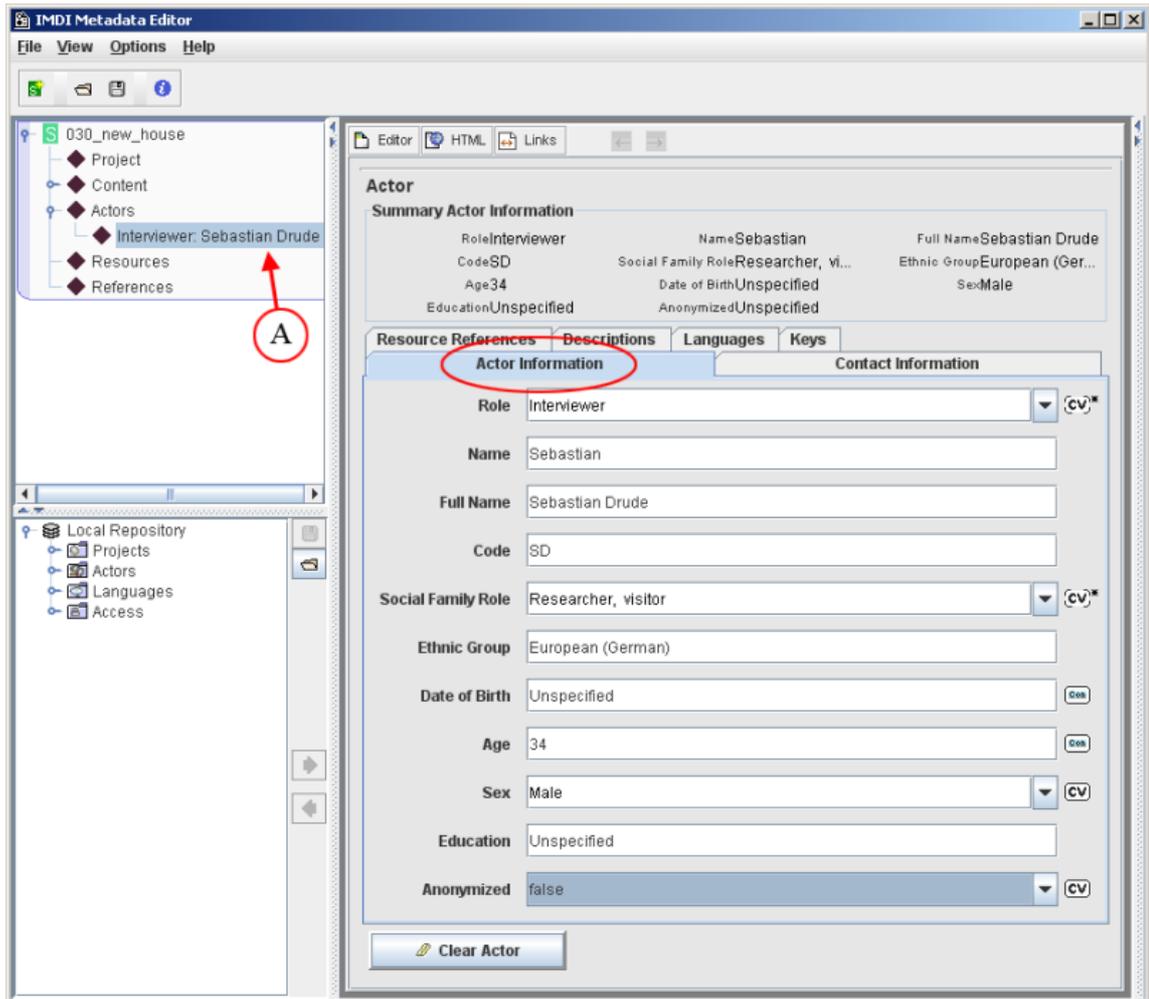


Figure 3.6. Actors / Actor information populated

When you populate the Actor Information element the new node in the Browse Panel now shows the Role and Full Name information (See Figure 3.6 A).

Once you have your actor information you will need to add their language(s) (Figure 3.7). There are two sub elements to Languages: Languages and Language Descriptions. Languages is where you list the languages of the Actor. Since one Actor can have more than one Language, you need to Add a language first before you can start populating the Language elements with data.

You can select the Name of the language from the controlled vocabulary. When you do so, the ID field will be automatically filled. When the Actor Language is not in the cv list, than you may fill the Language Name yourself. Is this case the ID field will not automatically be filled.

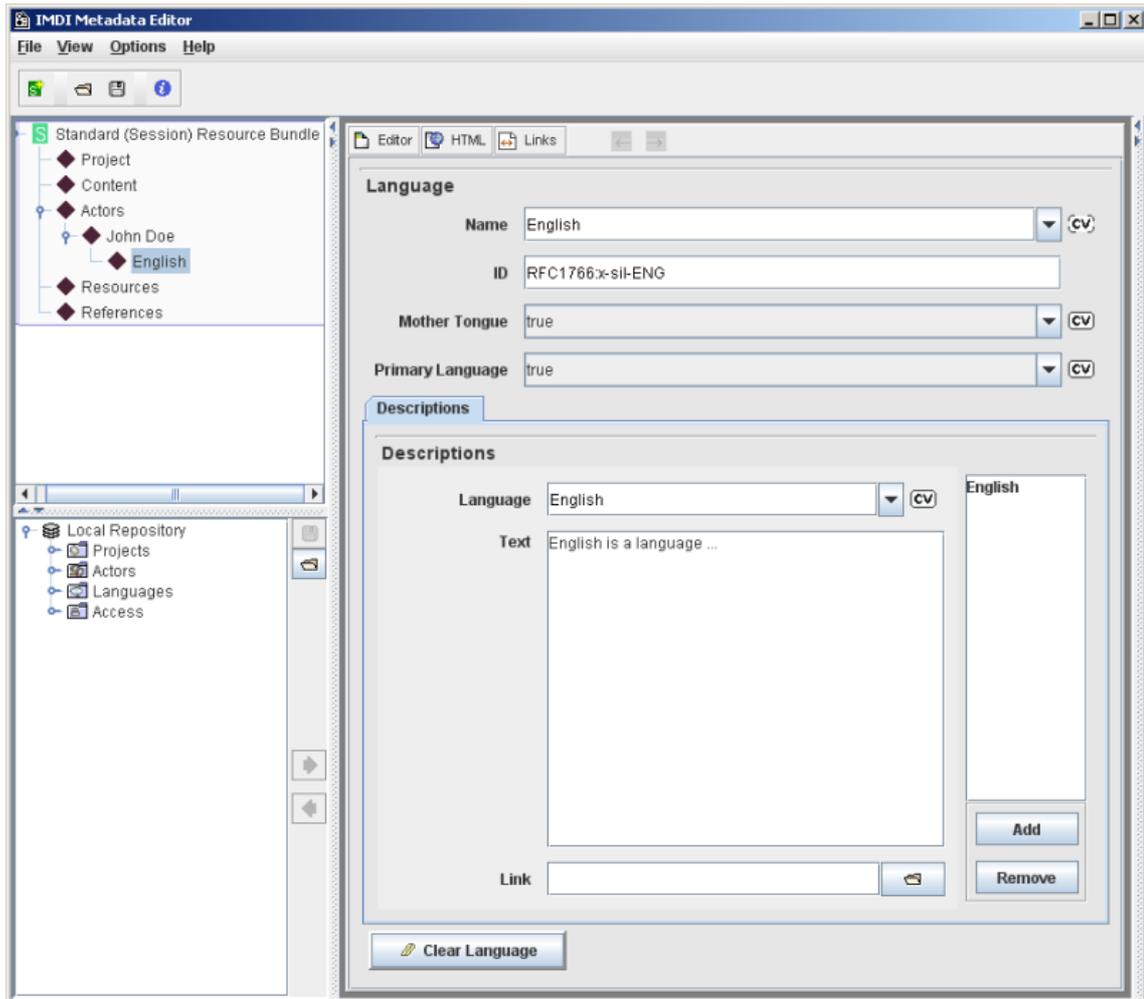


Figure 3.7. Actor / Languages / Language

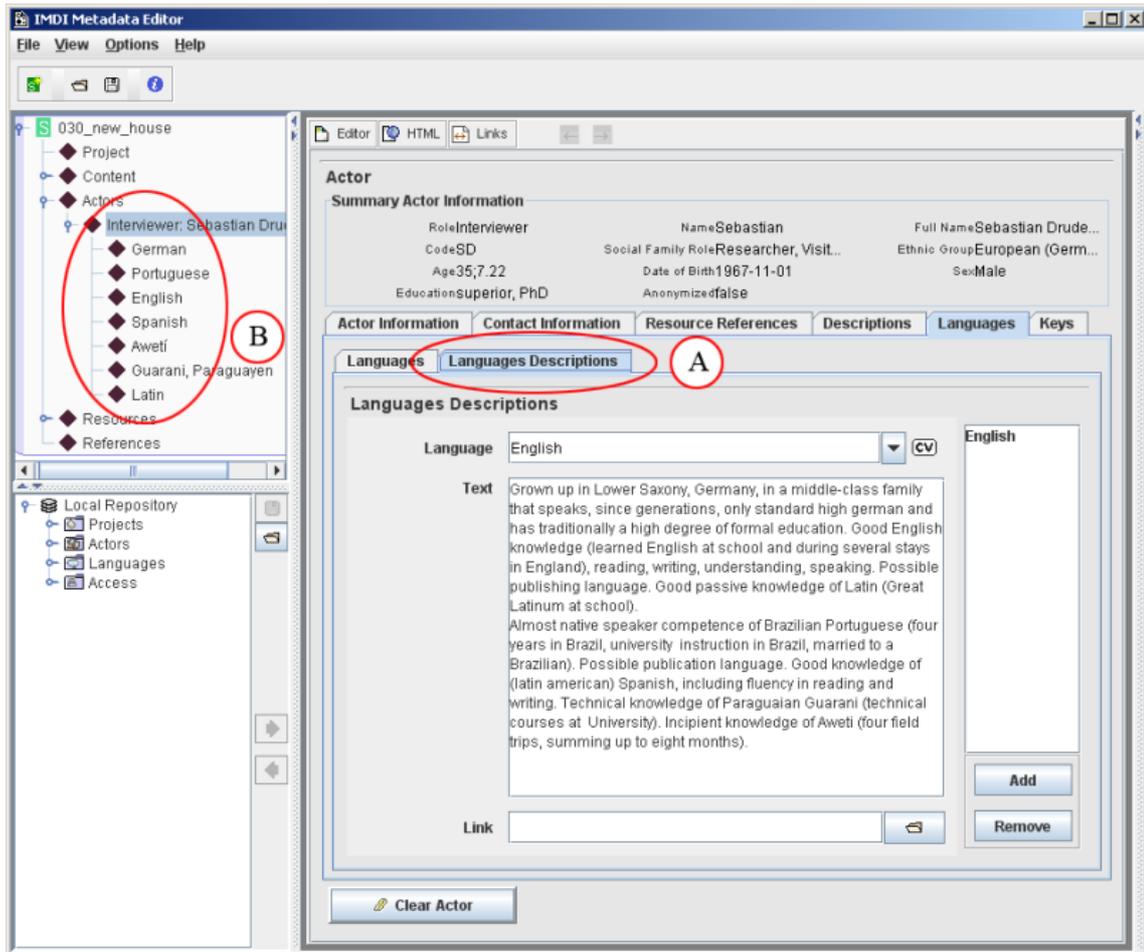


Figure 3.8. Actors / Languages / Language Descriptions

Figure 3.8 shows an actor's completed Language element viewed from their Language Descriptions tab (A). Remember, this Language Description describes the actor's languages as a whole.

To add another actor go back to the main Actors element and go through the same process.



Note

A brief note regarding Actor information: As has previously been mentioned, all archived metadata is open to web searches. Since the Actors node contains fields where you may name your subjects, you need to be 100% sure that it is OK for their names to be public. Otherwise you should leave these fields blank or encoded. Please be aware that the Code and Anonymized options do not currently protect the identity of the actor.

3.4. Resources

The Resources element is for information regarding resources (media files and annotations); it is also where you put information regarding the source file of the resources, for example reference to a DMF. The source element is used to supply information for the Digi-Team to make the media resource files. This user guide will only focus on the Source element, as that is where you put this information. To open the Source element click on the Sources tab and click on Add to open the metadata fields (Figure 3.9). When you add a source, as with the Actors element, you will initially have a new Unspecified node in the Browse Panel. By populating the Source ID and Format fields the new unspecified node now uses that information (B).

The view of the Content Panel for this new source is divided in two sections, the Source section in the top half and the tabbed section in the lower half. From the tabbed section we are only going to cover the Position and Descriptions element (Figure 3.9).

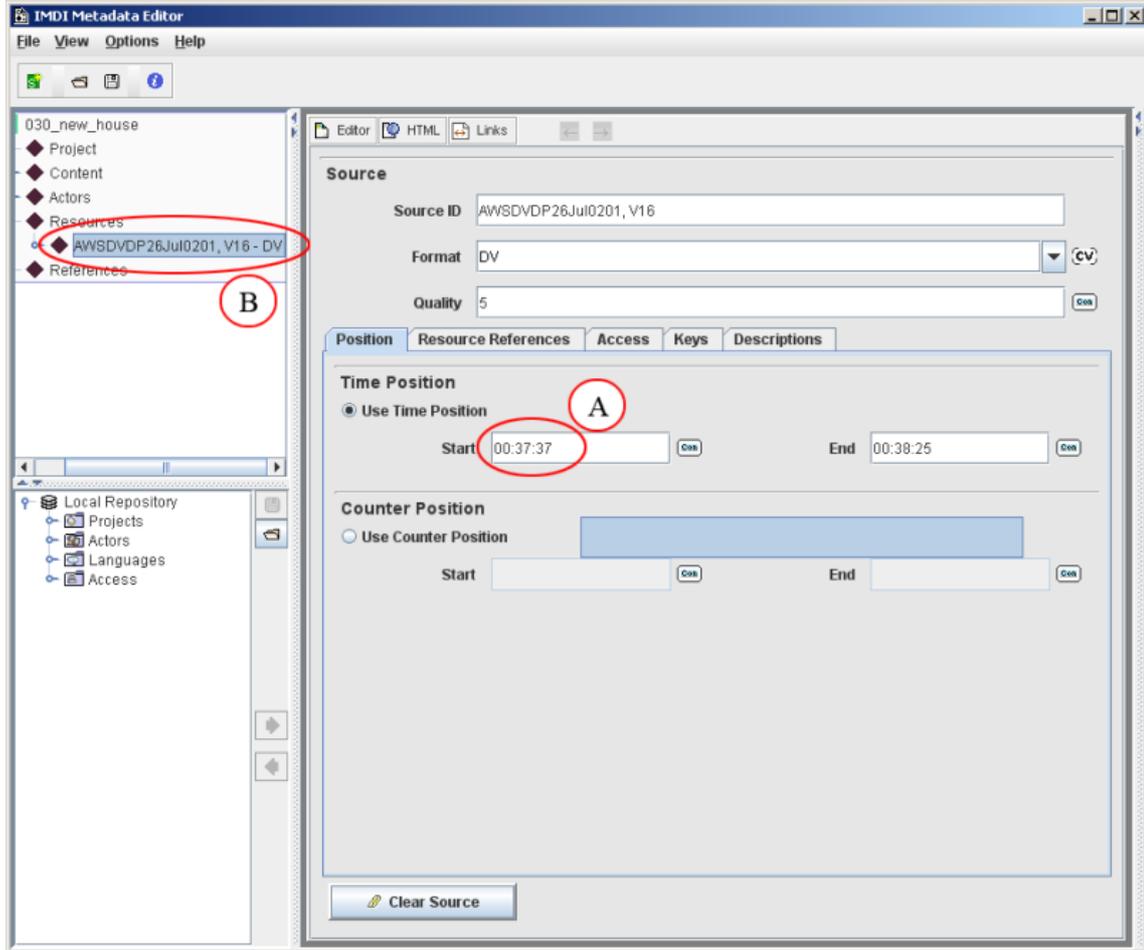


Figure 3.9. Adding source information

Source

Source ID:	This is the label of your source recording, which should be the tape label you gave the recording that was also used as the DMF label.
Format:	This is the format of the media you used to make the recording. There are a lot of cryptic initials and format codes in this drop-down selection, but if you place your mouse pointer over one you will get a pop-up giving a more detailed description of the format. Even though this field is unlocked controlled vocabulary, the selection from the drop-down list covers most formats, so free-text should not really be necessary.
Quality:	A very simple 1 to 5 grade, 1 being poor and 5 being flawless.

Position

The Position element is for you to enter the start and end time code information for the segment you want cut from the DMF to obtain the session resource media file.

There are two sections to this tab, Time Position and Counter Position. You use Time Position for this time code information.

<p>Time Position:</p>	<p>This is where you enter the start and end times for the clip you want. The preferred time code format is HH:MM:SS:mmm (Hours:Minutes:Secods:milliseconds), but you can also use HH:MM:SS:FF (Hours:Minutes:Seconds:Frames) (Figure 3.10) if your software doesn't work with milliseconds. Even if your time code information doesn't go to the detail of milliseconds or frames, you still need to recognize these segments of the time code, so use zeros. Milliseconds use three digits, frames use two. If you intend to use the whole DMF as the media session file you still need to populate the start and end times. Here you would use 00:00:00:000 in the start field and would have to calculate the end time using ELAN, or another program that uses detail greater then just seconds. Please do not use basic media players such as Windows Media Player and QuickTime to obtain time code information.</p>
<p>Counter Position:</p>	<p>Do not use this field for media resource file time code information. This field is used for alternate sources, such as page numbers, or non-time code sources in general.</p>

If your complete session involved more than one recording, for example two tapes were required to collect all the information, then you highlight the Resources node and add a new source. This is particularly useful if your DMF file is multiple recordings and you want different sections made into one recording. If you do use multiple files that need to be merged you must clearly note in the Resources Description tab the order the segments have to go in. Figure 3.10 shows a multi DMF Resource element viewed with the Resource node highlighted.

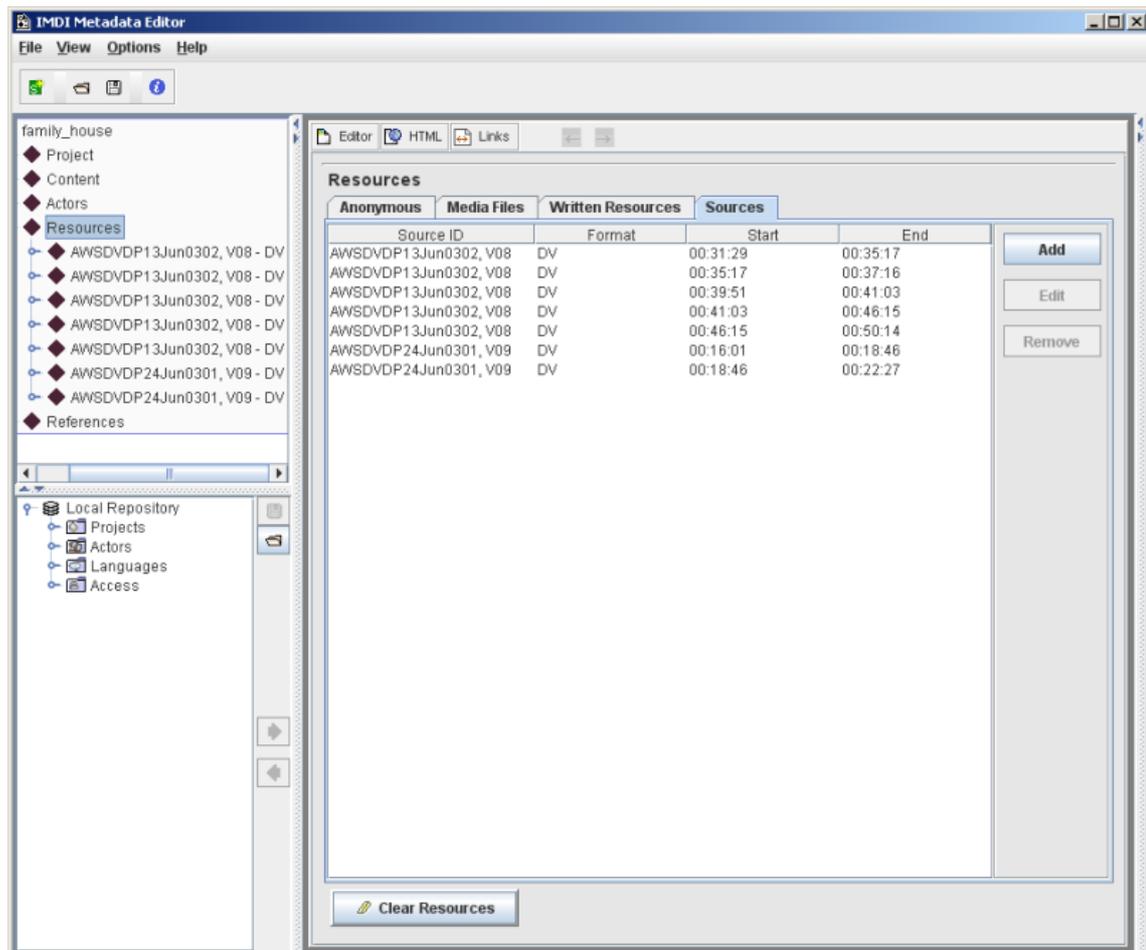


Figure 3.10. Multiple sources

You can also add multiple sources to make multiple media resource files (Figure 3.11). This is quite a complicated recording in that you could find in a gesture lab set-up. In this scenario you have three separate camera angles (Figure 3.11 A), a split-screen of the three separate angles (Figure 3.11 B) and an audio back up recording (Figure 3.11 C). On top of that, there is two of each recording that need to be merged.

In this scenario you would need to use the Details elements of each source to supply information on how you would like this session to be made. Also, as an IMDI file only uses one name to make all the output media resource files, you would also need to supply information on how you need the files named (Figure 3.12).

For extra detail on completing the Resource elements, please contact the Digi-Team (dobesarc@mpi.nl [mailto:dobesarc@mpi.nl]).

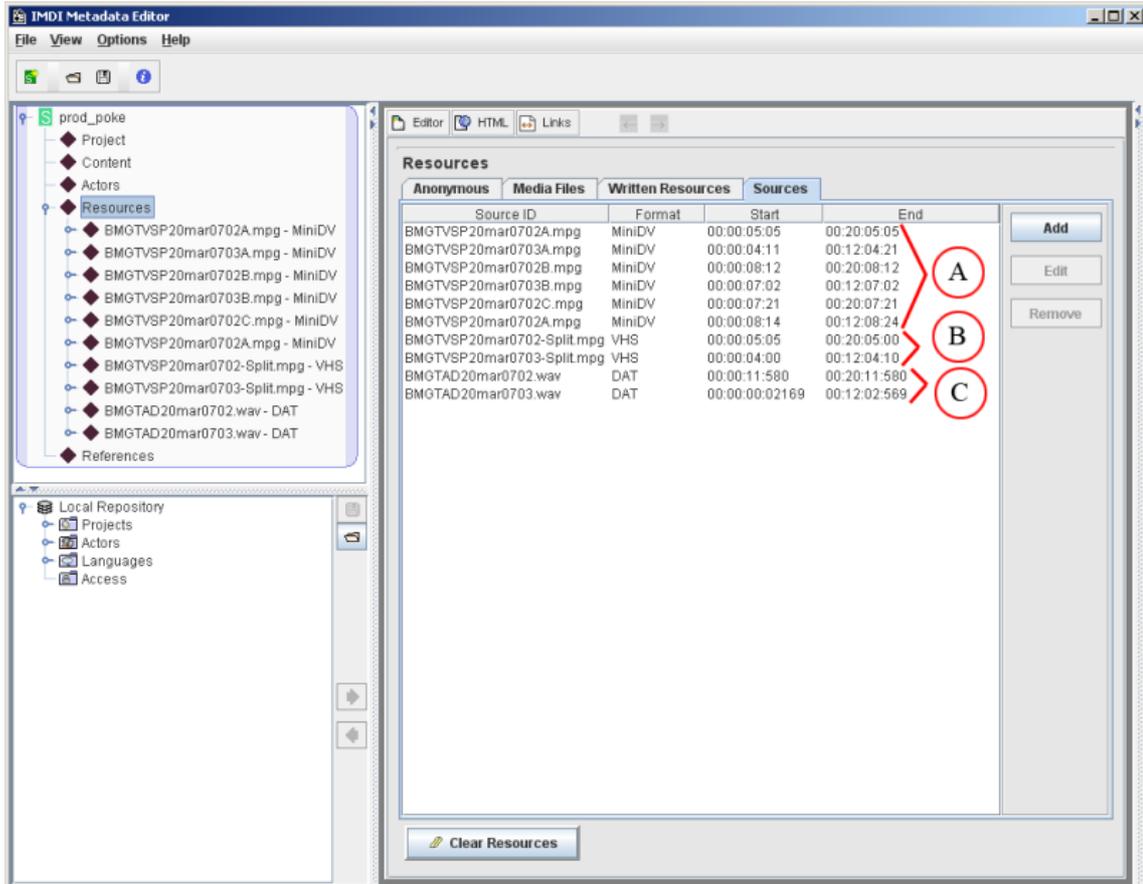


Figure 3.11. Resources multi Media Resource File session

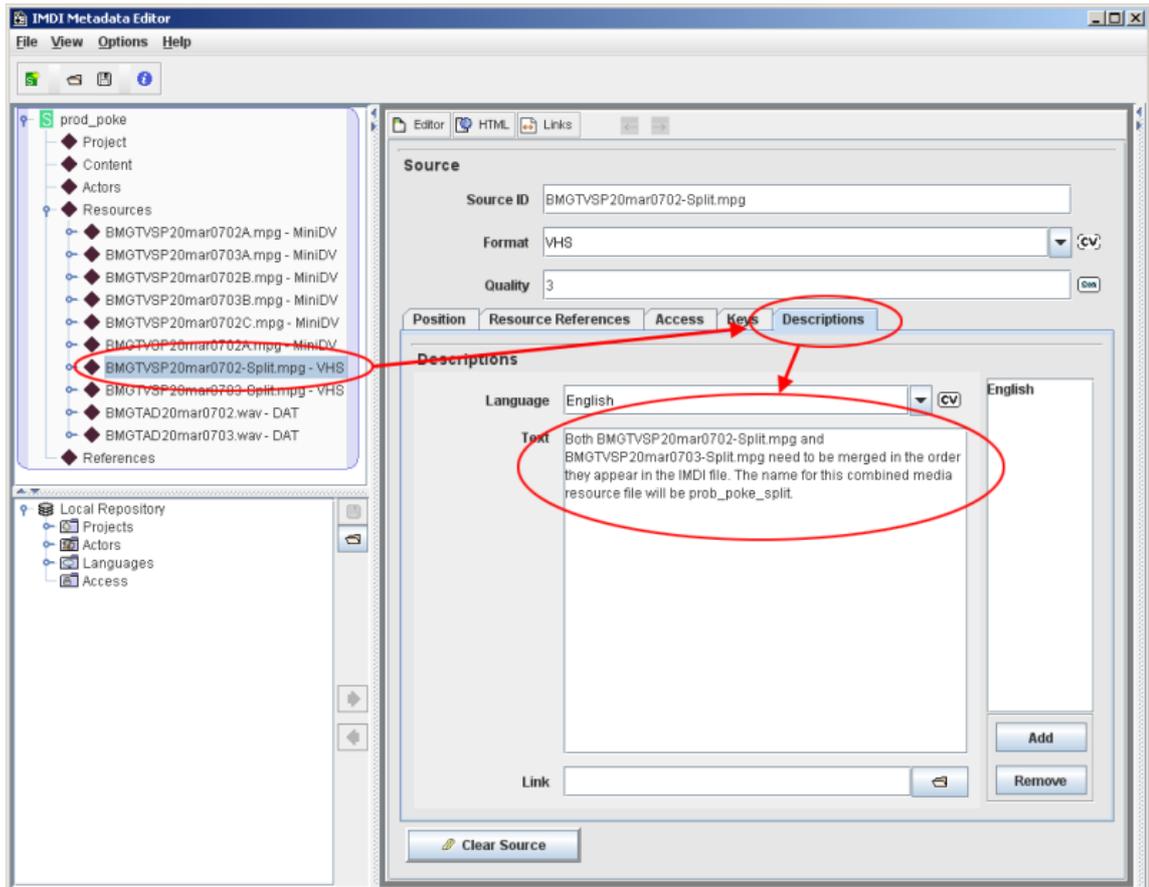


Figure 3.12. Resource / Source / Description

3.5. References

The References node is where you cross-reference to other sessions, field notes, publications or other external sources that are relevant to the content of the session. Please refer the Reference Manual.

Chapter 4. Repository panel

One of the main complaints from users regarding the IMDI editor has always been that you have to manually enter repeat data for every new IMDI file. These were complaints from users who hadn't mastered the IMDI editor yet, in particular what the Repository Panel was for. This is where you can drag and drop completed nodes from the Browse Panel into the relevant Repository Panel directory. Then when you create a new IMDI file you can drag and drop the information from the Repository Panel into the relevant node in the Browse Panel.

In Figure 4.1 is actor information that will be recurring throughout several IMDI files.

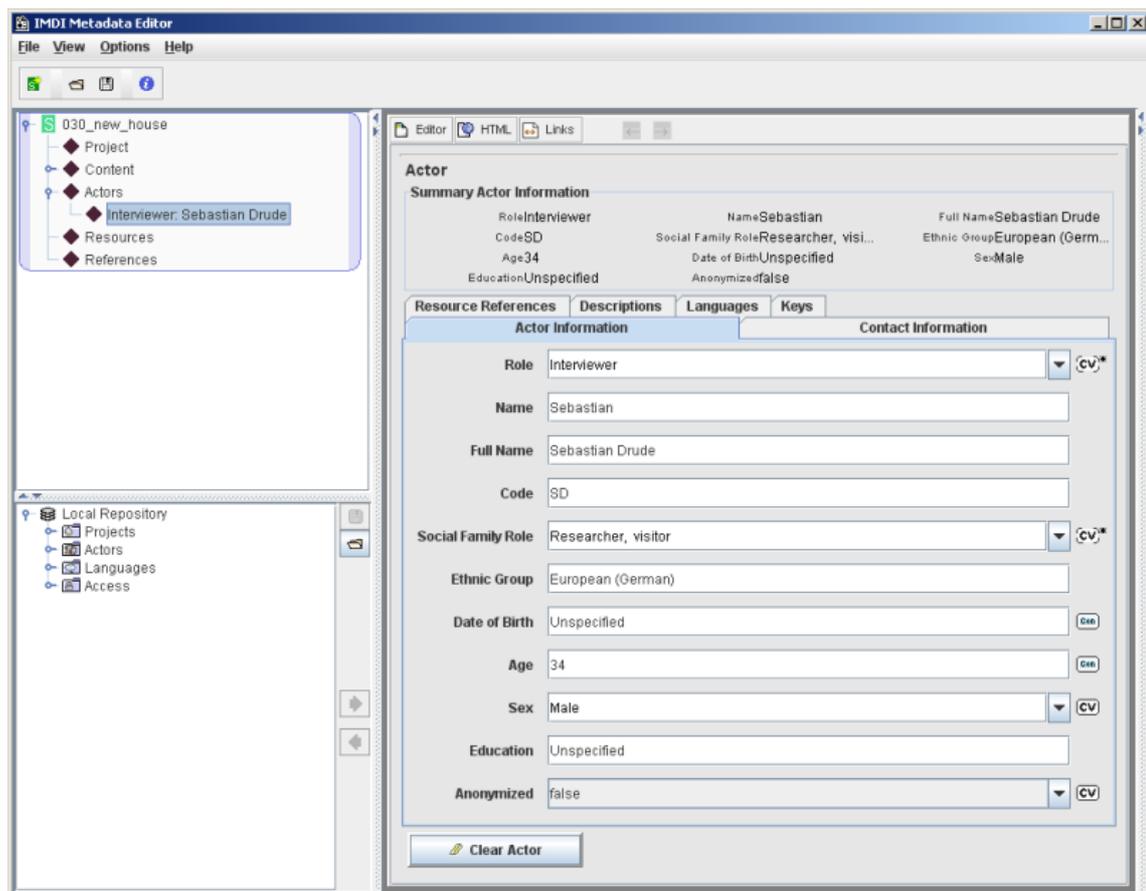


Figure 4.1. Repository Panel for Actor

By clicking on the 'Interviewer: Sebastian Drude' node in the Browse Panel and dragging it and dropping it into the Actors directory in the Repository Panel, you will have saved all the information from that element as a stub (see Figure 4.2).

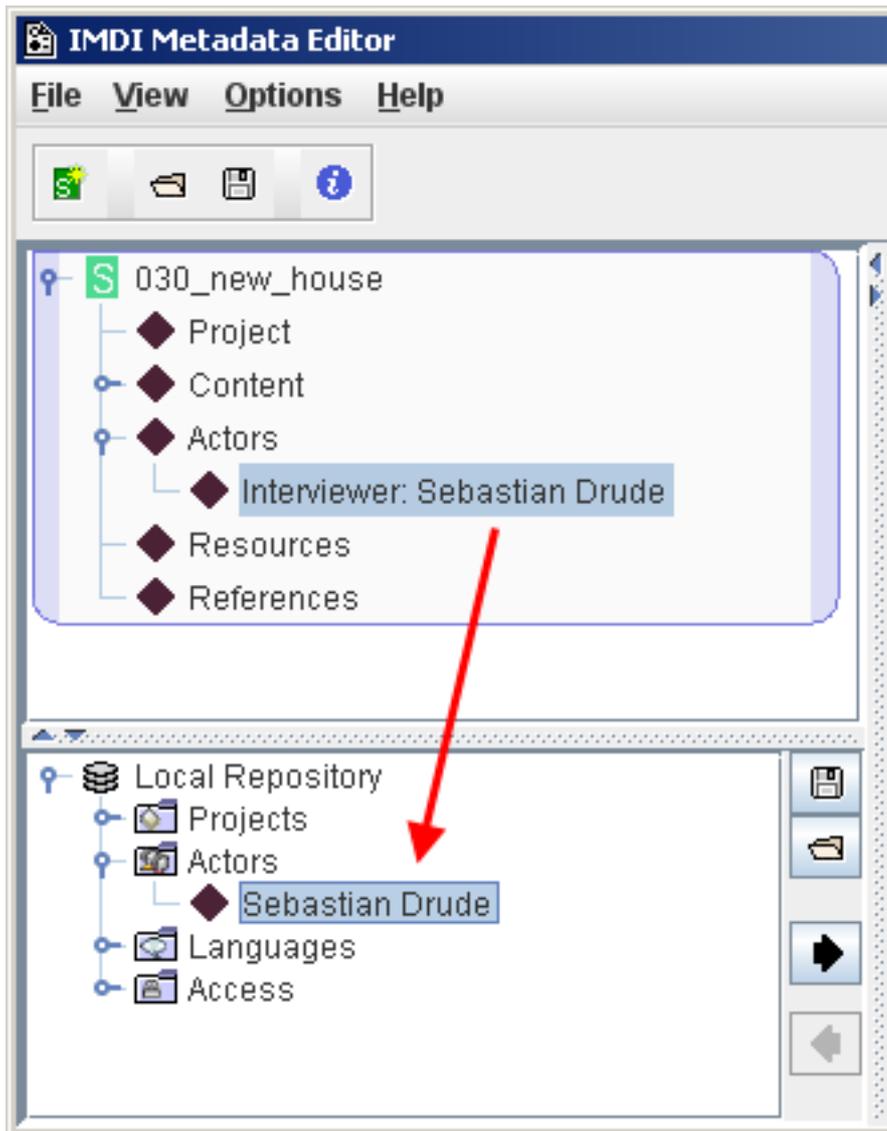


Figure 4.2. Storing information in the Repository

Now when you create a new IMDI file and Sebastian is again an actor in that session, all you need to do is drag the Sebastian Drude stub from the Actors directory in the Repository Panel back to the Actors node in the Browse Panel.

Actor information from a session can only be stored in the Actors Directory of the Repository and not in another directory. When dragging the information form the Browse Panel downward you will notice that next to the diamond you are moving you get a red disc with a white line horizontally across it. This means no entry, and appears only when you are over a directory, or an area where you are not allowed to drop the node.

If you have a lot of duplicate information, say for example that every IMDI file will be identical except for the file name and the DMF information in the Resources node then you need to populate all the fields that will have the repeat information and save this IMDI file. Then you can open this template IMDI populate the fields that will have the unique information and save this as the completed IMDI file.

Glossary

Corpus Manager	The manager of a user's corpus, who's main role it is to support the user in all archive related matters.
Digi-Team	Data managers, based at the MPI Nijmegen, who are responsible for digitizing recordings and cutting media resource files.
DMF	Digital Master File, a complete media recording in file format
IMDI file	The file you populate with metadata to describe your session.
Resource File	A generic term for files that link to the session
Media Resource File	A Resource File that is media, for example audio or video. Media Resource Files are cut from the DMF.
Written Resource file	A Resource File that is text, for example ELAN or Lexus files, but can also be any text file that links to the session.
Session Resource Bundle	The complete session. An IMDI file and all its linked Resource Files.
Session	A confusing term that has multiple uses by both Corpus Managers and the Digi-Team. It can mean: A Session Resource Bundle, a IMDI file without Resource Files or a Media Resource File on its own.
Source File	The IMDI editor term for DMF, but can also relate to any source that your resource files come from. So, if you resource file was a chapter from a book for instance, in this case the source file would be the title of the book.
Stub	An IMDI element saved in the Repository Panel.