Intera **e**Content

Intera Deliverable D2.2

Integrated Resource Domain

Project reference number	e-content EDC-22076 INTERA / 27924
Project acronym	*Intera*
Project full title	Integrated European language data Repository Area
Project contact points	Khalid Choukri Evaluations & Language Resources Distribution Agency S.A. 55-57 Rue Brillat Savarin, 75013 Paris, France Phone: +33 1 43 13 33 33 , Fax: +33 1 43 13 33 30 Email: choukri@elda.fr
Project web site	http://www.elda.fr/intera
EC project officer	Philippe Gelin
Document title	Integrated Resource Domain
Deliverable ID	D2.2-1 Second Version
Document type	Prototype, Demonstrator
Dissemination level	PP
Contractual date of delivery	M9 (01.010.2003)
Actual date of delivery	01.11.2003 (indication to ELRA that the prototype is operating)[1]
Status & version	first version (next and last version to come in summer)
Work package & task ID	WP2, all WP2 Tasks except 2.1
Work package, task & deliverable responsible	MPI, USAAR, LORIA, ELRA
Number of pages	13 except appendixes
Author(s) & affiliation(s)	Daan Broeder, Peter Wittenburg, MPI
Additional contributor(s)	
Keywords	Meta data, Language Resources
Abstract	
Additional notes & remarks	

---

[1] It was not clear to MPI that in addition to the working prototype also a report has to be written. This was created in February 2004.

Integrated Resource Domain

The Integrated Resource Domain has to be delivered in two steps as a prototype and demonstrator. The implementation of the prototype required much technology preparation work to adapt metadata technology to the new IMDI version and to validate incoming metadata descriptions, to integrate the metadata descriptions and to offer the integrated metadata domain via a well-defined portal. So all technology work is seen as preparatory to realize the demonstrator. Therefore, this report will touch the development work carried out and briefly describe the prototype.

1. Technology Work

The technology work is based on the new IMDI version that was discussed at the International IMDI Workshop in Nijmegen in November 2002. It was decided that the IMDI set will be changed and simplified slightly and that it will cover also metadata elements for written resources and lexicons. For details we refer to the appendix C where the new IMDI set is listed[2]. The discussions with the international experts took much longer then expected until the new standard was accepted by the IMDI community. In July 2003 the new version of the IMDI standard could be finished as IMDI 3.0.3. A complete set of XML-schemas was developed to define the set and the needed controlled vocabularies (see appendix D).

Further, the IMDI infrastructure was influenced by the experience of the users with the old tool versions. The new tools were adapted and new functionality was added. One essential extension was made to the IMDI definition: Special Profiles are now supported which allow projects or institutions to create their own special extensions to IMDI in addition to what is defined in the core set. Profiles are possible since the IMDI set from the beginning was designed as a set that is extendible. For example the Dutch Spoken Corpus project is using a number of additional elements taken from the TEI that should be searchable. For this extension a special profile was created.

In general we can state that there is no other metadata infrastructure world-wide that has such a level of support by a complete set of tools. The IMDI infrastructure is now even more interesting also for industry, since more and more management tools are being built around it. All tools are offered under Open Source licence and can be downloaded from the official IMDI site including manuals. For some of the tools the manuals have to be finished.

Editor

The most important tool is the editor that allows people to enter correct metadata descriptions. The new editor is a professional tool and certainly one of the best available world-wide. The functions are described in detail in the manual that is available via the IMDI web-site (see appendix A). The IMDI editor was completely revised to make it even more user friendly and more efficient. The editor supports a conversion from the old IMDI version to the new one, i.e. a user will be notified and asked whether a new version should be generated. At this place we will just stick with a screen shot that shows the new user interface (see figure 1).

From the users we were informed about errors in the new version of the editor. The code was corrected with respect to the errors we heard about. We also re-factured the code at some places, so that we can speak about a stable version again.

CV Editor

This editor allows people to add controlled vocabularies that are also important to guarantee useful search results. Also here we will only show a screen shot (see figure 2) and refer the interested user to the web-site. A manual is available (see appendix D).
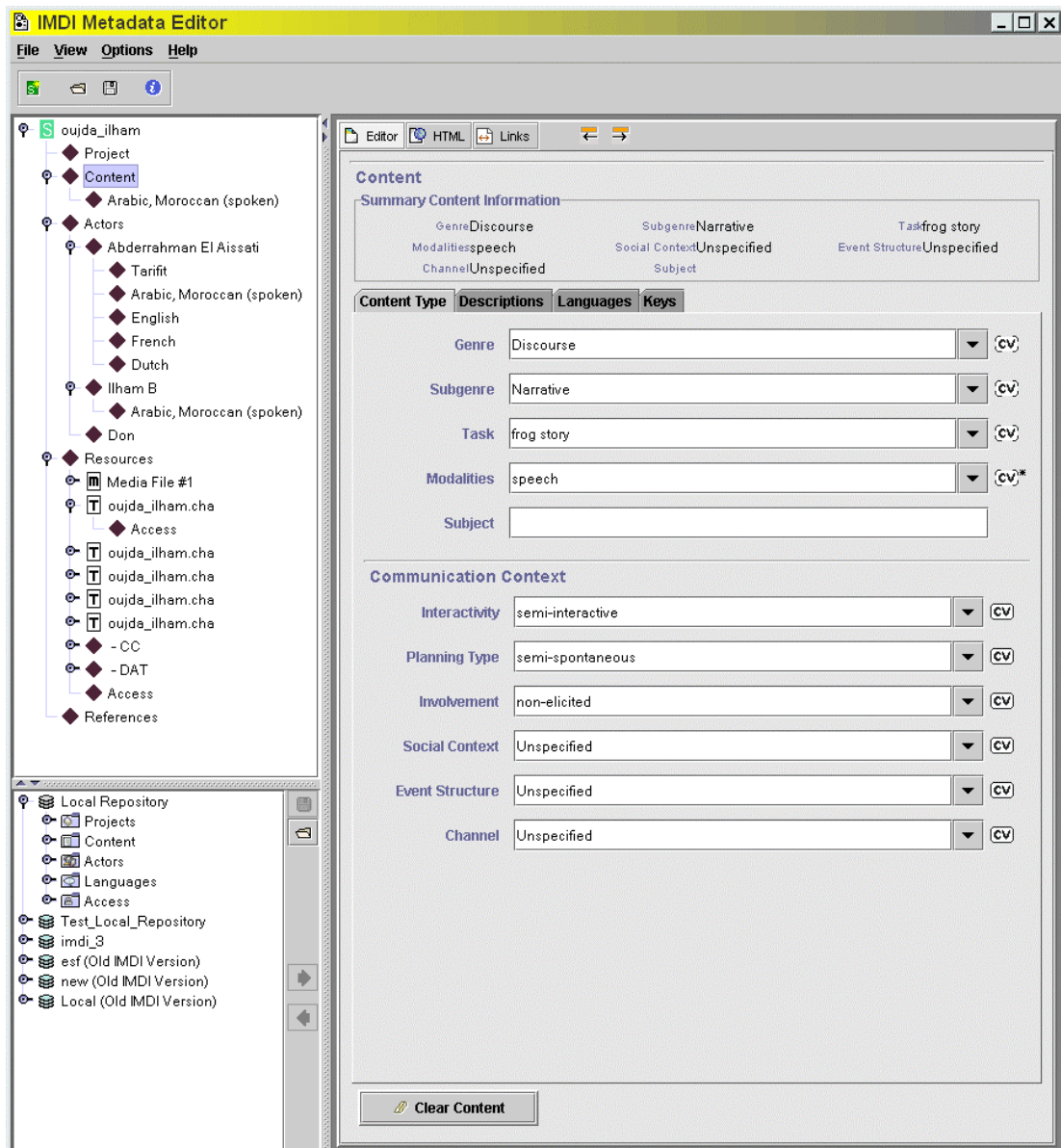
---

[2] We also refer to the Web-Sites: http://www.mpi.nl/ISLE and http://www.mpi.nl/IMDI for details.

Figure 1 shows a screenshot of the new editor version.

Figure 2 shows a screenshot of the CV Editor.

Corpus Structure Builder

Figure 3 shows a screenshot taken from the Excel program. This can be used to specify the IMDI hierarchy in an easy way. The output is transformed by a program into proper IMDI XML files.

The IMDI metadata infrastructure is not only focussing on searchable metadata descriptions but also on browsable hierarchies that can be used for organizational and management

purpose. Altogether the IMDI infrastructure is intended to form a corpus management framework to manage large collections of language resources.

In this respect the users confronted us with the need of an efficient way to generate corpus structures. We allow people to specify corpus structures with the help of for example Excel which allows fast and easy column operations. These Excel tables are then processed by a program to generate validated IMDI hierarchies existing of IMDI XML files (see a screenshot of a typical Excel Sheet with a corpus structure in figure 3). Also the editor was extended to allow people to generate corpus structure nodes.

Tree Builder

The Tree Builder is a tool that allows users to create their own virtual archive within the overall IMDI domain. Users can create arbitrary nodes and link metadata descriptions with them. All work can be done graphically by connecting existing nodes with others. In doing so researchers can create their own private working domain for a special study to be carried out for some time. This working domain much more easily allows the user to find and combine the needed resources. Figure 4 gives a screenshot of the Tree Builder.
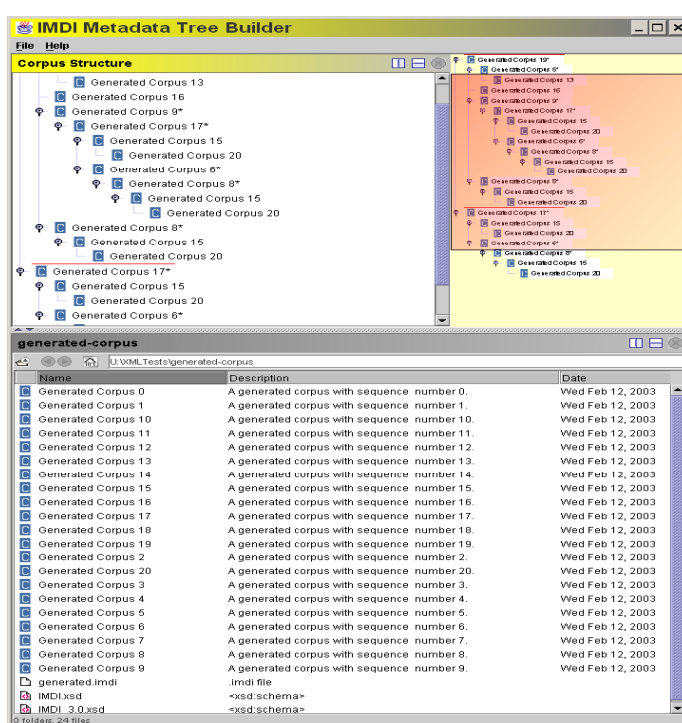
Figure 4 shows a screenshot of the Tree Builder that can be used to create new virtual collections from the whole IMDI domain by simple graphical means. This allows users to easily create temporary working spaces which leads to great efficiency.

Due to the necessary debugging of the editor code and to the lost of the programmer of the editor further work on the Tree Copier had to be postponed.

Browser

The metadata browser is another true IMDI tool operating in distributed IMDI environments and directly on the IMDI XML files. It can be used to easily navigate in the IMDI domain with the help of browsing and searching. It also allows to set bookmarks to nodes such that users can easily go to sub-corpora. The browser is also the basis of all advanced options such as the domain interlinking proposed for the INTERA project to immediately start tools when having found suitable resources. Of course, the browser has facilities to support special project files. The details are described in the manual which is attached as appendix B. Figure 5 gives an idea about the user interface.
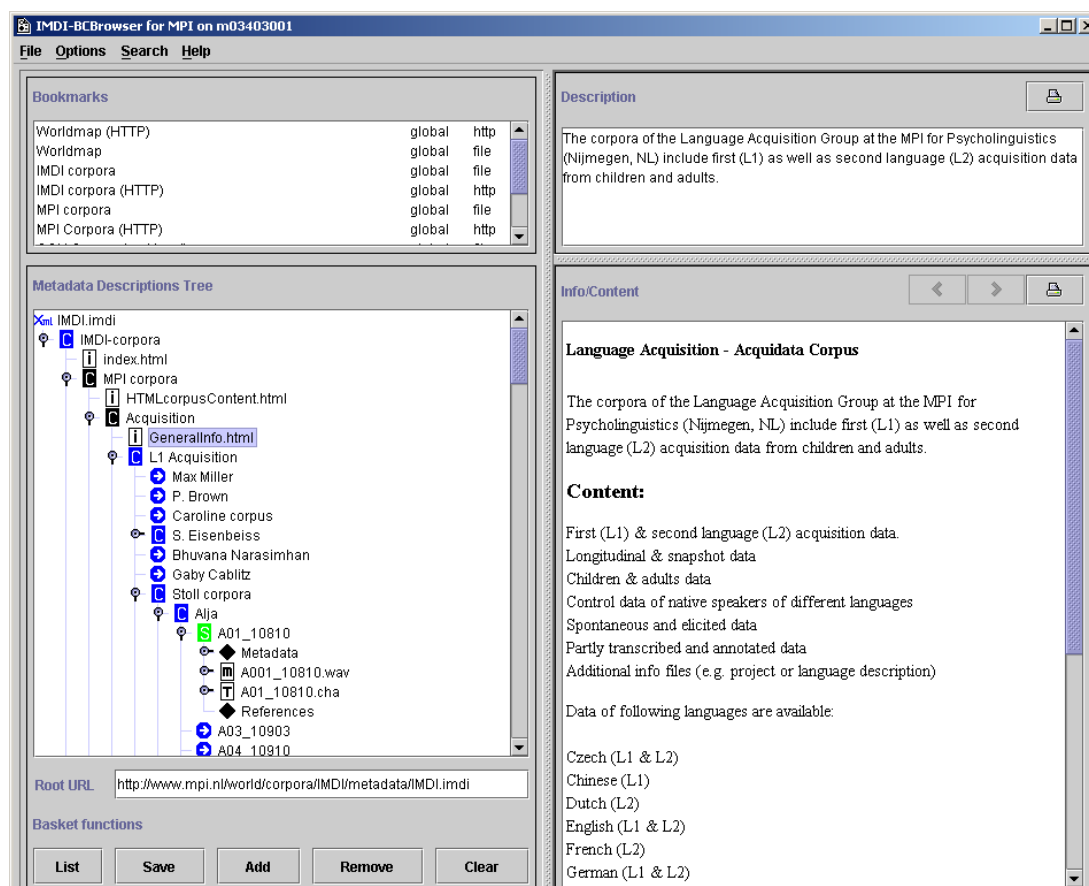
Figure 5 shows a screenshot of the new browser version that can process new as well as old IMDI files. It allows users to flexibly browse in the linked IMDI domain of XML files.

Structured Search Component

The structured search component is an integrated part of the browser, but could also be used in different frameworks. It offers to combine patterns with different elements that are part of the structured IMDI set and it also supports profiles. Its broad functionality is described in detail in the manual of the browser. To support search the browser supports creating a DB file for all IMDI files harvested. These are then queried to find hits. A user can use this facility to search through his private local corpora. Language archives have the possibility the possibility to make the created DB files available for remote querying by running a DB daemon by users through the same search interface. Again we just want to restrict ourselves in this document to give a screenshot of the component (see figure 6).
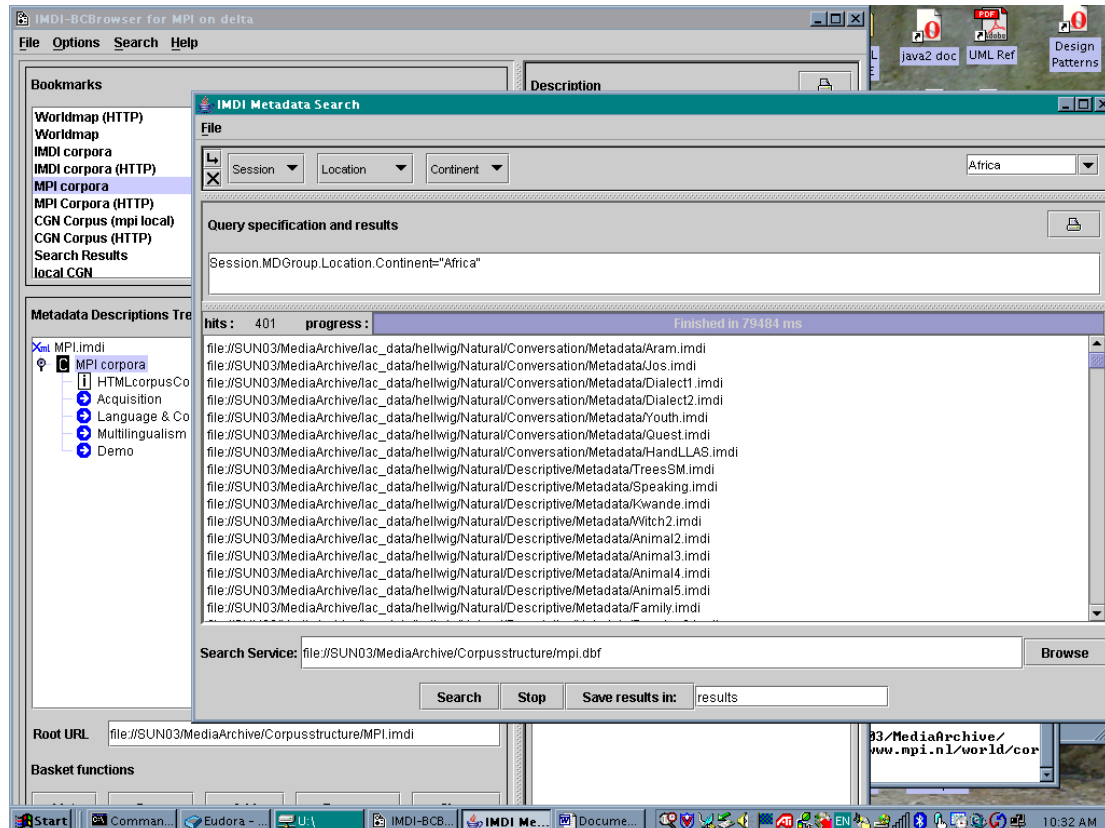
Figure 6 shows a screenshot of the search component that is integrated with the browser. It allows the user to enter queries supporting all structured elements of IMDI and also supports profiles.

HTML-Based Browsing

Knowing that many people have problems with downloading and installing special tools such as the IMDI browser we opened the possibility to navigate in the IMDI domain by using normal HTML browsers as indicated in the following figure. Of course, some restrictions have to be made for example with respect to the information about the corpus structure and the location of the user. Here only a path to the actual node is delivered. Nevertheless, the user can access all metadata information that is available. This mode is achieved by executing style-sheet based XSLT transformations on-the-fly, i.e. the HTML are created on the fly from the underlying XML representations (see figure 7).

It should be mentioned that these pages can be accessed by going to the following web-site: http://www.mpi.nl/corpus.

Figure 7 shows a screenshot of the way browsing is made possible via normal web browsers. XML files are transformed on-the-fly to HTML pages.

Unstructured Search

The same environment is used to test out an alternative search option. We also know that many users find it too difficult to formulate structured search queries, but want to have an unstructured search option that has the functionality of the Google search function. Although theoretically it does not sound convincing to execute unstructured search on structured data it offers some advantage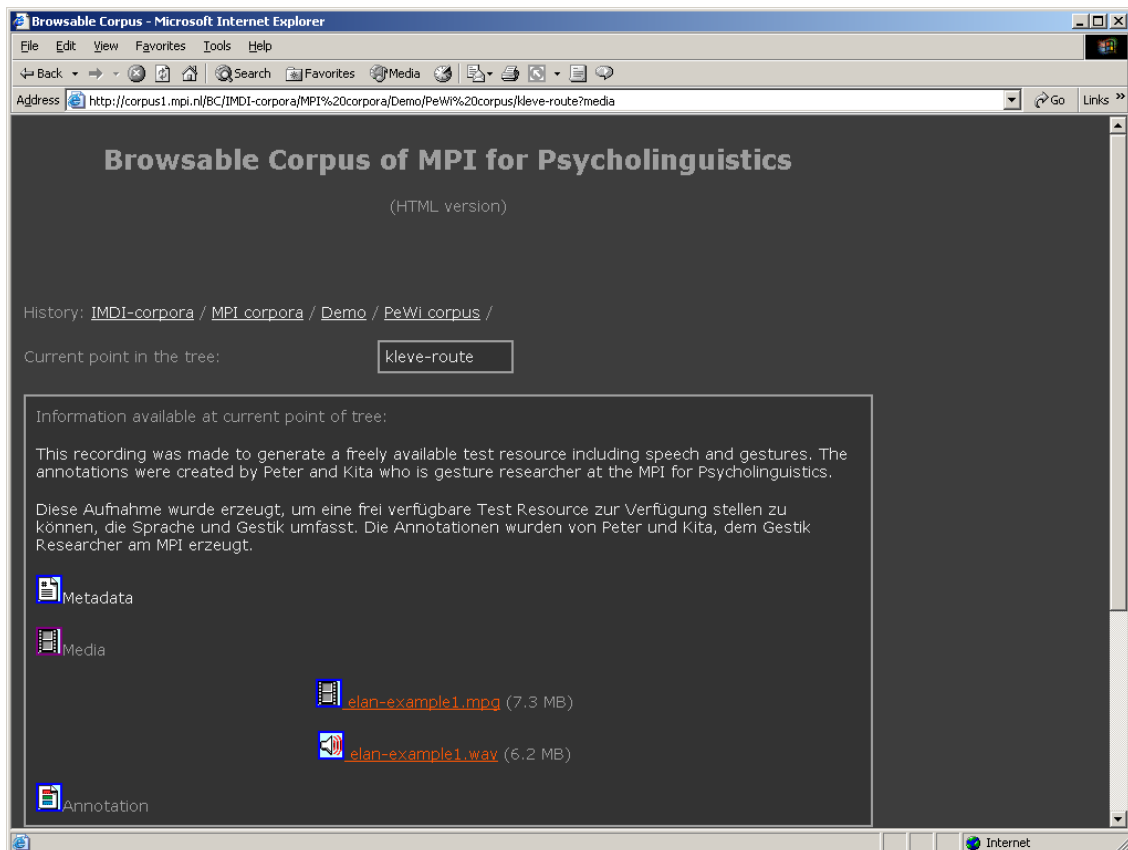s. Even people who don't know the semantics of the elements and those who don't understand how certain resources are characterized could execute searches on typical values and get hits. Programs are used to generate an index file covering all metadata elements including descriptions and other unconstrained text fields.

Figure 8 indicates this option which also can be tested out by going to the web-site: http://www.mpi.nl/corpus.
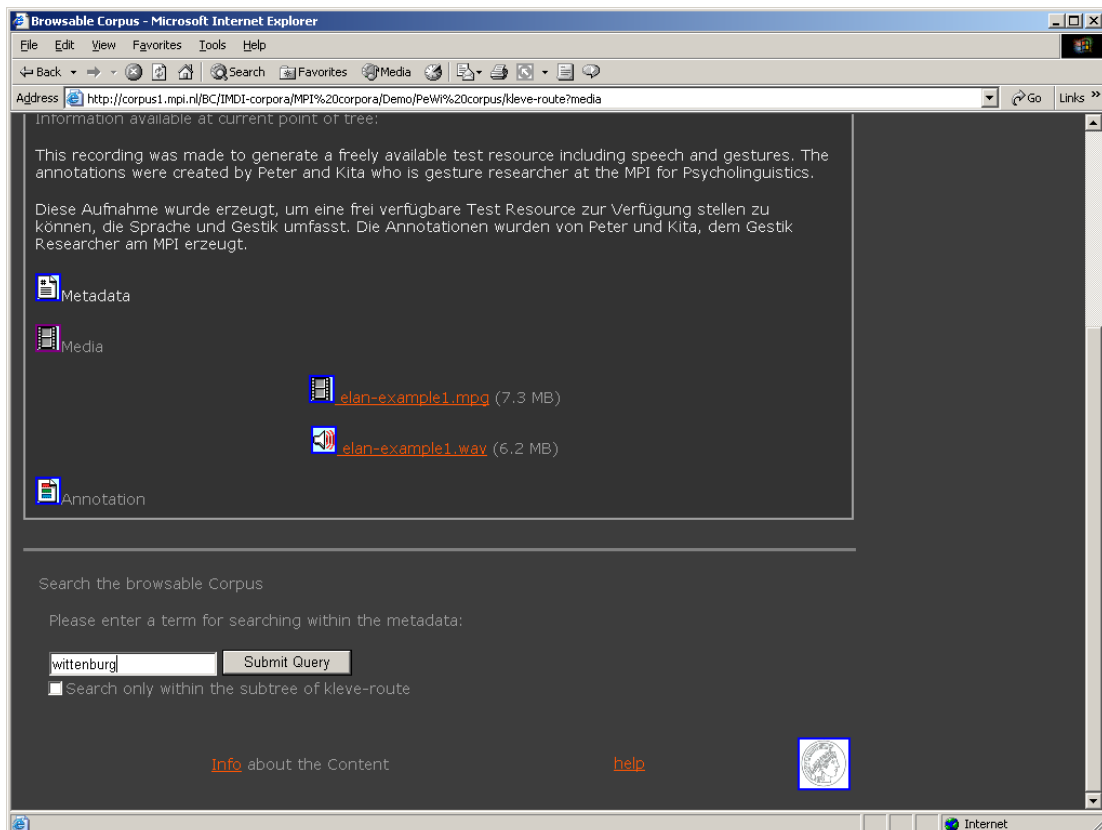
Figure 8 indicates the Google-like search field allowing the user to do unstructured search on the structured IMDI files.

Version Converter

A converter program that converts IMDI 2.x into IMDI 3.03 was created and was successfully tested. The IMDI conversion was carried out in March 2004. This conversion has many aspects that have to be considered carefully and many institutions that make use of IMDI have to be informed. So the step over to IMDI 3.03 is a major step and has consequences also to many management and index creation programs that have been developed during the past years. The adaptation of all programs is almost finished so that the move to the new IMDI version can be made soon. The converter is available also on the IMDI site. The editor was extended such that it can read in old IMDI files and transforms them to new ones. The user is warned that transformations will be done so that he/she can manually intervene if necessary.

9

IMDI-to-OLAC Gateway

Both relevant metadata initiatives for language resources have created new versions that can be seen as major changes. So the old mappings between OLAC and IMDI were not valid any longer. Together with the EMELD project the MPI created a new mapping version which was discussed and except one small adaptation in the mapping approved at the recent E-Meld conference in Michigan in July. First tests from OLAC as service provider resulted in the fact that they had problems with the many metadata records that the IMDI domain now contains. A discussion between the responsible IMDI and OLAC persons resulted in the conclusion that the OLAC service provider has to adapt its services to be able to cope with the IMDI records in a useful way. Given very general queries some users may formulate such as "give me all resources for German" there would be thousands of records, like one can expect for very unspecific web-queries. These thousands of hits would dominate the hits generated from small repositories for example. It was agreed that it is a task for the service provider to find a balance. The OLAC search engine will be adapted to find this balance in a way defined by the OLAC designers. When these adaptations have been carried out all IMDI descriptions will be searchable again via the OLAC interface as well.

Still it should be noticed that the IMDI metadata descriptions are much richer and that some relevant data will just be ignored. As background information we should add that IMDI and OLAC slightly operate from different foci. OLAC was designed as an interface layer with semantics close to Dublin Core where many different language resource archives act as data providers and create DC-like descriptions according to the Open Archives Initiative model. IMDI does not only tackle the aspect of resource discovery via searching but also wants to support browsing and resource management. Further, the IMDI set was designed bottom-up to meet the needs of the language resource community in more detailed types of descriptions.

It should be mentioned here that this work is carried out in relation to WP3 and the newly formed ISO TC37/SC4 committee. All categories from IMDI were entered into the Data Category Registry. It is the intention to also enter all OLAC categories and to make the relations that are now included in wrappers explicit and available via open repositories. Yet there is no proposal for open relation registries that could be used by search engines.

Access Rights Management

Another very important framework that has been developed recently and that is based on the IMDI infrastructure is a complete framework to define users and groups and associate access rights with the groups. The framework also allows managers to define new managers to delegate the right to define rights, since otherwise the access administration is not tractable. Also here for reasons of efficiency it is possible for the manager to select a node in the metadata domain and assign access rights to different types of resources that can be found under this node. A user manual was being written that describes the full functionality. Again screenshots are used here to indicate the functionality (see figures 9 and 10).

Also this tool was tested extensively, since a deployment could only be accepted if IPR questions are dealt with correctly. It was presented at LREC 2004 and is in operation. The manual can be found in Appendix C.

The ARM solution is based on a Postgres database, i.e. all commands entered via web-interfaces are stored in the database. Programs are generating appropriate entries in the HT Access file that determines the access rights for web-based users and entries in Access Control Lists for internal users. All software is available and can be re-used by other institutions. It is intended to extend these mechanisms to distributed access mechanisms within a coming European project such that the users will be confronted with one transparent domain of web-resources and have just one identity in this domain that will cover several language resource archives.

By having added an ARM component we have reacted on the comments of the reviewers and shown that we take IPR issues very serious.
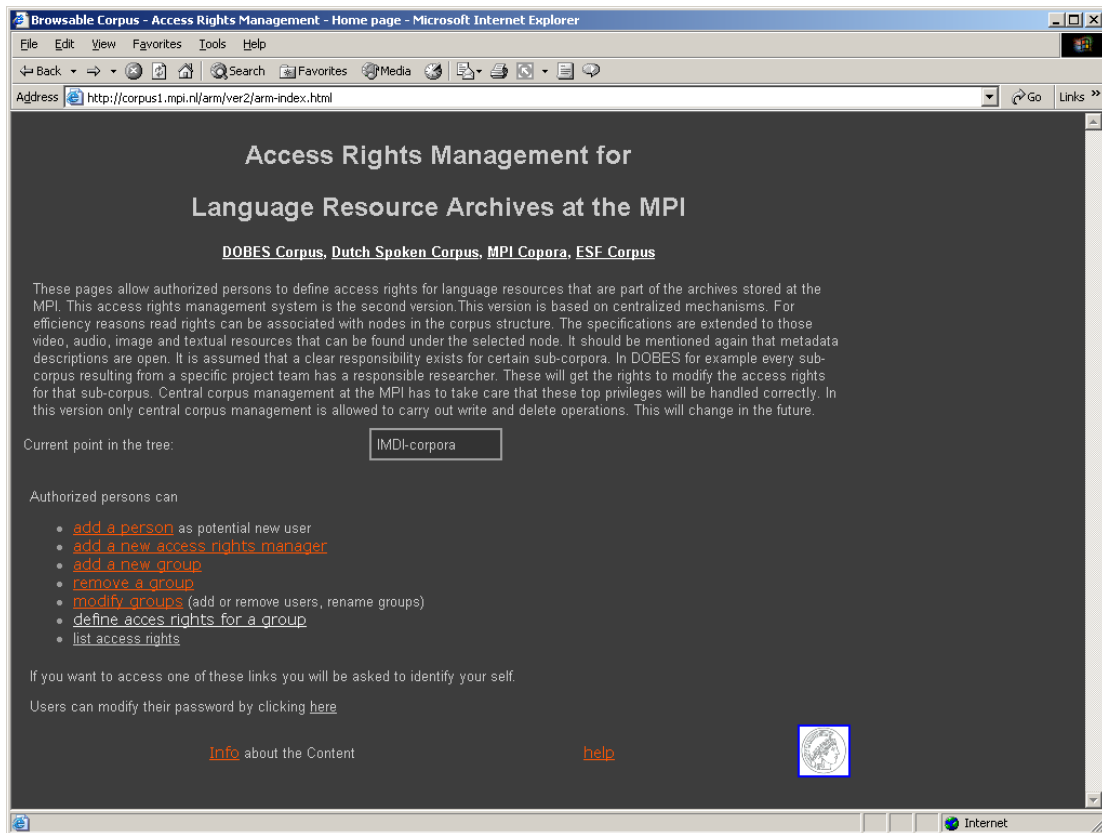
Access Rights Management for

Language Resource Archives at the MPI

DOBES Corpus, **Dutch Spoken Corpus**, **MPI Copora**, **ESF Corpus**

These pages allow authorized persons to define access rights for language resources that are part of the archives stored at the MPI. This access rights management system is the second version. This version is based on centralized mechanisms. For efficiency reasons read rights can be associated with nodes in the corpus structure. The specifications are extended to those video, audio, image and textual resources that can be found under the selected node. It should be mentioned again that metadata descriptions are open. It is assumed that a clear responsibility exists for certain sub-corpora. In DOBES for example every sub-corpus resulting from a specific project team has a responsible researcher. These will get the rights to modify the access rights for that sub-corpus. Central corpus management at the MPI has to take care that these top privileges will be handled correctly. In this version only central corpus management is allowed to carry out write and delete operations. This will change in the future.

Current point in the tree: IMDI-corpora

Authorized persons can

- add a person as potential new user
- add a new access rights manager
- add a new group
- remove a group
- modify groups (add or remove users, rename groups)
- define acces rights for a group
- list access rights

If you want to access one of these links you will be asked to identify your self.

Users can modify their password by clicking here

Info about the Content   help

Figure 9 shows a screenshot of the new access rights management system that is tightly connected with the IMDI metadata domain.

Figure 10 shows a screenshot of the "define access rights" option of the new access rights management system. It is linked with the current node in the metadata hierarchy and allows users to select options. At this moment only read access can be given. We are working on a semi-automatic ingest system that will also allow users to define write rights, however, for an archive this leads to complicated issues such as versioning, since nothing may be deleted.

2. Metadata Integration

In an earlier note we have described that it was much more difficult to convince data providers to participate in the metadata creation process and that at the beginning too little money was offered by us. Things have changed and first metadata is available. The following matrix gives information about the state.

Partner	Resource Provider	Resource Names	Resource Type	Metadata State
MPI	BAS (D)	Smartcom	Multimodal	integrated
MPI	BAS (D)	Verbmobil etc	Speech	integrated
MPI	Meertens Inst (NL)	GTPR corpus	texts	integrated
MPI	Florence (It)	Lablita Corpus	speech/text	integrated
MPI	Taalunie (NL)	Dutch Spoken Corpus	speech	integrated
MPI	MPI	Gesture Corpus	Multimodal	integrated
MPI	MPI	ESF Second Learner Corpus	speech/text	integrated
MPI	MPI	PMOLL Corpus	speech/text	integrated
MPI	MPI	various others	speech/text	integrated
USAR	DFKI	Negra, Tiger	text	MD ready
USAR	U Prag	?	text	to come
USAR	U Madrid	Coral Rom	text	MD in creation
USAR	U Budapest	?	text	to come
LORIA	ATILF	Frantext etc	text	MD in creation
ILC	parallel corp providers	various	text	MD in creation
ILSP	parallel corp providers	various	text	MD in creation
ELRA	ELRA	various	text	to come

According to this matrix the first metadata descriptions were delivered and a number have already been integrated into the domain (see figure 11).

13

In the left frame called "Metadata Description Tree" one can find the contributions that are marked as being integrated such as from CGN, BAS, IFA, GTRP (Meertens), C-ORAL (Florence) and all the resources covered by the MPI archive. Other resources will be added when they become available[3].

In addition to this list we can state that the following institutions are currently creating IMDI metadata descriptions (mostly on own money or via other funding sources):

- Univ Lund
- Sweden's National Dialect Archive
- U Uppsala
- U Stavanger
- U Groningen
- U Utrecht
- U Leiden
- U Helsinki
- Kotus Project Helsinki
- U Stockholm
- U London
- U Nijmegen
- U Kansas
- U Victoria (ca)
- EMU Michigan
- U Sydney
- U Melbourne
- U Kiel
- U Bochum
- U Bonn
- U Leipzig
- U Frankfurt
- U Bielefeld
- U Erfurt
- Federal U of Rio de Janeiro
- U Buenos Aires
- Phonogramm Archive Vienna
- MPI Nijmegen
- European Sign Language Communities in Se, UK, NL and D
- several others we don't know

Also these resources will become available under the IMDI portal when they are ready and registered. Partly they are already integrated as sub-parts of other projects. Given this list we can hope that an increasing number of institutions will join.

For every metadata provider the following steps are being carried out at the MPI:

- The metadata files provided are being validated, since most of them are created at least semi-automatically with the help of scripts. So a validation step is necessary.

- Some indexes have to be created to make the metadata searchable.

- The metadata hierarchies have to be created and integrated to offer browsing capability.

For all these steps the tools have been created such that they can be carried out efficiently. Often, some interaction has to be done to carry out the necessary corrections in particular on the semantic level.

---

[3] When downloading and installing the IMDI browser everyone could start looking at this data.

3. Prototype and Demonstrator

The prototype includes the software that was described in chapter 1 and the metadata domain where we indicate the state in chapter 2.

The prototype is running for some months already including all tools. All major tools are downloadable and can be used by others. Currently, some optimizations and debugging is carried out. This will lead to new releases that will be deployed from time to time. The growing metadata domain can be browsed and searched by going to the following web-site: http://www.mpi.nl/corpus.

At LREC 2004 a demonstration was given that covers already many of the resources that are mentioned in the matrix and that have to be integrated in the final deliverable. The intention was to create a sufficiently large metadata domain that makes it interesting for others to use it and hopefully to join. This would be in the line of the ENABLER declaration (ENABLER meeting in PARIS) which stated clearly that

- most language resources are still not visible on the Internet

- even less language resources are accessible via the Internet

- for all initiatives that are funded by governmental money it should be obligatory to create open metadata descriptions and to make the resources accessible via the Internet.

The IMDI infrastructure as further developed in INTERA allows people to create metadata in an easy way to have a low threshold for participation. The list given on the previous page shows that there is an increasing number of institutions that understand the relevance of metadata.

As was stated in the INTERA proposal, the delivered IMDI domain is not just a demonstrator but will be a service that will be further supported by the participating institutions. The availability of metadata is a must not only for resource discovery purposes, but is also a basis for efficient management.

The IMDI infrastructure is set up in a way that all interested institutions such as ELRA for example could easily create their own portal and offer the total or a selected domain. Of course, all IMDI tools are Open Source and free.

Appendix A

IMDI Editor, version 3.0

Manual

This manual was last updated: 22 Oct 2003
The latest version of the manual can be downloaded from the following webpage:
http://www.mpi.nl/IMDI/tools/
Author: Birgit Hellwig

Introduction

The IMDI (ISLE Metadata Initiative) Editor was developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. It is used to create metadata descriptions, both for sessions or resource bundles and for corpora. As such, it complements the other IMDI tools.

At the moment, the following IMDI tools exist (tools and manuals can be downloaded from http://www.mpi.nl/IMDI/tools/):

- IMDI CV-Editor: a tool for creating lists of controlled vocabulary items;

- IMDI Browser: a tool for displaying and accessing data and metadata;

- IMDI Metadata Search Tool: a tool for searching through metadata information;

- IMDI Tree Builder: a tool for creating browsable trees based on IMDI metadata descriptions (beta version).

This manual helps you to understand the structure of the IMDI Editor. It consists of the following parts:

I.   Navigating through the IMDI Editor (section 1), describing the structure of the IMDI Editor.

II. The IMDI Session Editor (section 2), illustrating the metadata categories used for describing sessions or resource bundles (i.e., audio/video and written resources).

III. The IMDI Lexicon Editor (section 3), illustrating the metadata categories used for describing lexicon resource bundles.

IV. The IMDI Corpus Editor (section 4), giving step-by-step instructions on how to manage a corpus with the help of the IMDI Editor.

V.   Appendices.

   ! Note: The IMDI Editor has undergone a number of changes since it was first created, both in terms of its user interface and in terms of the metadata categories supported. If you have metadata files created with previous versions of the Editor, please consult the following sections: section 1.6 (for a summary of changes to the user interface, introduced in version 2.0), and section 2.7 (for a summary of changes to the metadata categories, introduced in version 3.0).

! Note: This manual contains a number of references to corpora housed at the Max Planck Institute (MPI) for Psycholinguistics, Nijmegen. It uses these corpora as a means to illustrate the functionality of the IMDI Editor. Please note that any remarks only pertain to these corpora. Corpora stored at other archives or institutes follow different rules.

Notation Conventions

The following notation conventions are used:

- Screens, schemata and fields of the Editor are written in the font MS Sans Serif.

- (SHORTCUT) KEYS ARE WRITTEN IN SMALL CAPS.

- "Text in double quotes refers to menu items and screen displays."

- Information on troubleshooting starts as follows: **!**

Table of Contents

1 Navigating through the IMDI Editor

This part of the manual describes the structure of the IMDI Editor.

1.1 The IMDI Editor window

Starting the IMDI Editor opens up the "IMDI Metadata Editor" (or "IMDI Editor") window. Its initial display is as follows:
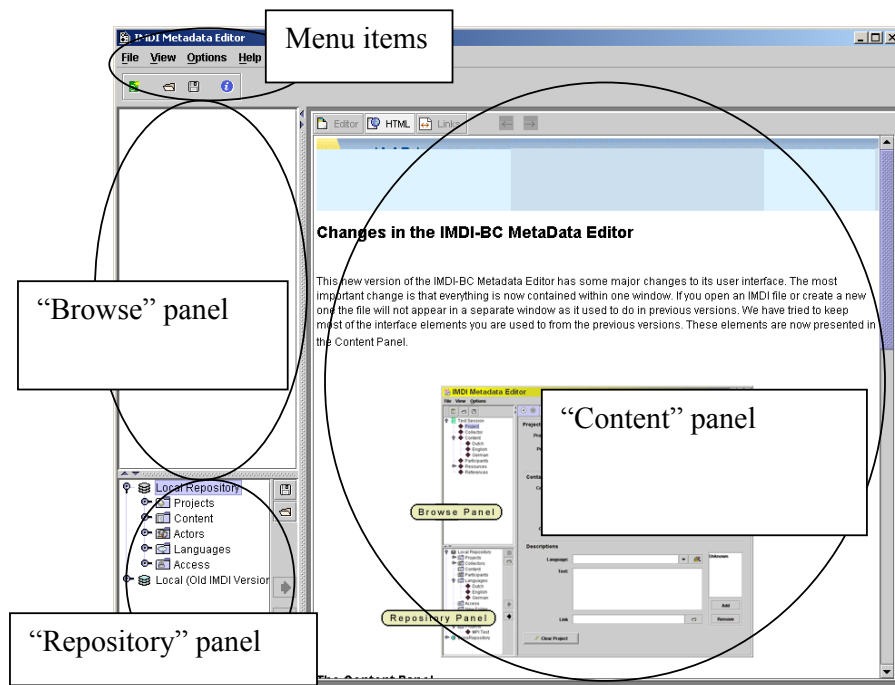
The "IMDI Editor" window has the following structure:

- a menu bar (see section 1.2);

- a "Browse" panel allowing you to browse through one (or more) IMDI files (see section 1.3);

- a "Repository" panel allowing you to work with recurring information (see section 1.4);

- a "Content" panel displaying screens, schemata and fields for entering metadata information (see section 1.5).

The size of any of the three panels ("Browse", "Repository", and "Content") can be changed relative to the size of the other panels. To change its size, do one of the following:
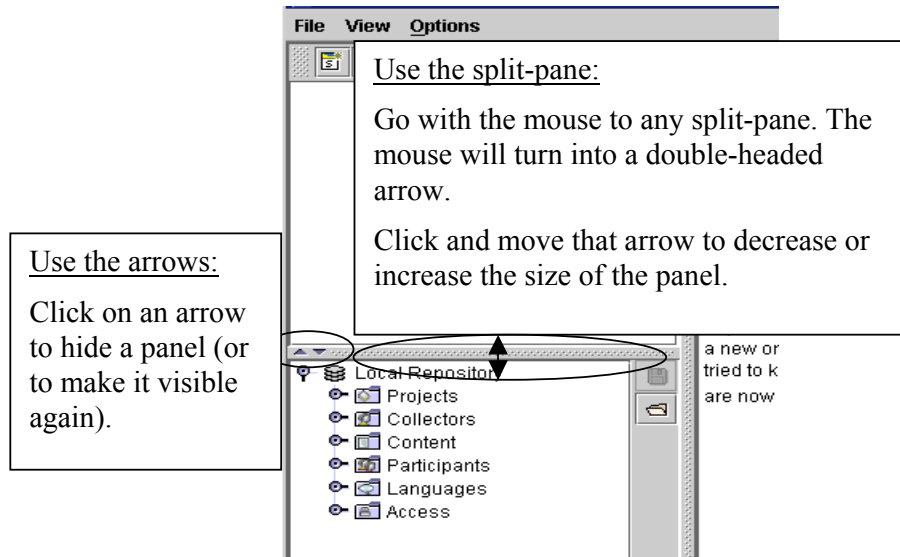
! Note: The structure of the "IMDI Editor" window has changed considerably in version 2.0 (see section 1.6 for a summary of the implemented changes). The structure of version 3.0 has only changed marginally, but there have been a number of changes to the metadata categories (see section 2.7 for details).

1.2 The menu items

The menu items are displayed in the top left corner of the "IMDI Editor" window.

To access any item, do one of the following:

(a) Click on the item.

(b) Or use the shortcut key ALT plus the underlined letter(s) (e.g., to access the item "File – New Session", press first the keys ALT+F and then N).

(c) Or use the shortcut key that is displayed to the right of the corresponding menu item (e.g., to access the item "New Session", press the keys CTRL+N). (This option is only available for some items.)

(d) Or click on the corresponding icon displayed below the menu bar. The following four icons are available:
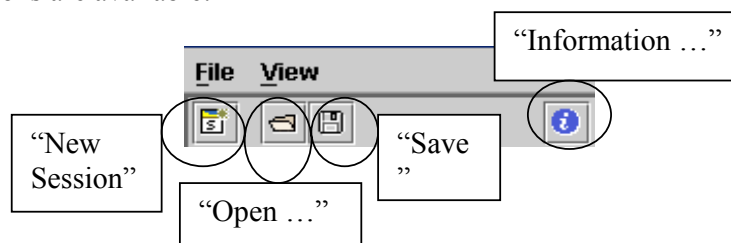
1.2.1 "File" menu

The "File" menu is used to create, open, close, save, print and exit IMDI file(s). It contains the following options:
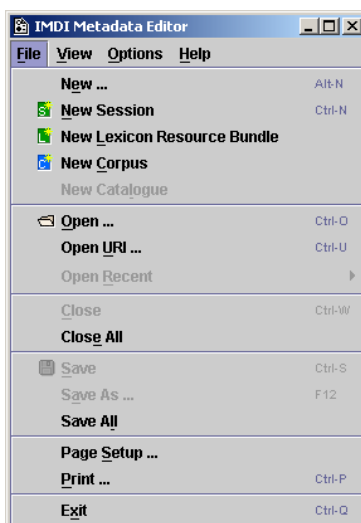
1.2.1.1 "New …"

"New …" allows you to create new IMDI files, and to assign them to a type and a profile (see section 1.2.3.1 for details).

Choose first a type (e.g., "Resource Bundles (Sessions)", "Lexicon Resource Bundles", or "Corpora") and then a profile (e.g., "Standard (Session) Resource Bundle"). Click on the profile, then click on "OK". The new file will be assigned to this type and profile.
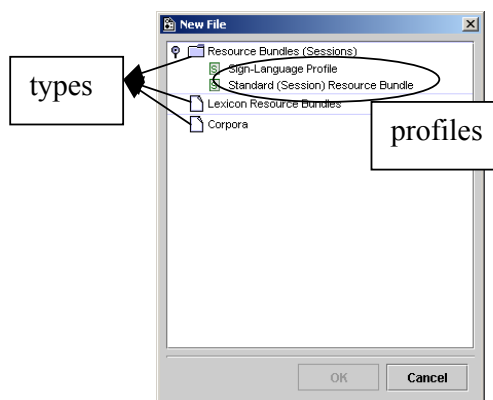
1.2.1.2 "New Session"

Click on "New Session" to create a new IMDI Session file. It will be automatically assigned to the default profile for Resource Bundles (Sessions) (see section 1.2.3.1; see section 2).

1.2.1.3 "New Lexicon Resource Bundle"

Click on "New Lexicon Resource Bundle" to create a new IMDI Lexicon file. It will be automatically assigned to the default profile for Lexicon Resource Bundles (see section 1.2.3.1; see section 3).

1.2.1.4 "New Corpus"

Click on "New Corpus" to create a new IMDI Corpus file. It will be automatically assigned to the default profile for Corpus bundles (see section 1.2.3.1; see section 4).

1.2.1.5 "New Catalogue"

Please ignore this option for the moment.

1.2.1.6 "Open …"

Click on "Open" to open an IMDI file. Browse to the folder that contains the IMDI file and double-click on it to open it.

> **!** Note: You can only open files of the IMDI Editor format. If you try to open a file of a different format, an error message will appear.

> **!** Note: If you open a file assigned to a profile that has been deleted, the default profile for that type will be used instead. E.g., you may have an IMDI Session file that was originally assigned to the "Sign-Language Profile", but later on you deleted this profile. If you now open this IMDI Session file again, it will be assigned to the default profile for session files (e.g., to the "Standard (Session) Resource Bundle" Profile) (see section 1.2.3.1 for details).

> **!** Note: If you open a file created with an old IMDI standard, it will be converted to the IMDI 3.0 standard (see section 2.7 for details).

1.2.1.7 "Open URI …"

Click on "Open URI" to download an IMDI file from the Web. A dialog window pops up. Enter the corresponding URI into the window and click on "OK" (see also section 1.2.1.6).

1.2.1.8 "Open Recent"

Go to "Open Recent" to view a list of the IMDI files that you have recently worked on. Click on any file from the list to open it (see also section 1.2.1.6).

1.2.1.9 "Close"

Click on "Close" to close the currently active IMDI file (i.e., the file that is highlighted in the "Browse" panel; see section 1.3). All other open IMDI files will remain open.

If the file to be closed has not been saved yet, a warning appears. Click on "Yes" to save the file before closing; click on "No" to not save the file; click on "Cancel" to return to the file.

1.2.1.10 "Close All"

Click on "Close All" to close all IMDI file (see also section 1.2.1.9).

1.2.1.11 "Save"

Click on "Save" to save the currently active IMDI file (i.e., the file that is highlighted in the "Browse" panel; see section 1.3).

> **!** Note: The "Save" option is only enabled if a file node (i.e., a session, lexicon, or corpus node) is highlighted in the "Browse" panel. If any other node is highlighted the following error message appears:
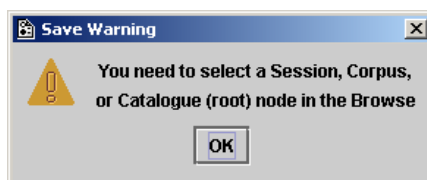
If you save an IMDI file for the first time, the "Save" dialog window appears. Browse to the folder where you want to save the file, assign a name to it, and then click on "Save".

> **!** Note: When you save an IMDI file, the IMDI Editor will check if the folder contains a file named **imdi.xsd**. If not, it will automatically create this file.

> **!** Note for researchers working at the MPI for Psycholinguistics, Nijmegen: The name of the IMDI file should be the same as the name of the corresponding annotation and media files (see section 2.5). The same name should also be entered in the field Session name (see section 2.1.1). Furthermore, the file name has to be Unix compatible: do not use file names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

1.2.1.12 "Save As …"

Click on "Save As" to save the currently active IMDI file (i.e., the file that is highlighted in the "Browse" panel; see section 1.3) under a different name (see also section 1.2.1.11).

1.2.1.13 "Save All"

Click on "Save All" to save all open IMDI files (see also section 1.2.1.11).

1.2.1.14 "Page Setup …"

Click on "Page Setup …" to specify the layout of the printed IMDI file (i.e., paper size, source, orientation, and margins).

1.2.1.15 "Print …"

Click on "Print …" to print the currently active IMDI file (i.e., the file that is highlighted in the "Browse" panel; see section 1.3).

1.2.1.16 "Exit"

Click on "Exit" to exit the "IMDI Metadata Editor" window. All open IMDI files will be closed. If some files are not yet saved, a warning appears that contains the following kind of information:

Checkmarks indicate the selected files (by default, all files are selected).

To remove a checkmark (i.e., to deselect a file), click in the box.

A list of all files that are not yet saved.

Click here to select or deselect all files.

Save All Selected Documents

☑ Ntii.imdi
☑ TreesSM.imdi
☐ Liit.imdi

Select All   Deselect All

Cancel   OK

Click "OK" to save all selected file. The deselected files will not be saved.

Click on "Cancel" to return to the Editor window.

1.2.2 "View" menu

The "View" menu displays information about the currently active IMDI file (i.e., the file that is highlighted in the "Browse" panel; see section 1.3). It contains the following options:

IMDI Metadata Edit...
File View Options Help

ⓘ Information ...   Alt-I
   File Prefixes ...
   Validate File ...
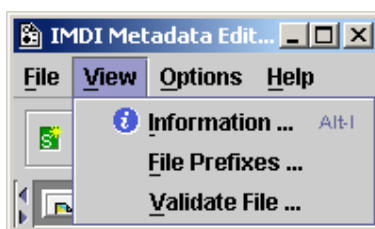
1.2.2.1 "Information …"

Click on "Information …" to view information about the currently active IMDI file. The following kind of information is displayed:

properties of the IMDI file (such as location, size, last modified, IMDI standard etc.)

information about whether the file is valid with respect to the XML schema

the prefixes that are used in the IMDI file (see section 1.2.3.4)

Information

Ntii.imdi

Path:

D:\WINNT\Personal\Software\Manuals\IM

| Information | Prefixes | Validation |

File Size: 12 kB

Last Modified: Dec 05 2002 09:17

Creation Date: Nov 08 2002

Originator:

Format ID: IMDI 1.8

...ype: SESSION

...sion: 2

1.2.2.2 "File Prefixes …"

Click on "File Prefixes …" to view the prefixes that are used in the IMDI file (see section 1.2.3.4).

1.2.2.3 "Validate File …"

Click on "Validate File …" to view information about whether the file is valid with respect to the XML schema.

1.2.3 "Options" menu

The "Options" menu allows you to set a number of preferences. It contains the following options:

IMDI Metadata Editor

File View Options Help

Profile Manager

Anonyms ...

Vocabularies ...

Editor

Preferences ... Alt-P
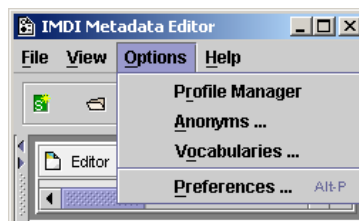
1.2.3.1 "Profile Manager"

Every IMDI file is assigned to a type and to a profile. These types and profiles are managed through the menu item "Profile Manager".

By default, there are three different types of IMDI files:

- Resource Bundles or Sessions that are used to describe audio/video and written resources (see section 2);

- Lexicon Resource Bundles that are used to describe lexicon resources (see section 3);

- Corpora that are used to describe larger corpora (see section 4).

Types are associated with one or more profiles. These profiles specify the available metadata categories together with their constraints. For example, members of the sign language research community have developed a "Sign-Language Profile". IMDI files that are based on this profile contain some sign-language-specific metadata categories and constraints, e.g.:

IMDI files based on the "Sign-Language Profile" contain the metadata category "Speaker/Signer".

This "Speaker/Signer" category contains keywords that are of relevance to sign-language research (e.g., "Deafness.Status" etc.).

For each keyword, there are constraints on its values. These constraints are visualized in the form of pull-down menus of controlled vocabulary items (e.g., "hearing", "hard-of-hearing" etc.).

To manage types and profiles, click on "Profile Manager. The following options are available:

Add a new profile:

> Click on the type where you want to add the profile. Then click on "Add" (to add a locally-stored file) or on "Add URL" (to add a file stored on the Web). A dialog window appears, allowing you to browse to the file or to enter its URL.

Delete a profile:

> Click on the profile that you want to delete. Then click on "Delete".

Select the default profile for a type:

> Click on the profile that you want to make the default profile for a type. Then click on "Make Default". From now on, this profile will automatically be used whenever you create a new IMDI file based on this type, i.e., for new Session files (see section 1.2.1.2), new Lexicon files (see section 1.2.1.3) or new Corpus files (see section 1.2.1.4). The non-defaults can still be accessed through the menu item "File / New …" (see section 1.2.1.1).

Click on "OK" to exit the "Profile Manager" window again.
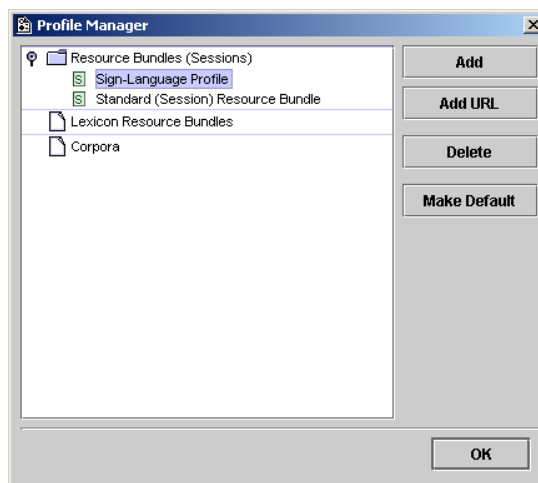
> **!** Note: Profiles are stored in the folder "user-directory \ IMDI-TOOLS \ Profiles \ local). For the moment, profiles have to be programmed by hand. In future versions of the IMDI tools, it will be possible to create profiles with the help of the IMDI Editor.

1.2.3.2   "Anonyms …"

"Anonyms …" is used whenever you have chosen to make the name of an actor anonymous, i.e., whenever you have chosen to enter a code (instead of the full name) into the field Full Name on the screen Actor (see section 2.4.1.1). If you have done so, the option "Anonyms …" allows you to create a file that maps this code to a full name. You (and everybody whom you grant access to this file) will be able to see the full name, while everybody else will only see the code.

To create such a file, do the following:

1. Click on "Options".
2. Click on "Anonyms …".

The "Anonyms overview" dialog window appears. It allows you to create, modify and remove mappings, and it displays all available mappings, e.g.:
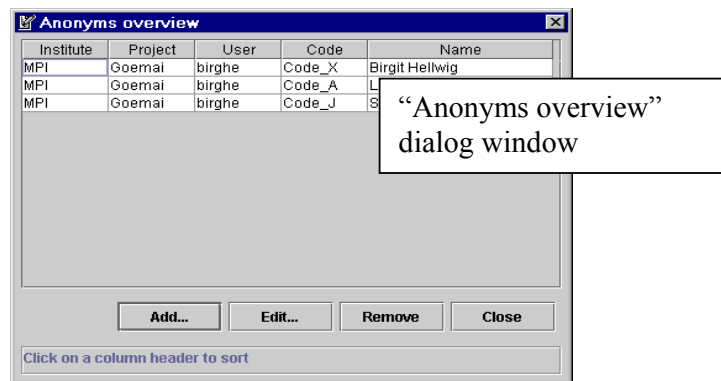
The following options are available:

- You can sort the available mappings. Click on any of the column headers to sort the corresponding column alphabetically.

- To add a mapping, click on "Add …". The "Anonyms input" dialog window appears. Please fill in the institute (e.g., MPI), the project (e.g., Goemai), the code of the actor and his/her full name.
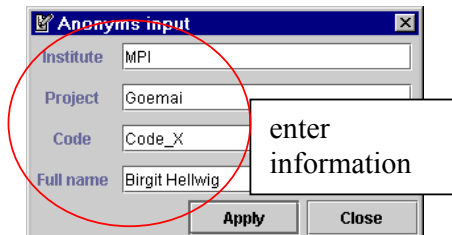
To save the mapping, click first on "Apply" and then on "Close".

The new mapping will be displayed in the "Anonyms overview" dialog window.

! Note: The code that you enter into the "Anonyms input" dialog window should be different from the code that you enter in the field Code on the Actor screen (see section 2.4.1). If the codes are identical, the IMDI Editor will also display the full name in the field Code on the Actor screen – not the intended code.

! Note: The column "User" (in the "Anonyms overview" dialog window) does not appear in the "Anonyms input" dialog window. This column is automatically filled in by the program.

- To edit a mapping, click on it to highlight it. Then click on "Edit …". The "Anonyms input" dialog window appears and displays the highlighted mapping. Type in your changes.
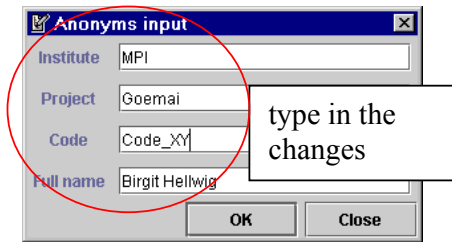
Click on "OK" to save the changes.

- To remove a mapping, click on it to highlight it. Then click on "Remove". The mapping is removed without further warning.

3. Click on "Close" to exit the "Anonyms overview" dialog window.

   All mappings are saved in the file **user-name.ano** in the folder "user-directory \ IMDI-TOOLS \ Anonyms", e.g.:
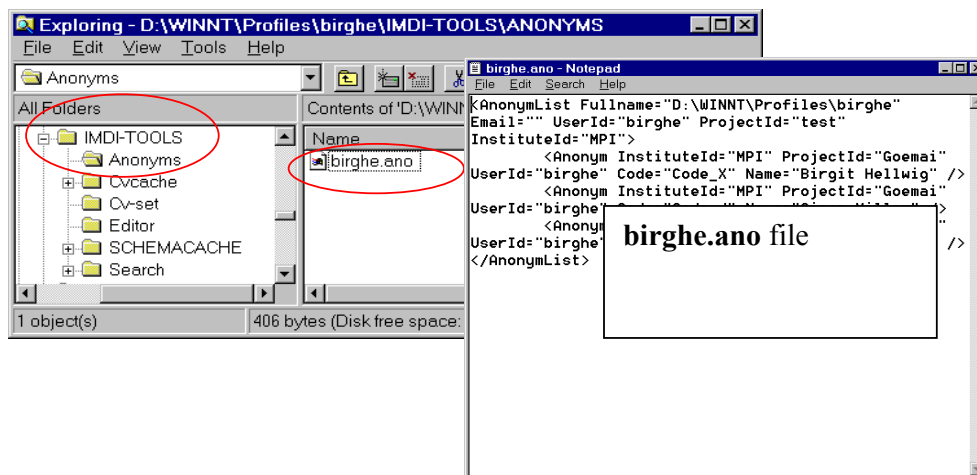
   If you want other people to have access to these mappings, send them a copy of the file **user-name.ano**. They should then save this file in their folder "user-directory \ IMDI-TOOLS \ Anonyms".

1.2.3.3 "Vocabularies …"

The IMDI tools make use of so-called "controlled vocabularies" to ensure consistency across users (i.e., they provide lists of values for specific metadata categories or keys). These vocabularies are stored on the Web, and a copy of them is stored in the cache of your desktop computer (in the folder "user-directory \ IMDI-TOOLS \ CVCACHE") (see the IMDI CV-Editor manual for details).

You can view these vocabularies through the menu item "Vocabularies …". Click on it to access the "Vocabs overview" window, e.g.:
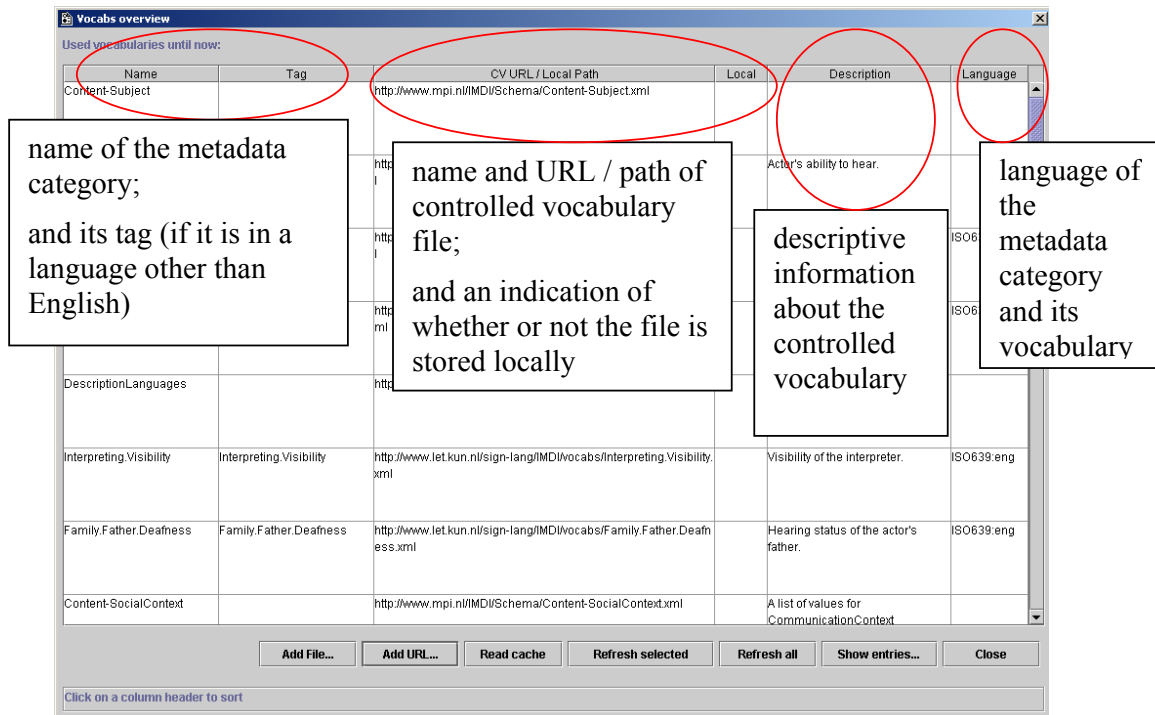
The "Vocabs overview" window offers you the following options:

- You can sort the available controlled vocabularies. Click on any of the column headers to sort the corresponding column alphabetically.

- "Add File …"

   Click on "Add File …" to add a controlled vocabulary file that is stored locally. The "Browse" window appears, allowing you to browse to the folder and to select the file. The vocabulary will then be added to the "Vocabs overview" window.

- "Add URL …"

   Click on "Add URL …" to add a controlled vocabulary file that is stored on the Web. The "Vocab add" window appears. Enter the URL of the vocabulary, click on "Apply" and then click on "OK". The vocabulary will be added to the "Vocabs overview" window.
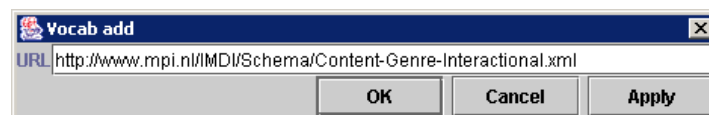
- "Read cache"

  Click on "Read cache" to view all available controlled vocabularies (i.e., all vocabularies that are stored in the folder "user-directory \ IMDI-TOOLS \ CVCACHE").

- "Refresh selected"

  Click on a vocabulary to select it, then click on "Refresh selected" to have the IMDI tools check for a newer version of that vocabulary on the Web, and to update the cache accordingly.

- "Refresh all"

  Click on "Refresh all" to have the IMDI tools check for newer versions of all vocabularies on the Web, and to update the cache accordingly.

- "Show entries …"

  Click on a controlled vocabulary to select it, then click on "Show entries …" to view its content., e.g.:
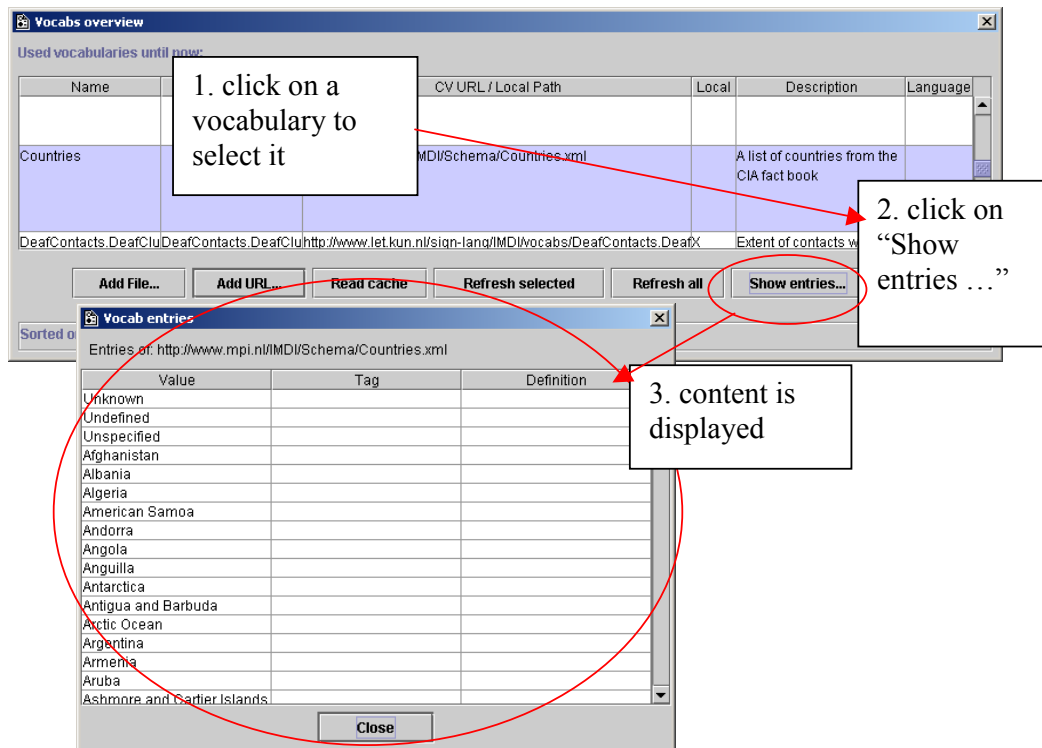
- "Close"

  Click on "Close" to close the "Vocabs overview" window.

1.2.3.4 "Preferences …"

Click on "Preferences …" to access the "Preferences" dialog window.

In the "Preferences" dialog window, the following four options are available:

(1) "Prefixes"

(2) "Resources"

(3) "Update Policy"

(4) "Font"

Click on any of the headings in the left panel to activate the corresponding option.
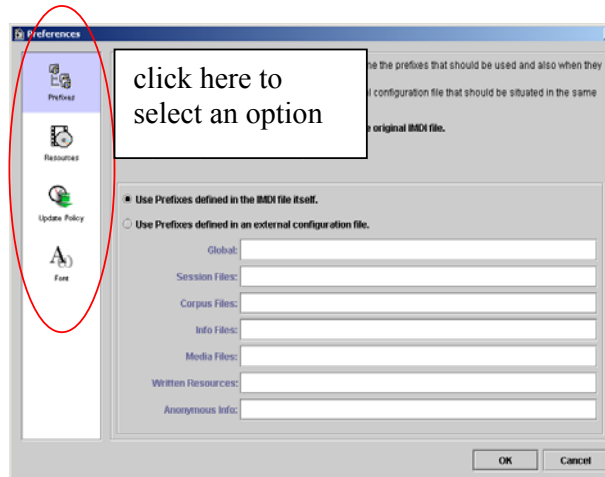
"Prefixes"

This option allows you to specify the path to the different types of files (see section 4 for details).

"Resources"

This option allows you to specify the browse options (see section 4 for details).

"Update Policy"

The IMDI tools make use of so-called "controlled vocabularies" (see section 1.2.3.3). These vocabularies are stored on the Web, and a copy of them is stored in the cache of your desktop computer (in the folder "user-directory \ IMDI-TOOLS \ CVCACHE"). The Web version is changed sometimes, but the IMDI tools will continue to use the locally-stored version until the cache is updated again. By default, the cache is updated whenever you run an IMDI tool. However, you can change this policy, and determine if and when the vocabularies should be updated.

Click on "Update Policy" to determine this policy. The following options are available:
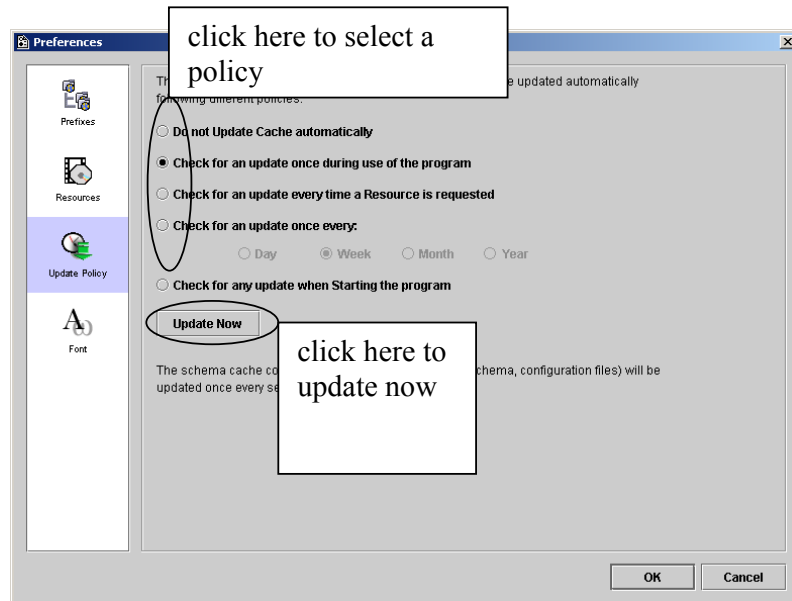
"Font"

This option allows you to display special character sets and fonts. To select a set and a font, do the following:

1.   Under "Unicode Block" (in the left panel), select a unicode block by clicking on it. The available fonts are then displayed under "Font" (in the right panel).

2.   Under "Font", select a font by clicking on it.

3.   Click on "OK" to implement the new font.
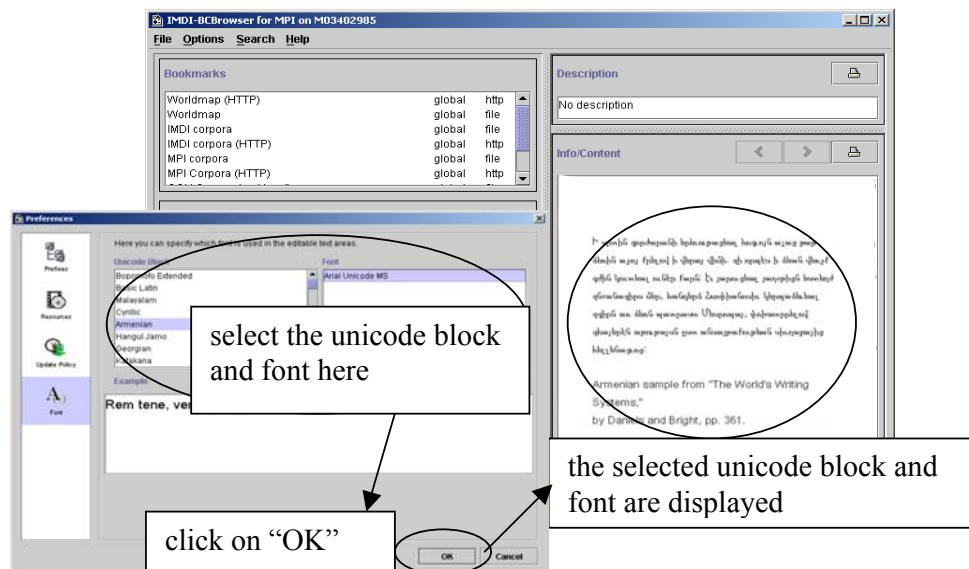
1.2.4 "Help" menu

Click on "Help" to view information about the IMDI Editor, e.g.:
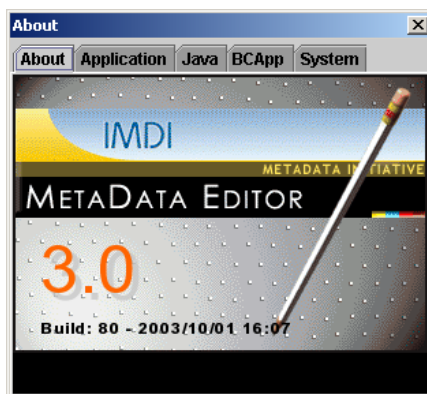
1.3 The "Browse" panel

The "Browse" panel allows you to browse through one (or more) IMDI (Session, Lexicon or Corpus) files. It displays all currently open IMDI files, e.g.:
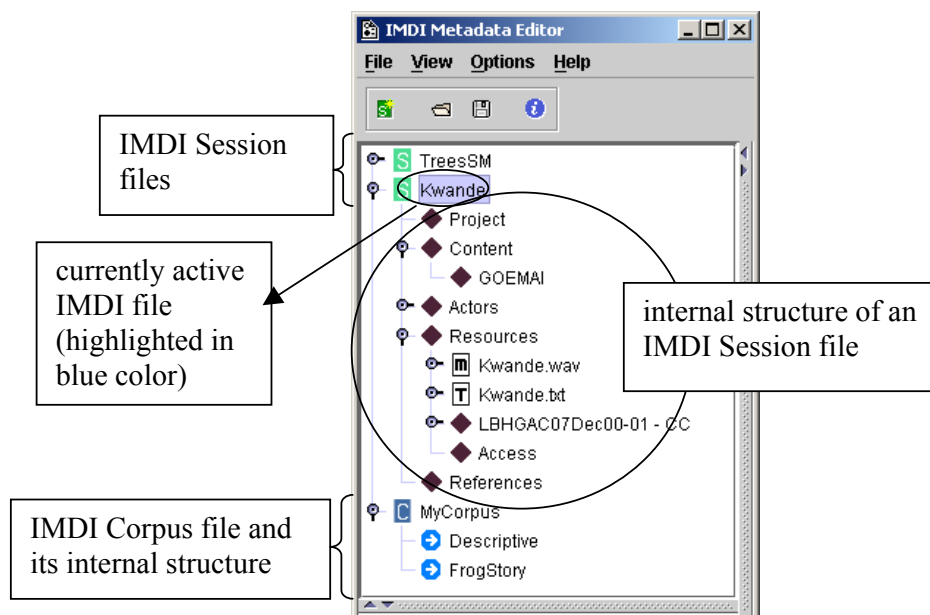
The "Browse" panel makes use of the following icons:

[S]   session node (corresponding to an IMDI Session or Lexicon file, see sections 2 and 3)

[C]   corpus node (corresponding to an IMDI Corpus file, see section 4)

◆   metadata node

[m]   media file node

[T]   written resource (or "transcription") file node

[L]   lexicon node

➜   corpus link node

Under each session or corpus node, the structure of the corresponding IMDI (Session, Lexicon or Corpus) file is displayed in the form of metadata nodes, media files nodes, written resource file nodes, lexicon file nodes and/or corpus link nodes.

The nodes of the "Browse" panel interact in the following ways with the screens of the "Content" panel (see section 1.5):

(a) Click on any node in the "Browse" panel to make the corresponding IMDI file the currently active IMDI file. It will be highlighted in blue color, and its metadata information will be displayed in the "Content" panel, e.g.:

"Browse" panel:

"Project" node of the file "TreesSM.imdi"

"Content" panel:

Project screen (of the file "TreesSM.imdi")
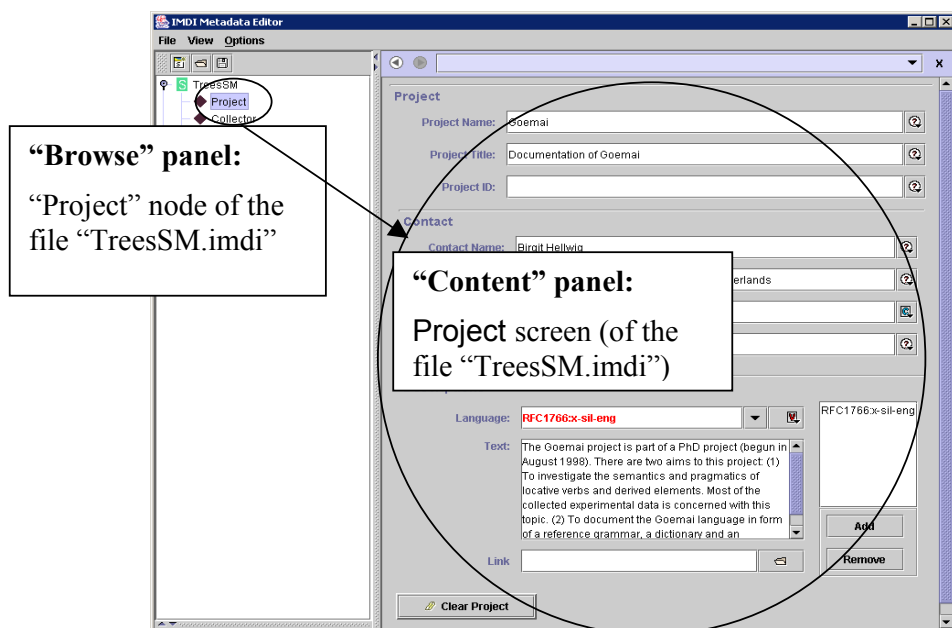
(b) Some nodes in the "Browse" panel allow you to add further sub-nodes, which are then displayed as screens in the "Content" panel. For example, the node "Actors" allows you to add a sub-node for each actor. To add a sub-node, do one of the following:

Click on a node in the "Browse" panel. It is highlighted in blue color.

Then click with the right-mouse button on the highlighted node. A pull-down menu appears.

Select "Add" from the pull-down menu.

Alternatively, click on an "Add" button in the "Content" panel.

The added sub-node is displayed in the "Browse" panel

When you click on the added sub-node, the "Content" panel displays the additional screen
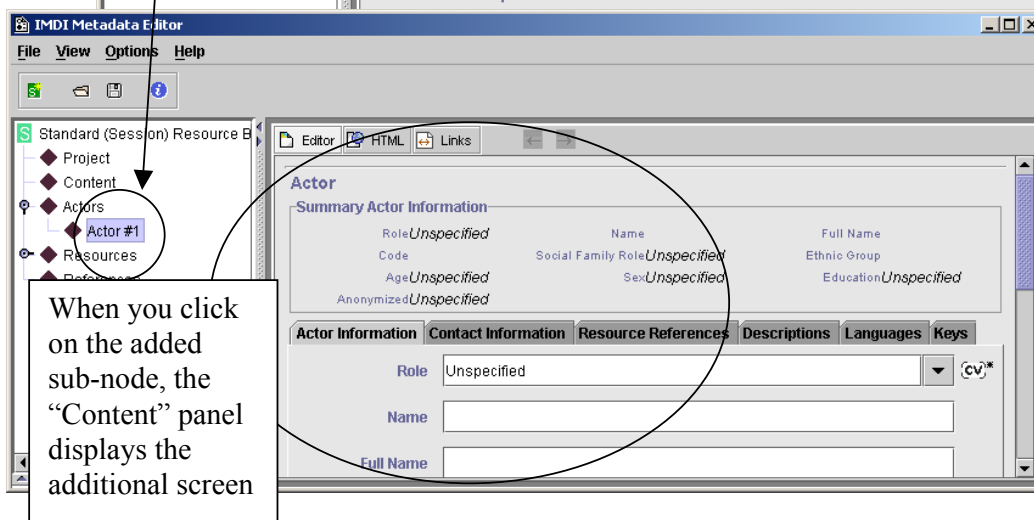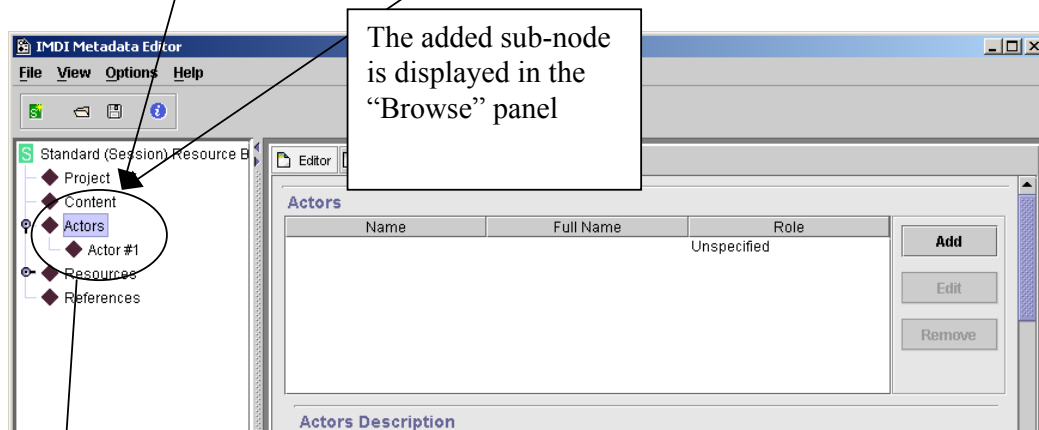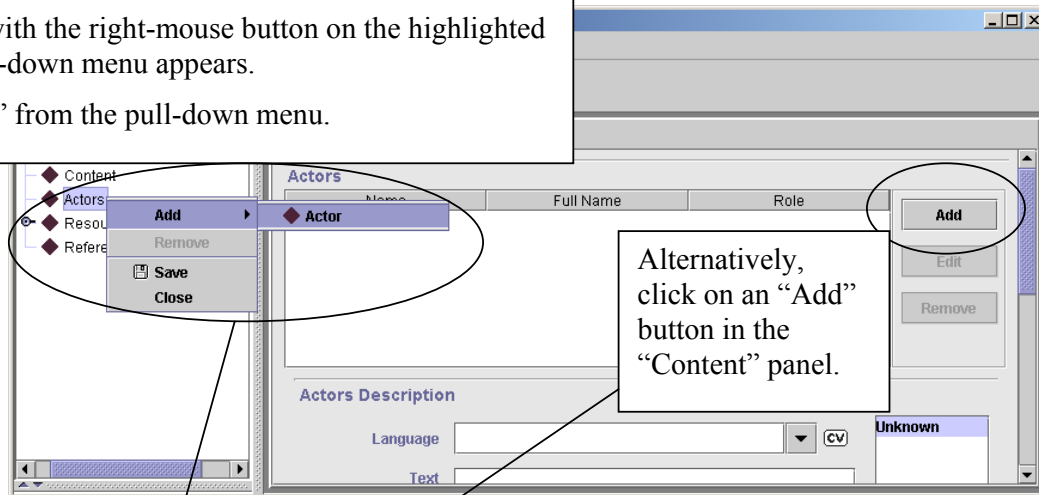
You can insert the new sub-node anywhere under an appropriate top node. To select its position, do one of the following:

- Add the sub-node (see above), and click on it in the "Browse" panel, keep the mouse button clicked and drag it to its new position. A line will appear at this position. Release the mouse button, and the sub-node will be inserted in its new position, e.g.:
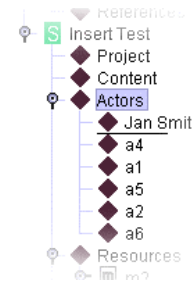
- Alternatively, do the following:

(1) In the "Content' panel, click on the sub-node before which you want to insert the new sub-node. It is highlighted in blue color.

(2) Click on "Add".

(3) The new sub-node is added before the selected sub-node.
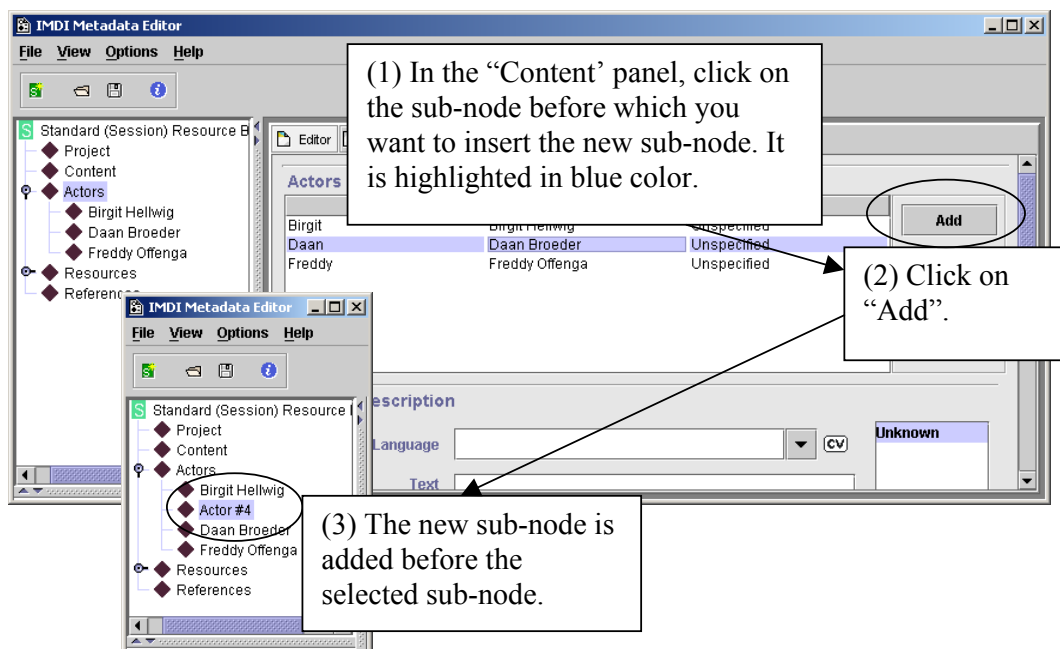
! Note: The IMDI Editor makes use of different "Add" buttons for different purposes, and not all of them add new sub-nodes. This function is only available for those nodes that allow for the three options "Add …", "Edit …", and "Remove".

(c) Added sub-nodes can be removed again. Do one of the following:

In the "Browse" panel, click on the sub-node that you want to remove. It will be highlighted in blue color. Then click with the right-mouse button on the highlighted node. A pull-down menu will appear. Select "Remove" from the pull-down menu.

Alternatively, in the "Content" panel, click on the sub-node that you want to remove. Then click on the "Remove" button.

1.4 The "Repository" panel

The "Repository" panel allows you to work with recurring metadata information. Such information is entered once into the "Content" panel (see section 1.5), and is then stored in the "Repository" panel for future usage.

For example:

information about the language "GOEMAI" is entered into the "Content" panel

this information is then stored in the "Repository" panel
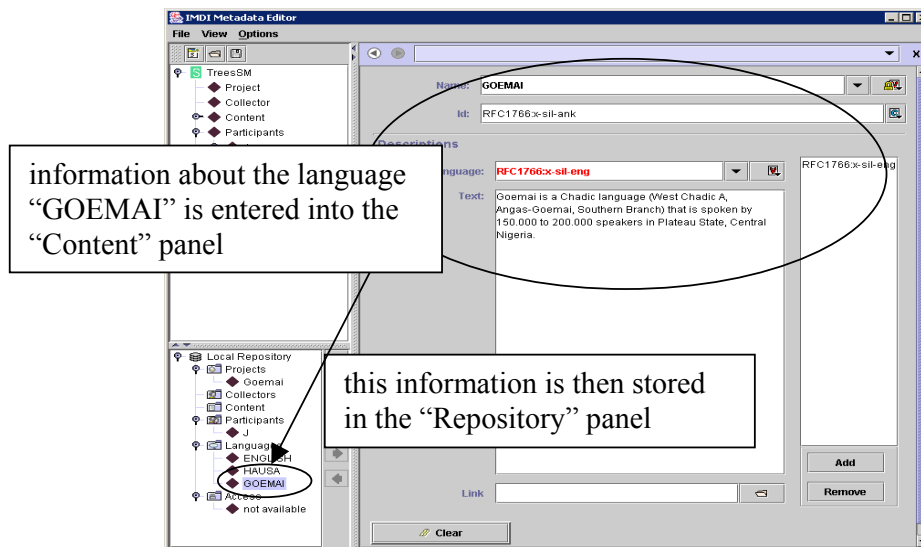
! Note: The repository is stored under the name **local.repository.xml** in the folder "user-directory \ IMDI-TOOLS \ SDRCache". The files that contain the actual metadata information are automatically assigned a name and are then stored in the same directory, e.g.:

local repository, e.g.:

files that store the metadata information, e.g.:
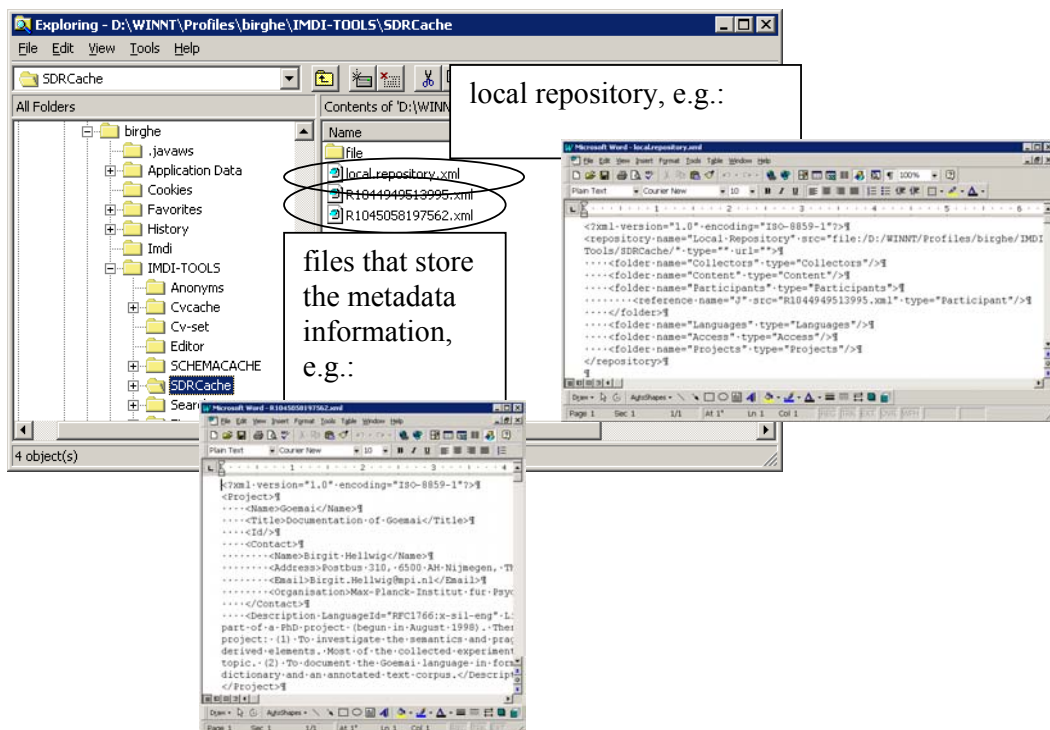
The "Repository" panel displays a number of buttons and a pull-down menu (which is accessed by clicking on a node, first with the left mouse button to select it and then with the right mouse button to open the pull-down menu). The buttons and the pull-down menu are used for the following purposes:

- Storing metadata information in the repository (see section 1.4.1).

- Deleting metadata information from the repository (see section 1.4.2).

- Managing the structure of the repository (see section 1.4.3).

- Exporting nodes from the repository (see section 1.4.4).

- Importing nodes into the repository (see section 1.4.5).

- Updating the information in the repository (see section 1.4.6).

- Using metadata information from the repository (see section 1.4.7).

1.4.1   Storing metadata information in the repository.

To store metadata information in the repository, do one of the following:

- Drag and drop a node from the "Browse" panel to the appropriate node in the "Repository" panel. For example, to add the language "GOEMAI" to the repository, do the following:

1. Click on the node in "Browse" panel.

2. Keep the mouse button clicked and drag it to the appropriate node in the "Repository" panel.

3. Release the mouse button. The node is added to the repository.
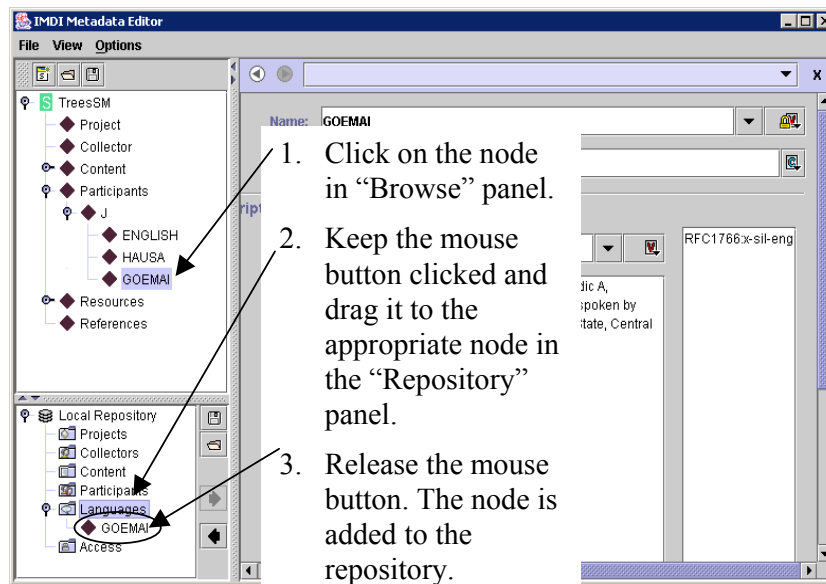
- Alternatively, make use of the "Get Data" option in the "Repository" panel. For example, to add the actor "J" to the repository, do the following:

1. Click on the sub-node in "Browse" panel.

2. Click on the appropriate top node in the "Repository" panel.

3. Make use of the "Get Data" option:

   (a) Either click on the "Get Data" button.

   (b) Or right-click on the highlighted node, and then click on the "Get Data" item in the pull-down menu.

In both cases, the node is added to the repository.
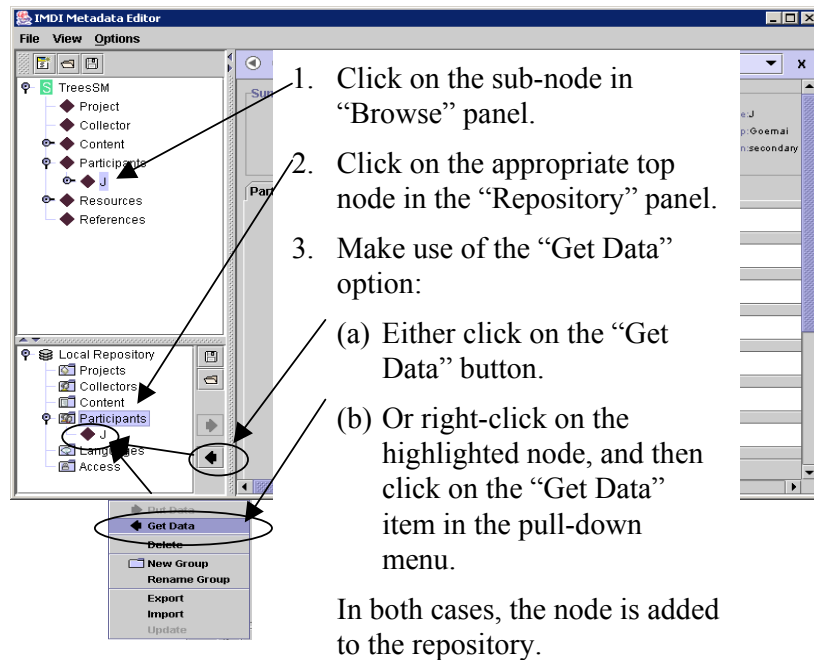
! Note: You can only add a sub-node from the "Browse" panel to an appropriate corresponding top node in the "Repository" panel, e.g., a language can only be added to a language-type node, and an actor can only be added to an actor-type node. If the two nodes (of the "Browse" and "Repository" panels) do not match, the program will automatically disable the "Drag and drop" and the "Get Data" functions.

1.4.2 Deleting metadata information from the repository.

To delete metadata information, do the following:

1. In the "Repository" panel, click on the node that you want to delete. The node will be highlighted in blue color.

2. Click with the right mouse button on the highlighted node. A pull-down menu appears.

3. In the pull-down menu, click on "Delete". After a warning, the highlighted node is deleted.

1.4.3 Managing the structure of the repository.

The following two management options are available:

- Add a new repository group. Do the following:

1. Click on the node where you want to add the new group.

2. Right-click on that node.

3. Choose "New Group" from the pull-down menu.

4. Enter the name of the new group. The group will automatically be assigned to the right type/node.

5. Click on OK.

The new group is added to the repository.
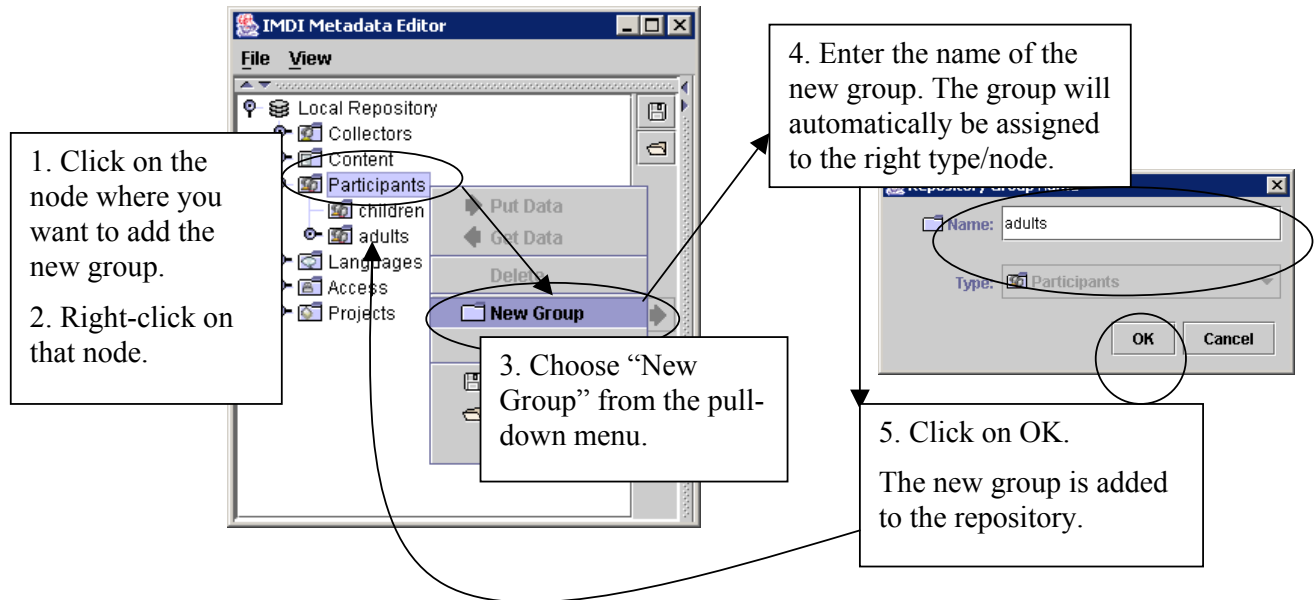
- Rename a repository group. Do the following:

1. Click on the node that you want to rename.

2. Right-click on that node.

3. Choose "Rename Group" from the pull-down menu.

4. Enter the new name.

5. Click on "OK".

The group is renamed.
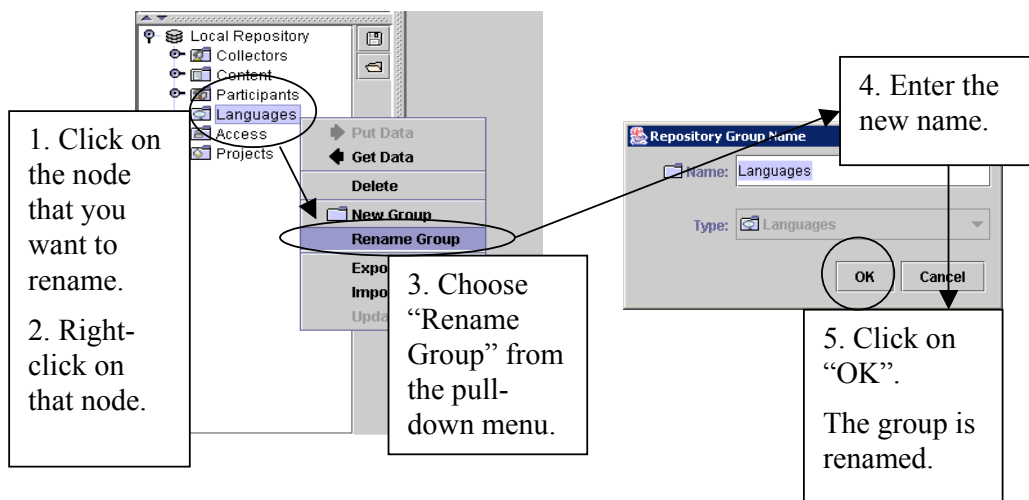
1.4.4 Exporting nodes from the repository.

You can export a node from the repository. Do the following:

1. Click on the node that you want to export. It will be highlighted in blue color.

2. Make use of the "Export" options:

(a) Either click on the "Export" button, e.g.:

"Export" button.
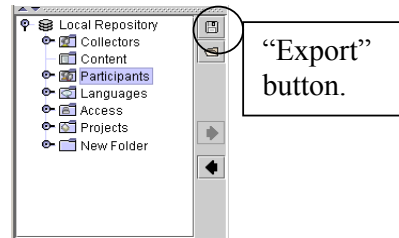
(b) Or right-click on the highlighted node, and then click on the "Export" item of the pull-down menu, e.g.:

"Export" item.

The "Export Repository" dialog window appears.

3. Give the name and location of the file that should contain the exported node. Then click on "OK" (or press the key ENTER).

1. Enter a name.

2. Browse to the location.

3. Click on "OK".
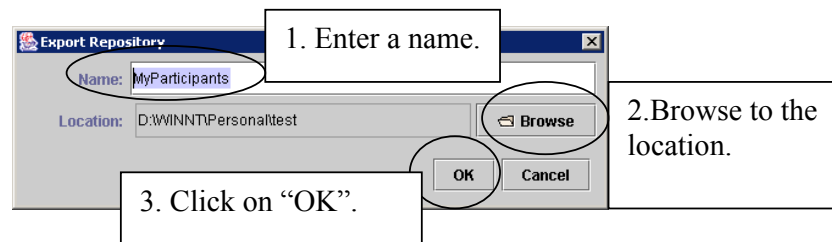
The file is saved in the specified folder. Furthermore, the exported repository is displayed as an additional node in the "Repository" panel, e.g.:

The files are saved in the specified folder.

The exported file is added to the "Repository" panel.
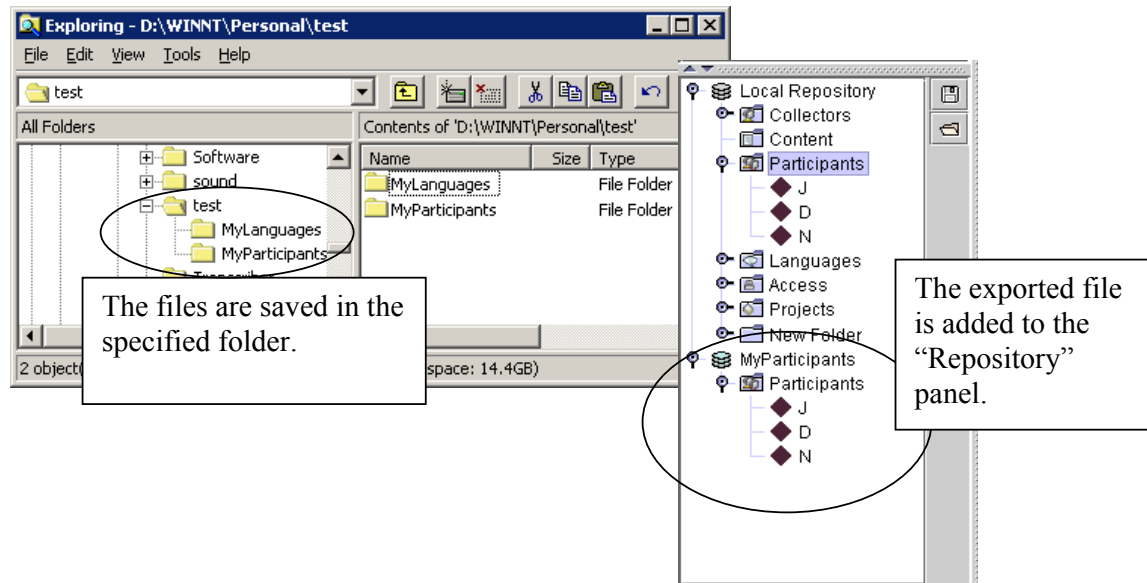
1.4.5   Importing nodes into the repository.

You can import a node into the repository. In this case, the IMDI Editor creates a copy of the original file and stores this copy in the folder "user-directory \ IMDI-TOOLS \ SDRCache".

> **!** Note: If you have old templates that were created with an earlier version of the IMDI Editor (i.e., earlier than version 2.0), you need to import them into the "Repository" panel.

To import a node, do the following:

1. Make use of the "Import" options:

    (a) Either click on the "Import" button, e.g.:

"Import" button.
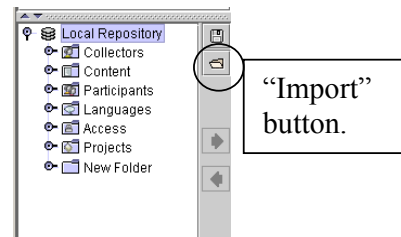
(b) Or right-click somewhere in the "Repository" panel, and then click on the "Import" item of the pull-down menu, e.g.:

| Put Data |
| Get Data |
| **Delete** |
| 🗀 **New Group** |
| **Rename Group** |
| Export |
| **Import** |
| Update |

"Import" item.
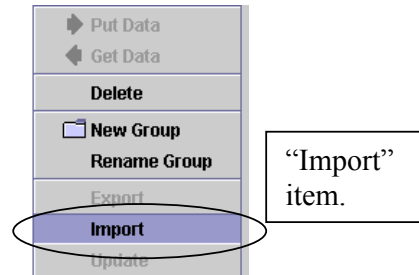
The "Import Repository" dialog window appears.

2. Browse to the file that contains the repository, and click on "OK".

🌺 **Import Repository** ⊠

Location: WINNT/Personal/test/MyParticipants/repository.xml 🗁 **Browse**

Enter the name and location of the repository.

OK   Cancel
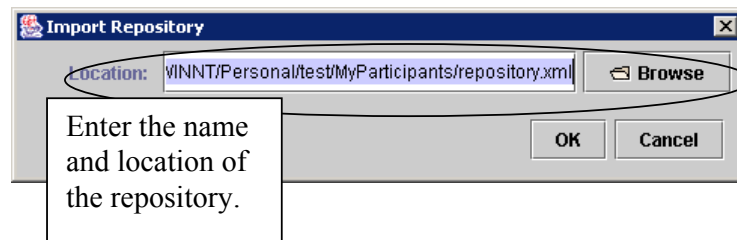
The file is displayed as an additional node in the "Repository" panel, e.g.:

Local Repository
Collectors
Content
Participants
Languages
Access
Projects
New Folder
MyParticipants

The imported file is added to the "Repository" panel.
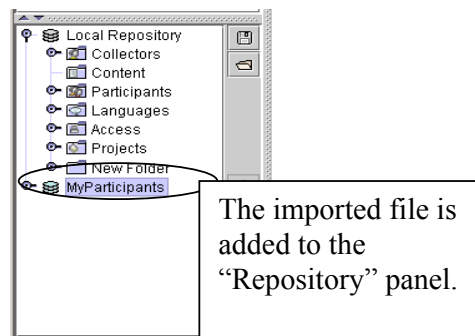
1.4.6 Updating the information in the repository.

This option allows you to update an imported repository (i.e., the copy of the file, see section 1.4.5) after the original file has been modified.

Do the following:

1.   Click on the imported repository. It will be highlighted in blue color.

2.   Right-click on that repository. A pull-down menu appears.

3.   Click on "Update". The information in the repository is updated.

1.4.7 Using metadata information from the repository.

To use metadata information from the repository, do one of the following:

- Drag and drop a sub-node from the "Repository" panel to the appropriate top node in the "Browse" panel. For example:
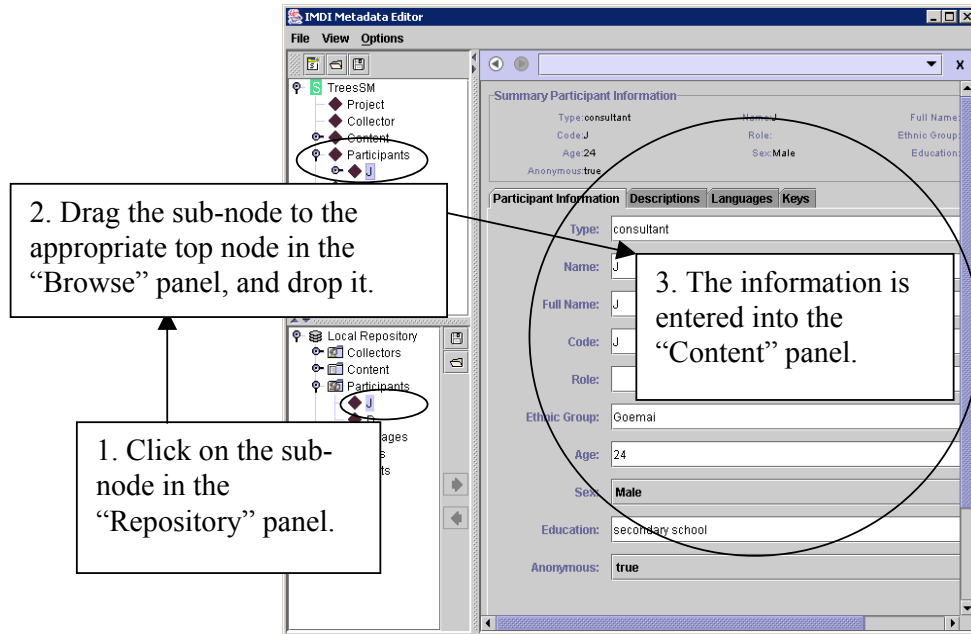
- Make use of the "Put Data" option in the "Repository" panel. For example, to insert information about access rights, do the following:
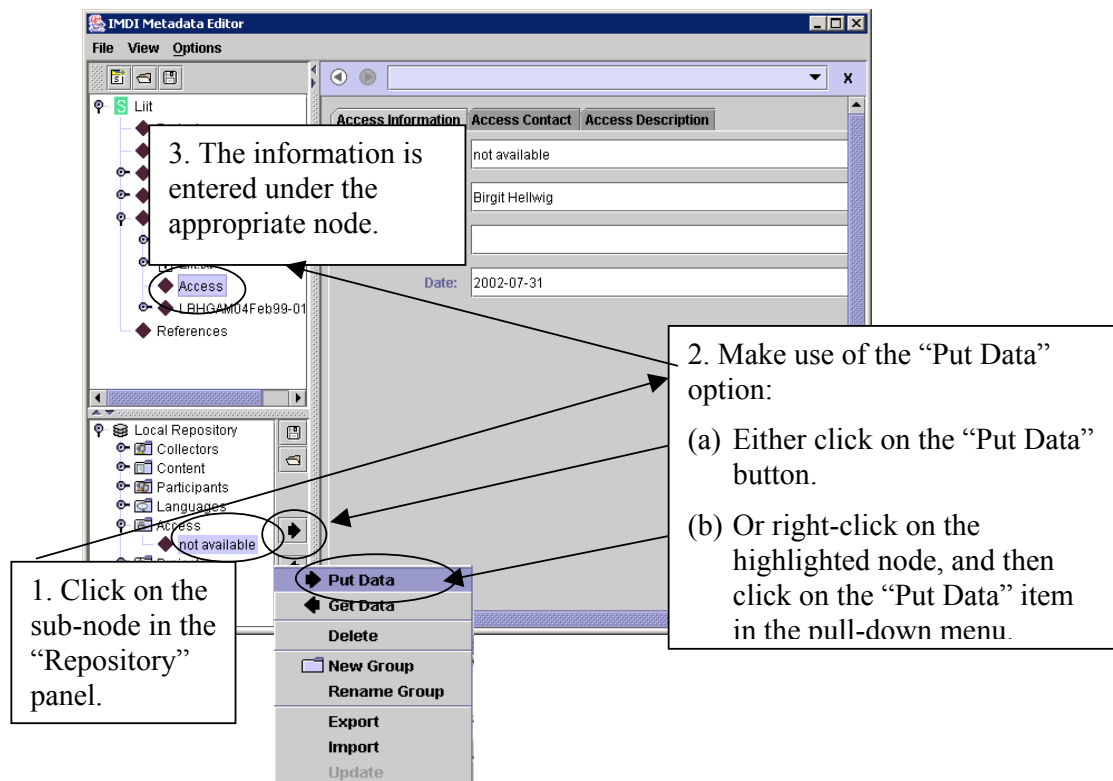
! Note: You can only insert information into the appropriate corresponding node of the "Browse" panel (and the corresponding screen of the "Content" panel), e.g., information about an actor can only be inserted in the "Actors" node of the "Browse" panel (and the screen Actors of the "Content" panel). If a non-matching node is active in the "Browse" panel, the program will automatically disable the "Drag and drop" and the "Put Data" functions.

! Note: You can add a group of nodes to the "Browse" panel, e.g., a group of actors. In this case, you have to "drag and drop" or "put" the higher node from the "Repository" panel into the "Browse" panel (e.g., the node "Actors").

! Note: You can overwrite an exiting node by dragging a node from the "Repository" panel onto a node of the "Browse" panel (e.g., by dragging actor X onto the node actor Y). In this case, the following warning will appear:
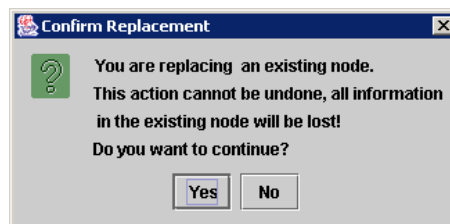
Click "Yes" if you want to replace the node, otherwise click "No".

1.5 The "Content" panel

The "Content" panel displays the screens (see section 1.5.1) and the schemata and fields (see section 1.5.2), which are used for entering metadata information. And it allows you to specify links to media, written resources, lexicon, info and metadata files (see section 1.5.3).

After you have entered metadata information, you can choose to view this information either in the IMDI Editor format (the default view) or as an HTML document. Click on the buttons at the top of the "Content" panel to switch between the two views.
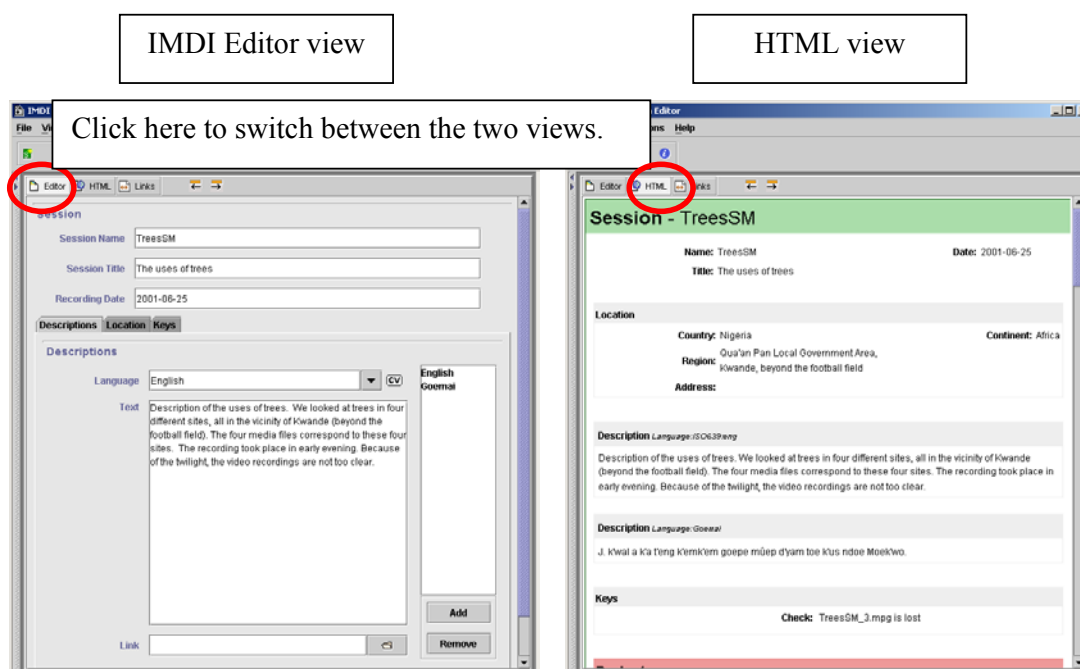
1.5.1 Screens

The "Content" panel displays a number of screens. The following main screens are available:

IMDI Session Editor (see section 2) and IMDI Lexicon Editor (see section 3):

- Session   general information about the session (see section 2.1);
- Project   information about the project within which the session data was collected (see section 2.2);
- Content   information about the content of the session (see section 2.3);
- Actors   information about the actors participating in the session (see section 2.4);
- Resources   information about written resource files, media files, and non-digital sources associated with the session (see sections 2.5), or about lexicon files (see section 3.1);
- References   cross-references to any material that is relevant to the content of the session (see section 2.6).

IMDI Corpus Editor (see section 4):

- **Corpus** links to IMDI Session, IMDI Lexicon and IMDI Corpus files.

These main screens are accessed through clicking on the corresponding node in the "Browse" panel (see also section 1.3).

Click on a node in the "Browse" panel.

The corresponding screen is displayed in the "Content" panel.
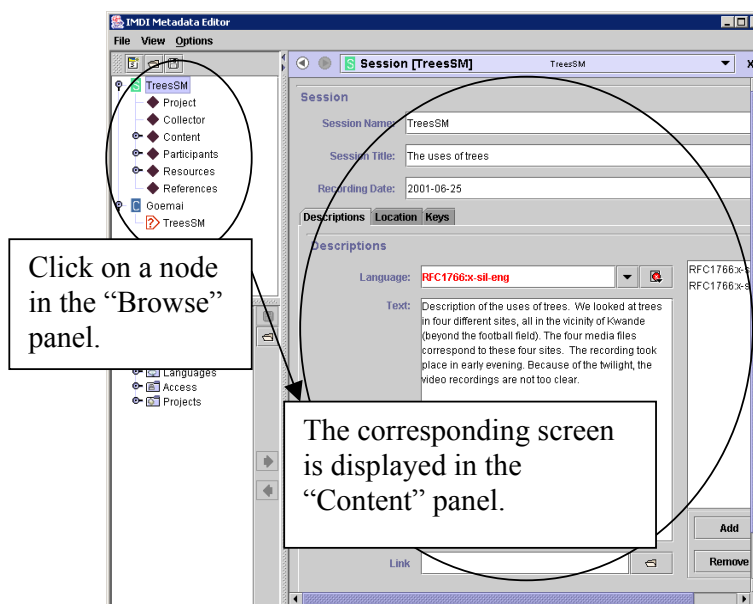
The IMDI Editor screens are structured in the following ways:

- All screens display schemata and fields into which the metadata information is entered (see section 1.5.2).

- Some screens contain sub-screens that are displayed like filing cards. These sub-screens give access to additional schemata and fields.

schema

fields

sub-screen headers
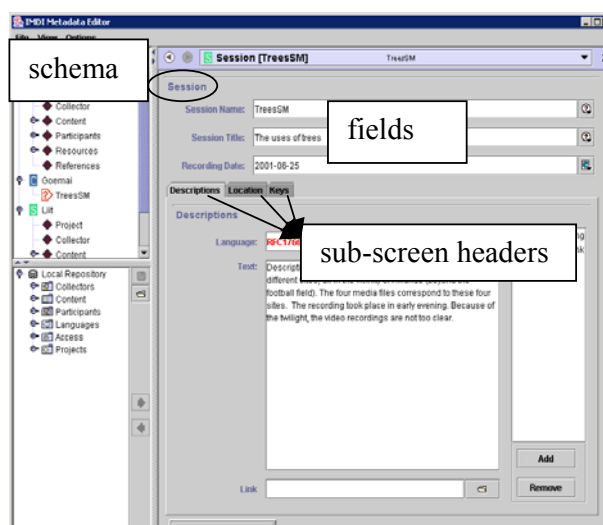
- Some screens have "Add …" buttons (and corresponding "Remove" and "Edit …" buttons) that allow you to enter multiple types of information of the same kind, e.g., one sub-screen added for each language spoken by an actor. These additional sub-screens are then displayed as separate sub-nodes in the "Browse" panel (see also section 1.3), e.g.:

To add a new sub-screen, click on "Add …".

To modify a sub-screen, click on it to highlight it, and then double-click on it, or click on "Edit …".

To remove a sub-screen, click on it to highlight it, and then click on "Remove"

3 sub-screens displaying 3 languages

There are two options for navigating through the "Content" panel:

(1) You can use the mouse: click with the mouse into a field (to enter information), on a sub-screen header (to activate the corresponding sub-screen) or on a button (to execute the corresponding command).
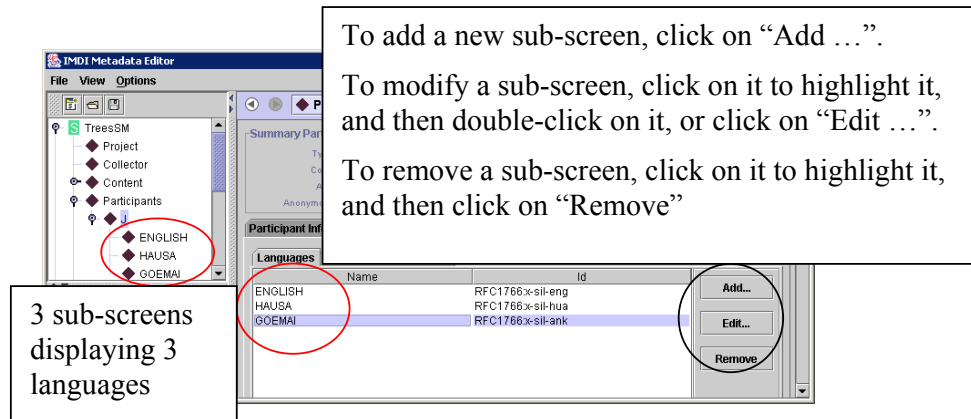
(2) Alternatively, you can use the following shortcut keys.

- Press TAB to move to the next field or button.

- Press SHIFT+TAB to move to the previous field or button.

- When a button is highlighted, press SPACE to execute the corresponding command.

- Press the left or right arrow key to move from sub-screen to sub-screen header.

1.5.2 Schemata and fields

Screens are made up of different units, so-called 'schemata'. Each schema contains a number of boxes into which the actual information is entered. These are called 'fields', e.g.:
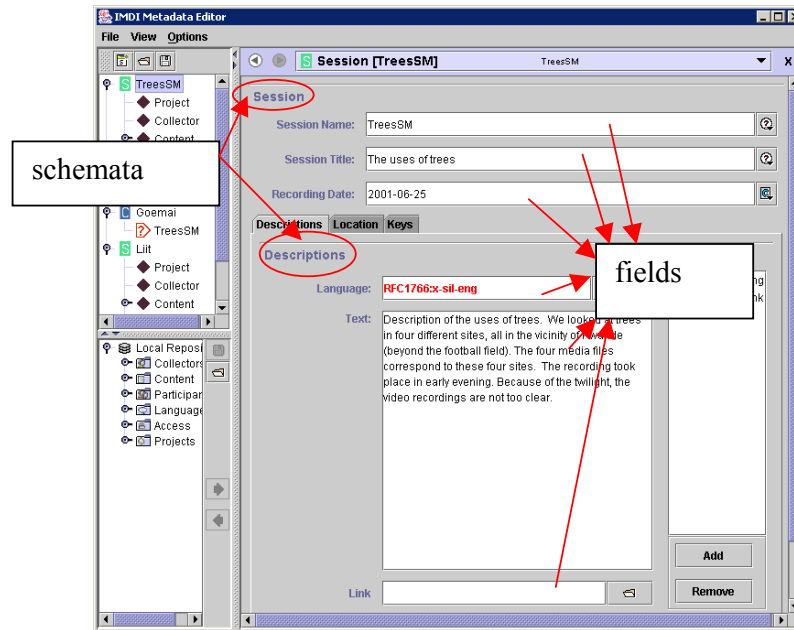
Most of the fields are more or less standardized.[4] The following symbols and options are available:

[CV]	'closed controlled vocabulary'	You can enter only one value, and this value must be selected from the pull-down menu.
[CV]*	'closed controlled vocabulary list'	You can enter more than one value (separated by commas), but all values must be selected from the pull-down menu.
(CV)	'open controlled vocabulary'	You can enter only one value. You can either select this value from the pull-down menu, or type in an alternative value.
(CV)*	'open controlled vocabulary list'	You can enter more than one value (separated by commas). You can either select these values from the pull-down menu, or type in alternative values.

In all four cases above, you can choose a value directly from the pull-down menu. Alternatively, you can start typing, in which case the pull-down menu will automatically open to display the available values.

[Con]	'constrained format'	The value must be entered in a certain format (e.g., a date must be entered in the format **YYYY-MM-DD**). As soon as you start typing, the format is displayed in the field (highlighted in blue color). Please type over this format.

---

[4] Please see the document "IMDI Metadata Elements for Session Descriptions" under http://www.mpi.nl/IMDI/ for updated lists of controlled vocabularies, their values and their definitions.

For all fields, there are tool-tips available. Point with the mouse to the label of the metadata category and leave it there for one second. The tool-tip appears, giving you information about this category, e.g.:
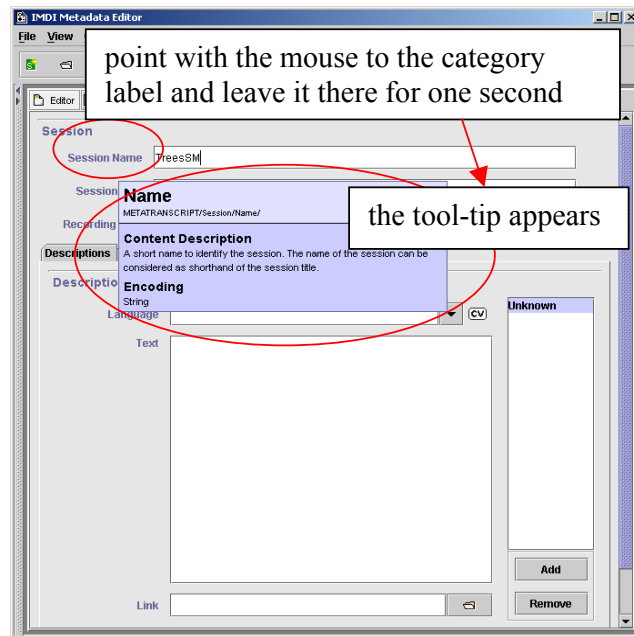
In addition to these standardized fields, the IMDI Editor allows for the possibility to enter project-specific information. Such information is entered into a Keys schema (see "Appendix 1: Recurring Schemata (Keys)" on instructions of how to fill in a Keys schema), e.g.:
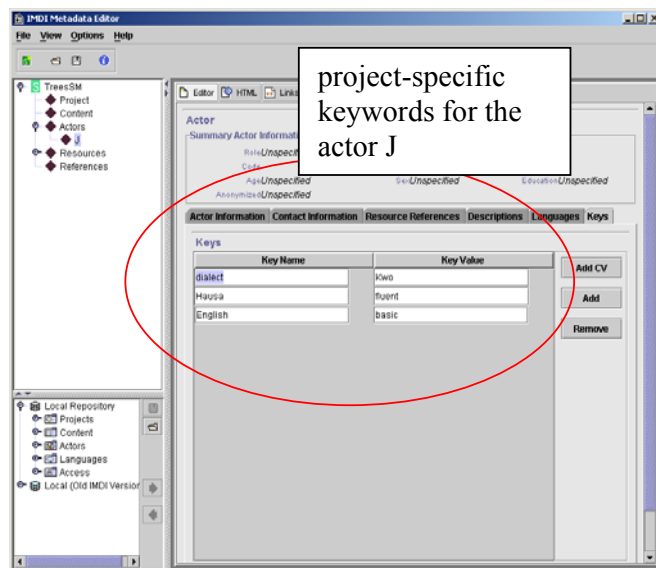
! Note: It is possible to use (project-specific) controlled vocabularies in a Keys schema, i.e., to have a Keys schema display a pull-down menu containing predefined values. See "Appendix 1: Recurring Schemata (Keys)" for details.

In addition to the standardized fields and the keyword fields, there are Descriptions schemata (see "Appendix 1: Recurring Schemata (Descriptions)" on instructions of how to fill in a Descriptions schema). Such a schema contains a prose description that could serve as a reminder (to you or others) of the circumstances of data collection.

1.5.3 Links

In addition to the information that you enter directly into an IMDI file, you can create links to other files, i.e., to media files (see section 2.5.2), written resource files (see section 2.5.3), lexicon files (see section 3.1), info files (see "Appendix 1: Recurring Schemata (Descriptions)") and other IMDI files (see section 4). Such links are specified in the IMDI file itself, and the corresponding files can then be accessed through the IMDI Browser, e.g.:
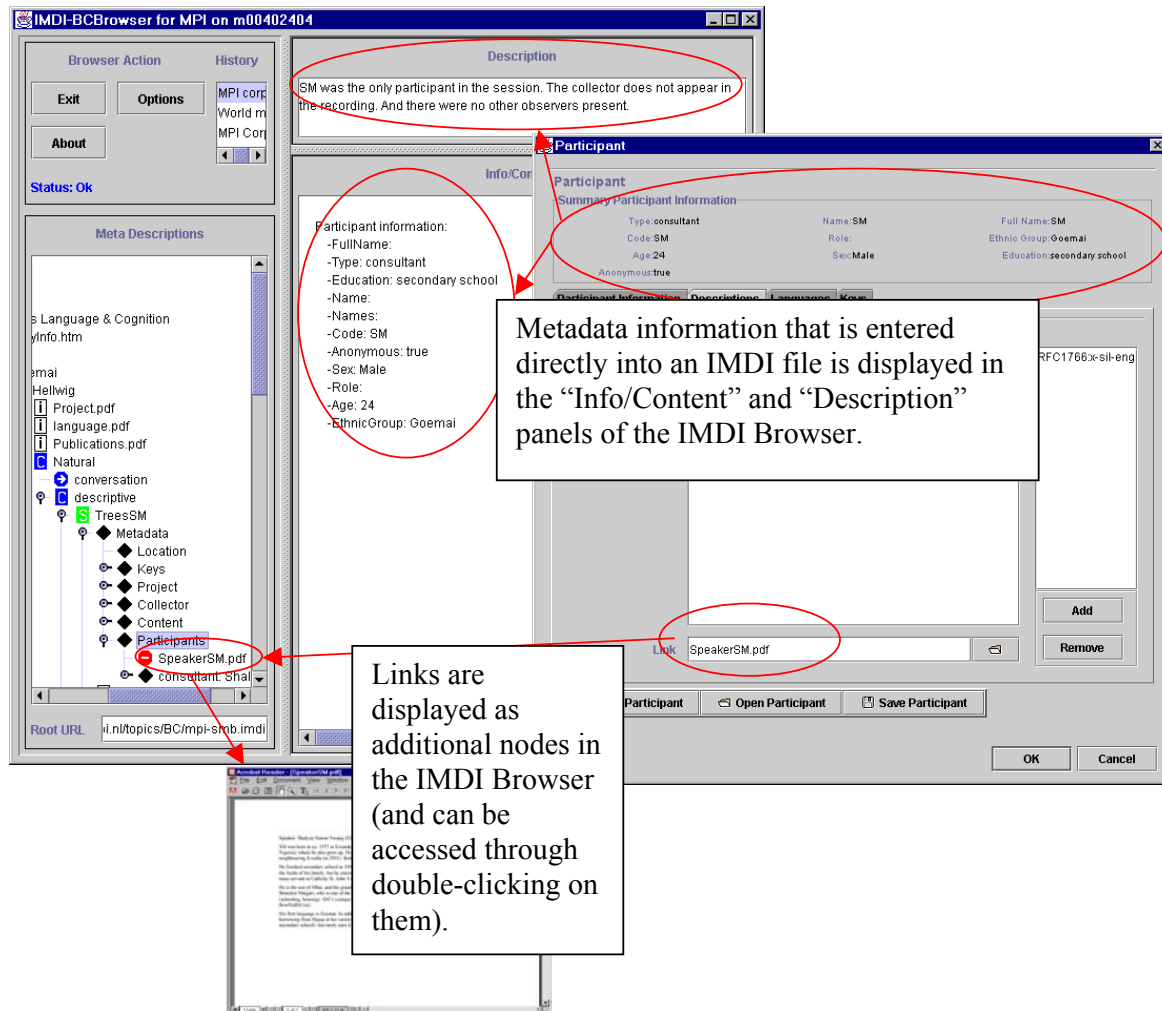
! Note for researchers working at the MPI for Psycholinguistics, Nijmegen: Please specify only the name of such a file – do **not** make use of the "open folder" icon to the right of the *Link* field.

1.6 Changes to the user interface (version 2.0)

The following changes to the user interface of the IMDI Editor were implemented in version 2.0:

| Version 2.0	Functionality	corresponding item in Version 1.5 (and earlier)
"Browse" panel	displays all currently open files	"Windows" menu item
	displays the internal structure of an IMDI file	screen headers
"Repository" panel	stores recurring information	templates

2 The IMDI Session Editor

This part of the manual guides you through the metadata categories of the IMDI Session Editor, as defined in the IMDI 3.0 standard. Note that the IMDI Editor also allows you to create project-specific profiles, containing additional categories. However, these additional categories are entered into Keys schemata, and they do not form part of the IMDI 3.0 standard (see section 1.2.3.1 for details on profiles; and see "Appendix 1: Recurring Schemata (Keys)" for details on Keys schemata). They are thus not discussed in this section.

The IMDI Session Editor creates IMDI Session files (***.imdi**) that describe audio/video and written resources. Each file describes a single unit, termed 'session' or 'resource bundle'.[5] This unit usually corresponds to a meaningful unit of analysis, e.g., to a piece of data having the same overall content, the same set of actors, and the same location and time (e.g., one elicitation session on topic X, or one folktale, or one 'matching game', or one conversation between several speakers).

Every session or resource bundle consists of one metadata file (i.e., an IMDI Session file) plus an unlimited number of other resources (i.e., media files, annotation or written resource files, and/or info files). The metadata file contains all the information that is entered into the metadata categories – categories that are displayed in the form of screens, schemata and fields in the "Content" panel of the IMDI Editor (see also section 1.5), e.g.:

---
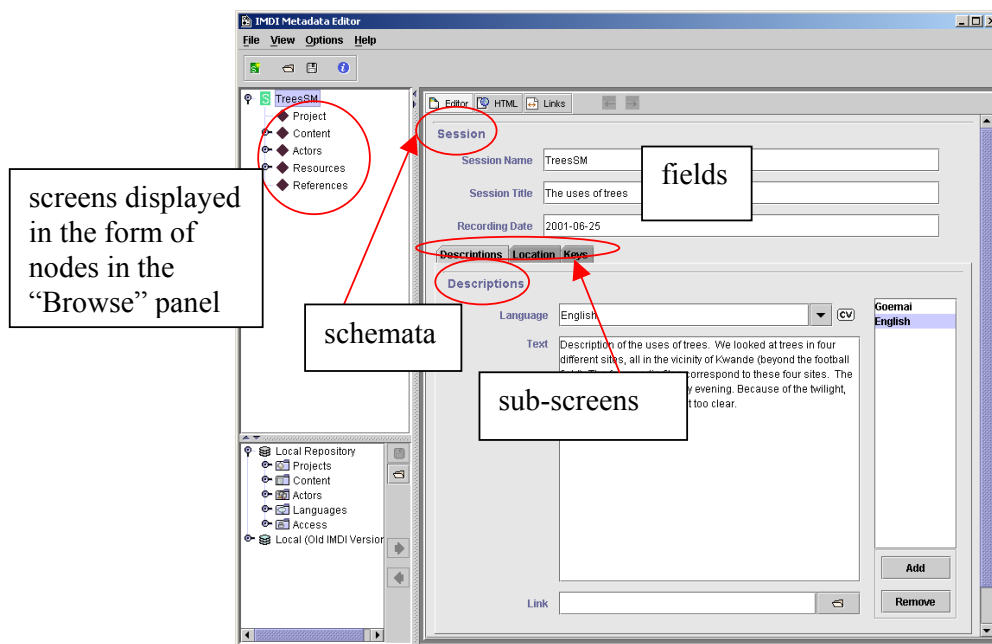
[5] Throughout the manual, these two terms are used interchangeably, with 'session' being the preferred term for audio/video resources, and 'resource bundle' for written resources.

The following sections illustrate the purpose of the various screens, schemata and fields. It is organized on a screen-by-screen basis, following the structure as it is displayed in the "Browse" panel:

- **Session** (see section 2.1);
- **Project** (see section 2.2);
- **Content** (see section 2.3);
- **Actors** (see section 2.4);
- **Resources** (see section 2.5);
- **References** (see section 2.6).

When entering information into the IMDI Editor, please keep the following points in mind:

(1) What kind of information would you want to search for? Make sure that such information is entered into standardized fields or **Key** schemata.

(2) All information entered on a screen is only relevant to that screen. This point is especially important for recurring schemata such as **Descriptions**, **Keys**, **Language**, or **Access**. These schemata occur on a number of different screens, and although they always look the same, they ask for different information – information that is relevant to the corresponding main screen.

(3) Remember to make use of the "Repository" panel to store recurring information (see section 1.4).

> **!** Note for researchers working at the MPI for Psycholinguistics, Nijmegen: The metadata information entered into the IMDI Editor is meant to be visible to the outside world via the Web. This only concerns the metadata information – **not** the audio, video and annotation files. But please be aware that other people will have access to your metadata files. Keep this in mind when you enter metadata information into the Editor, and make sure to exclude all sensitive information.
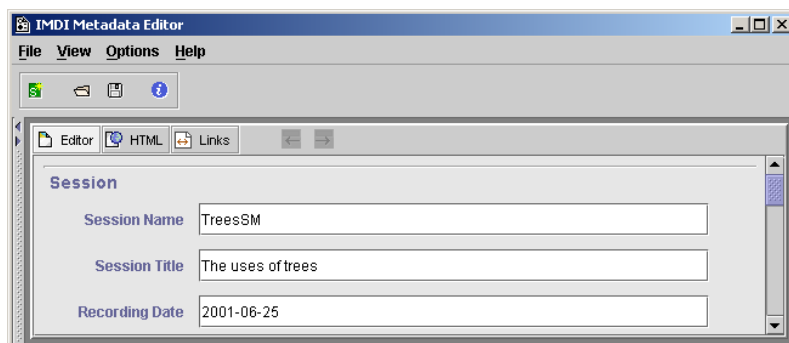
2.1 Session

This screen contains general information about the session or resource bundle. All the information that is entered into the **Session** screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):
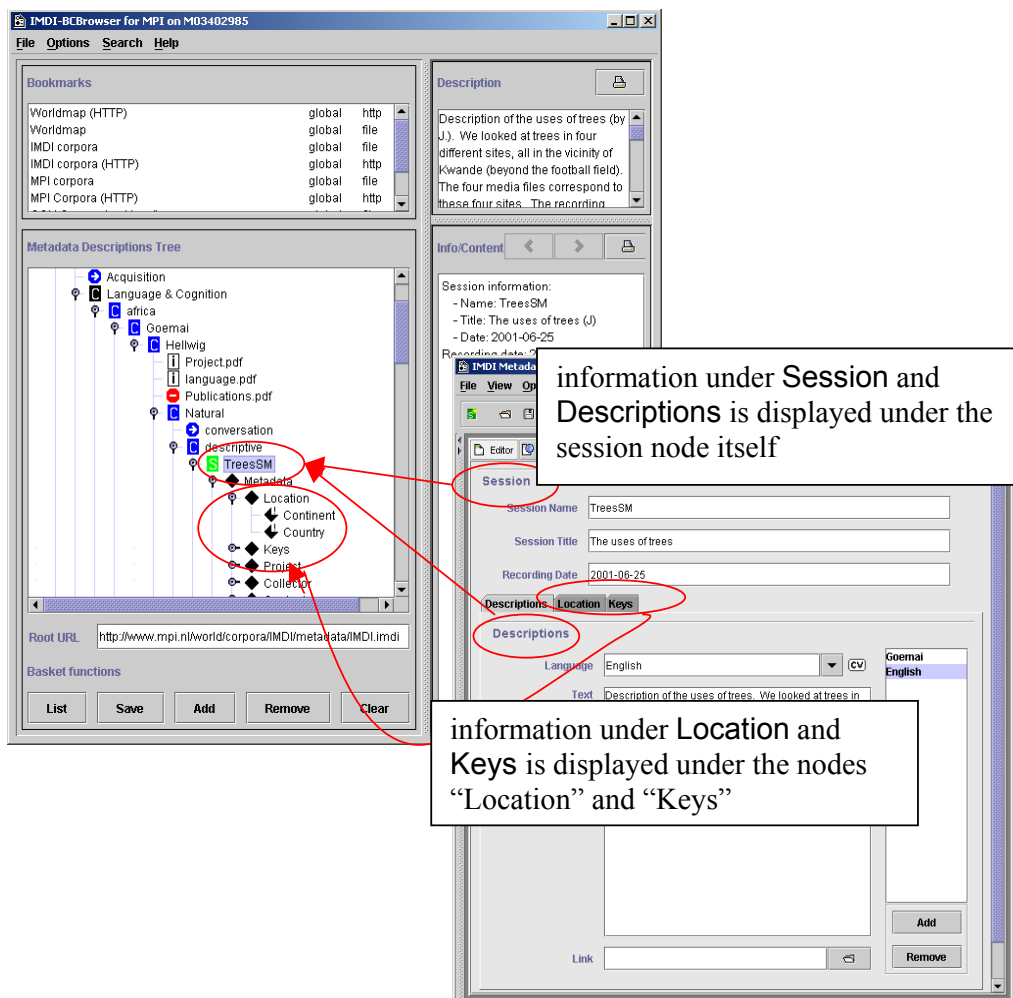
2.1.1 Session

The name, title and recording date of the session. For example:

Session Name

A short name or abbreviation that uniquely identifies the session.

> **!** Note for researchers working at the MPI for Psycholinguistics, Nijmegen: This name should be the same as the name of the IMDI file (see section 1.2.1.11) and as the name of the corresponding annotation and media files (see section 2.5). Furthermore, the file name has to be Unix compatible: do not use file names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

Session Title

The complete title of the session. Usually, it is the spelled-out version of the abbreviated Session Name.

Recording Date

The date at which the session data was collected. The date has to be entered in the format **YYYY-MM-DD**. Once you start typing into this field, the IMDI Editor will automatically display the required format (highlighted in blue color). Type over it to enter the date.

2.1.2 Descriptions

A description of the circumstances under which the data was collected.

Whenever you access a session through the IMDI Browser, this description is the first piece of information that you will see. It should therefore contain all the information that is necessary to quickly remind you about this session. However, it should **not** contain an elaborate description of the content (reserve this information for the Content screen, see section 2.3.2). E.g.:

62

See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a **Descriptions** schema. Remember: The field **Language** refers to the language in which the description is written – **not** to the language under investigation.

2.1.3 Location

The location at which the data was collected. For example:

Continent

The continent where the data was collected. Choose the continent from the pull-down menu, or type it in. It is a "closed vocabulary", i.e., only values listed in the pull-down menu are accepted by the IMDI Editor.

Country

The country where the session data was collected. Choose the country from the pull-down menu, or type it in. It is a "closed vocabulary", i.e., only values listed in the pull-down menu are accepted by the IMDI Editor.

Address

The address where the session data was collected (e.g., at school XY).

Region

The region (province, town, suburb, etc.) where the session data was collected. To add a region, click on the "Add" button; to remove a region, click on it to highlight it and then click on the "Remove" button.

2.1.4 Keys

Keywords that are relevant to (a) the collection of the data and (b) the creation of the IMDI file. For example:

the person who created the IMDI file

screens of the IMDI file that are not yet filled in

Name	Value
IMDI_file_creator	birghe
Missing_information	annotation unit
Check	TreesSM_3.mpg is lost

things that need to be checked (e.g., a lost file)

Clear Session

See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.2 Project

This screen contains information about the project within which the session data was collected. All the information that is entered into the Project screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):
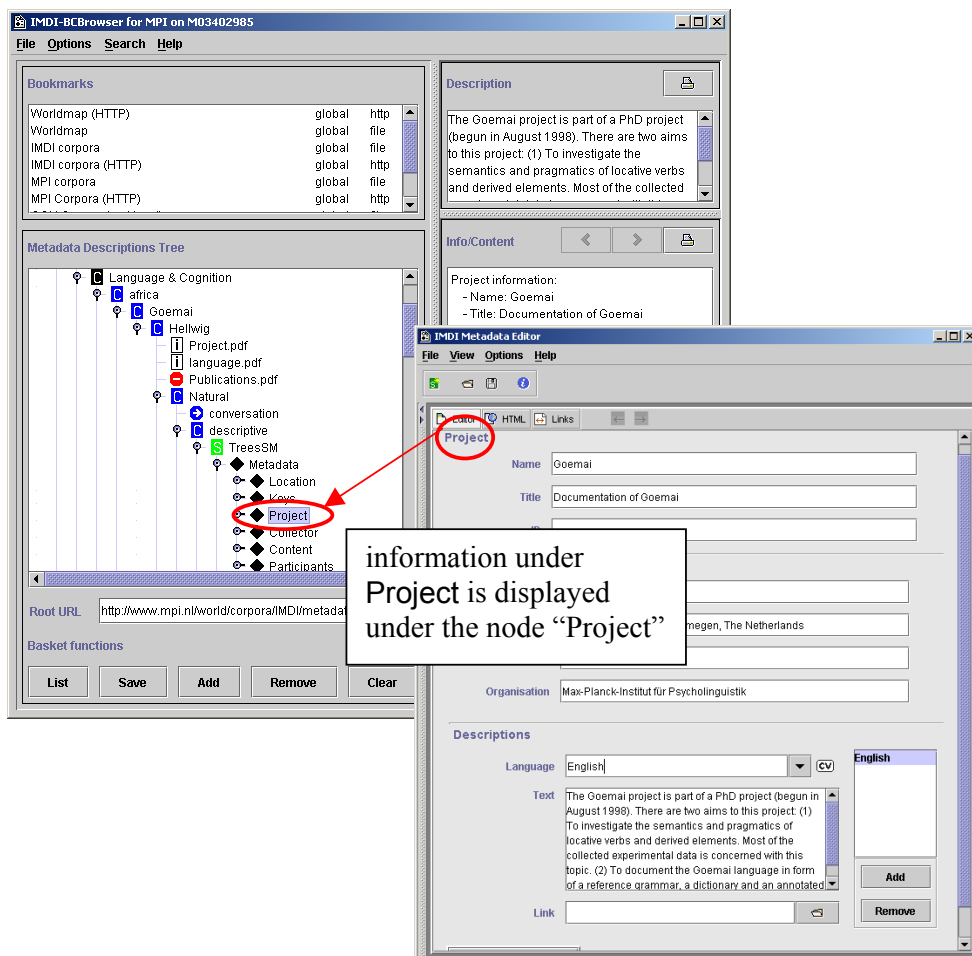
Enter the information into the IMDI Editor as follows:

Name

A short name or abbreviation that uniquely identifies the project.

Title

The full title of the project.

ID

A unique identifier for the project (if any), e.g. "IST-1999-10651".

Contact Information

Contact information about the person or institution responsible for the project (the **Name**, the **Address**, the **E-mail** address and the **Organisation** he/she belongs to).

Descriptions

A description of the scope and goals of the project. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a **Descriptions** schema. Remember: The field **Language** refers to the language in which the description is written - **not** to the language under investigation.

2.3 Content

This screen contains information about the content of the session. All the information that is entered into the Content screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):

information under Descriptions is displayed under the "Content" node itself

information under Content Type is displayed under the nodes of each category (e.g., "Task", "Modalities" etc.)

information under Languages and Keys is displayed under the nodes "Languages" and "Keys"
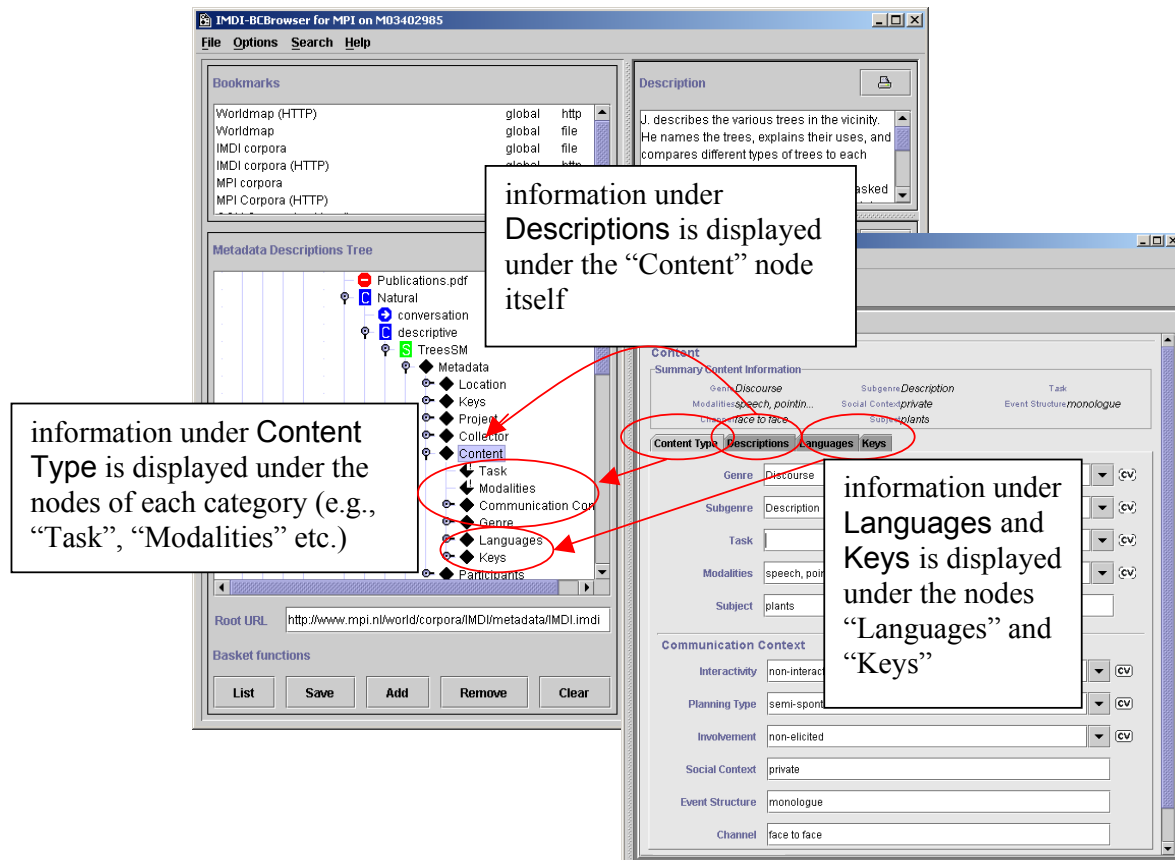
2.3.1 Content Type

Information about the genre, the task, the modalities, the subject, and the communication context. For example:

Many of the fields under Content Type offer predefined values that are displayed in the form of pull-down menus:[6] some of them only accept values from the pull-down menu, while others allow you to enter alternative values (see below). Furthermore, it is sometimes possible to enter more than one value – in this case, please separate the values with a comma. See section 1.5.2 for the different options and how they are symbolized in the IMDI Editor.

Genre and Subgenre

Information about the genre and the subgenre, whereby Subgenre is dependent on Genre. I.e., the value that you enter under Genre will determine the values that are available under Subgenre. Both Genre and Subgenre are "open vocabularies", i.e.,

---

[6] Please see the document "IMDI Metadata Elements for Session Descriptions" under http://www.mpi.nl/IMDI/ for updated lists of values and their definitions.

68

you do not have to select a value from the pull-down menu, but can enter an alternative value instead. The following values have been suggested:

Genre **Subgenre**	Examples
"Discourse"	
"Narrative"	a folktale, a historical narrative, a personal experience narrative
"Oratory"	a summing-up speech by a legal counsel, a political speech, a church sermon
"Procedural"	a recipe, an instruction on how to build a house
"Formulaic"	proverbs, greetings/leavetakings
"Language Play"	riddles, humor
"Interview"	
"Conversation"	
"Description"	a description of the layout of a compound
"Unintelligible speech"	
"Ritual/religious texts"	a prayer, a healing ritual, a catechism
"Fiction"	a detective novel, a science fiction story
"Newspaper article"	a political essay, a scientific report
"Radio/TV feature"	a political discussion on the radio, a documentary on animal life on the TV
"Drama"	
"Film"	
"Play"	
"Opera"	
"Musical"	
"Singing"	
"Individual Song"	
"Chant"	a psalm, a slogan during a demonstration
"Chorus"	
"Instrumental music"	
"Poetry"	a ballad, an oral epic
"Literature"	a short novel, a tragedy
"Secondary document"	
"Personal notes"	
"Stimuli"	a picture book elicitation, a story retelling, a matching game

! Note: If there are several **Subgenres**, you can enter all of them into the **Subgenre** field. Separate them with a comma.

Task

The name of the experimental or stimuli-based task that was carried out (if any). Choose a value from the pull-down menu (e.g., "info-kiosk", "wizard-of-oz", "travel-planning", "room reservation", or "frog story"), or type in an alternative value.

Modalities

The modalities under investigation. Choose a value from the pull-down menu (e.g., "speech", "writing", "gestures", "pointing-gestures", "signs", "eye-gaze", "facial-expressions", "emotional-state", "haptic"), or type in an alternative value.

Subject

The subject(s) or topic(s) of the session. There are no constraints on this field.

Communication Context

Information about the communication context, i.e., levels of participant interaction, the degree of planning through the speaker(s), the involvement of the researcher(s), the social context, the event structure, and the channel. All six are "closed vocabularies", i.e., a value from the pull-down menu must be chosen – other values will not be accepted. Enter the following information:

- Interactivity

  The level of participant interaction, i.e.:

value	comments
"interactive"	an interaction between at least two participants. It may or may not include an investigator, e.g.: • conversation • many narratives • matching game
"non-interactive"	a monologue, produced without expecting extended verbal responses from the hearer(s), e.g.: • many oratory texts and songs • some narratives
"semi-interactive"	primarily a monologue, but punctuated by repeated interjections from the hearer(s), e.g.: • a child interrupting a narrative • hearer(s) repeatedly prompting a narrator

- Planning Type

  The degree of planning through the consultant, i.e.:

value	comments
"spontaneous"	an unprompted speech whose topic is not determined by the investigator or an observer, e.g.: • conversation • chatting • joke-telling • singing while harvesting
"semi-spontaneous"	a prompted speech whose topic is determined in some way by an investigator or a community member, but whose participants speak freely within this context, e.g.: • interview • queries (e.g., 'Tell me about the history of your village.' 'Show me how to make tortillas.') • retellings (e.g., the speaker is asked to re-tell a story from a picture book, or to describe a task in his/her own words) • promptings (e.g., children answering a teacher's questions)
"planned"	the structure and content of the speech is planned in advance by the consultant/performer, e.g.: • political or ritual speech • poem recitation Note: This entry does not (necessarily) refer to an elicitation session, where a consultant is given a framework but does not plan his/her answer.

- Involvement

  The involvement of the researcher, i.e.:

value	comments
"elicited"	the investigator asks the speaker(s) to produce isolated phonemes, words, utterances or grammatical structures, e.g.: • production of sounds in different phonological environments • responses to (morphological, lexical, etc.) questionnaires
"non-elicited"	the investigator does not interfere verbally with the speech event (other than with his presence)
"no-observer"	no outside observer is present (only a tape recorder)

- Social Context

  The social context of the event, i.e.:

value	comments
"family"	
"private"	
"public"	
"controlled environment"	

- **Event Structure**

  The event structure, i.e.:

value	comments
"monologue"	
"dialogue"	
"conversation"	
"not natural format"	

- **Channel**

  The channel of communication, i.e.:

value	comments
"face to face"	
"experimental setting"	
"broadcasting"	
"telephone"	
"human-machine dialogue"	

2.3.2   Descriptions

A description of the content of the session. For example:

a description of the content

See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.3.3 Languages

Information about the language(s) **used** in the session, e.g.:

Language screen:

information about the language

(see section 2.3.3.1)

description of the how the language(s) are used in the session

(see section 2.3.3.2)

2.3.3.1 Language

Each language that is used in the session has to be listed on the Language screen. To add a language (see also section 1.3), do the following:

the list of languages

click on "Add" to add another language

to edit or remove a language, click on the language to highlight it, then click on "Edit" or "Remove"

The "Add" and "Edit" buttons give you access to the **Language** screen that contains the following kind of information:

- **Name**

  The name of the language. This name is standardized. Choose it from the pull-down menu, or type it in. (Please use capital letters.)

- **ID**

  The identifier of the language (based on the Ethnologue identifiers, see

  http://www.ethnologue.com/web.asp

). The IMDI Editor automatically enters the correct identifier once you have filled in the **Name** field above.

- **Mother Tongue**

  Please ignore this field. It has been disabled for **Content . Language**, and is only available for **Actor . Language** (see section 2.4.1.5).

- **Primary Language**

  Please ignore this field. It has been disabled for **Content . Language**, and is only available for **Actor . Language** (see section 2.4.1.5).

- **Dominant Language**

  Specifies whether or not this language is the language that is used most frequently in the session. From the pull-down menu, select either "true" (if it is the dominant language) or "false" (if it is not).

- Source Language

  In a second language acquisition context, this field specifies whether or not this language is the source language. From the pull-down menu, select either "true" (if it is the source language) or "false" (if it is not).

- Target Language

  In a second language acquisition context, this field specifies whether or not this language is the target language. From the pull-down menu, select either "true" (if it is the target language) or "false" (if it is not).

- Descriptions

  A description that gives background information about the language in general. Note that the description is **not** about the role of the language in that particular session (reserve this information for the Languages Descriptions sub-screen, see section 2.3.3.2). See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

- Resource References

  The IMDI Editor automatically assigns a unique resource identifier to each media file and written resource (see section 2.5). The sub-screen Resource References lists all available resource identifiers: it displays the identifier together with the type of resource (e.g., "N10155 (Media File)" in the screenshot above), and it displays an empty box to the left.

  This sub-screen allows now you to link the language to a specific resource. E.g., a language may only be used in resource "N10155 (Media File)", but not in resource "N10198 (Media File)". In this case, the language should only be linked to "N10155 (Media File)". To link a resource, click into the empty box to its left. A checkmark will appear there.

2.3.3.2 Languages Descriptions

Enter a description of the role of each language as it is used in the session (e.g., language of elicitation, main language, code-switching, etc.). Note that the description should **not** contain background information about the language in general (reserve this for the Language sub-screen, see section 2.3.3.1). See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.3.4 Keys

This sub-screen contains keywords that are relevant to the content of the session. For example:

See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.4 Actors

This screen contains information about the actors involved in the session. All the information that is entered into the **Actors** screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):

information about each **Actor** is displayed under the node "Actors" (or "Participants" in the older version)

see section 2.4.1

see section 2.4.2

2.4.1 Actor

Each actor involved in the session has to be listed on the **Actors** screen. If an actor occurs in two different **Roles** (e.g., once as "collector" and once as "annotator") (s)he has to be listed twice. To add an actor (see also section 1.3), do the following:
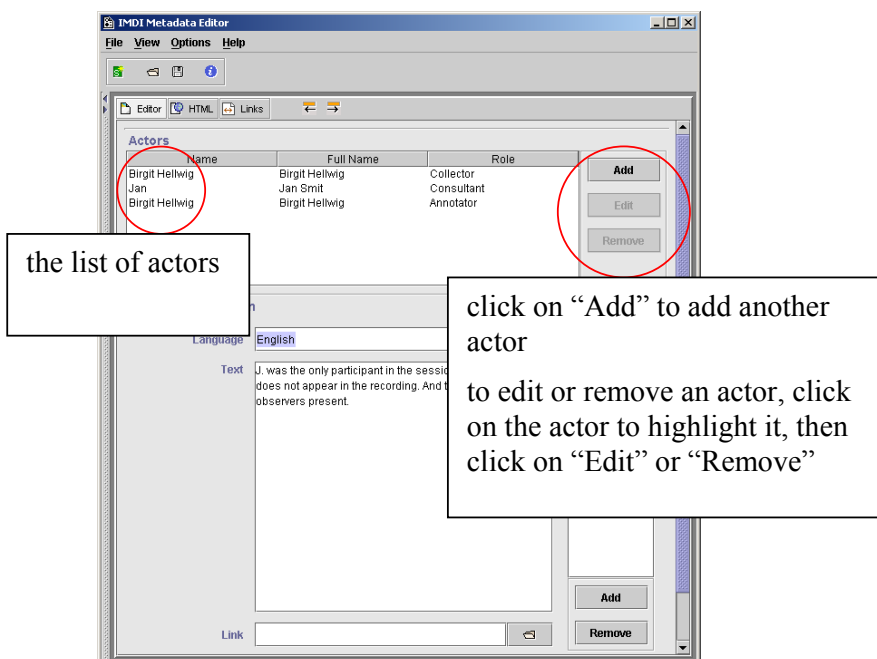
The "Add" and "Edit" buttons give you access to the **Actor** screen that contains the following kind of information:
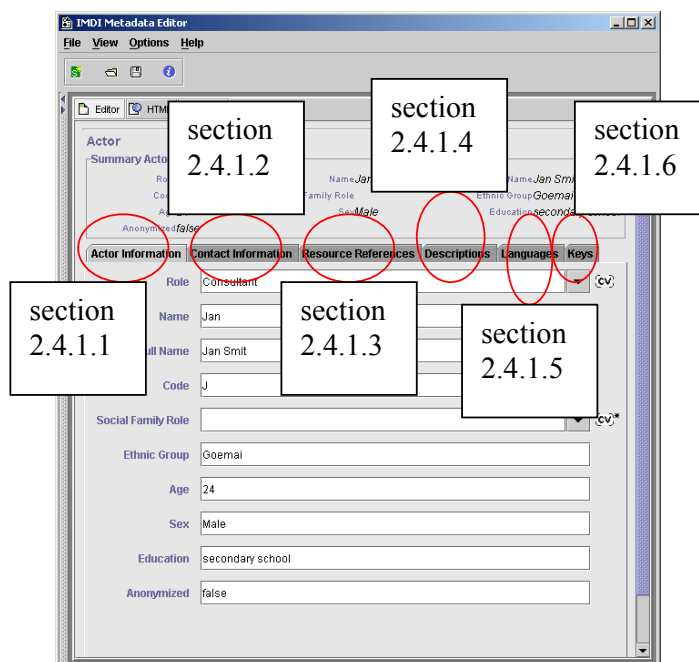
2.4.1.1　　　　Actor Information

Information about the individual actor. For example:

Role

The function of the participant in the session. Choose a value from the pull-down menu, or type in an alternative value. For example:

value	comments
"Annotator"	
"Author"	
"Collector"	
"Consultant"	
"Computer"	
"Depositor"	
"Editor"	
"Filmer"	
"Illustrator"	
"Interviewer"	
"Photographer"	
"Publisher"	
"Recorder"	
"Referent"	
"Researcher"	
"Speaker/Signer"	
"Translator"	

Name

The name of the actor, i.e., the name that other actors in the session use to identify him/her. It is usually not the same as his/her full name. (See also the field Anonymized below.)

Full Name

The full name of the actor. (See also the field Anonymized below.)

Code

Short unique code to identify the actor. It usually corresponds to the code that is used in transcriptions and annotations to identify parts that were uttered by him/her. (See also the field Anonymized below.)

Social Family Role

The social or family role of the actor, i.e., his/her relationship to other actors participating in the session. Choose a value from the pull-down menu, or type in an alternative value. For example:

value	comments
"Father"	
"Mother"	
"Sibling"	
"Boss"	
"Partner"	
"Student"	
"Teacher"	
"Shaman/Priest"	
"Mayor"	
"Doctor"	

Ethnic Group

The ethnic group of the actor.

Age

The age of the actor. Please enter the age in the following format: **YY** or **YY;MM** or **YY;MM.DD**.

> **!** Note: If the exact age is not known, it is nevertheless useful to enter an approximate age. This will allow you later to conduct searches on all actors who are in the age range between, e.g., 20 and 30 years of age.

Sex

The sex of the actor. It has to be chosen from the pull-down menu.

Education

Type in the education or literacy level of the actor. For example: "primary school", "secondary school", "literate", "illiterate", etc. There are no constraints on this field

Anonymized

Specifies whether or not the actor's name was made anonymous.
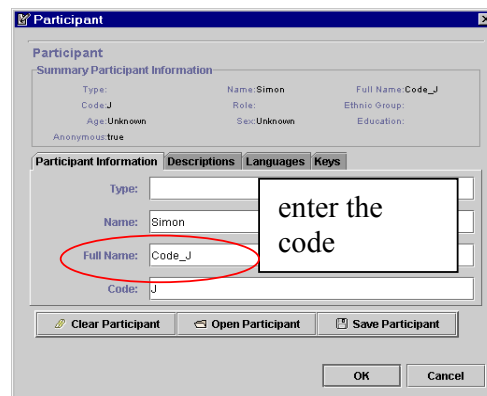
> Choose "true" if you have entered a code in the field **Full Name**.

> Choose "false" if you have entered the full name in the field **Full name**.

> **!** Note: If you have chosen the option "true", you can create a conversion file that maps the code onto the full name. Make use of the menu item "Options / Anonyms …" (see section 1.2.3.2). If you have created such a file, the actor information is entered, saved and displayed as follows:

1. Enter the chosen code into the field **Full Name** (instead of the full name).

   ! Note: The code entered into the field **Full Name** should **not** be the same as the code entered into the field **Code**.
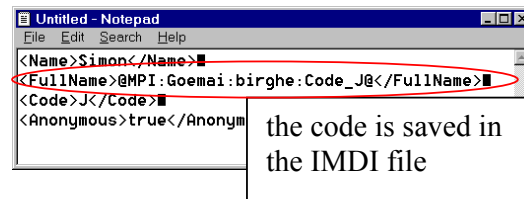
   enter the code

2. The code that you have entered into the field **Full Name** is saved in the IMDI file as follows:

   @Institute: Project: User: Code@
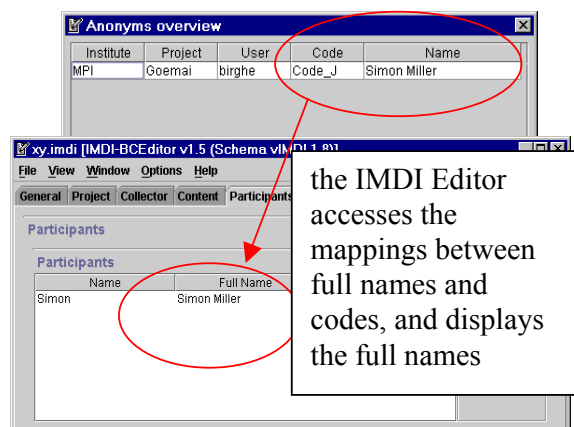
   the code is saved in the IMDI file

3. The IMDI Editor accesses the mappings that you have defined in the "Anonyms overview" dialog window (see section 1.2.3.2), and displays the full name.

   the IMDI Editor accesses the mappings between full names and codes, and displays the full names

   All people who have access to the file that specifies these mappings will be able to see the full name, all others will see the code.

! Note: Previous versions of the IMDI Editor supported a different method for rendering full names anonymous. If you have old IMDI files and if you have

doubts about how the full names are displayed, please contact your corpus manager.

2.4.1.2   Contact Information

Contact information about the actor (the **Name**, the **Address**, the **E-mail** address and the **Organisation** he/she belongs to).

2.4.1.3   Resource References

The IMDI Editor automatically assigns a unique resource identifier to each media file and written resource (see section 2.5). The sub-screen **Resource References** lists all available resource identifiers: it displays the identifier together with the type of resource (e.g., "N10155 (Media File)" in the screenshot below), and it displays an empty box to the left.
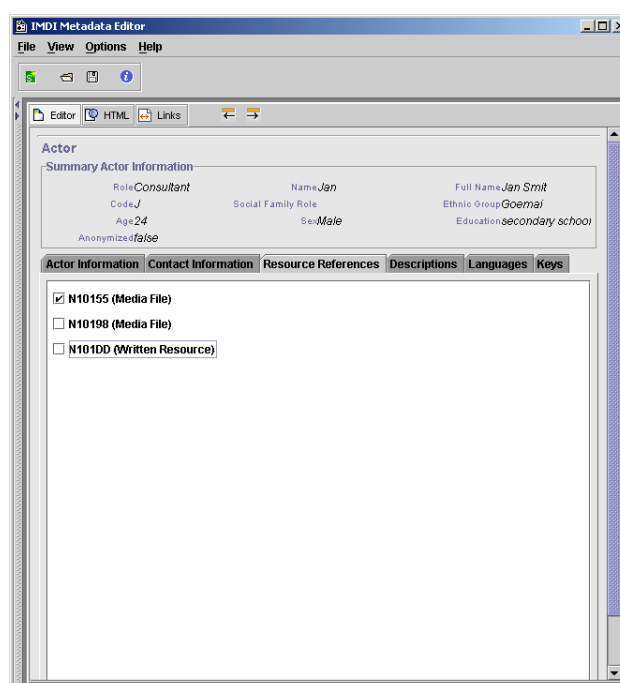
The sub-screen **Resource References** now allows you to link the actor to a specific resource. E.g., an actor may only appear in the specified **Role** in resource "N10155 (Media File)", but not in resource "N10198 (Media File)". In this case, the actor should only be linked to "N10155 (Media File)". To link a resource, click into the empty box to its left. A checkmark will appear there.

2.4.1.4   Descriptions

A general description of the individual actor (independent of his/her Role in the session). For example:

See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.4.1.5 Languages

Information about all the languages that **the actor is familiar with** (independent of whether or not the actor uses them in the session).

Language screens:

general information about each language

description of the actor's use of and familiarity with the languages
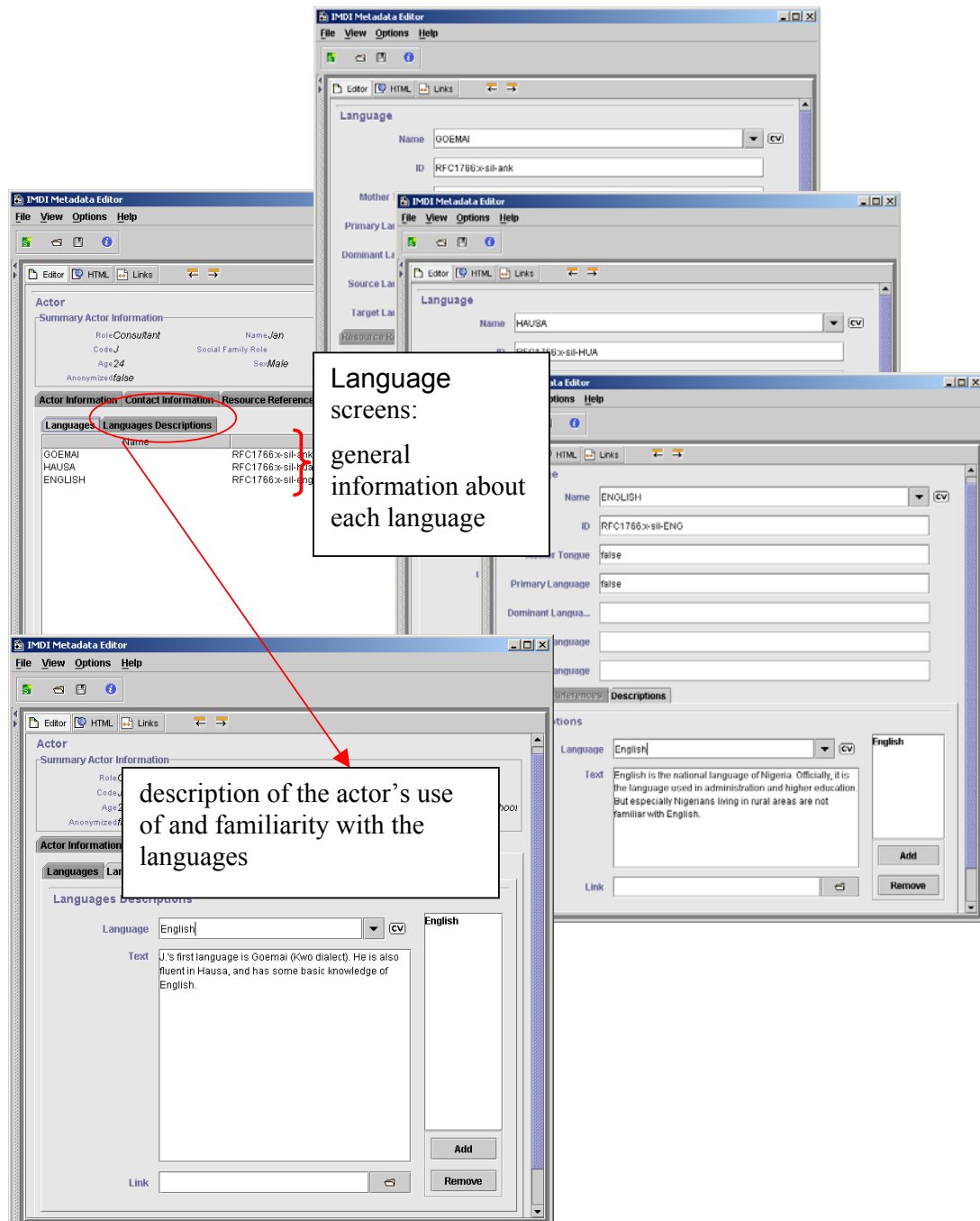
Languages

Each language that is spoken by the actor has to be listed on the **Language** screen. To add a language (see also section 1.3), do the following:

the list of languages

click on "Add" to add another language

to edit or to remove languages, click on the language to highlight it, then click on "Edit" or "Remove"

The "Add" and "Edit" buttons give you access to the **Language** screen that contains the following kind of information:

- **Name**

  The name of the language. This name is standardized. Choose it from the pull-down menu, or type it in. (Please use capital letters.)

- **ID**

  The identifier of the language (based on the Ethnologue identifiers, see **http://www.ethnologue.com/web.asp**). The IMDI Editor automatically enters the correct identifier once you have filled in the **Name** field above.

- **Mother Tongue**

  Specifies whether or not this language is the mother tongue of the actor. From the pull-down menu, select either "true" (if it is the mother tongue) or "false" (if it is not).

- **Primary Language**

  Specifies whether or not this language is the language that the speaker is most fluent in. From the pull-down menu, select either "true" (if it is the most fluent language) or "false" (if it is not).

- Dominant Language

  Please ignore this field. It has been disabled for Actor . Language, and is only available for Content . Language (see section 2.3.3.1).

- Source Language

  Please ignore this field. It has been disabled for Actor . Language, and is only available for Content . Language (see section 2.3.3.1).

- Target Language

  Please ignore this field. It has been disabled for Actor . Language, and is only available for Content . Language (see section 2.3.3.1).

- Descriptions

  A description that gives background information about the language in general. Note that the description is independent of the actor's familiarity with it (reserve such information for the Languages Descriptions sub-screen below). See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **<u>not</u>** to the language under investigation.

- Resource References

  Please ignore this sub-screen. It has been disabled for Actor . Language, and is only available for Content . Language (see section 2.3.3.1).

Languages Descriptions

A description of the set of languages that the participant is familiar with. Note that the description does **<u>not</u>** contain background information about the language in general (reserve this for the Language . Descriptions schema above). See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **<u>not</u>** to the language under investigation.

2.4.1.6 Keys

This sub-screen contains keywords that are relevant to the individual actor (independent of his/her Role in the session).

See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.4.2 Actors Descriptions

A description of the interactions and interrelations among different actors of the session. Note that this description concerns the set of actors as a whole – information about specific actors should be described by means of Actor . Descriptions (see section 2.4.1.4).

description of the actor constellation

See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.5 Resources

This screen contains information about the resources belonging to the session. All the information that is entered into the Resources screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):
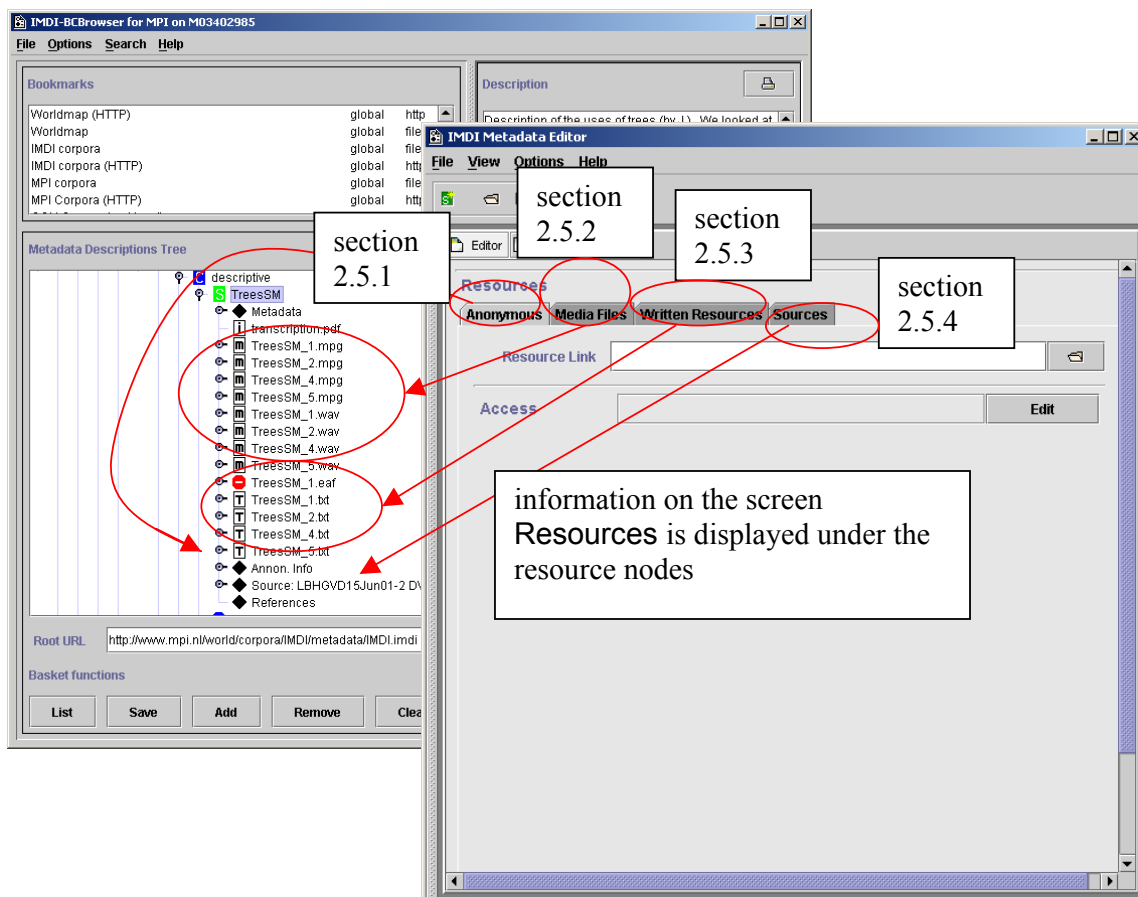
The IMDI Editor automatically assigns a unique resource identifier to each media file and written resource. The resource identifiers are then used on the sub-screen Resource References to link a source (see section 2.5.4.3), an actor (see section 2.4.1.3) or a language (see section 2.3.3.1) to a specific resource. E.g., it may be the case that an actor only plays a role in one of the resources, but not in the others – in this case, you can use the resource identifiers to link the actor to this one resource, but not to the others.

Click on the button "Links" at the top of the "Content" panel to view all links, e.g.:
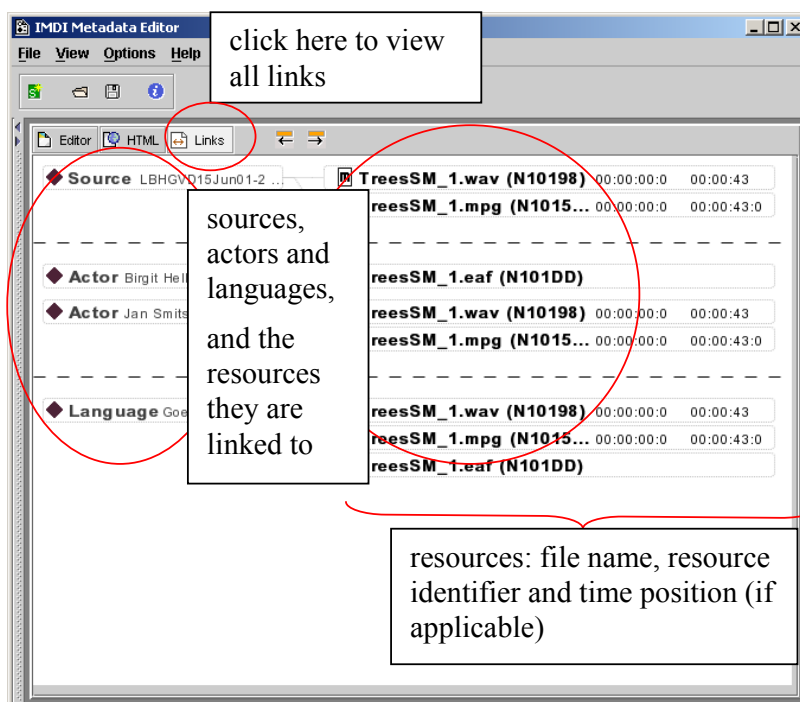
Within this "Link" window, you can click on any source, actor, language or resource to jump to the corresponding screen.

> ! Note: If you remove a resource, the specified links are automatically removed as well.

2.5.1 Anonymous

If you have made the names of actors anonymous (on the screen Actors, see section 2.4.1.1) and if you have created a file that maps these codes onto full names (by using the menu item "Options / Anonyms …", see section 1.2.3.2), you can use the screen Anonymous to specify access rights to this mapping file. As illustrated in section 1.2.3.2, the mapping file is stored under the name **user-name.ano** in the folder "user-directory \ IMDI-TOOLS \ Anonyms" on your desktop computer – i.e., it is only available to you. However, you might want to store a copy of it in the corpus itself (e.g., to protect it against data loss, in case of a system error). In this case, store the file in your corpus (or ask your corpus manager to do it for you), specify the name and directory of the file, and the access rights to it. See "Appendix 1: Recurring schemata (Access)" for instructions on how to fill in an Access schema.

2.5.2 Media Files

Each digitized media file (e.g. audio, video or image file) associated with the session has to be listed on the Media Files screen. To add a media file (see also section 1.3), do the following:
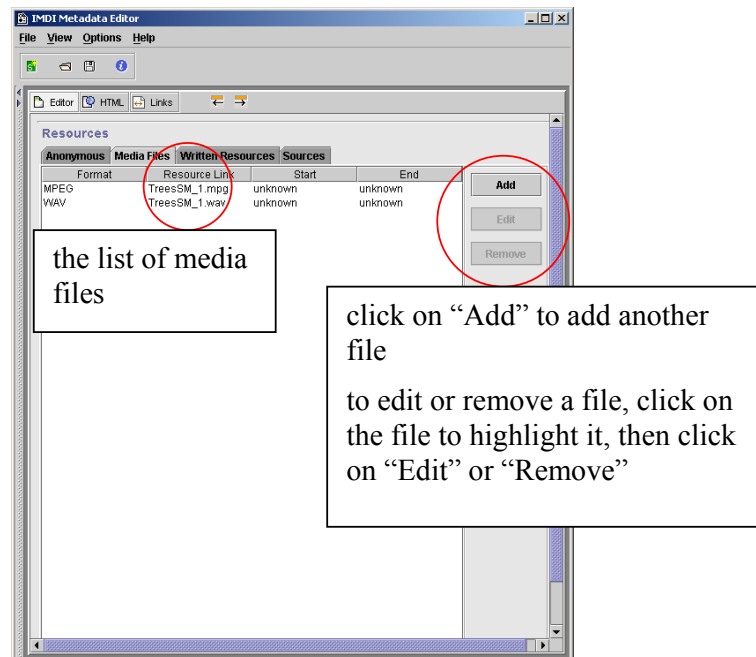
The "Add" and "Edit" buttons give you access to the Media File screen that contains the following kind of information:
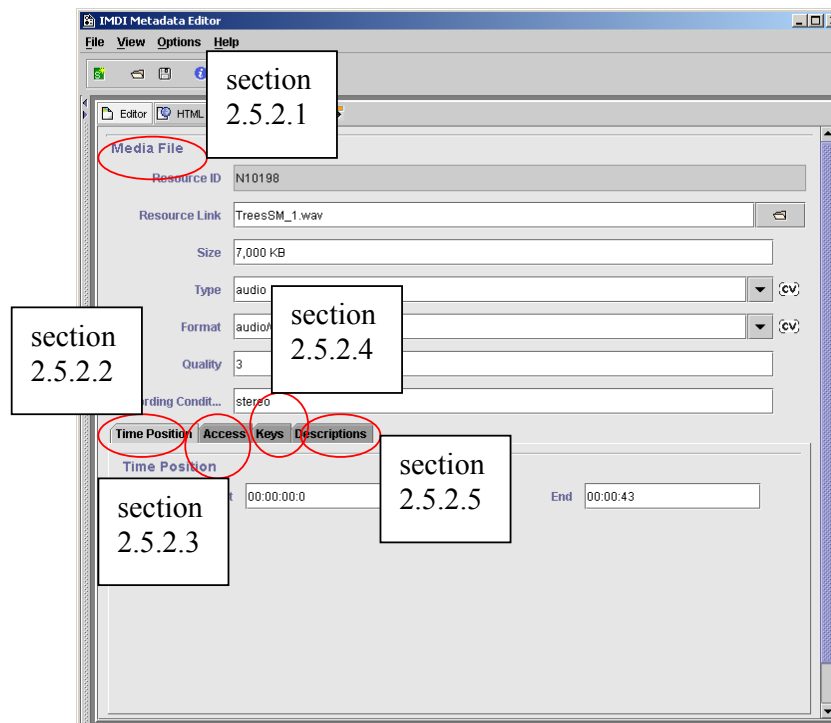
2.5.2.1　　　　Media File

This screen contains general information about the media file, i.e.:

- Resource ID

  The IMDI Editor automatically assigns a unique resource identifier to each media file. This assignment is automatic, and cannot be influenced by the user (see section 2.5).

- Resource Link

  The link to the corresponding media file, i.e., its name and location.

  ! Note for researchers working at the MPI for Psycholinguistics, Nijmegen: Please specify only the name of the file – do **not** make use of the "open folder" icon to the right of the Link field. The name of the media file should be the same as the name of the IMDI file (see section 1.2.1.11), and as the name entered in the field Session Name (see section 2.1.1). Furthermore, the file name has to be Unix compatible: do not use file names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

- Size

  The (approximate) size of the media file in bytes. Note that this field is not standardized: it is meant to be read by humans (to give them an indication of the size before they, e.g., download it), not to be processed through the computer.

- Type

  The type of the media file. Choose a value from the pull-down menu, or type in an alternative value.

value	comments
"audio"	
"video"	
"image"	
"document"	
"drawing"	
"text"	

- Format

  The format of the media file. Choose a value from the pull-down menu, or type in an alternative value.

value	comments
"video/mpeg1"	
"video/mpeg2"	
"video/mpeg4"	
"video/quicktime"	
"audio/wav"	
"audio/aiff"	
"image/jpeg"	
"text/pdf"	
"text/html"	

- Quality

  The quality of the recording. Choose an option from the pull-down menu (1 stands for low and 5 for high quality).

- Recording Conditions

  The technical conditions under which the media file was recorded, e.g., the equipment used in the recording (e.g., microphone type, amplifier type, mono/stereo recording, etc.). There are no constraints on this field.

2.5.2.2 Time Position

The start/end position of the session on the media file. Please enter the start/end position in the following format: **hh:mm:ss:f** (i.e., hours:minutes:seconds:frames).

2.5.2.3 Access

Information about the access rights to the **media file**. See "Appendix 1: Recurring schemata (Access)" for instructions on how to fill in an Access schema.

2.5.2.4 Keys

Keywords that are relevant to the media file. See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.5.2.5 Descriptions

A description of the media file. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.5.3 Written Resources

Each written resource associated with the session has to be listed on the **Written Resources** screen. To add a written resource (see also section 1.3), do the following:
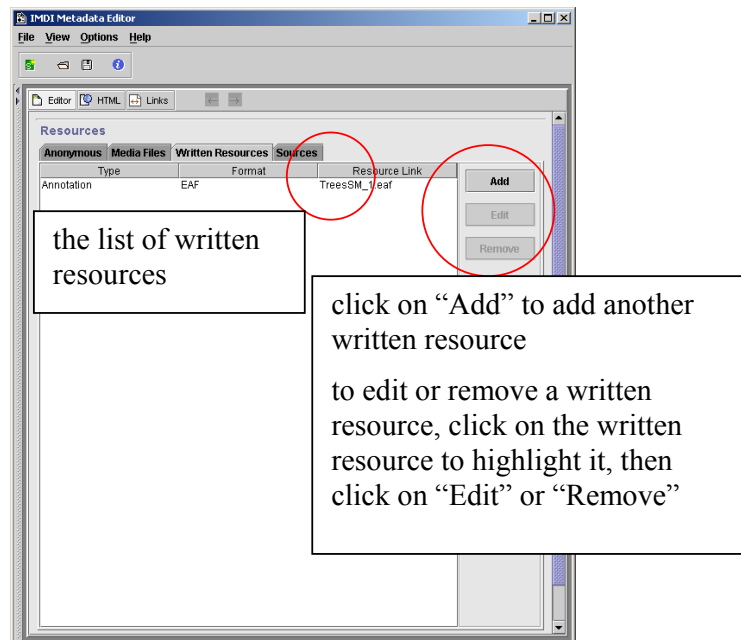
The "Add" and "Edit" buttons give you access to the Written Resource screen that contains the following kind of information:

2.5.3.1 Written Resource

This screen contains general information about the written resource, i.e.:

- Resource ID

  The IMDI Editor automatically assigns a unique resource identifier to each written resource. This assignment is automatic, and cannot be influenced by the user (see section 2.5).

- Date

  The date when the written resource was created. Please enter the date in the following format: **YYYY-MM-DD**, e.g. 2000-12-30.

- Resource Link

  The link to the file containing the written resource, i.e., its name and location.

  ! Note for researchers working at the MPI for Psycholinguistics, Nijmegen: Please specify only the name of the file – do **not** make use of the "open folder" icon to the right of the Link field. The name of the written resource should be the same as the name of the IMDI file (see section 1.2.1.11), and as the name entered in the field Session Name (see section 2.1.1). Furthermore, the file name has to be Unix compatible: do not use file names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

- Media Resource

  The link to the media file from which the written resource originated (if applicable).

- Anonymous

  Specify whether or not you have made actors' names anonymous in the written resource.

  Choose "true" if you have used a code.

  Choose "false" if you have used the real name.

2.5.3.2 Information

- **Type** and **Subtype**

    The type and subtype of the written resource, whereby **Subtype** is dependent on **Type**. I.e., the value that you enter under **Type** will determine the values that are available under **Subtype**. Both **Type** and **Subtype** are "open vocabularies", i.e., you do not have to select a value from the pull-down menu, but can enter an alternative value instead. The following values have been suggested:[7]

Type Subtype	Examples

"Primary Text"
 	
    "Documentary"
 	
    "Fiction"
	any material that is the object of study

"Annotation"
 	
    "gesture"
 	
    "orthography"
 	
    "phonetic"
 	
    "phonology"
 	
    "morphology"
 	
    "morphosyntax"
 	
    "syntax"
 	
    "semantics"
 	
    "pragmatics"
 	
    "typology"
	an annotation (e.g., transcription or grammatical analysis) of the material under study

"Lexical analysis"
 	
    "dictionary"
 	
    "terminology"
 	
    "wordlist"
 	
    "lexicon"
	a lexical analysis of the material under study

"Ethnography"	an ethnographic analysis of the material under study
"Study"	the written resource is used for a specific subfield of linguistic science

---

[7] Please see the document "IMDI Metadata Elements for Session Descriptions" under **http://www.mpi.nl/IMDI/** for updated lists of values and their definitions.

! Note: If there are several **Subtypes**, you can enter all of them into the **Subtype** field. Separate them with a comma (e.g., "orthography, morphology").

- Format

  The file format of the file that contains the written resource. Choose a value from the pull-down menu, or type in an alternative value.

value	comments
"CHAT"	
"Shoebox"	
"RDBMS"	
"TRS"	
"EAF"	
"AIF"	
"BAS"	
"text/plain"	
"text/pdf"	
"text/html"	

- Size

  The (approximate) size of the written resource in words.

- Derivation

  The relation of the written resource to other documents. It is a "closed vocabulary", i.e., the value has to be chosen from the pull-down menu, i.e.:

value	comments
"Analysis"	
"Translation"	
"Commentary"	
"Criticism"	
"Annotation"	

- Content Encoding

  The name of the encoding scheme used for creating the written resource (if applicable). For example:

  > Eurotype (i.e., following the Eurotype guidelines)

- Character Encoding

  The name of the character encoding used for creating the written resource, e.g., UTF-8.

- Language ID

  The identifier of the language that is used in the written resource, e.g. 'English' for an English translation. It has to be entered in a standard format. Please enter either the ISO identifier (i.e., enter "ISO639:eng" for English), or the Ethnologue identifier (i.e., enter "RFC1766:x-sil-eng" for English) – please look up the ISO identifiers under

  http://lcweb.loc.gov/standards/iso63

9-2/langhome.html, and the Ethnologue identifiers under http://www.ethnologue.com/web.asp

.

2.5.3.3 Validation

Gives information about the validation state of the written resource.

- Type

  The type of validation. It is a "closed vocabulary", i.e., you have to choose a value from the pull-down menu, i.e.:

value	comments
"formal"	
"content"	

- Methodology

  The methodology of validation. It is a "closed vocabulary", i.e., you have to choose a value from the pull-down menu, i.e.:

value	comments
"hand"	
"automatic"	
"semi-automatic"	

- Level

  Gives an estimation of how much of the resource was validated (enter a value between 0 and 100%).

- Description

  A description of the validation. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.5.3.4 Access

Information about the access rights to the **file containing the written resource**. See "Appendix 1: Recurring schemata (Access)" for instructions on how to fill in an Access schema.

2.5.3.5 Keys

Keywords that are relevant to the written resource. See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.5.3.6 Descriptions

A description of the written resource. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.5.4 Sources

Each source (e.g., video or audio tapes, or books) associated with the session has to be listed on the Sources screen. To add a source (see also section 1.3), do the following:

the list of sources

click on "Add" to add another source

to edit or remove a source, click on the source to highlight it, then click on "Edit" or "Remove"
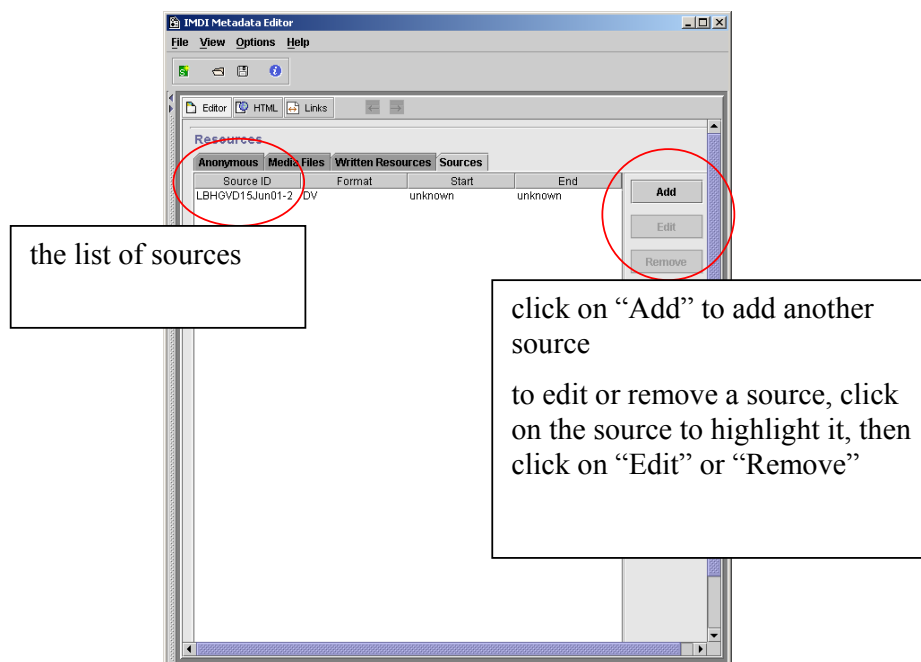
The "Add" and "Edit" buttons give you access to the Source screen that contains the following kind of information:

section 2.5.4.1

section 2.5.4.5

section 2.5.4.2

section 2.5.4.3

section 2.5.4.4

section 2.5.4.6
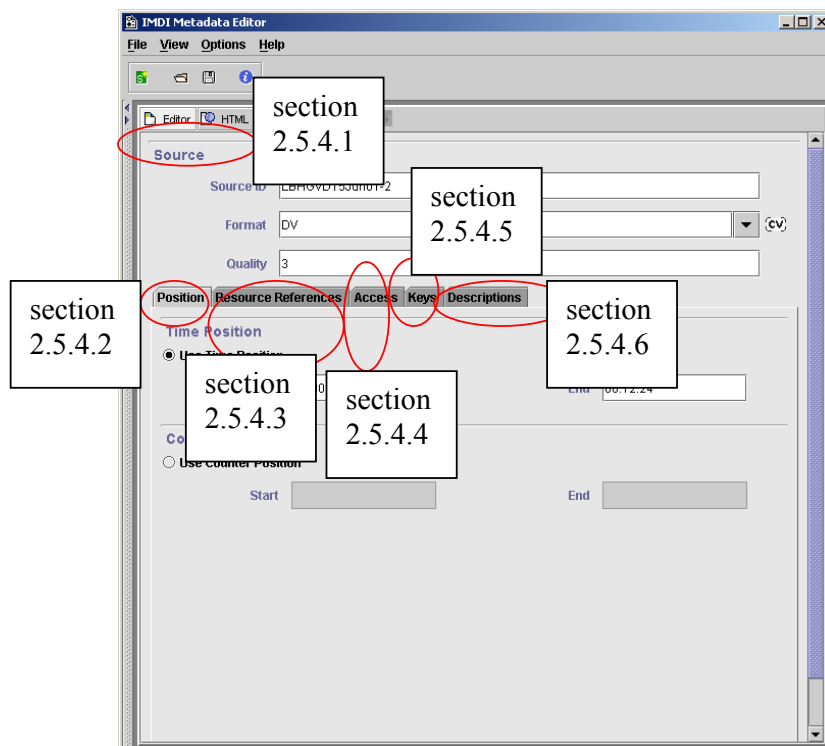
2.5.4.1   Source

This screen contains general information about the source, i.e.:

- Source ID

  A short code to identify the source.

  ! Note for researchers working at the MPI for Psycholinguistics, Nijmegen: Please consult the tape labeling conventions.

- Format

  The physical storage format of the source. Choose a value from the pull-down menu, or type in an alternative value.

value	comments
"CC"	Compact Cassette (i.e., normal audio cassette)
"CD"	Compact Disc
"CDROM"	Compact Disc - Read-Only Memory
"DAT"	Digital Audio Tape
"MD"	Mini Disc
"Reel"	Reel-to-reel tape
"DVD"	Digital Video Disc
"DVDROM"	Digital Video Disc - Read-Only Memory
"Hi8"	Hi8 Video Tape
"VHS"	VHS Video Tape
"DV"	Digital Video
"U-matic"	U-matic Tape
"Book"	A publication on paper
"Microfiche"	A film card

- Quality

  The quality of the tape. Choose an option from the pull-down menu (1 stands for low and 5 for high quality).

2.5.4.2   Position

The start/end position of the session in the corresponding source. Please enter this position in the following formats:

- In the case of digital audio tapes use the field Time Position: enter **hh:mm:ss** (i.e., hours:minutes:seconds).

- In the case of digital video tapes use the field Time Position: enter **hh:mm:ss:f** (i.e., hours:minutes:seconds:frames).

- In the case of non-digital tapes and books, use the field Counter Position: enter a sequence of digits to represent the counter position (of a tape) or the page numbers (of a book).

  ! Note for researchers working at the MPI for Psycholinguistics, Nijmegen: If you have asked the digitization team to digitize your tape, the team will create a digital master file (DMF) from it. You then need to go through this DMF (using either Windows Media Player or ELAN) to identify the relevant

session: enter the start and end position of this session (as it appears on this DMF) under Position. The digitization team will then segment the DMF further according to these specifications, and create the appropriate media file.

2.5.4.3 Resource References

The IMDI Editor automatically assigns a unique resource identifier to each media file and written resource (see section 2.5). The sub-screen Resource References lists all available resource identifiers: it displays the identifier together with the type of resource (e.g., "N10155 (Media File)" in the screenshot below), and it displays an empty box to the left.
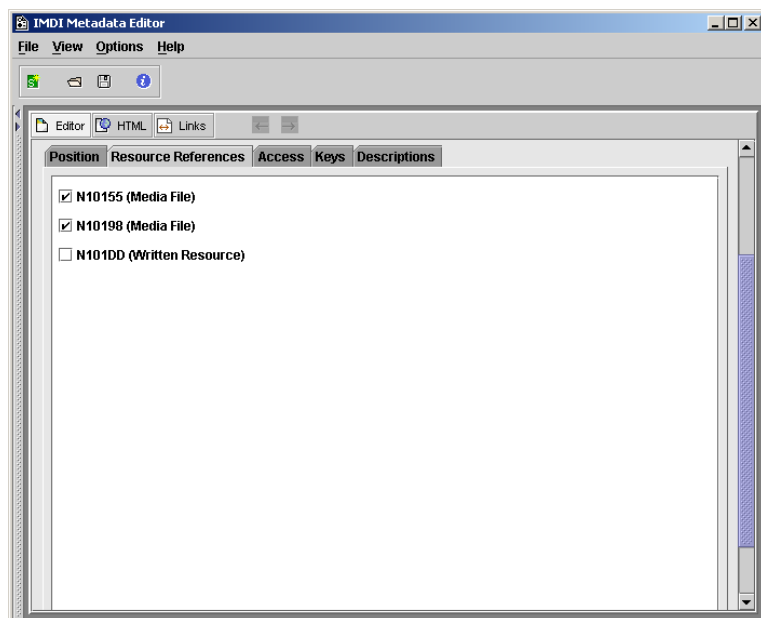
The sub-screen Resource References now allows you to link the source to a specific media file or written resource. E.g., a source may only be connected to the media files (e.g., "N10155 (Media File)" and "N10198 (Media File)"), but not to a written resource. To link a resource, click into the empty box to its left. A checkmark will appear there.

2.5.4.4 Access

Information about the access rights to the **source**. See "Appendix 1: Recurring schemata (Access)" for instructions on how to fill in an Access schema.

2.5.4.5 Keys

Keywords that are relevant to the source. See "Appendix 1: Recurring Schemata (Keys)" for instructions on how to fill in a Keys schema.

2.5.4.6 Descriptions

A description of the source. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

2.6 References

This schema contains cross-references to other sessions, fieldnotes, or publications that are relevant to the content of the session. All the information that is entered into the References screen will be displayed in the IMDI Browser in the following way (see the IMDI Browser manual for details):
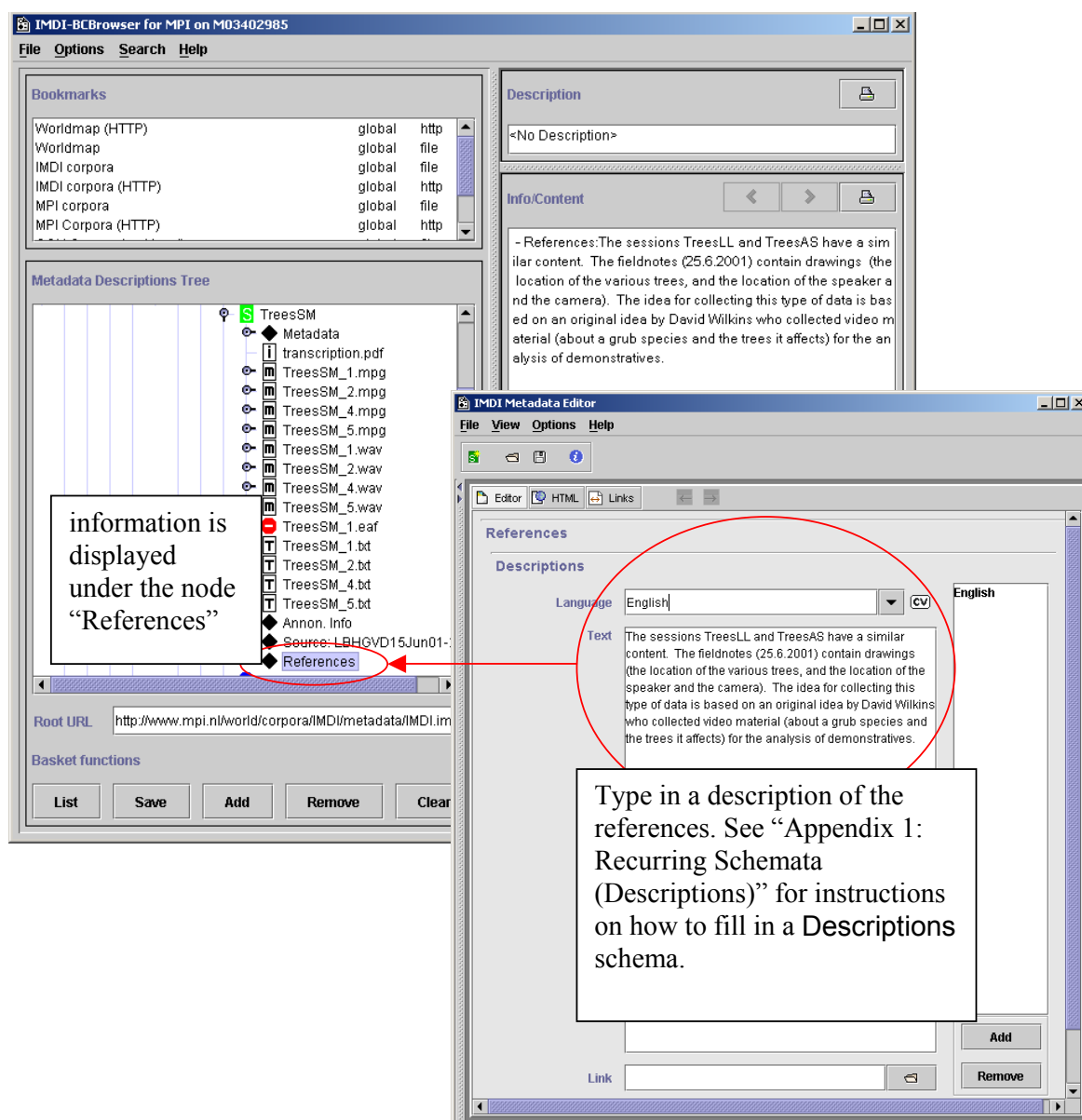
2.7 Changes to the metadata categories (IMDI 3.0 standard)

The most important change implemented in the IMDI Editor version 3.0 concerns the supported metadata categories: the Editor now supports the IMDI 3.0 standard, and all IMDI files created with this version of the Editor will automatically conform to this new standard. IMDI files created with an older version of the Editor need to be converted to the new standard.

This section first summarizes the implemented changes (see section 2.7.1), and then describes the conversion of old IMDI files (see section 2.7.2) and old repositories (see section 2.7.3) to the 3.0 standard.

2.7.1 Changes to the metadata categories

The following changes in metadata categories were implemented in the IMDI 3.0 standard:

IMDI 2.0 (and earlier) Standard	IMDI 3.0 Standard
Collector	this screen is now subsumed under Actor, with "collector" being one possible value for the field Role (see section 2.4.1.1)
Content . Genre . Interactional Content . Genre . Discursive Content . Genre . Performance	the three subtypes of genre are replaced by one Genre and one Subgenre, whereby the possible values of Subgenre are dependent on the value of Genre (see section 2.3.1)
Language	an additional panel allows you to specify resource references (see section 2.3.3.1)
Participant	this screen is replaced by Actor; Actor now includes Contact information; an additional panel allows you to specify resource references (see section 2.4.1.3)
AnotationUnit	this screen is replaced by Written Resource (see section 2.5.3)
Resources	a Keys schema is added; Media File and Written Resource are assigned Resource Ids; an additional panel under Source allows you to specify resource references (see section 2.5)
Lexicon Resource	this screen is added to create lexicon resource bundles (see section 3)

2.7.2 Conversion of old IMDI files to the 3.0 standard

The IMDI Editor version 3.0 supports the IMDI 3.0 standard. Files that were created with an older version of the Editor can still be opened, but they need to be converted to the new standard. The conversion process is as follows:

1. Open the old file in the IMDI Editor 3.0 (see section 1.2.1.6 for how to open a file). A warning of the following kind will appear:
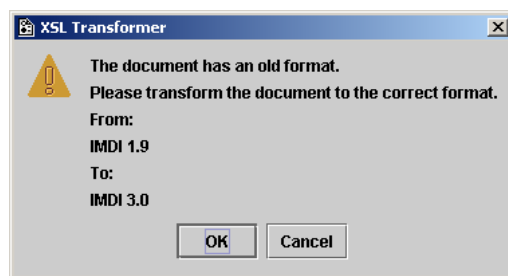
2. Click on "OK" to convert the file. (Otherwise click on "Cancel" to not convert it.) The file will be converted to the 3.0 standard, using the default profile for that type of file (see section 1.2.3.1 for details on profiles). A backup copy of the file in its old standard will be saved with the extension **.old.imdi**.

A window appears, informing you about this process, e.g.:
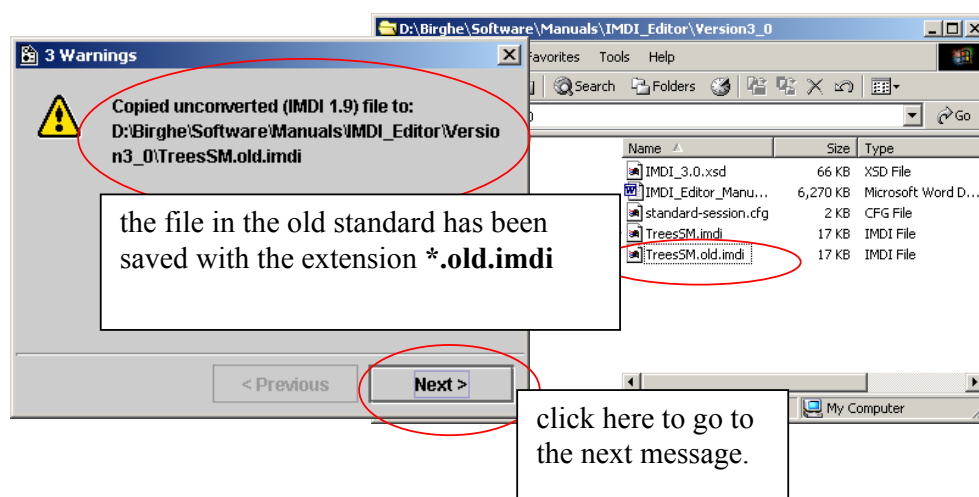
Furthermore, the converter informs you about any errors that took place during the conversion process. Click on "Next" to read this information.

The following two types of errors can occur:

(1) Errors concerning the Genre element: In earlier versions, Genre consisted of three separate subtypes. These were removed and were mapped onto one Genre and one Subgenre instead (see section 2.7.1). While some of these mappings are known to the converter, others are not. If the mapping is not known, the converter will generate an error message, informing you that the old Genre has been mapped onto the Keys schema of the Content screen instead (see section 2.3.4), e.g.:
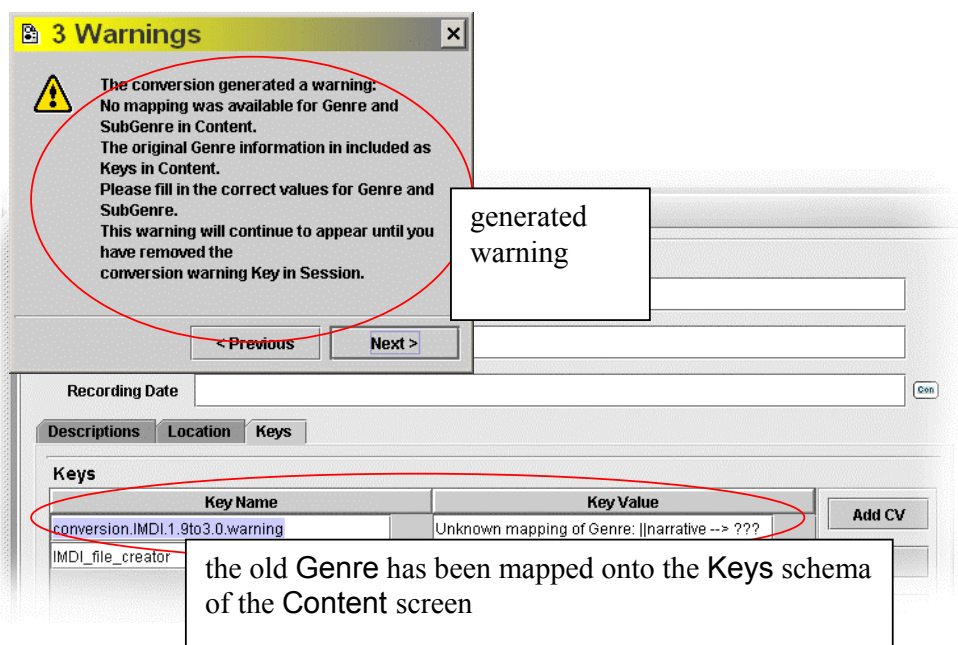
Please remove this keyword manually, and fill in the Genre and Subgenre fields by hand.

(2) Errors concerning then usage of wrong values, e.g., the usage of a wrong format (such as a wrong format for a date or an age) or the usage of a value that is not specified in a closed controlled vocabulary (see section 1.2.3.3 for controlled vocabularies). In all these cases, you will receive the following message, informing you that such errors were detected:
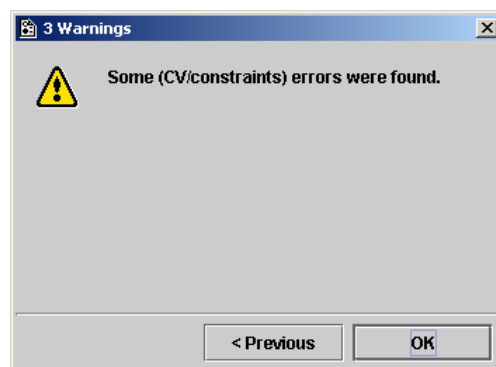
In the IMDI file itself, such errors are highlighted in orange color, and a description of the error is added to the tool-tip (which is activated by pointing with the mouse to the category label). You can navigate to these errors by using the two arrow buttons at the top of the "Content" panel, e.g.:
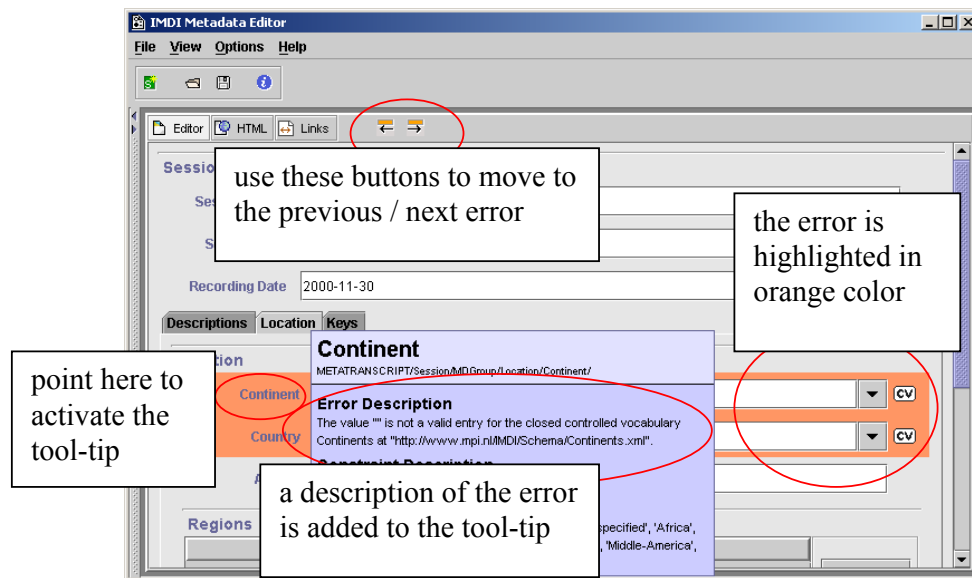
Please fill in the correct values by hand.

> **!** Note: Unless you manually change the items that do not conform to the IMDI 3.0 standard, the error messages will continue to appear (in the case of the Genre errors), and the offending items will continue to be highlighted in orange color (in the case of the value errors).

! Note: If the converter was able to map the old Genre onto the new Genre, it will inform you of this fact: it will add a description (under Content . Description) and it will additionally display the old Genre as keywords (under Content . Keys), e.g.:

a description informing you of the successful conversion

the old Genre is additionally displayed in the form keywords
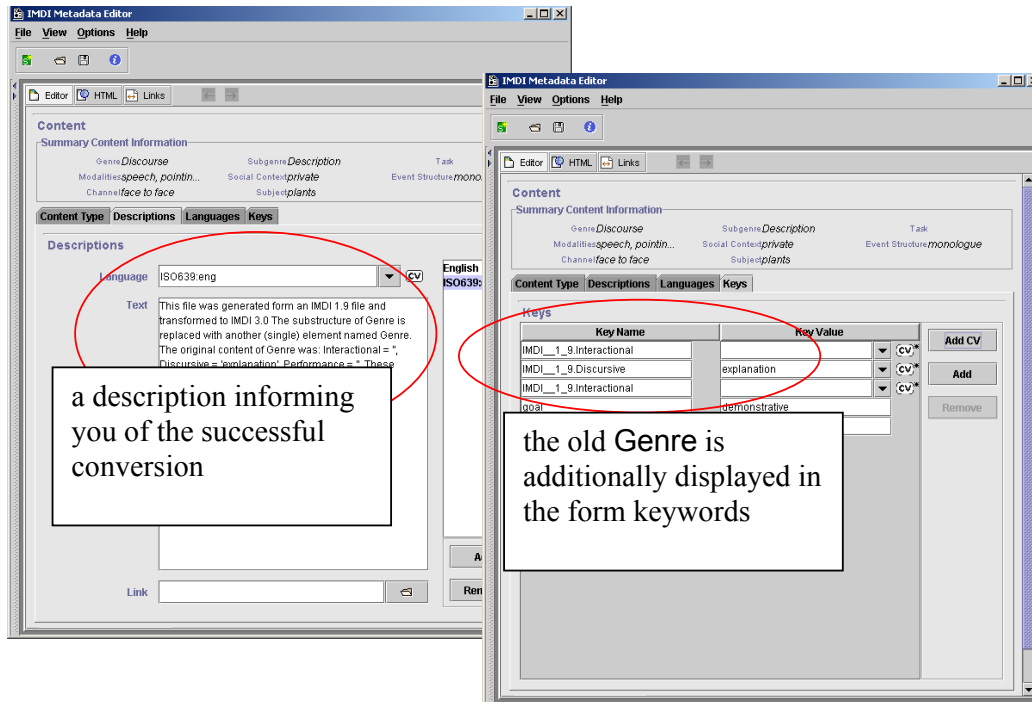
2.7.3 Conversion of old repositories to the 3.0 standard

When you start the IMDI Editor version 3.0 for the first time, it will automatically convert all old repositories to the new IMDI 3.0 standard (see section 1.4 for details on repositories). The converted repository is **not** added to the node "Local Repository", but to the node "Local (Old IMDI Version)". The old repositories are not deleted – if you need to revert back to an older version of the IMDI Editor, the unconverted repositories will be used again.

the old repository is automatically converted and added to the node "Local (Old IMDI Version)"
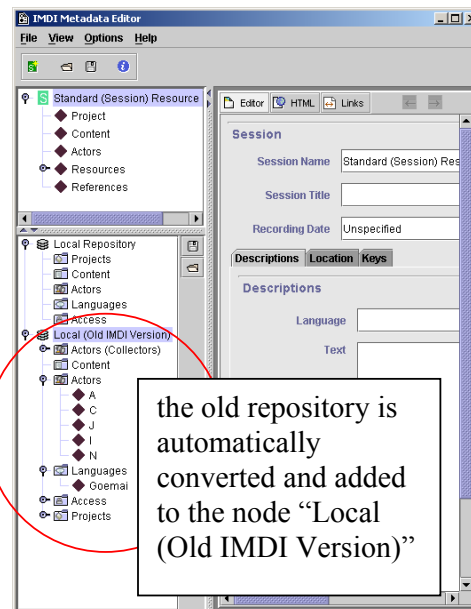
3 The IMDI Lexicon Editor

The IMDI Lexicon Editor creates IMDI Lexicon files (***.imdi**) that describe lexicon resources. Its structure is almost identical to the IMDI Session Editor, i.e., it displays the screens Session (see section 2.1), Project (see section 2.2), Content (see section 2.3), Actors (see section 2.4), and References (see section 2.6). The only difference concerns the screen Resources: the IMDI Lexicon Editor does not display Anonymous, Media Files, Written Resources and Sources – instead it displays Lexicon Resource. This section of the manual describes only the screen Lexicon Resource. For the other screens, please consult section 2.

3.1 Lexicon Resource

The lexicon resource associated with the session is listed under Lexicon Resource, which allows you to give general information about the resource, about its lexical entries, about the metalanguage used, and about access rights and restrictions to the resource. You can furthermore give a prose description and add keywords.
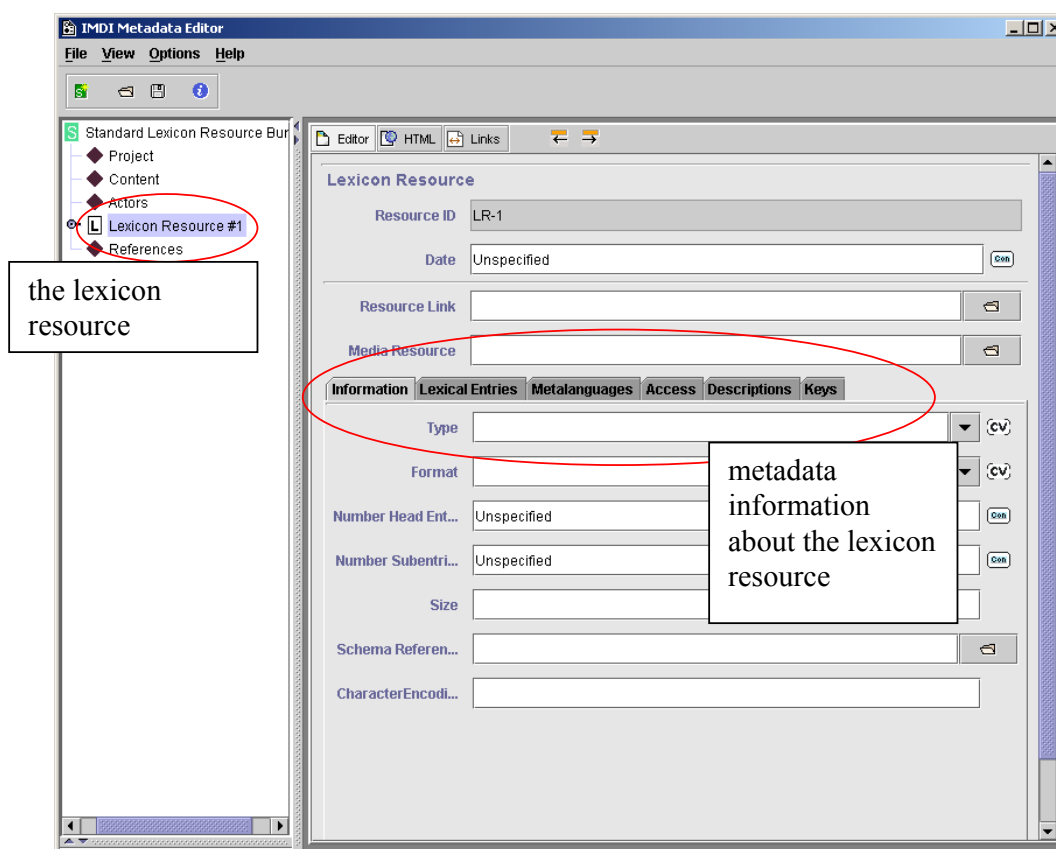
! Note: This part of the IMDI Editor is still under development, and the controlled vocabularies for the metadata categories have not yet been specified. Please consult the web page http://www.mpi.nl/IMDI/ for further information.

111

4 The IMDI Corpus Editor

This part of the manual guides you through using the IMDI Corpus Editor to create a corpus hierarchy. All sessions or resource bundles (see section 2) and lexicon resource bundles (see section 3) belong to a corpus, e.g., to the corpus of the language XY, which, in turn, is subdivided according to various criteria (e.g., it may be subdivided on the basis of genre, or on the basis of the age of the participants, etc.).

> **!** Note: In many cases, the corpus manager takes care of structuring your corpus, i.e., you will **not** need the IMDI Corpus Editor. You only need the Corpus Editor in case you manage your own corpus.

> **!** Note for researcher working at the MPI for Psycholinguistics, Nijmegen: If you plan on using the IMDI Corpus Editor, please contact corpus.manager@mpi.nl first.

The IMDI Corpus Editor creates an IMDI Corpus file (***.imdi**).

The following steps are necessary to create, structure and manage a corpus:

1. Specify the directory information of all info, media, written resource and lexicon files that belong to a session, and of all IMDI files that belong to a corpus (see section 4.1).

2. Select the sessions that belong to a subcorpus. Select the subcorpora that belong to a corpus (see section 4.2).

3. Create a corpus node (see section 4.3).

These steps are explained in the following three sections.

4.1 Specify the directory information of files

The IMDI Editor creates links to different types of files. These links are entered into fields that are labeled Link (see also section 1.5.3):

*   In the IMDI Session Editor, Link fields exist for info files (see "Appendix 1: Recurring Schemata (Descriptions)"), media files (see section 2.5.2), and written resource files (see section 2.5.3).

*   In the IMDI Lexicon Editor, Link fields additionally exist for lexicon files (see section 3.1).

*   In the IMDI Corpus Editor, Link fields exist for IMDI Session, IMDI Lexicon and IMDI Corpus files, i.e., files created with the IMDI Session, Lexicon or Corpus Editor (see section 4.2).

In order to link files, you need to provide information about (a) their names and (b) their directory location.

Usually, if your corpus is managed by a corpus manager, you only need to enter the file name. In this case, type the file name directly into the Link field of the IMDI Session, Lexicon or Corpus Editor. Do **not** make use of the browsing option (i.e., the "open folder" icon to the right of the Link field). Please make sure that **no** prefixes are defined in the "View" / "Preferences …" / "Prefixes" menu (see section 4.1.2).

However, if you manage your own corpus, you need to enter both the file name and the directory location. You can enter this information either directly into the Link field (see section 4.1.1) or you can make use of prefixes (see section 4.1.2).

4.1.1 Link fields

You can enter the directory information directly into the Link field. Do the following:

1. Make sure that the browsing option is set correctly. Do the following:

    1. In the IMDI Editor, click on "Options" menu.

    2. Click on "Preferences …".

    3. In the "Preferences" dialog window, click on "Resources".

    4. In the "Resources" dialog window, make sure that there is **no** checkmark in the box next to "Retain only the filename when browsing".

        ! Note: If there is a checkmark in this box, the directory information will not be saved, e.g.:
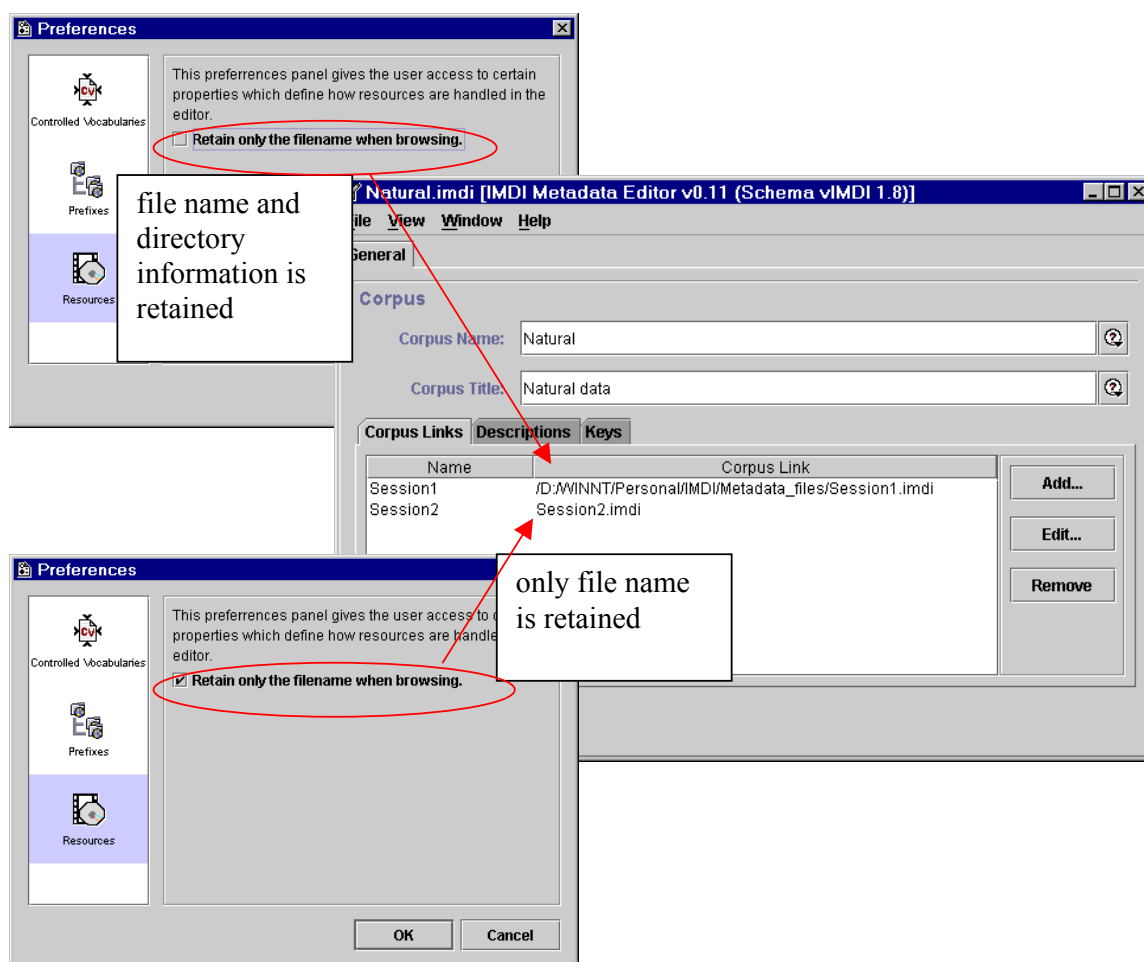
2. Make sure that the preferences are set correctly. Do the following:

 1. In the IMDI Editor, click on "Options" menu.

 2. Click on "Preferences …".

 3. In the "Preferences" dialog window, click on "Prefixes".

 4. In the "Prefixes" dialog window, make sure that there is a checkmark in the box next to "Use these prefixes only when none were defined in the original IMDI file" (see section 4.1.2).

3. In the IMDI Editor, click on the "open folder" icon to the right of the Link field. The "Open" dialog window appears.
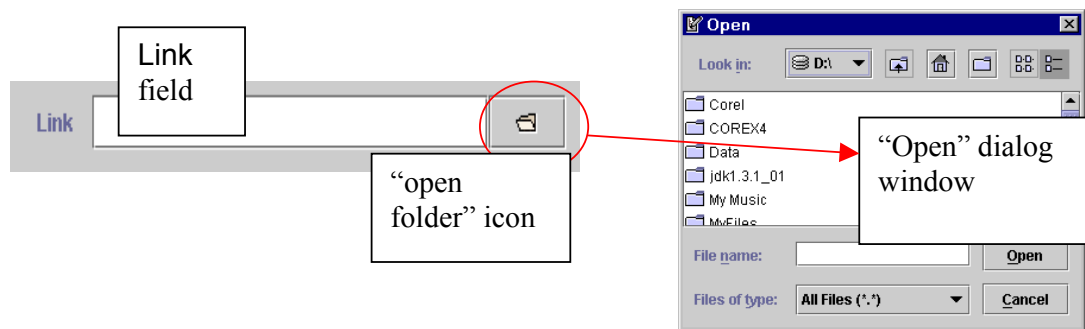
4. In the "Open" dialog window, navigate to the folder that contains the file.

5. Double-click on the file to insert its name and directory location into the Link field.

4.1.2 Prefixes

You can enter the directory information by means of prefixes. Do the following:

1. In the IMDI Editor, click on "Options" menu.

2. Click on "Preferences …".

3. In the "Preferences" dialog window, click on "Prefixes". The "Prefixes" dialog window opens.

4. In the "Prefixes" dialog window, specify when the prefixes should be used. Choose one of the following two options (by clicking in the box next to this option):

   (1) "Use these prefixes only when none were defined in the original IMDI file."

   I.e., if the directory in formation is already specified in the original IMDI file, the prefixes defined in the "Prefixes" dialog window will not be used.

   (2) "Always use the prefixes defined in this panel."

   I.e., the prefixes are always used – even if a different directory is specified in the original IMDI file.

5. In the "Prefixes" dialog window, specify where you have defined the prefixes. Choose one of the following two options (by clicking in the box next to it):

(1) "Use Prefixes defined in the IMDI file itself."

Choose this option if the prefixes are defined in the "Prefixes" dialog window. These prefixes will then be saved within the IMDI file itself. To define prefixes, enter the directory information for each type of file, e.g.:

- "Global": the directory that is shared by all files, e.g., the path "D:\WINNT\Personal\IMDI\" in the illustration below.

- "Session Files": the location of your IMDI Session files (relative to Global), e.g., the folder "Metadata_files\" in the illustration below.

- "Corpus Files": the location of your IMDI Corpus files (relative to "Global").

- "Info Files": the location of your info files (relative to "Global").

- "Media Files": the location of your media files (relative to "Global").

- "Written Resources": the location of your written resource files (relative to "Global").

- "Anonymous Info": the location of your file that specifies the mapping of codes onto full names (relative to "Global").

    ! Note: Please add a backslash after each folder.

    ! Note: This option is especially useful whenever files of the same type are stored in the same directory – because then you only need to specify the directory information once in the "Prefixes" dialog window.
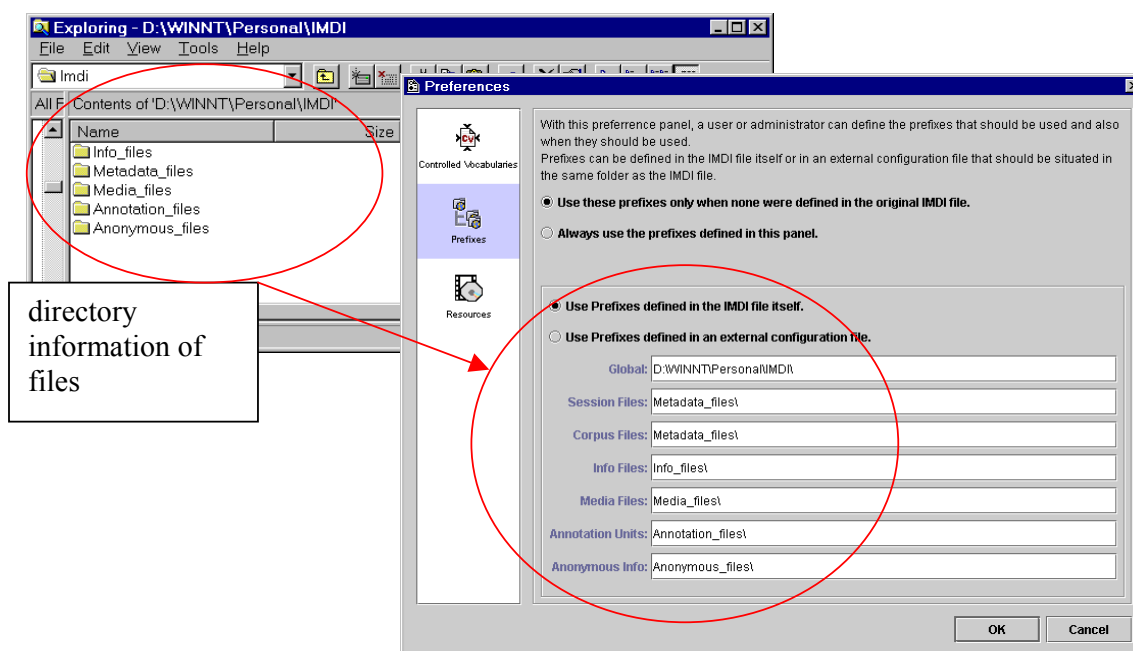
(2) "Use Prefixes defined in an external configuration file."

Choose this option if the prefixes are defined in an external configuration file.

This option is more flexible than option (1) above because the prefixes are not saved in the IMDI file itself. I.e., if you decide to move files to a different directory at a later stage, you need to change the prefixes only once in the external configuration file – not in each IMDI file. It is therefore recommended that you make use of this option whenever you manage a large corpus.

If you choose this option, you are prompted to enter the name and directory of the configuration file. It should be in the same folder as your IMDI Session files.

enter name and directory of the configuration file

configuration file, e.g.:

```
<?xml version="1.0" encoding="ISO-8859-1"?><!DOCTYPE
METATRANSCRIPT
[<!ENTITY annotationunitPrefix "Annotation_files\">
<!ENTITY corpusPrefix "Metadata_files\">
<!ENTITY infolinkPrefix "Info_files\">
<!ENTITY globalPrefix "D:\WINNT\Personal\IMDI\">
<!ENTITY anonymousPrefix "Anonymous_files\">
<!ENTITY mediafilePrefix "Media_files\">
<!ENTITY sessionPrefix "Metadata_files\">]>
```

! Note: All preferences that have been set in the "Options / Preferences …" menu are automatically used for each new IMDI file – unless you change them manually. Whenever you create a new IMDI file, please make sure that the preferences are set correctly, so that you do not accidentally insert wrong prefixes.

There is a tool-tip available that reminds you of the defined prefixes. To access it, point with the mouse into the Link field. A blue box appears informing you about the prefixes, e.g.:

point here

Link protocol.pdf

defined prefixes

PREFIX= D:\WINNT\Personal\IMDI\Info_files\
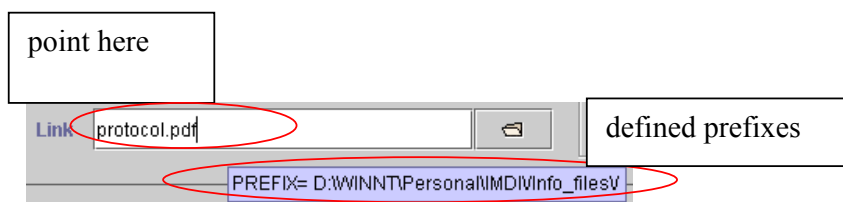
4.2 Select the sessions and subcorpora that belong to a corpus

A corpus consists of subcorpora and sessions. For example, the following illustration shows a corpus labeled "Goemai corpus (preliminary version)", together with its two subcorpora "Natural data" and "Elicited data". The subcorpus "Elicited data" contains further subcorpora (labeled "Matching games" and "Picture books"). Each subcorpus consists of sessions (labeled "Session1" to "Session6") that contain the actual session data (i.e., IMDI Session files with metadata information and links to info, media, written resource, and lexicon files).

Goemai corpus (preliminary version):
    Natural data
        Session1
        Session2
        Session3
    Elicited data
        Matching games
            Session4
            Session5
        Picture books
            Session6
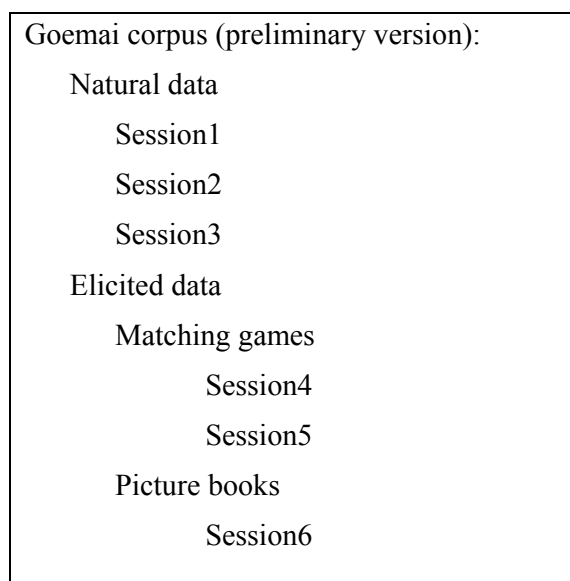
To specify such a hierarchical corpus structure, you need to create an IMDI Corpus file for each corpus and subcorpus. You are asked to provide the following information:

(1) Corpus Name:

A short name or abbreviation that uniquely identifies the corpus or subcorpus.

> **!** Note for researchers working at the MPI for Psycholinguistics, Nijmegen: The Corpus Name should be the same as the name of the corresponding IMDI Corpus file. Furthermore, the file name has to be Unix compatible: do not use file names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

(2) Corpus Title:

The complete title of the corpus or subcorpus. Usually, it is the spelled out version of the abbreviated Name.

(3) Descriptions:

Descriptive information about the corpus or subcorpus. See "Appendix 1: Recurring Schemata (Descriptions)" for instructions on how to fill in a

Descriptions schema. Remember: The field Language refers to the language in which the description is written - **not** to the language under investigation.

(4) Corpus Links:

Specify links to (a) all IMDI Session files or (b) all IMDI Corpus files that belong to the subcorpus or corpus.

(a) Specify the IMDI Session files that belong to the subcorpus. E.g., in the IMDI Corpus file **Natural.imdi** (i.e., "Natural data"), specify that it contains the IMDI Session files **Session1.imdi**, **Session2.imdi** and **Session3.imdi**.
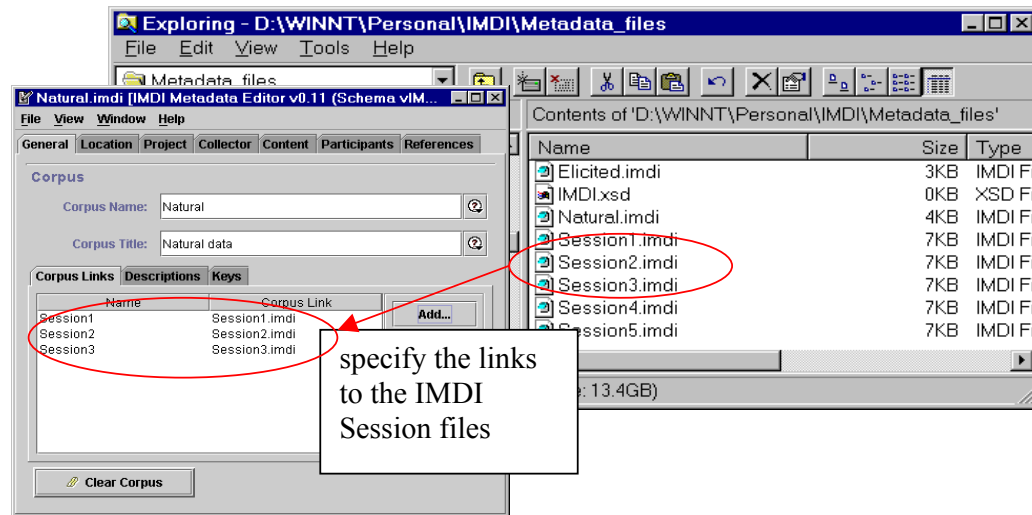
(b) Specify the IMDI Corpus files that belong to the corpus or subcorpus. E.g., in the IMDI Corpus file **Goemai.imdi** (i.e., "Goemai corpus (preliminary version)"), specify that it contains the IMDI Corpus files **Natural.imdi** (i.e., "Natural data") and **Elicited.imdi** (i.e., "Elicited data").
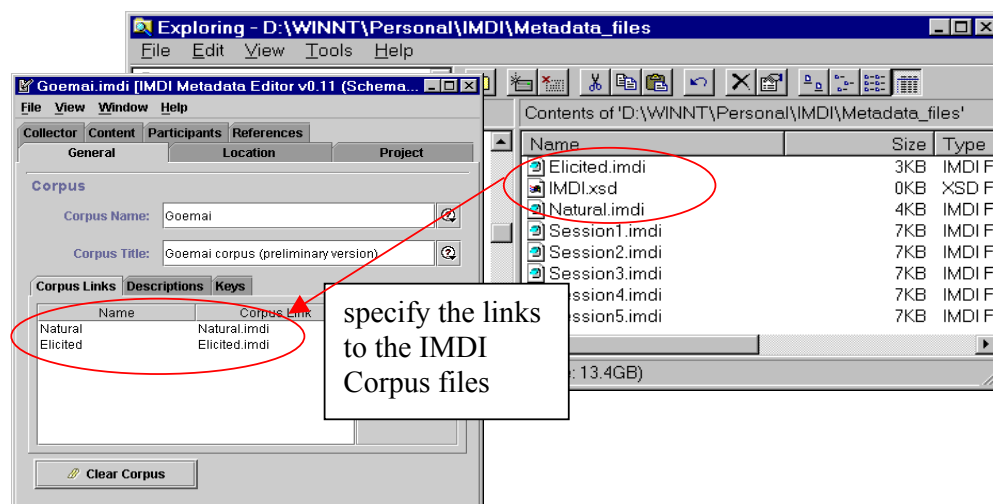
To specify a link, do the following:

1. Click on the "Add" button. The "CorpusLink" dialog window appears.
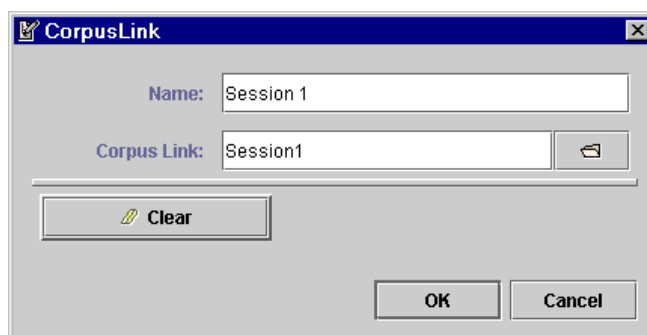
2. Enter the name of the session or corpus in the field Name.

3. Specify the directory location of the corresponding IMDI Session/Corpus file in the field Corpus Link.

4.3 Create a corpus node

As the last step, you need to create a corpus node in the IMDI Browser. To create such a node, do the following:

1. Open the IMDI Browser.

2. At the bottom of the "Metadata Descriptions Tree" panel, enter the location of the IMDI Corpus file into the field labeled Root URL. Press ENTER.

    The corpus node, together with its subcorpus nodes and sessions, is displayed in the "Metadata Descriptions Tree" panel, e.g.:
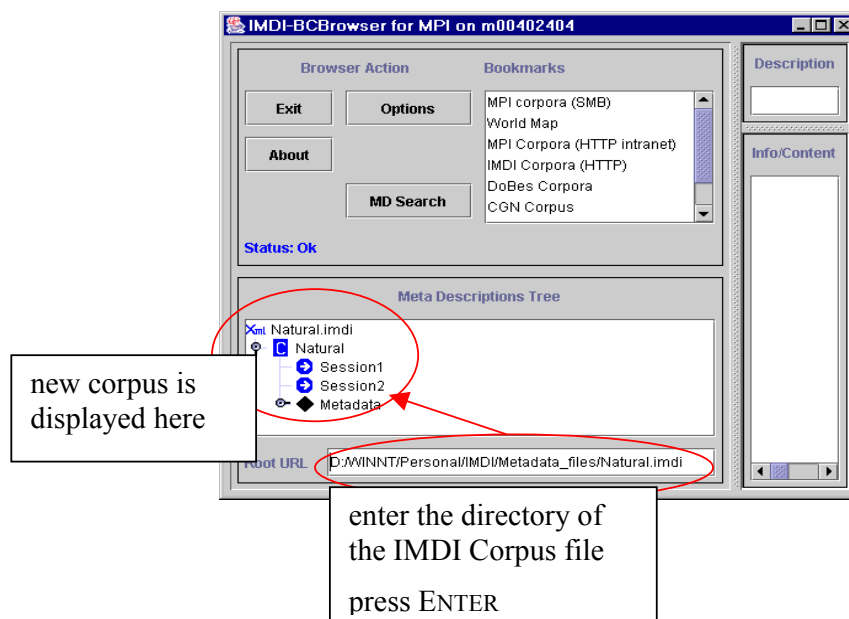
3. To save the new corpus node permanently, do the following:

1. In the "Metadata Descriptions Tree" panel, click on the corpus node to select it.

2. Right-click on the selected corpus node.

    A pull-down menu appears.

3. Choose "Add to Bookmarks" from the pull-down menu.

    The "Input" dialog window appears.

4. Enter a name for the bookmark and click "OK".

The new corpus node is saved as a bookmark in the "Bookmarks" panel. Every time you restart the IMDI Browser, you can access your corpus via the "Bookmarks" panel (see the separate manual "IMDI Browser" for details).
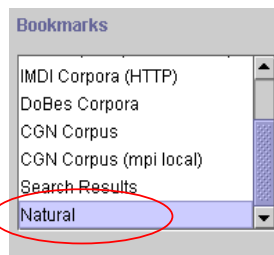
Appendix 1: Recurring schemata

This appendix gives information about the recurring schemata Descriptions, Keys and Access.

Appendix 1: Recurring Schemata (Descriptions)

A Descriptions schema contains a prose description that could serve as a reminder (to you or others) of the circumstances of data collection. Descriptions schemata are not searchable.

The Descriptions schema is structured as follows:
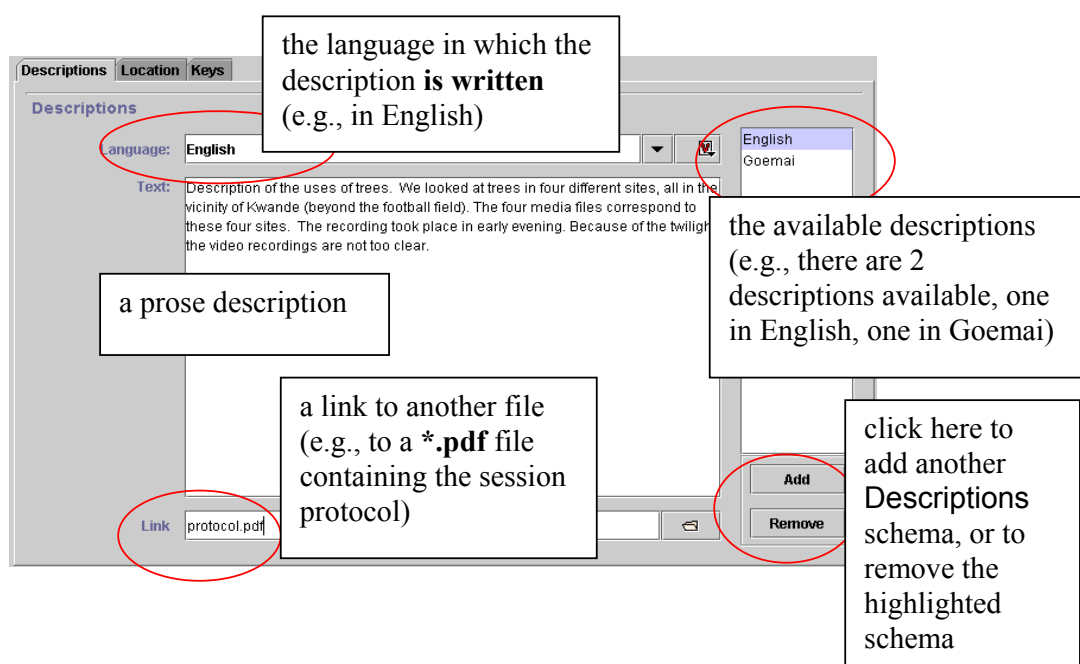
Language   The language in which the **description is written**.

Either type in the language, or select the language from the pull-down menu.

! Note: In earlier versions of the IMDI Editor, a code was entered into this field. Whenever you open an IMDI file created with such an earlier version, this code is displayed (e.g., "RFC1766:x-sil-eng" instead of English). In newer versions of the Editor, the code is still stored in the file itself (e.g., "RFC1766:x-sil-eng") – but the user interface will display the code as a name (e.g., as "English").

Text   A prose description.

Link   A link to an info file that contains further relevant information,
              e.g., a session protocol.

> **!** Note for researchers working at the MPI for
> Psycholinguistics, Nijmegen: Please specify only the name of
> the file – do **<u>not</u>** make use of the "open folder" icon to the right
> of the Link field. And please save the info file as either a ***.txt**
> file, a ***.pdf** file or an ***.html** file.

You can create descriptions in different languages, e.g., one description in English
(for the research community) and one in the language under investigation (for the
language community). In this case, you have to add another Descriptions schema.
Do the following:

- Click on the "Add" button to add a new schema.

- All schemata are displayed at the right of the Descriptions schema. They are
  listed under the language name (as defined in the Language field).

  - To access a schema, highlight it by clicking on it. It appears in blue color,
    and its content is displayed. The content of all other schemata is hidden.

  - To delete a schema, highlight it by clicking on it. Then click on the
    "Remove" button to delete it.

Appendix 1: Recurring Schemata (Keys)

Keys contain user-defined information that is (a) project-specific, (b) not taken care of in the standardized fields, and (c) meant to be searched. For example:
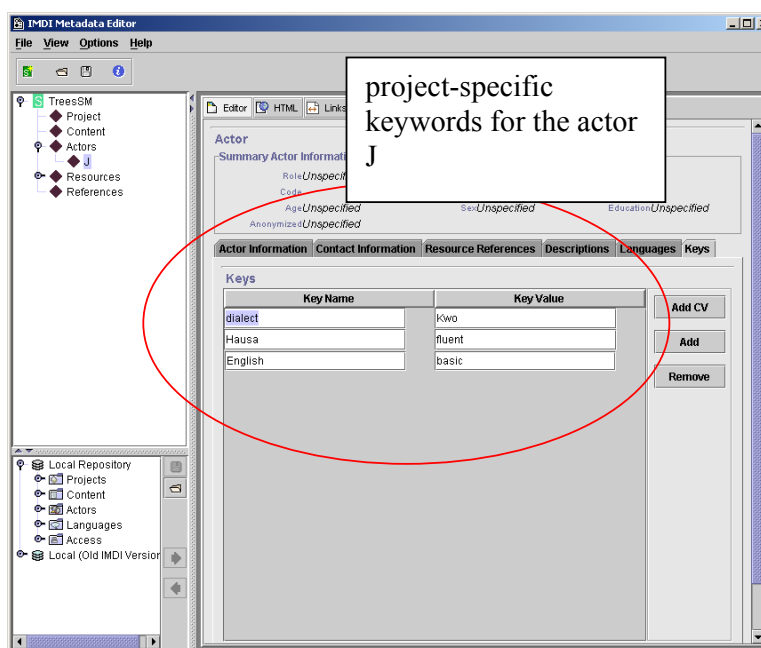
To add a key, do the following:

1. Click on the "Add" button. A new key will be added.

2. Specify a Key name. The name has to be a single word (e.g., 'MetaDescriptionCreator'). Do not use blank spaces.

3. Specify a Key value. Multiple words are allowed (e.g., 'Student Assistant XY').

To delete a key again, highlight it by clicking on it. Then click "Remove".

It is furthermore possible to add (project-specific) controlled vocabularies to a Keys schema, i.e., to have a Keys schema display a pull-down menu containing predefined values (see also section 1.2.3.3; and see the IMDI CV-Editor manual for instructions on how to create controlled vocabularies).

123

To add a controlled vocabulary to a **Keys** schema, do the following:

1. Click on "Add CV".

2. The "Vocabs overview" window opens and displays all available vocabularies. If the needed vocabulary is not displayed, add it by clicking on "Add File …" (for a locally-stored vocabulary) or "Add URL …" (for a vocabulary stored on the Web).

3. Click on a vocabulary to select it.

4. Click on "OK" to add the selected vocabulary to the **Keys** schema.

5. A dialog window appears, asking you to choose the type of vocabulary (see also section 1.5.2):

   - "vocabulary" (only one value can be entered) vs. "vocabulary list" (more than one value can be entered);

   - "closed" (only values from the pull-down menu are accepted) vs. "open" (alternative values can be added)

6. Click on "OK".

7. The controlled vocabulary is added to the **Keys** schema.

Appendix 1: Recurring schemata (Access)

Access schemata group information about access rights to a resource.

> **!** Note: The different archives handle access rights differently. For example, many of the corpora housed at the MPI for Psycholinguistics, Nijmegen, automatically deny access to media files, written resource files, lexicon files and media tapes. But other archives may follow a different policy.

Specify the access rights as follows:

Information	General information about access rights to the resource.
Availability	Information about the availability of the resource. This field is not standardized yet – for the moment, please enter a prose text (e.g., not available, available in 5 years, available to person XY).
Date	Date at which the access rights were set. Please enter the date in the following format: **YYYY-MM-DD**, e.g. 2000-12-30.
Owner	Name of the person/institution that owns the resource.
Publisher	Name of the publisher who is responsible for the distribution of the resource.

Contact   Information about whom to contact in order to gain access to the resource: Name, Address, E-mail address, and Organisation.

Descriptions   Prose description of the access restrictions.

Appendix B

IMDI Browser, version 2.0

Manual

This manual was last updated: 30 Jun 2004
The manual can be downloaded from the following webpage: http://www.mpi.nl/IMDI/tools/
Author: Birgit Hellwig (Updates: Dieter Van Uytvanck)

Introduction

The IMDI (ISLE Metadata Initiative) Browser was developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. It is used to manage metadata and data files. As such, it complements the other IMDI tools, notably the IMDI Editor.

At the moment, the following IMDI tools exist (tools and manuals can be downloaded from http://www.mpi.nl/IMDI/tools/):

- IMDI (Metadata) Editor: a tool for creating metadata descriptions, both for so-called sessions or resource bundles[8] and for subcorpus nodes;

- IMDI CV-Editor: a tool for creating lists of controlled vocabulary items;

- IMDI Metadata Search Tool: a tool for searching through metadata information;

- IMDI Tree Builder: a tool for creating browsable trees based on IMDI metadata descriptions (beta version).

The IMDI Browser supports the following features:

- displaying and accessing data and metadata through a hierarchical tree structure and a world map;

- direct access to media, annotation and metadata files;

- searching of metadata files (by means of the IMDI Metadata Search Tool).

This manual explains and exemplifies the features of the IMDI Browser. It is organized around the following three chapters:

1. Basic information about the IMDI Browser (section 5).

2. Displaying and accessing data (section 6).

3. Searching data (IMDI Metadata Search Tool) (section 7).

> **!** Note: This manual contains a number of references to corpora housed at the Max Planck Institute for Psycholinguistics (MPI). It uses these corpora as a means to illustrate the functionality of the IMDI Browser. Please note that any remark (e.g., about the structure of the corpora, their availability, the contact

---

[8] Please note: throughout this manual, the terms 'session' and 'resource bundle' are used interchangeably (see also section 1.2).

person etc.) only pertain to these corpora. Corpora stored at other archives or institutes follow different rules.

Notation Conventions

The following notation conventions are used:

- Menu items, icons and screen displays are written in the font MS Sans Serif.

- (SHORTCUT) KEYS ARE WRITTEN IN SMALL CAPS.

- Information on troubleshooting starts as follows: **!**

Table of Contents

5 Basic information

This section of the manual introduces you to the basic concepts and possibilities of the IMDI Browser. It introduces the concepts of corpora (section 5.1) and of sessions or resource bundles (section 5.2). For illustration purposes, it makes use of the corpora stored at the Max Planck Institute for Psycholinguistics; for corpora stored at other institutes, different rules may apply.

5.1 Corpora

By default, the IMDI Browser gives access to all corpora that are stored at the Max Planck Institute for Psycholinguistics (MPI), Nijmegen, The Netherlands.

In addition – or alternatively – the Browser can display corpora created by you or by other archiving initiatives (see section 6.1.1).

Currently, the following corpora are housed at the MPI:
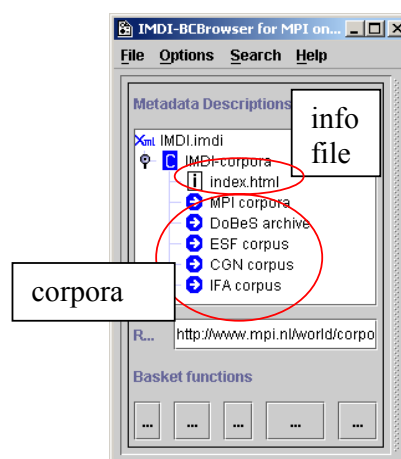
MPI corpora	Corpora collected by researchers affiliated with the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands.
DoBeS archive	Corpora collected by researchers affiliated with the DoBeS (Dokumentation Bedrohter Sprachen) project funded by the Volkswagen foundation.
ESF corpus	European Science Foundation Second Language Acquisition Database.
CGN	Corpus Gesproken Nederlands.[9]
IFA corpus	Corpora collected by researchers affiliated with the Instituut voor Fonetische Wetenschappen (Institute of Phonetic Sciences), Amsterdam, The Netherlands.

> **!** Note: Depending on the version of the IMDI Browser, it may not always be possible to access all corpora.

---

[9] If you access the Corpus Gesproken Nederlands via the IMDI Browser, content searches are not possible. To conduct content searches in the CGN corpus, you need the corresponding CDs (see the separate manual "Corpus Gesproken Nederlands (COREX)" for details).

Each corpus contains further subcorpora collected by individual researchers or project teams. The internal structure of each subcorpus varies according to the purposes and needs of the project. You can access information about their content and structure through the info files displayed in the IMDI Browser (see section 6.3.4.3).

Each corpus contains metadata, media, annotation and info files (see section 6.3.4). These files can be accessed through the following two mechanisms (but please note that data files may not always be accessible to the general public; see section 5.2):

(a) By navigating through the hierarchical tree structure (see section 6.3.1).

Each corpus is organized hierarchically in the form of tree structures (see the illustration to the right).

Tree structures consist of nodes that group files together on the basis of, e.g., the geographical region, the discourse genre, the sex or age of the speaker, the dialect of the speaker, the target/source language etc.

All nodes are displayed in the IMDI Browser. Double-click on a node to access the next level in the hierarchy.

At the lowest level of the hierarchy, the actual metadata, media, annotation and info files are displayed.
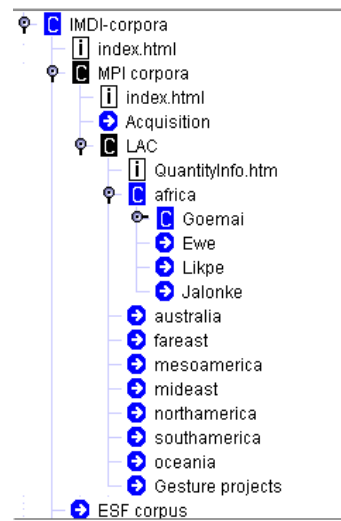
> **!** Note: The tree structure does **not** display the physical location of files. This means that the same file can easily be displayed in different parts of the corpus, e.g., the same file may be displayed under the nodes 'folktale', 'male speaker', and 'age-group 20 to 30 years'.

(b) By searching the metadata files (see section 7).

All resources or data files are accompanied by metadata files, i.e., files that give information about the data. These metadata files are searchable. For example, as illustrated below, you could search for all texts by female speakers above 60 years of age. The IMDI Browser displays the search results, and allows you direct access to the corresponding data files.
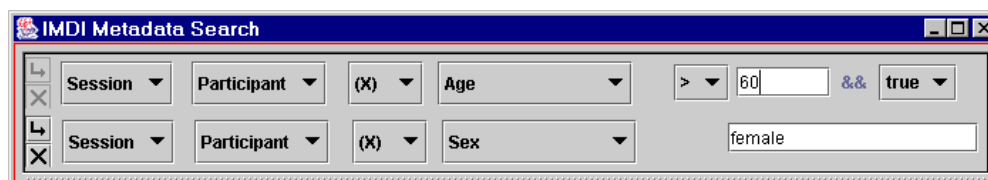
5.2 Sessions or resource bundles

The lowest level of the corpus hierarchy contains the session or resource bundle. (These two terms are used interchangeably, 'session' being the preferred term for spoken corpora, and 'resource bundle' for written corpora.) A session or resource bundle corresponds to a meaningful unit of analysis, usually to a piece of data having the same overall content, the same set of participants, and the same location and time, e.g., one elicitation session on topic X, or one folktale, or one 'matching game', or one conversation between several speakers.

A session or resource bundle contains different files. The following four types of files exist:

(1) Metadata files (see section 6.3.4.1).

These files contain information <u>about</u> the session, e.g., its date and location, its content and its participants. They are of the IMDI Editor format.

(2) Media files (see section 6.3.4.2).

These files contain the audio or video recordings. They are usually digitized in one of the following formats: MPEG (***.mpg**), Cinepak-Quicktime-Movies (***.mov**), WAVE (***.wav**).

(3) Annotation files (see section 6.3.4.2).

These files contain the transcripts, codings and annotations. Their format varies (e.g., ELAN, Media Tagger, Shoebox, CHAT, etc.)

(4) Info files (see section 6.3.4.3).

These files contain further background information on specific topics. They are in PDF or HTML formats.

With a few exceptions, metadata and info files are publicly accessible. This policy enables researchers to search metadata and to thereby gain an overview of the available material.

Annotation and media files are handled differently in the different corpora: some corpora allow for general access (e.g., the CGN corpus), while others do not (e.g., the MPI corpora). In the latter case, if you have not been explicitly granted access by the

responsible researcher(s), you will not be able to access the files. To gain access, ask the responsible researcher(s) for permission. Contact details are displayed in each corpus, e.g.:
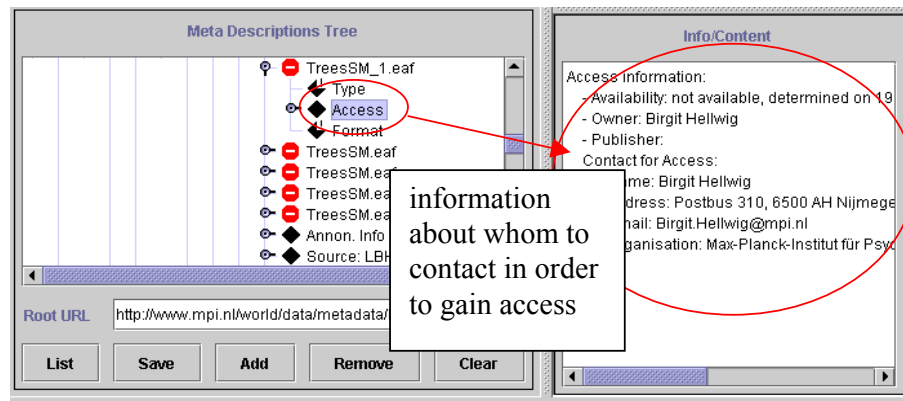
Responsible researcher(s), by contrast, have unrestricted access to all files. This includes the following possibilities:

- Access to metadata and info files, including (a) all suppressed information and (b) the possibility to update the information. See sections 6.3.4.1 and 6.3.4.3.

- Access to media files. See section 6.3.4.2.

- Access to annotation files, including the possibility to continuously update them. See section 6.3.4.2.

6 Displaying and accessing data

Starting the IMDI Browser opens up the window IMDI-BCBrowser for MPI (referred to in this manual as the IMDI Browser window). In the IMDI Browser window, you can view and access the corpora, i.e., you can read information about the available data, access the files, and initiate searches. The IMDI Browser window contains the following information:

1. Menu items (see section 6.1)

2. Bookmarks panel (see section 6.2)

3. Metadata Descriptions Tree panel (see section 6.3)

4. Info/Content panel (see section 6.4)

5. Description panel (see section 6.5)
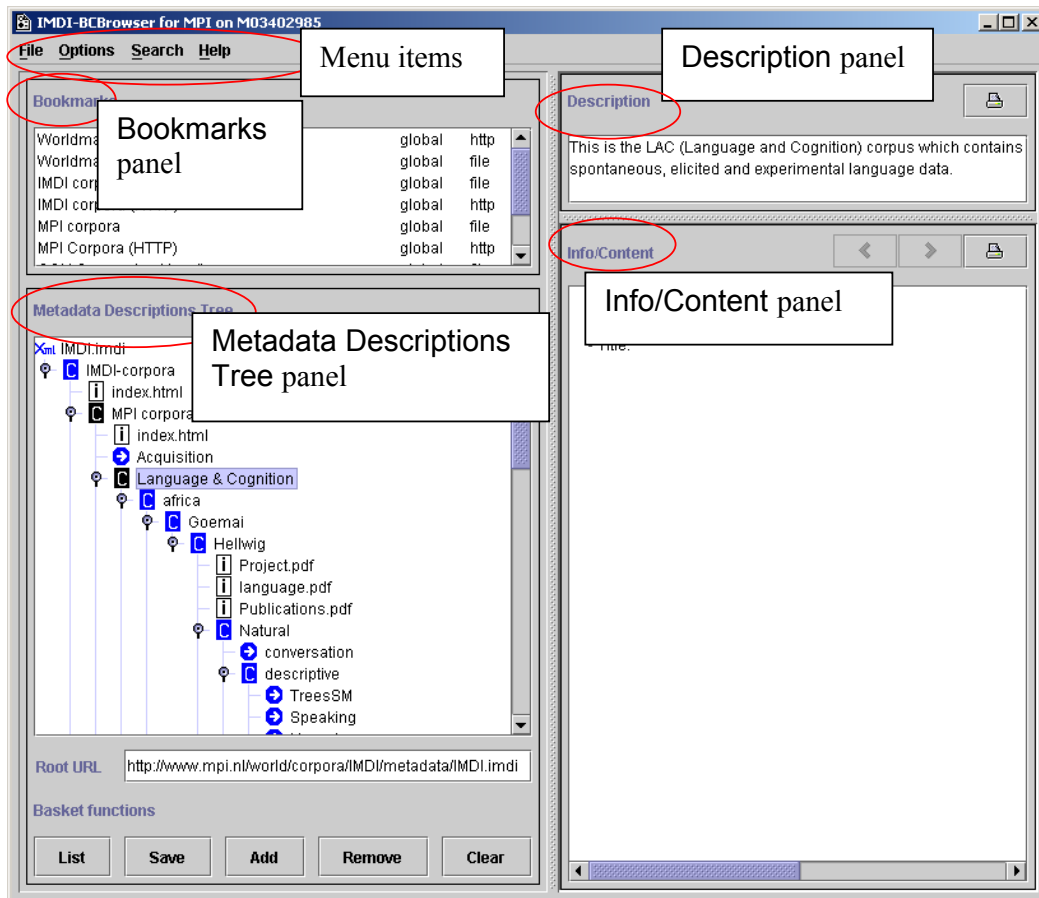
6.1 Menu items

The following four menu items are available: File (see section 6.1.1), Options (see section 6.1.2), Search (see section 6.1.3), Help (see section 6.1.4).

To access any menu item, do one of the following:

(e) Click on that item.

(f) Or use the shortcut key ALT plus the underlined letter(s) (e.g., to access the item File, press the key ALT+F).

6.1.1   File menu

The File menu supports the following two options:

Open File

Click on Open File to open a corpus file, i.e., an *.imdi corpus file that has been created with the IMDI Editor and that contains information about the corpus.

The Open dialog box appears, allowing you to browse to the folder that contains the *.imdi corpus file. In the dialog box, click on Open, and the IMDI Browser window will display the selected corpus node. E.g.:

selected file

file is displayed in the IMDI Browser window
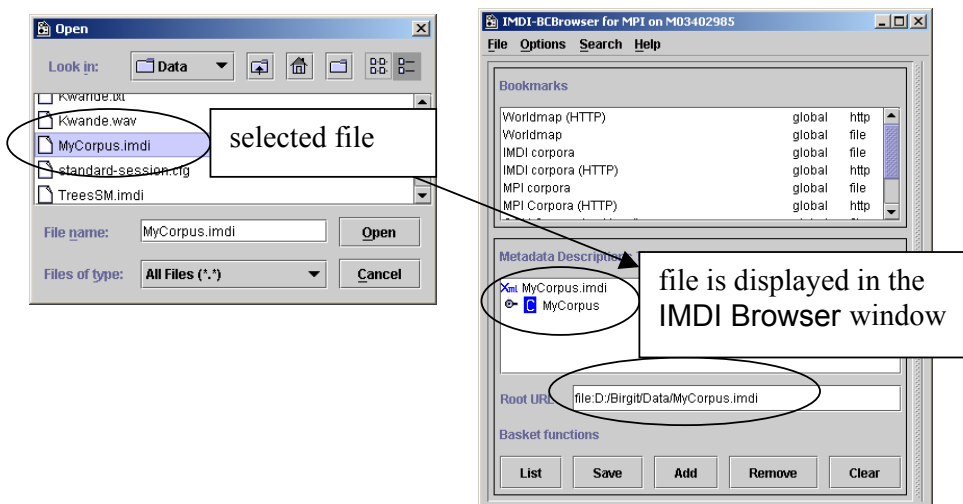
Exit

Click on Exit to exit the IMDI Browser window.

Note that the Exit item turns into a Close item when you open a second IMDI Browser window, either through the option Clone Node (see section 6.3.1) or through accessing your search results (see section 7.4). If you click on Close, only the current IMDI Browser window will be closed.

6.1.2 Options menu

The Options menu supports the following four options:
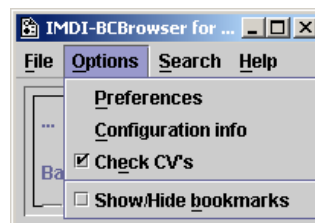
Preferences

Click on Preferences to select the update policy and the fonts. The Preferences dialog window appears, offering you the following two options:

Update Policy

Click on Update Policy to determine the policy for updating controlled vocabularies stored in the cache. The following policies are available:
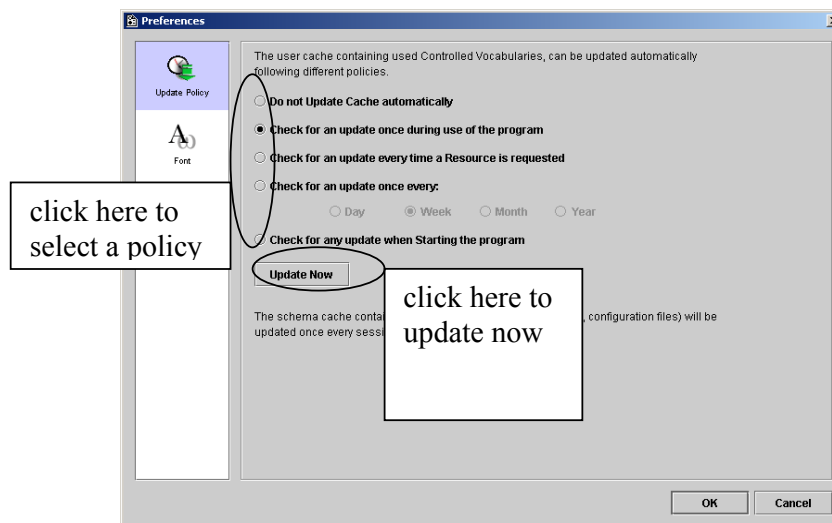
! Note: The IMDI tools make use of so-called "controlled vocabularies" (i.e., lists of values for a specific key) to ensure consistency across users. These

vocabularies are stored on the Web, and a copy is stored in the cache of your desktop computer (in the folder "user-directory \ IMDI-TOOLS \ CVCACHE"). The Web version is changed sometimes, e.g., new values are added. However, the IMDI tools will continue to use the locally-stored version until the cache is updated again. By default, it is updated whenever you run the IMDI Browser. However, you can change this policy, and determine if and when it should be updated.

Font

This option allows you to display information from metadata files that contains special character sets. To select a set and a font, do the following:

4. Under **Unicode Block** (in the left panel), select a unicode block by clicking on it. The available fonts for that block are displayed under **Font** (in the right panel).

5. Under **Font**, select a font by clicking on it.

6. Click on **OK** to implement the new font in the Browser.
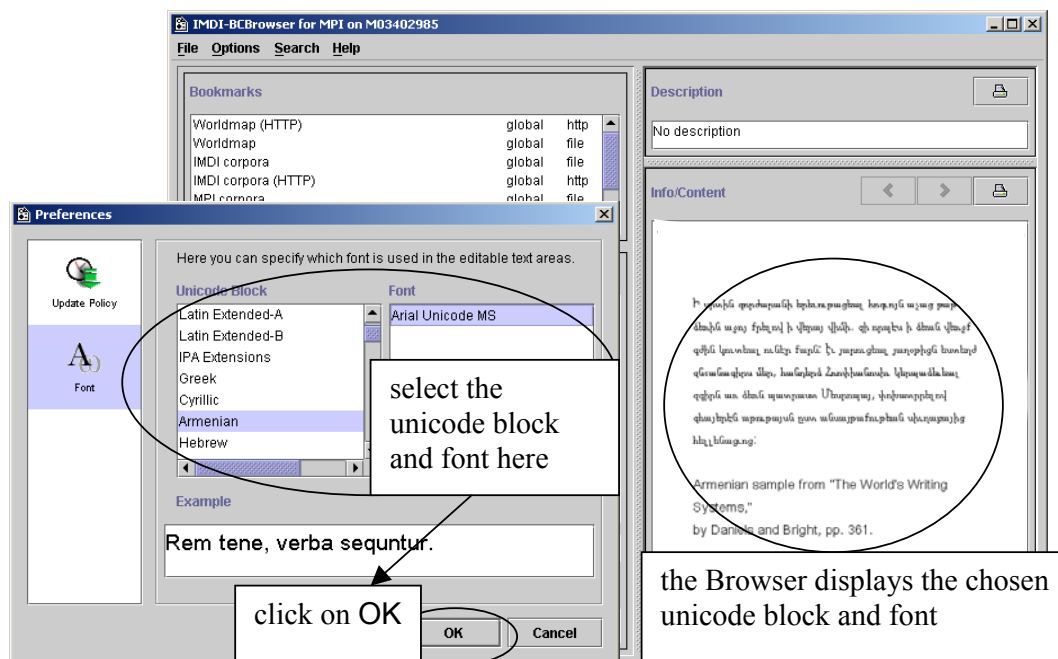
Configuration Info	Click on Configuration Info to view information on the configuration of the tool (you might need this information when reporting a bug to the developers), e.g.:
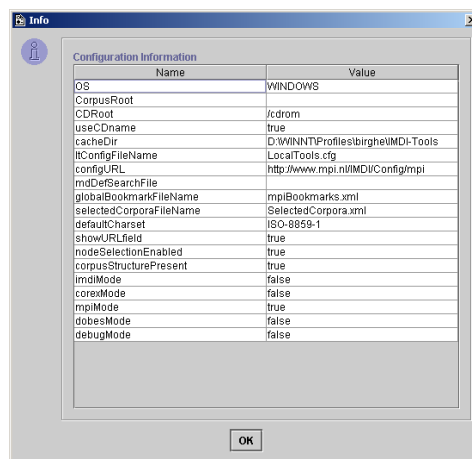
Check CV's	Click in the box to the left of Check CV's to enable the IMDI Browser to check if the values used in a metadata file are correct, i.e., are specified in the controlled vocabulary.
Show/Hide bookmarks	Click in the box to the left of Show/Hide bookmarks to show or hide the Bookmarks panel in the IMDI Browser window.

6.1.3 Search menu

The Search menu supports the following option:
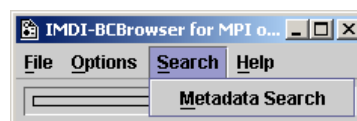
Metadata Search	Click on Metadata Search to access the IMDI Metadata Search tool (see section 7).

6.1.4 Help menu

The Help menu supports the following option:
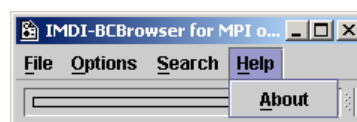
About	Click on About to view the copyright and version information.

138

6.2 Bookmarks panel

In the Bookmarks panel, you can save shortcuts, i.e., "bookmarks", to parts of the corpus. Such bookmarks have the advantage that you do not need to navigate through the entire corpus hierarchy in order to access that part of the corpus.

Depending on the version and the configuration of the IMDI Browser, a number of bookmarks are displayed by default. For example, the following default bookmarks give access to the corpora housed at the Max Planck Institute for Psycholinguistics:

World Map (HTTP)	Access (via the web server) to a world map that displays the locations of documented languages.
World Map	Access (via the local MPI network) to a world map that displays the locations of documented languages.
IMDI Corpora	Access (via the local MPI network) to the MPI, DoBeS, ESF, CGN and IFA corpora.
IMDI Corpora (HTTP)	Access (via the web server) to the MPI, DoBeS, ESF, CGN and IFA corpora.
MPI Corpora	Access (via the local MPI network) to the MPI corpora.
MPI Corpora (HTTP)	Access (via the web server) to the MPI corpora.
CGN Corpus (mpi local)	Access (via the local MPI network) to the CGN corpus.
CGN Corpus (HTTP)	Access (via the web server) to the CGN corpus.
Search Results	Access to your search results.

Whenever you double-click on a bookmark, the corresponding node is displayed in the Metadata Descriptions Tree panel (see section 6.3).

In addition to the predefined bookmarks, you can create your own bookmarks. Do the following:

1. In the Metadata Descriptions Tree panel (see section 6.3), navigate to the corpus or session node for which you want to create a bookmark.

2. Click on the node to select it. It will be highlighted in blue color.

3. Click with the right mouse button on the node to open a pull-down menu.

4. Select Add to Bookmarks from the pull-down menu. The Input dialog box appears.

5. Specify a name for the bookmark, and click on OK. The new bookmark is added to the Bookmarks panel, and will be marked as personal.

E.g.:

navigate to a node, select it, right-click on it and select Add to Bookmarks from the pull-down menu

specify a name for the bookmark, and click on OK

the bookmark is added and marked as personal
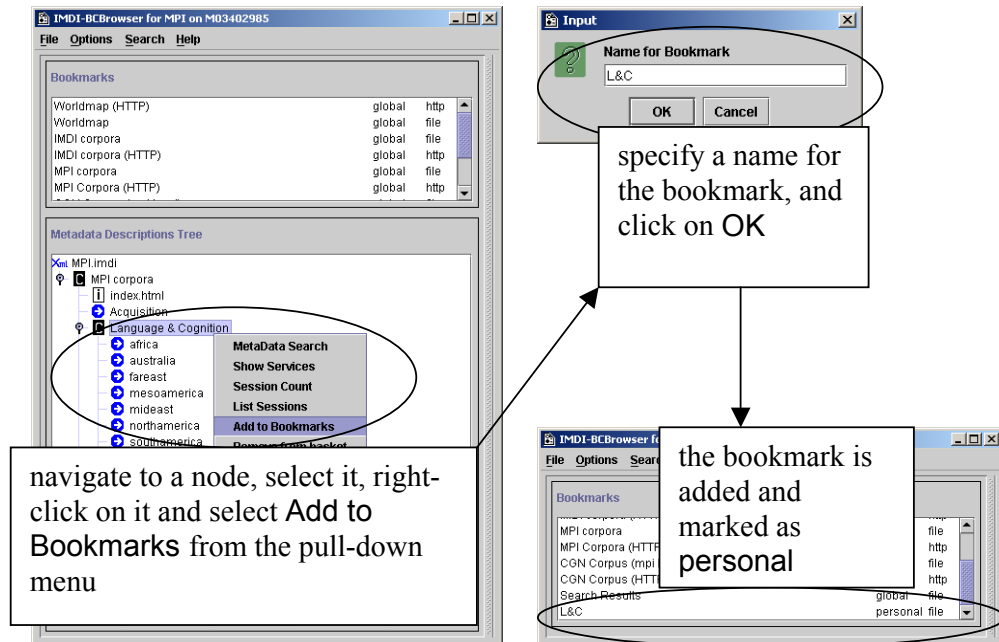
The bookmarks are saved permanently by the IMDI Browser, i.e., they remain available every time you restart the Browser.

To remove a bookmark, do the following:

1. In the Bookmarks panel, click on the bookmark that you want to remove. It will be highlighted in blue color.

2. Click with the right mouse button on the bookmark to open a pull-down menu.

3. Select Delete Bookmark from the pull-down menu. The bookmark is deleted without further warning

> **!** Note: You can only delete bookmarks created by yourself (i.e., those marked personal in the Bookmarks panel), but not the predefined bookmarks (i.e., those marked global).

6.3 The Metadata Descriptions Tree panel

The Metadata Descriptions Tree panel allows you to navigate through the corpus hierarchy. It serves the following three purposes:

1. The hierarchy of nodes allows you to easily navigate through the corpus to the session data (see sections 6.3.1 and 6.3.2).

2. Pull-down menus and the five buttons at the bottom of the panel allow you to select parts of the corpus for purposes of conducting searches (see section 6.3.3).

3. The linking of metadata, media and annotation files at the session level allows you to immediately access all relevant data (see section 6.3.4).

The following screenshot illustrates the **Metadata Descriptions Tree** panel.

corpus directory

corpus information

corpus nodes

session nodes

URL of the corpus directory file

session data

buttons for selecting parts of the corpus

IMDI-BCBrowser for MPI on M03402985

File Options Search Help

Metadata Descriptions Tree

IMDI.imdi
IMDI-corpora
index.html
MPI corpora
index.html
Acquisition
Language & Cognition
africa
Goemai
Hellwig
Project.pdf
language.pdf
Publications.pdf
Natural conversation
Aram
Jos
Dialect1
Dialect2
Youth
Quest
descriptive
TreesSM
Speaking
Metadata
Speaking.wav
Speaking.txt
Annon. Info
Source: LBHGAC02Oct00

Root URL   http://www.mpi.nl/world/corpora/IMDI/metadata/IMDI.imdi

Basket functions

List   Save   Add   Remove   Clear

6.3.1 Navigating through the corpus (tree structure)

Whenever the **IMDI Browser** window opens up, the node **IMDI Corpora** is displayed in the **Metadata Descriptions Tree** panel.

If you want to access any other predefined part of the corpus, do the following:

1. Go to the **Bookmarks** panel (see section 6.2).

2. Double-click on one of the bookmarks. The corresponding node is displayed in the **Metadata Descriptions Tree** panel.

In the **Metadata Descriptions Tree** panel, make use of the following options to navigate through the corpus:

- Double-click on any node to open it and display the next level in the hierarchy.

  ! Note: The **Metadata Descriptions Tree** panel distinguishes between open and closed corpus and session nodes. These are represented through the following icons:

  C S   ➡

  icon of an open   icon of a closed
  corpus or session   corpus/session

  Some program commands do not work when a node is closed. If any of the commands do not seem to work, make sure that the node is open.

! Note: Because of the large amount of data that is loaded, it may take some time until the IMDI Browser responds to your command and opens a node.

- Click on any closed node, and then point with the mouse to it to activate the tool-tip that displays brief information about the node.

- Click on any open node to select it. It will be highlighted in blue color. The information relevant to the selected node is displayed in the Info/Content (see section 6.4) and Description panels (see section 6.5).

- Right-click on any selected (i.e., highlighted) node to open a pull-down menu that displays the available options. Click on any item in the pull-down menu to select the corresponding option.

Depending on the kind of node, the following options are available:

Add to basket	Adds the node to the list of nodes to be searched (see section 6.3.3).
Add to Bookmarks	Adds the node to the Bookmarks panel (see section 6.2).
Bookmark Info	Displays information about the bookmark in the Info/Content panel (see section 6.4).
Clone Node	Opens a second IMDI Browser window that displays only the corresponding node.
Create Service Files	Creates a file that is needed for searching through the metadata (see section 7.1.2).
Delete Bookmark	Deletes the node from the Bookmarks panel (see section 6.2).
Delete Local Service Files	Deletes the locally-stored file that is needed for searching through the metadata (see section 7.1.2).
ELAN	Opens the media file or the session annotation in ELAN (see section 6.3.4.2).
IMDI-BCEditor	Opens the metadata file in the IMDI Editor (see section 6.3.4.1).
List Sessions	Lists all sessions contained under this node in the Info/Content panel (see section 6.4).
Media Synchronized Viewers	Opens the session annotation in a separate Viewer, synchronized with the media file (see section 6.3.4.2).
MetaData Search	Opens the IMDI Metadata Search window (see section 7).
Open with Acrobat Reader 5.0	Opens the file with Acrobat Reader 5.0 (see section 6.3.4.3).

Open with Winamp	Opens the file with Winamp (see section 6.3.4.2).
Remove	Removes the node from the Metadata Descriptions Tree panel.
Remove from basket	Removes the node from the list of nodes to be searched (see section 6.3.3).
Save File Content	Downloads the file in a non-compressed format.
Session Count	Displays the number of sessions contained under this node in the Info/Content panel (see section 6.4).
Show	Displays the file content in the Info/Content panel (see sections 6.3.4.2 and 6.4).
Show As HTML	Displays the file in an HTML format in the Info/Content panel (see sections 6.3.4.1 and 6.4).
Show Content-Type	Displays the file format in the Description panel (see section 6.5).
Show Description	Displays the file description in the Description panel (see section 6.5).
Show File Content	Displays the file in an XML format in the Info/Content panel (see section 6.4).
Show format	Displays the file format in the Info/Content panel (see section 6.4).
Show Info	Displays the file content in the Info/Content panel (see section 6.4).
Show LR's	Displays the directory information for all files contained under this node in the Info/Content panel (see section 6.4).
Show Services	Displays the name and directory information of the file needed for searching through the metadata in the Info/Content panel (see section 7.1).
Show URL	Displays the directory information for the metadata file in the Description panel (see section 6.5).
Windows Media Player	Opens the media file in Windows Media Player (see section 6.3.4.2).

! Note: Depending on the configuration of your IMDI Browser, other tools may be available. Furthermore, under Windows, the default option for starting media files of specific types is also supported (e.g., if Windows is set to open

*.wav files in Windows Media Player, the option Windows Media Player will be available).

! Note: For some nodes, one of the options is marked with the symbol #. In this case, the corresponding option will start automatically whenever you double-click on the open node.

! Note: Not all options are available for all nodes.

6.3.2 Navigating through the corpus (world map)

The IMDI Browser displays a world map with the locations of all documented languages. To access the world map, do the following:

1. In the Bookmarks panel double-click on the bookmark Worldmap.

    The info file **world.html** is displayed in the Metadata Descriptions Tree panel.

2. In the Metadata Descriptions Tree panel click on the info file **world.html**.

    The world map is displayed Info/Content panel.

3. To navigate, you have the following options:

    - Click on any part of the world map to enlarge the corresponding region.

    - Make use of the Backward and Forward buttons to move to previous and subsequent views.

    - Click on any language to open the corresponding node in the Metadata Descriptions Tree panel.

For example:
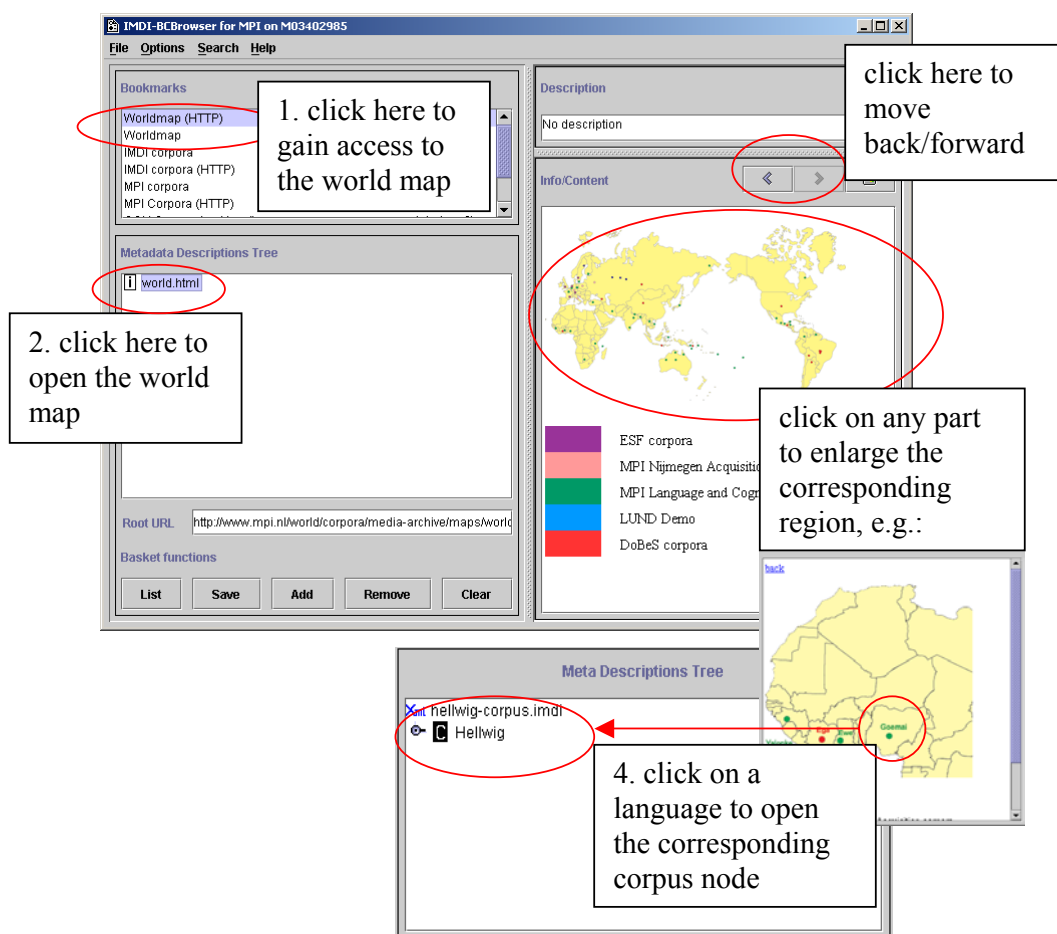
6.3.3 Selecting parts of the corpus for purposes of searching (basket function)

The IMDI Browser allows you to search through the metadata files. By default, this search is done throughout an entire corpus (see section 7), but it is possible to limit it to one (or several) selected corpus and/or session nodes, i.e., to limit it to nodes that have been put in the "basket".

To put a corpus or session node into the basket, do the following:

1. In the Metadata Descriptions Tree panel, double-click on the node to open it.

2. Click on the open node to select it. It will be highlighted in blue color.

3. Do one of the following:

   (a) Click the Add button at the bottom of the Metadata Descriptions Tree panel.

   (b) Or click with the right mouse button on the highlighted node; then select Add to basket from the pull-down menu.

145

The icon of any selected node will change its color to gray, e.g.:
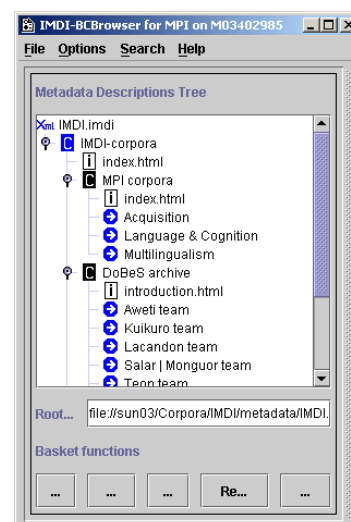

non-selected node   selected node

Once an item is selected, the **List** button at the bottom of the panel will be highlighted in red color.

> **!** Note: The red color was chosen to alert you to the fact that a list has been compiled. The existence of such a list has consequences for the search process: metadata search will only search the listed nodes, i.e., the nodes that were put into the basket (see section 7.1). If the search process does not yield the expected results, please make sure that the list contains all relevant nodes. And if you want to search through an entire corpus (not just through the listed nodes), you have to delete the list again (see below).

4. Repeat this process to add other nodes to your selection.

> **!** Note: Please select only nodes that are listed below a corpus node for which there is a search file available. Such corpus nodes are visualized in the IMDI Browser by means of a white "C" on a black background (see section 7.1). As indicated in the illustration to the right, you can select and search nodes contained under **MPI corpora** *or* under **DoBeS archive**. But you cannot select and search one node from **MPI corpora** *and* one from **DoBeS archive** – since they do not have a common search file.

To view a list of the selected nodes, click the **List** button at the bottom of the **Metadata Descriptions Tree** panel. It will display all selected nodes, e.g.:
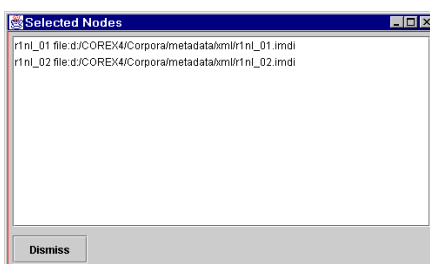
To remove a node from the list, do the following:

1. In the Metadata Descriptions Tree panel, double-click on the node to open it.

2. Click on the open node to select it. It will be highlighted in blue color.

3. Do one of the following:

(a) Click the Remove button at the bottom of the Metadata Descriptions Tree panel.

(b) Or click with the right mouse button on the highlighted node; then select Remove from basket from the pull-down menu.

4. Repeat this process to remove other nodes from your selection.

You can remove all selected nodes from this list by clicking the Clear button at the bottom of the Metadata Descriptions Tree panel.

You can save the selected list for future uses. Click the Save button at the bottom of the Metadata Descriptions Tree panel. The following message informs you that your list has been saved:
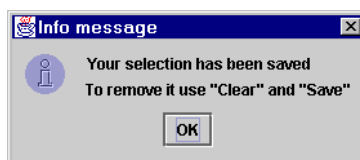
! Note: Once you have saved a selected list, you can only remove it by first clicking the Clear button (to remove all selected nodes) and then the Save button (to save the empty list).

When you are satisfied with your selection, you can initiate the search (see section 7).

6.3.4 Accessing data

The IMDI Browser contains links to files or to information from files. These links are symbolized through the following icons:

◆   link to information from metadata file(s) (i.e., information from IMDI files).

m   link to digitized media file(s), containing audio/video data.

T   link to transcription and annotation file(s).

i   link to info file(s) providing general background information.

⬤   link not available (either because the file does not yet exist, or because access is denied).

Click on any of these icons, and the IMDI Browser will display the relevant information.

6.3.4.1 Metadata information

Metadata information is displayed under each session node. It allows you to (a) read about the circumstances under which the session data was collected and (b) search for relevant data (see section 7).

It contains the following kind of information:

- Information about the date and location.

- Information about the project within which the data was collected.

- Information about the person who collected the data.

- Information about the content.

- Information about the participant(s).

- Information about the source (i.e., the audio/video tape), the digitized media file(s) and the annotation file(s).

- Cross-references to other relevant sessions and publications.

To view the metadata information of a session, click on the corresponding metadata icon. The information is displayed in the Info/Content panel (see section 6.4) or the Description panel (see section 6.5).
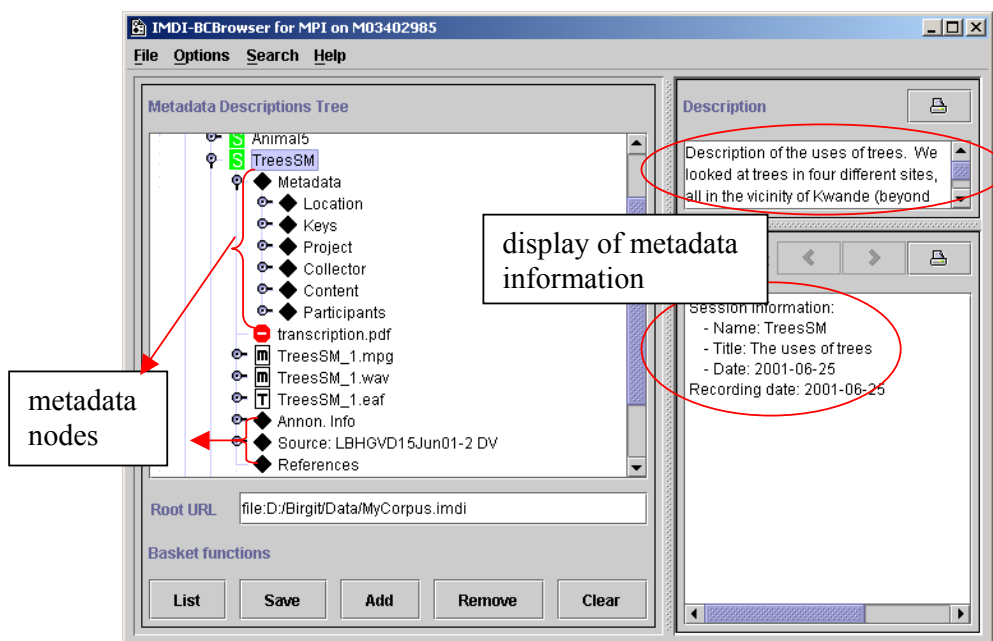
Furthermore, you can access all metadata information through the session node. Double-click on a session node to open it, then click on it to select it, and right-click to open a pull-down menu. The pull-down menu offers you the following options:

- Choose Show As HTML from the menu. The metadata information is displayed in an HTML format in the Info/Content panel, e.g.:
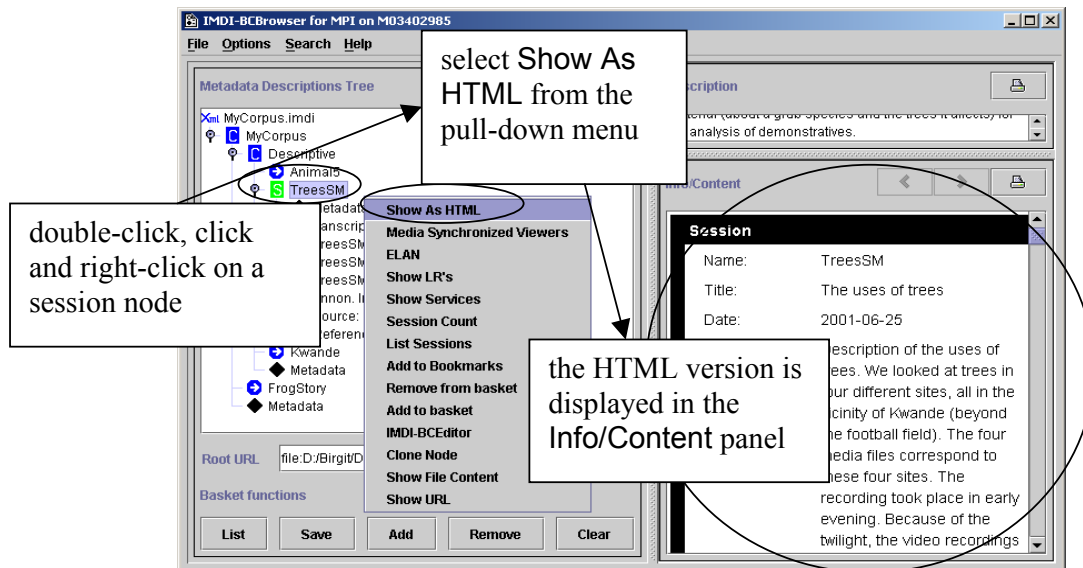
- Choose IMDI-BCEditor from the menu. The IMDI Editor opens with the corresponding metadata file, e.g.:
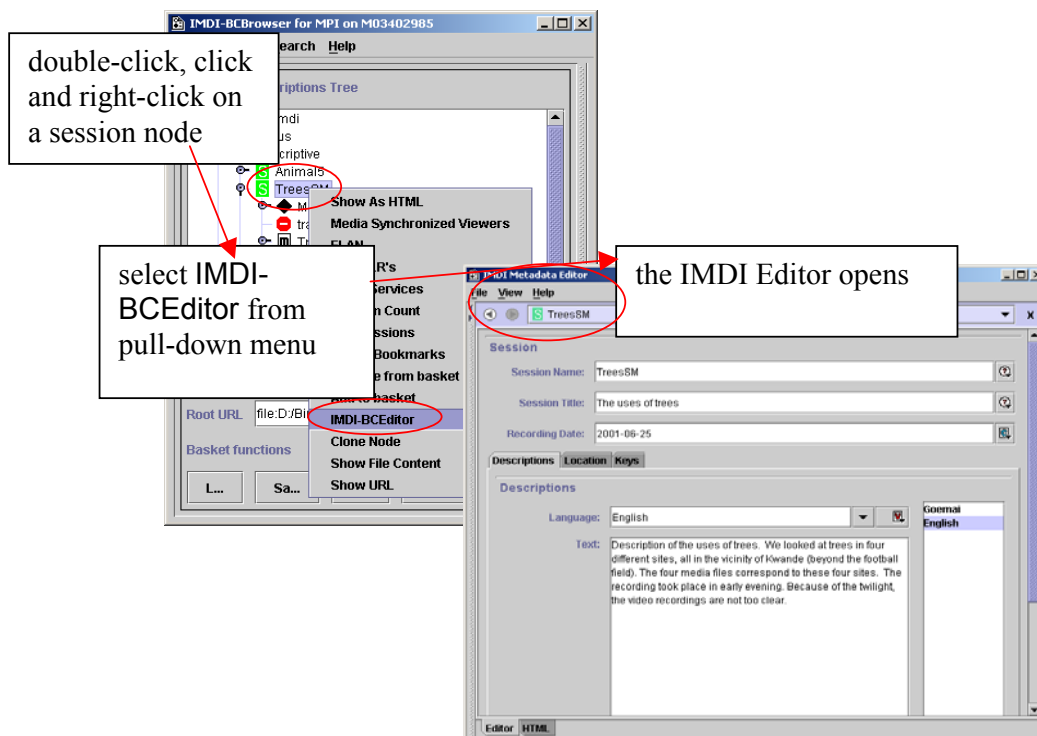
! Note: You can update the metadata information in the IMDI Editor. However, in most archives, it is not possible to save the updated file directly to the server. Please save it to the local disk, and contact your corpus manager. In any case, please contact your corpus manager before you make any substantial changes to make sure that your changes do not interfere with the overall corpus management.

6.3.4.2 Digitized media files and transcription/annotation files

A session node also contains links to digitized media files (i.e., to audio and/or video files), and to transcription/annotation files. These files can be accessed either through the session node, or the media file node or the transcription/annotation file node.

The following options are available:

(a) Right-click on a selected media file node to open the pull-down menu, then choose ELAN, Windows Media Player or Open with Winamp from the menu. The media file will be opened in ELAN,[10] Windows Media Player or Winamp, e.g.:

---

[10] ELAN is an annotation tool developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. Tool and manual can be downloaded from http://www.mpi.nl/tools/.

In order to view digital images (*.gif, *.jpg) right-click on the corresponding node and select # Show #. The image will be displayed in the right window, titled *Info/Content*:
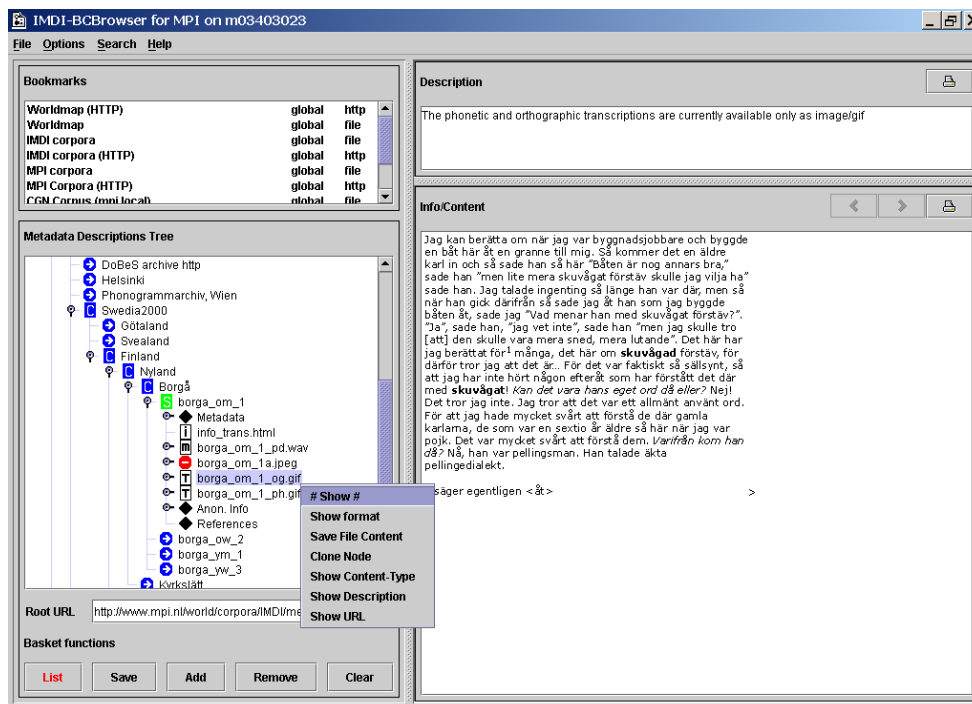
(b) Right-click on a selected session node to open the pull-down menu. If the annotation of that session was done in ELAN (i.e., if there is an *.eaf file), the pull-down menu will display the options ELAN and Media Synchronized Viewers. Select one of these options to view the media file together with its annotations.

ELAN

media synchronized viewers
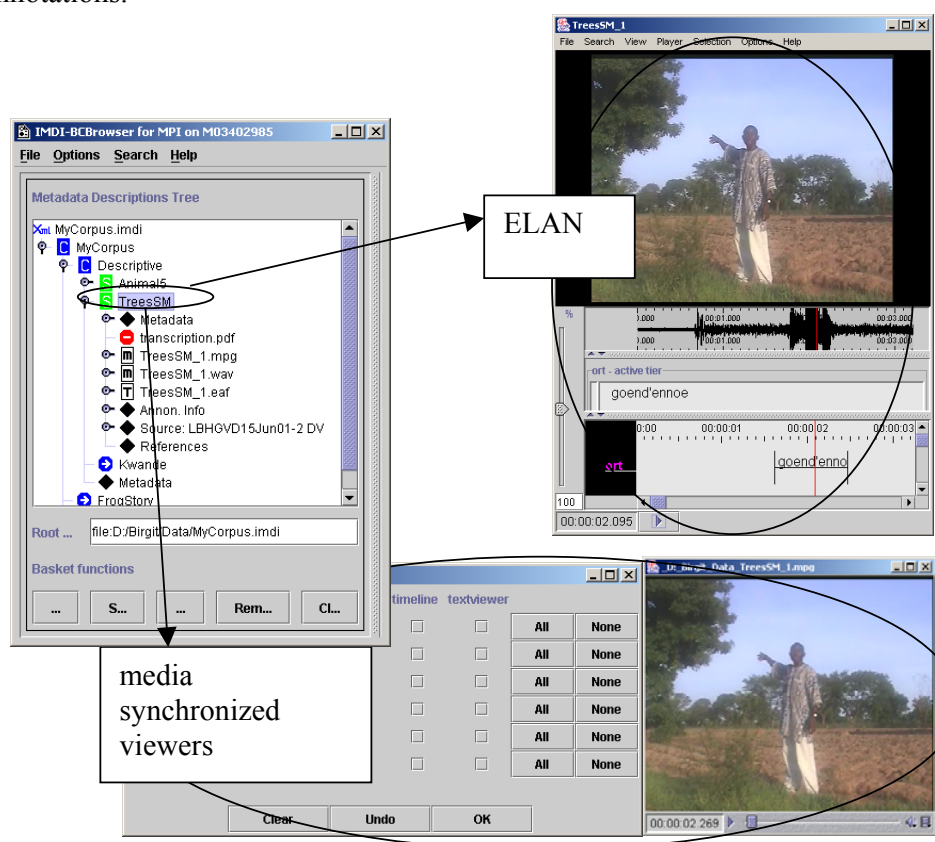
By default, the Media Synchronized Viewer option displays the video image together with a dialog window. This dialog window allows you to select the annotation tier that you want to view, and the Viewer in which you want to view it.

- To select an option, click in the corresponding box (to add a checkmark), or click on the button All (to select all Viewers for one tier). Then click on OK to activate the corresponding viewer.

- To deselect an option, click in the corresponding box (to delete a checkmark), or use the buttons None (to deselect all Viewers for a tier) or Clear (to deselect all Viewers). Then click on OK to deactivate the corresponding viewer.

- Click on Undo to undo the last selection.

All Viewers are synchronized in time, i.e., when you play the video, the other Viewers will display the corresponding annotation, highlighted in red color.

For example:

Subtitle Viewer

Timeline Viewer

available Viewers

available tiers

Grid Viewer

Taglist Viewer

Text Viewer
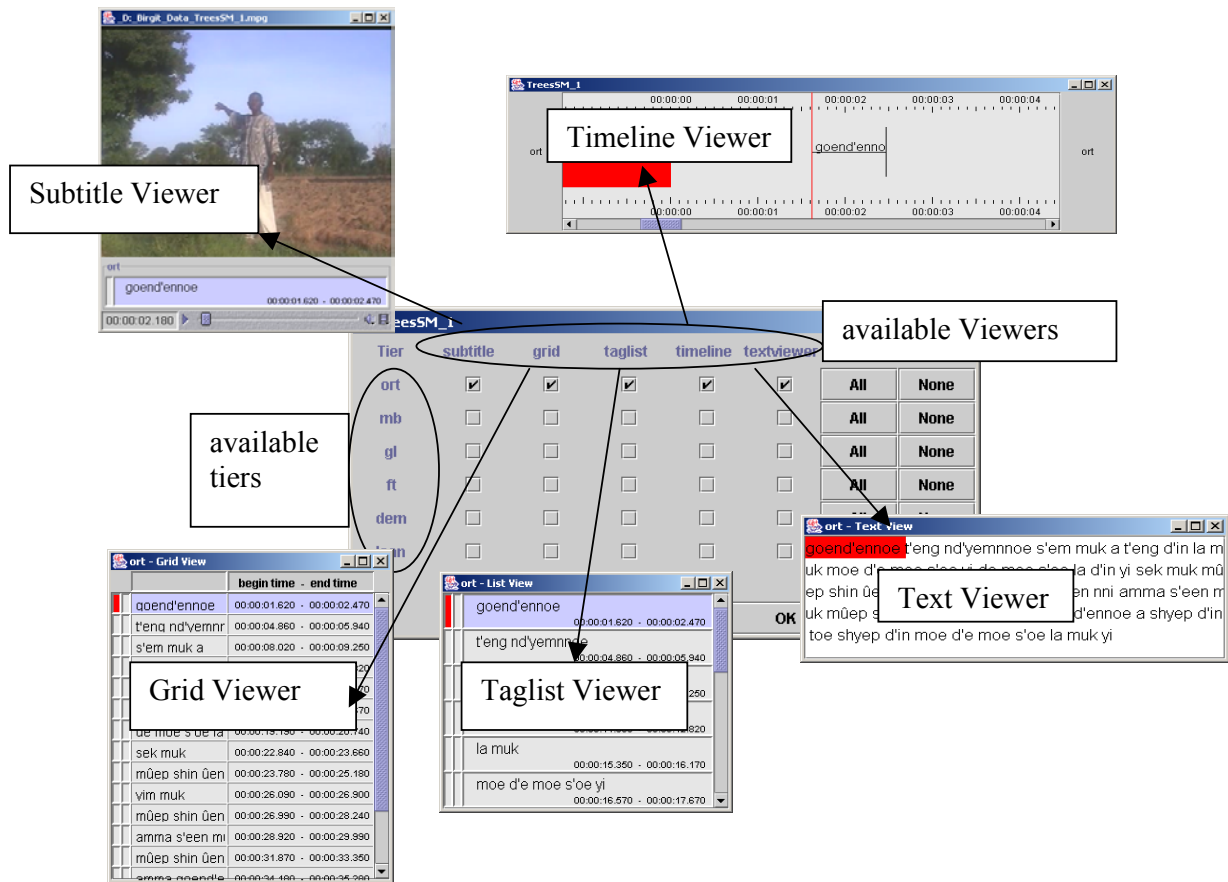
(c) Right-click on a selected annotation file node to open the pull-down menu, then choose Show from the menu. The file content is displayed in the Info/Content panel, e.g.:
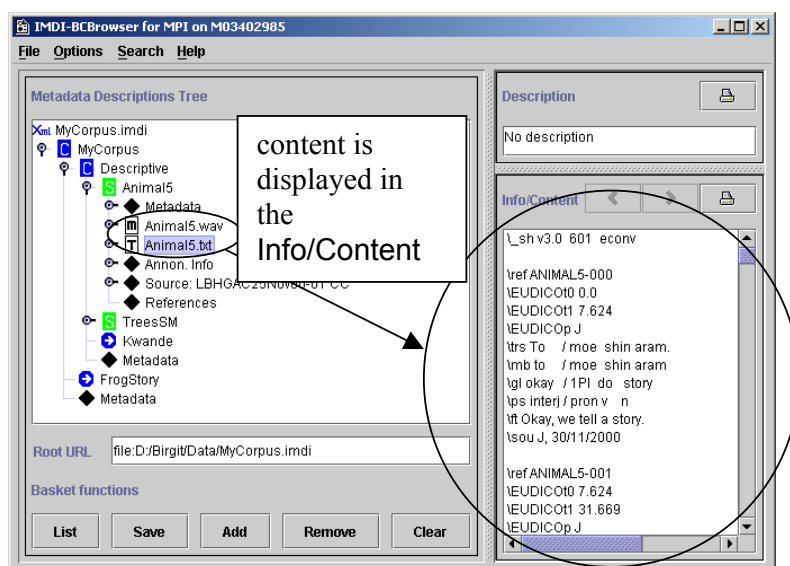
! Note: There are no restrictions on the format of your annotation files (ELAN, Media Tagger, Shoebox, CHAT, etc.).

! Note: You can constantly update your annotation files (provided you do not change their file names), which allows you to reflect changes in your analysis. But note that, for the moment, only ELAN can be directly accessed via the IMDI Browser. If you want to edit annotation files of a different format, please navigate to the physical location of that file (by using, e.g., Windows Explorer).

154

6.3.4.3   Info files

Both corpus and session nodes can contain links to info files. These files provide general background information about the corpus or the individual session.

If such a file is an HTML file, click on it: its content will automatically be displayed in the Info/Content panel. If the file is a PDF file, click on it and then right-click on it to open the pull-down menu. Choose Open with Acrobat Reader 5.0 from the pull-down menu. Acrobat Reader will open up to display the file.

For example:
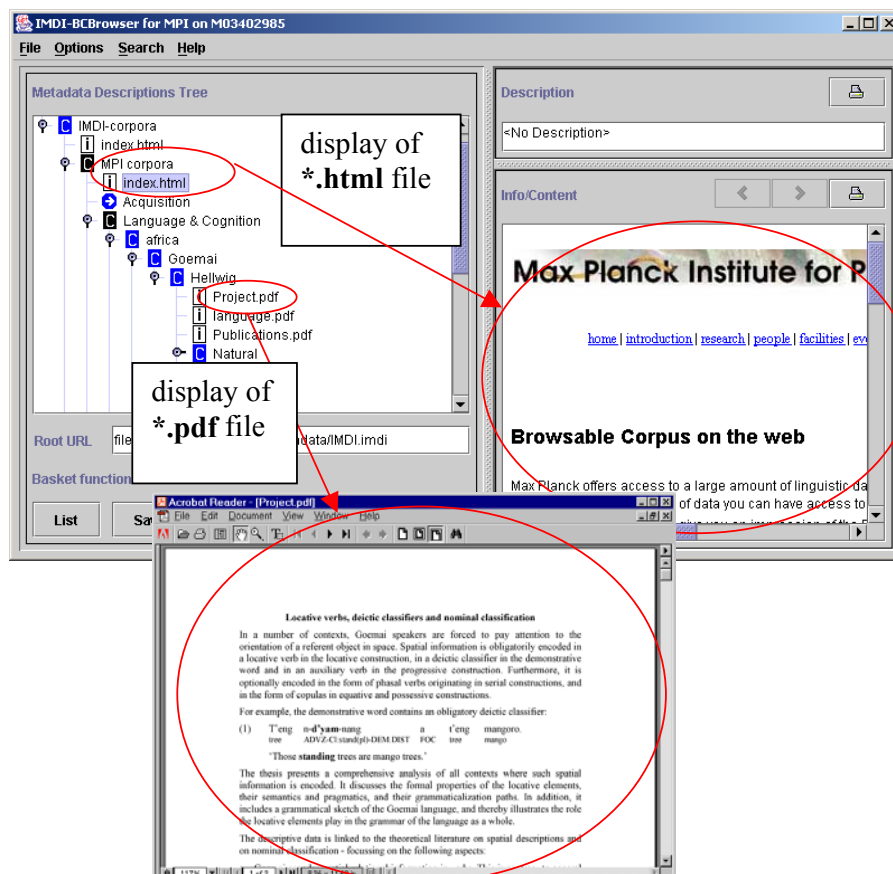
6.4 Info/Content panel

The Info/Content panel displays information about corpus, session and file nodes. To read the information, click on the node in the Metadata Descriptions Tree panel, e.g.:

> ! Note: The information is displayed in form of searchable keywords. For prose information, look at the Description panel (see section 6.5).
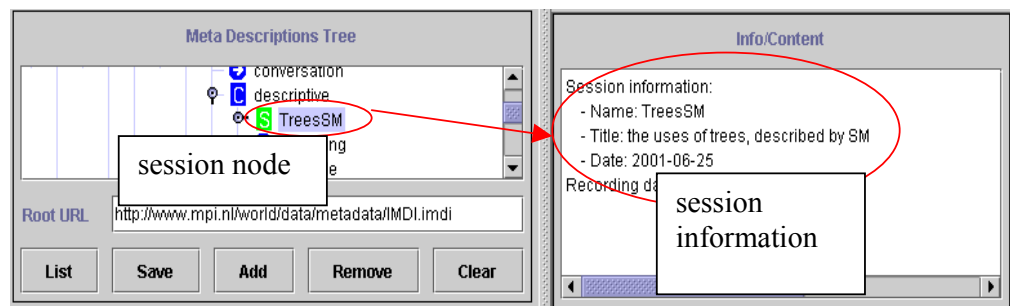
The Info/Content panel can display additional information. To view this additional information, do the following:

1. In the Metadata Descriptions Tree panel, click on the node whose information you want to view. It will be highlighted in blue color.

2. Right-click on the highlighted node.

3. Select one of the following options from the pull-down menu (see section 6.3.1):

- Bookmark Info: Displays information about the bookmark.

- List Sessions: Lists all sessions contained under this node.

- Session Count: Displays the number of sessions contained under this node.

- Show: Displays the file content.

- Show As HTML: Displays the file in an HTML format.

- Show File Content: Displays the file in an XML format.

- Show format: Displays the file format.

- Show Info: Displays the file content.

- Show LR's: Displays the directory information for all files contained under this node.

- Show Services: Displays the name and directory information of the file needed for searching through the metadata.

The corresponding information is displayed in the Info/Content panel.

The Info/Content panel also displays the world map (see section 6.3.2). And it displays Backward and Forward buttons (that allow you to move to the previous or subsequent view) and a Print button (that allows you to print the information displayed in the panel).

6.5 Description panel

The Description panel displays a description of corpus, session and file nodes. To read the description, click on the node in the Metadata Descriptions Tree panel, e.g.:

> ! Note: The Description panel displays prose information. For searchable keywords, look at the Info/Content panel (see section 6.4).

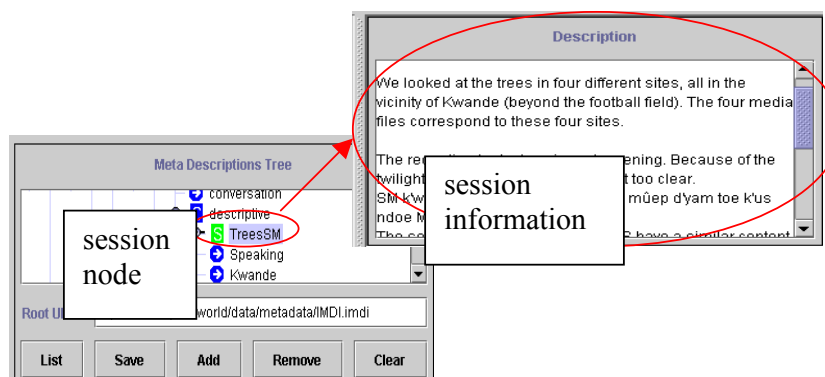The Description panel can display additional information. To view this information, do the following:

1. In the Metadata Descriptions Tree panel, click on the node whose information you want to view. It will be highlighted in blue color.

2. Right-click on the highlighted node.

3. Select one of the following options from the pull-down menu (see section 6.3.1).

   - Show Content-Type: Displays the file format.

   - Show Description: Displays the file description.

   - Show URL: Displays the directory information for the metadata file.

   The information is displayed in the Description panel.

The Description panel also displays a Print button (that allows you to print the information displayed in the panel).

7 Searching data (IMDI Metadata Search Tool)

To start the IMDI Metadata Search Tool, do one of the following:

(a) Double-click on the IMDI Metadata Search Tool icon.

(b) Or use the IMDI Browser interface:

    (1) Use the menu items (see section 6.1.3).

        1. Click on Search.

        2. Click on Metadata Search.

    (2) Use the Metadata Descriptions Tree panel (see section 6.3).

        1. Go to the Metadata Descriptions Tree panel.

        2. Double-click on a node to open it.

        3. Click on it to select it. It will be highlighted in blue color.

        4. Right-click on it to open a pull-down menu.

        5. Select MetaData Search from the pull-down menu.

The following screenshot illustrates the IMDI Metadata Search window:
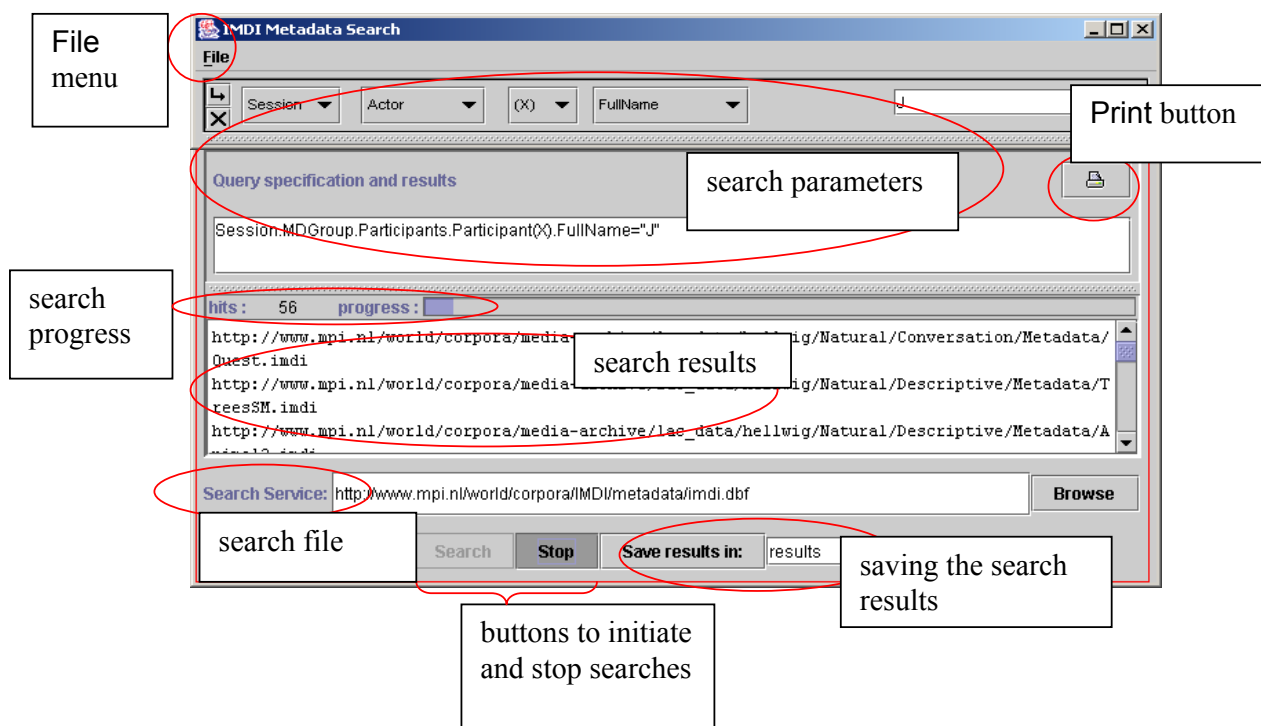
This section introduces the search options. The following options are available:

1. Specifying the corpus to be searched (section 7.1).

2. Specifying the search parameters (section 0).

3. Changing the profile (section 7.2).

4. Initiating and stopping the search (section 7.3).

5. Displaying the search results (section 7.4).

6. Saving the search results (section 7.5).

7. Printing (section 7.6).

8. Closing the IMDI Metadata Search window (section 7.7).

7.1 Specifying the corpus to be searched

The corpus to be searched is displayed in the box to the right of Search Service:. For metadata search to work, you need to enter the name and location of the ***.data** file that contains the metadata information of the corresponding corpus (see section 7.1.1). Many corpora provide you with predefined ***.data** files, but you also have the possibility to create your own files (see section 7.1.2).

7.1.1 Selecting a ***.data** file

To enter a ***.data** file in the box to the right of Search Service:, do the following:

(a) Either: Type in its name and location.

(b) Or: Make use of the Browse button to the right of this box.

(c) Or: Make use of the IMDI Browser interface. In this case, do the following:

1. In the Metadata Descriptions Tree panel (see section 6.3), double-click on the corpus node to open it. Then click on the node to select it. It will be highlighted in blue color.

2. Do one of the following:

    (a) In the Search menu, click on Metadata Search (see section 6.1.3).

    (b) Or in the Metadata Descriptions Tree panel (see section 6.3), right-click on the selected node; then select MetaData Search from the pull-down menu.

    In the IMDI Metadata Search window, the ***.data** file associated with the selected corpus node is automatically entered into the box, e.g.:
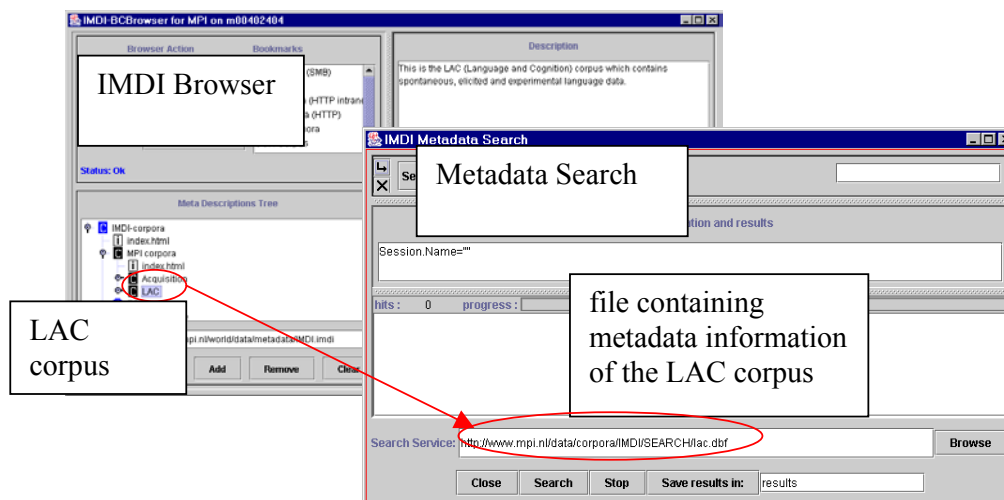
! Note: The name and location of a ***.data** file can be seen in the IMDI Browser. In the Metadata Descriptions Tree panel, double-click on a corpus node (to open it), click on it (to select it), and then click with the right-mouse button to open a pull-down menu. From the pull-down menu, select Show Services. The name and location of the corresponding ***.data** file is displayed in the Info/Content panel (see section 6.3.1).

! Note: Corpora can only be searched if there is a ***.data** file available. Such corpora are visualized in the Metadata Descriptions Tree panel by means of a white "C" on a black background. Corpora that have no corresponding ***.data** file are visualized by means of a white "C" on a blue background. If you click on a 'non-searchable' subcorpus that is part of a larger 'searchable' corpus, the whole corpus will be searched – not just the subcorpus. E.g.:

By default, [...] corpus. However, it is possible [...] this case, you have to manually [...] wser. To select nodes, make us[...] ttom of the Metadata Descriptions Tree panel (see section 6.3.3). Once you have selected node(s), metadata search is automatically restricted to the selected node(s). Again, you can only search parts of corpora that have a corresponding ***.data** file.
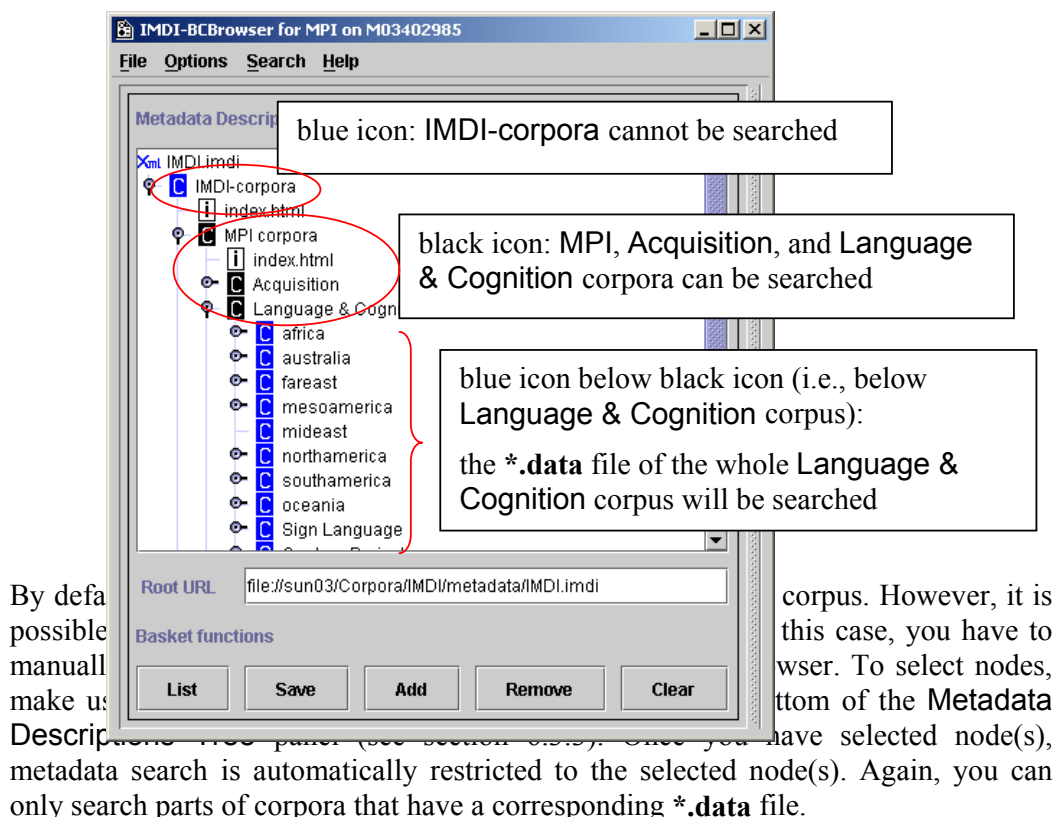
! Note: If the search process does not yield the expected results, please check the following possibilities:

(1) Make sure that the ***.data** file displayed next to Search Service: is the correct file.

(2) For searches through an entire corpus: make sure that you have not accidentally selected a part of the corpus. If you have selected a part, the List button at the bottom of the Metadata Descriptions Tree panel is highlighted in red color.

(3) For searches through selected parts of a corpus: make sure that the list (displayed under the List button at the bottom of the Metadata Descriptions Tree panel) contains all relevant nodes.

7.1.2 Creating a new *.data file

Many corpora provide you with predefined ***.data** files. Such corpora are visualized in the Metadata Descriptions Tree panel by means of a white "C" on a black background (see section 7.1.1 above). However, it is also possible to create your own ***.data** file. To do so, do the following:
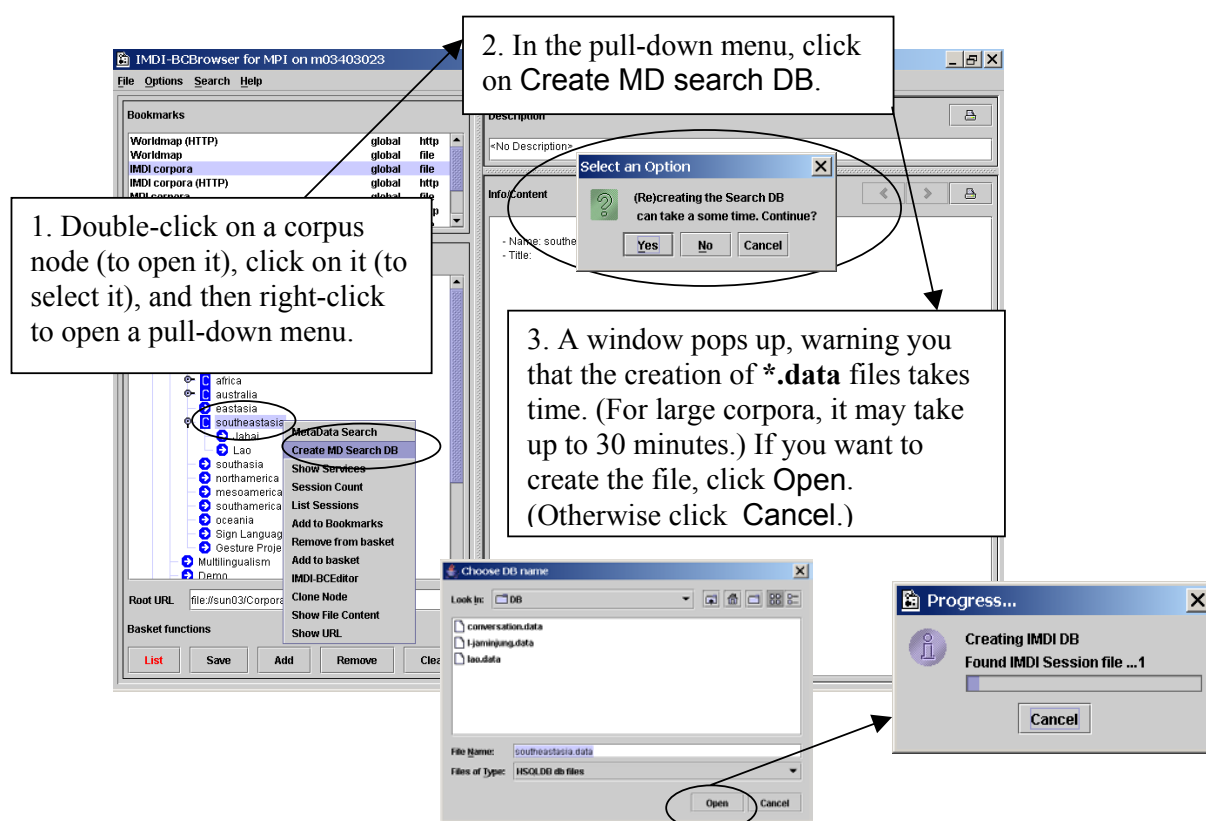
5. After the ***.data** file has been created, the corpus node changes its color from blue to black, and you can now search this corpus (see section 7.1.1).

! Note: These ***.data** files are stored permanently in your local directory "user-directory \ IMDI-TOOLS \ DB". Specifying the search parameters

The search parameters are specified in the topmost part of the IMDI Metadata Search window, e.g.:
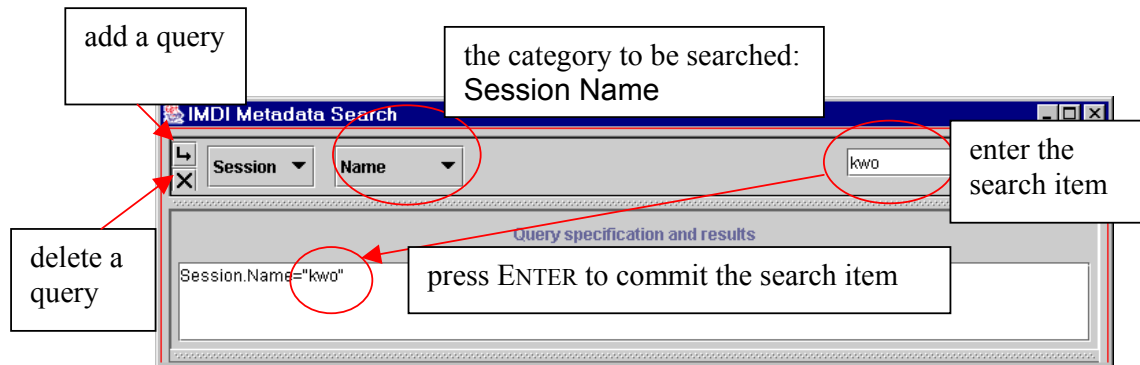
The following search options are available:

1. Selecting the category to be searched (section 7.1.3).

2. Entering the search item (section 7.1.4).

3. Adding or deleting a search query (section 7.1.5).

7.1.3 Selecting the category to be searched

The categories that can be searched are displayed in form of pull-down menus. By default, Session and Name are displayed. These two options allow you to search for all sessions with a particular name (as in the illustration above).

Click on the pull-down menus to select other categories:

- The first category is always Session (i.e., you will always search for a session)

- The second category is chosen from the following list: Name, Date, Location, Key, Participant, or Content (i.e., you will search for a session that has a certain name, or a certain recording date, etc.).

- Depending on your choice for the second category, further pull-down menus will appear to narrow down the category.

  ! Note: If too many categories and pull-down menus are added, the IMDI Metadata Search window cannot display them all. To increase its size, click on the full screen icon (in the top right corner of the window):

Some examples of metadata searches are given below:

(1) The screenshot below illustrates a search for all sessions that were recorded in the Netherlands.

In this case, Session and Location were chosen as the first two categories. A third pull-down menu appeared subdividing the category Location into Continent, Country, and Region. When Country was chosen, a fourth pull-down menu appeared displaying the names of all countries.

The selected categories and subcategories were then automatically added to the box Query specification and results.
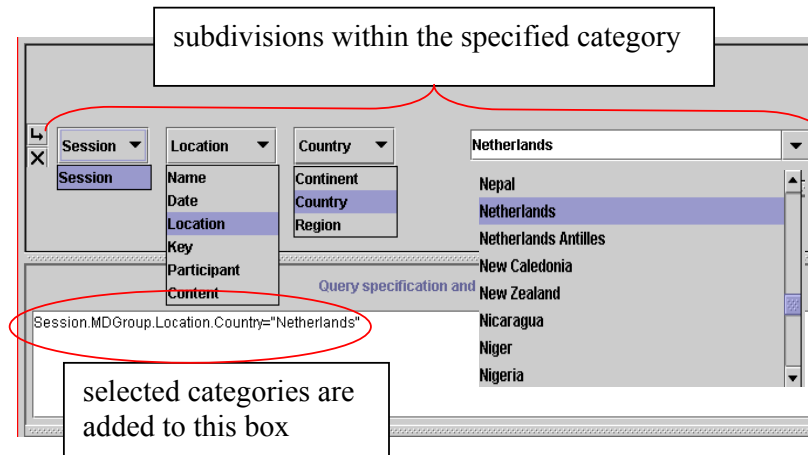
(2) The screenshot below illustrates a search for all sessions that were recorded after 1.1.2000 and before 1.1.2001.
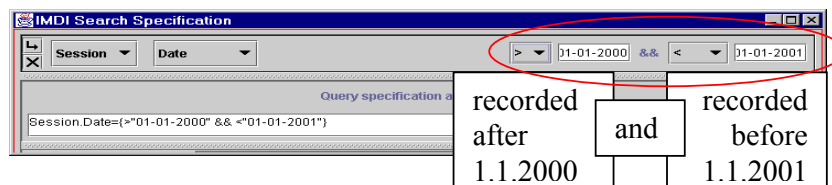
In this case, Date was chosen as the second category (i.e., the date of the recording). Further pull-down menus appeared that allowed to specify the exact time period, e.g.:
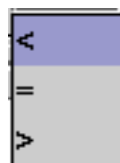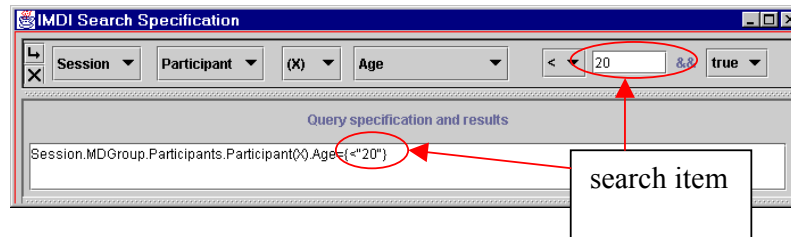

search for a date before the specified date

search for the exact date

search for a date after the specified date

7.1.4 Entering the search item

After you have selected the search categories, enter the search item into the box in the top right corner, and then press the key ENTER, e.g.:

! Note: Some search items have to be entered manually, others can be chosen from a pull-down menu (see section 7.1.3). If you enter the search item manually, you always need to press the key ENTER to commit the item.

! Note: Alternatively, you can type the search item directly into the box **Query specification and results**. Type it between the two quotation marks following the category specification, e.g. *"20"* in the illustration above.

7.1.5 Adding or deleting a search query

You can add and delete queries as follows:

       use this button to add a query

       use this button to delete the query

When two (or more) queries are specified, the search process will find all sessions that contain query 1 *and* query 2 (*and* all subsequent queries). For example, in the following illustration, all sessions will be found that contain utterances from (1) a male participant X *and* (2) a female participant Y.
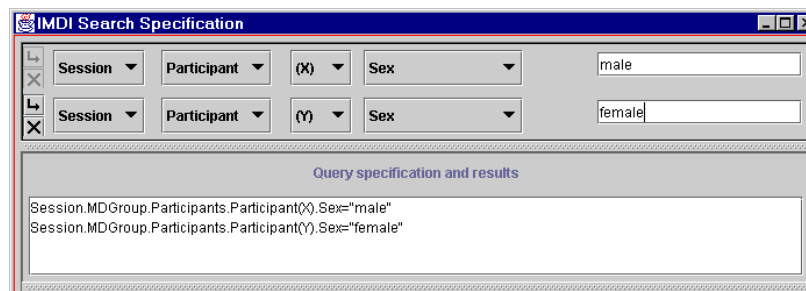
! Note: Variables such as X and Y are only relevant if you specify more than one query. In this case, they allow you to specify that the participant of query 1 (e.g., X) should either be different from the participant of query 2 (e.g., Y)

(as in the illustration above), or that (s)he should be identical to the participant of query 2 (e.g., X). Please ignore these variables if you specify one query only.

7.2 Changing the profile

Section 0 above has shown that the categories to be searched are displayed in form of pull-down menus. These categories are specified in a so-called "profile" (stored in the folder "user-directory \ IMDI-TOOLS \ Profiles \ local). To change the profile, do the following:

1. Click on File menu.

2. Click on Change Profile. The Open dialog window appears and displays all available profiles, e.g.:

3. Double-click on the profile you want to use. The IMDI Metadata Search window will display the chosen profile. E.g., the following illustration displays some categories that are specific to sign language research and that are defined in the profile "SignLanguage.Profile.xml":
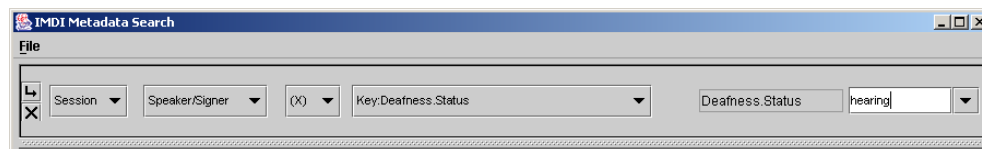
! Note: For the moment, profiles have to be programmed by hand. In future versions of the IMDI tools, it will be possible to create profiles with the help of the IMDI (Metadata) Editor.

7.3 Initiating and stopping the search

Click the Search button at the bottom of the IMDI Metadata Search window to start the search. During the search the number of "hits" (matches) to your search criterion is shown, as well as the progress towards completion of the search. Once the search process has started, you can use the Stop button to stop the search.

! Note: Because of the large amount of data that is searched, it may take some time before the search is completed.

7.4 Displaying the search results

The box Query specification and results displays the results, i.e., the name and location of each session. You can access any search result by double-clicking on it. This will open up an IMDI Browser window containing the selected session, its metadata descriptions and its media and annotation files.

7.5 Saving the search results

You can save the search results in form of a corpus node. This corpus node is displayed in the Metadata Descriptions Tree panel of the IMDI Browser window. Do the following:

1. In the IMDI Metadata Search window, enter a name into the text box following the button Save results in:.

2. Click the button Save results in:. The results are saved under the specified name.

You can access the results through the IMDI Browser interface. Do the following:

1. In the IMDI Browser window, go to the Bookmarks panel (6.2).

2. Double-click on Search Results.

The Metadata Descriptions Tree panel (see section 6.3) displays the new corpus node below the node labeled Results.
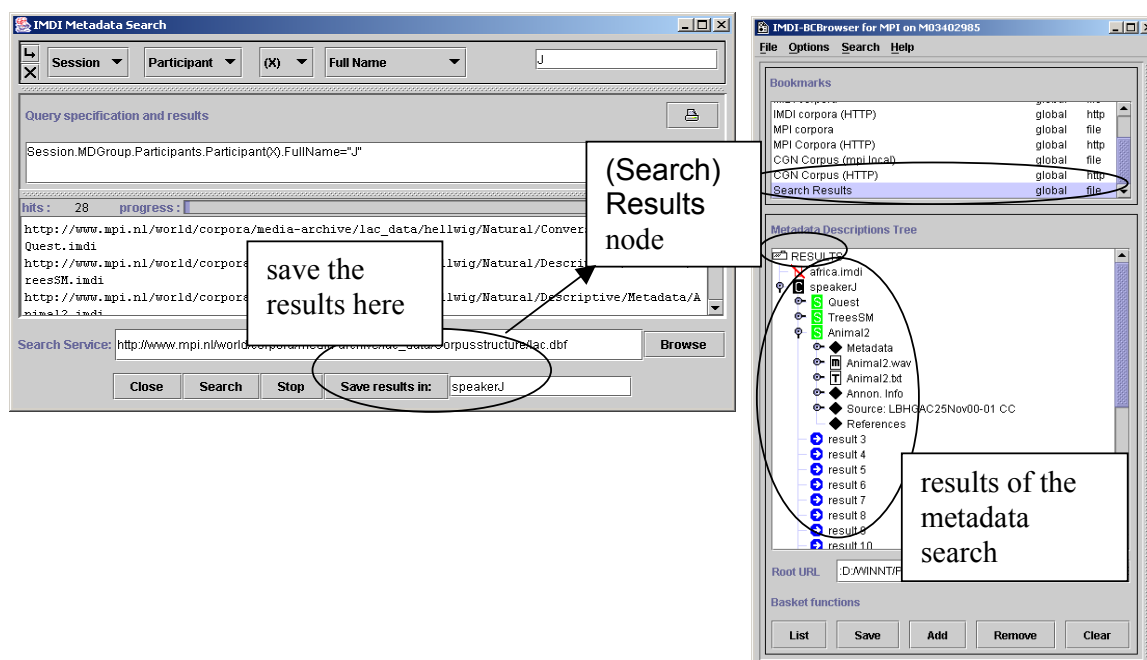
The new corpus node is treated like any other corpus node, i.e., it contains sessions, and the sessions contain metadata, media and annotation files.

You can remove the saved corpus node again. Do the following:

1. In the Metadata Descriptions Tree panel (see section 6.3), click on the node to select it. It will be highlighted in blue color.

2. Right-click on the selected node to open a pull-down menu.

3. Select Remove from the pull-down menu.

   The node will be removed.

   ! Note: After selecting Remove, the icon of the removed node remains visible although it cannot be accessed anymore. The icon will be removed after exiting the node Results.

7.6 Printing

You can print the search parameters (see section 0) together with the search results (see section 7.4). To do so, click on the Print icon, e.g.:
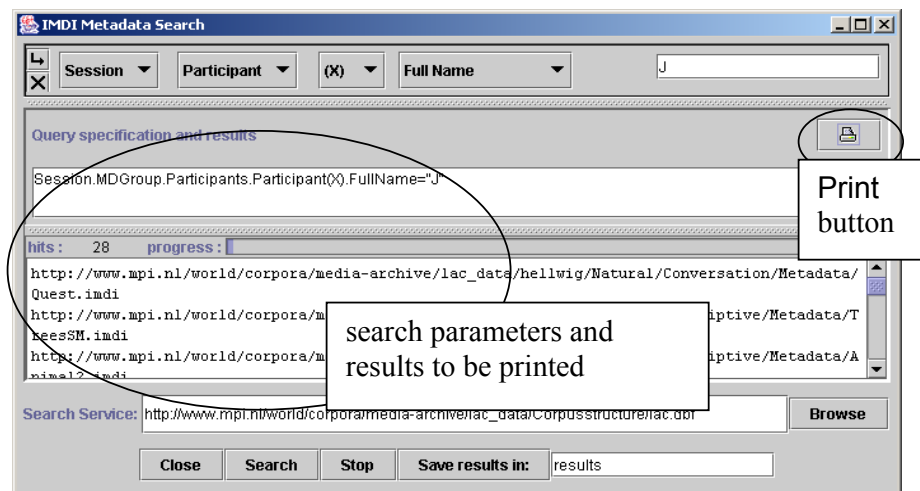
7.7 Closing the IMDI Metadata Search window

Click on File menu, then click on Close to exit the window.

Appendix C

Access Rights Management for

Language Resource Archives at the MPI

Manual

This manual was last updated: March 23, 2004

Introduction

The Access Rights Management-web pages are developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. They are used to allow authorized persons to define access rights for language resources that are part of the archives stored at the MPI (i.e. DOBES Corpus, Dutch Spoken Corpus, MPI Corpora and ESF Corpus). At the moment, only the right to read resources can be assigned. The corpora are arranged in a hierarchical tree structure and access rights can be assigned to nodes in that structure. All Nodes below a certain Node in the hierarchy inherit the rights of this Node. Access rights assigned to nodes higher in hierarchy are overruled by access rights assigned to nodes lower in hierarchy.
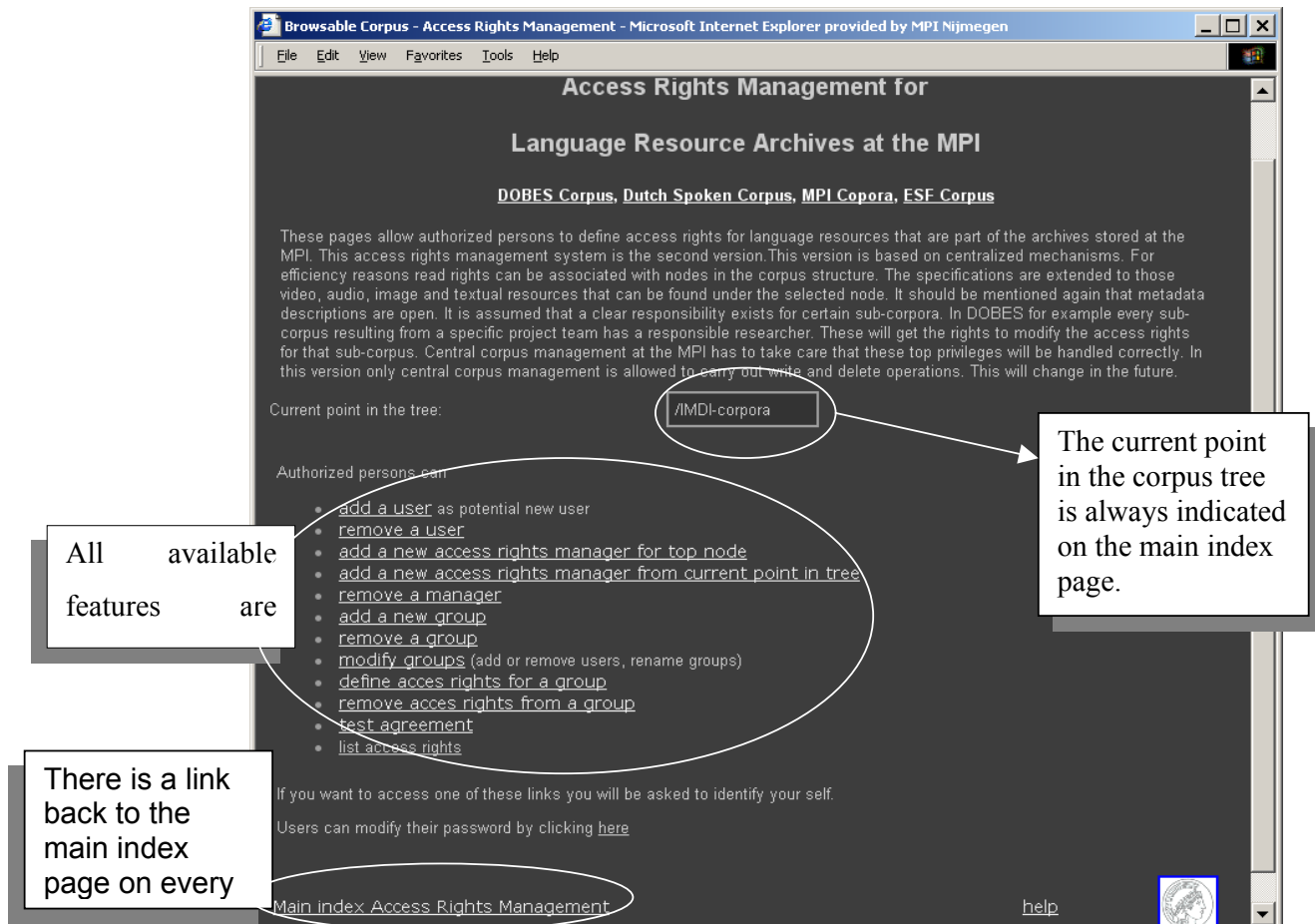
The Access Rights Management-web pages support the following features:

- Adding and removing users;
- Adding and removing access rights managers
- Adding and removing groups;
- Modifying groups;
- Defining and removing access rights for a group;
- List access rights.

The Access Rights Management-web pages can be accessed from the Browsable Corpus web-site (http://corpus1.mpi.nl/BC/IMDI-corpora/). At the bottom of the page is a link "set access rights for this node". Clicking this link opens the Access Rights Management main index page. The current node in the corpus tree is the starting point for the Access Rights Management.

> **!** Note: you need username and password to login.

The main index page-display is as follows:

This manual explains and exemplifies the features of the Access Rights Management-web pages. It is organized around the following 5 parts:

VI.   Adding and removing users (chapter 1).

VII.   Adding and removing access rights managers (chapter 2).

VIII.   Adding, removing and modifying groups (chapter 3).

IX.   Defining and removing access rights for groups (chapter 4).

X.   List access rights (chapter 5).

Notation Conventions

The following notation conventions are used:

- Screens, schemata and fields of the web pages are written in the font MS Sans Serif.

- (SHORTCUT) KEYS ARE WRITTEN IN SMALL CAPS.

- "Text in double quotes refers to menu items and screen displays."

- Information on troubleshooting starts as follows: **!**

Table of Contents

1   Adding and removing users

This part of the manual describes how authorized persons can add potential new users and remove users. There are two kinds of users:

- definers
- normal users

Only definers can give access rights to other normal users and can create other definers. Definers are always connected to a certain level in the hierarchical tree structure. Therefore, definers can only lend access rights for the current level in the tree, or any level below.

1.1 Adding a new users

From the main index page, click the "add a user" link. In the opened form, the definer has to provide the following date concerning the potential user:

- first name or initials
- prefix (optional)
- username
- password: the user can later modify this password
- user class: class is "normal" (for normal users) or "definer". The created definer can create new users, but is not able to set access rights. To solve this a Node must be assigned additionally to the definer (see chapter "Adding and removing access rights managers").
- postal address (optional)
- email address: potential new users have to accept an agreement, sent to them by email (see chapter "add a new group")

1.2 Removing users

From the main index page, click the "remove a user" link.

In the opened page, all users are listed, with username and creator.

If you want to list only the users that you have created, check the "Creator's only" box.

You can select how you want the list to be ordered from the pull-down menu. The list can be ordered by user first name, user last name, user username, creator first name or creator last name.

Click the "Del" button in front of the user you want to remove. You can only remove a user that you have created.

2 Adding and removing access rights managers

Definers can create access rights managers (also definers) and give persons the permission to assign access rights to other users for a (sub-) corpus that belongs to your domain of authority. The created access right manager inherits your access rights for the specified (sub-) corpus, i.e. both can overwrite settings the other made, but only the creator has the right to withdraw the given authority.

2.1 Add a new access rights manager from top node

From the main index window, click the "add a new access right manager from top node" link. Select the correct top node from the Top Node pull-down menu. Next, select the person (an existing user of the user class "definer") you want to authorize from the User name pull-down menu and click the "submit" button.

2.2 Add a new access rights manager from the current point in tree

Instead of creating a new access rights manager from a top node, authorized persons can also create a new access rights manager from the current point in the tree. From

the main index page, click the "add new access rights manager from current point in tree" link.

To add a new access rights manager, do the following:

In the opened page, the current node is indicated, under Current node.

Select the correct (sub-) corpus node that will be the starting point from the Node name pull-down menu. You can click a field in the list and scroll the list.

Next, select the definer from the User name pull-down menu and click the "Submit" button.

2.3 Removing access rights managers

To remove an access rights manager, click the "remove a manager" link from the main index page. In the opened page, all managers are listed, with login name, creator and path to the node that is set as the starting point for the (sub-) corpus node.

If you want to list only the managers that you have created, check the "Creator's only" box. You can also select how you want the list to be ordered (by manager first name, manager last name, manager login name, creator first name, creator last name or path) from the pull-down menu.

Next, click the "Del" button in front of the manager you want to remove. You can only remove an access rights manager that you have created.

3   Adding, removing and modifying groups

3.1 Add a new group

From the main index window, click the "add a new group" link. A new page is opened with a form that allows you to create a new group and add users to that group. Only the creator of a group can modify that group.

To create a new group, do the following:

Add the name for the new group here.

Select the users you want to add to this new group. Hold down the ctrl-key to select more than one user, or hold down the shift-key to select a block of users

Select an agreement that the user has to accept. The agreement is sent to the user by email. This feature is optional, if you don't want your users to accept any agreement select "(none)".

Click the "AddGroup" button to create the group.

Click the "Reset" button to clear all fields.

3.2 Remove a group

From the main index window, click the "remove a group" link. In the opened page, all groups are listed, with the creator. If you want to list only the groups that you have created, check the "creator's only" box. You can select how you want the list to be ordered (by group name, creator first name or creator last name) from the pull-down menu.
Next, click the "Del" button in front of the group you want to remove. You can only remove groups that you have created.

3.3 Modify groups

From the main index window, click the "modify groups" link. The opened page allows you to add and remove members from a group. You can select more members at a time to add or to remove. Only creators are allowed to modify their groups. To add or remove users, do the following:

Select the group you want to modify from the pull-down menu. Listed are all groups and their creators in brackets.

If you want to list only the groups that you have created, check the "creator's only" box.

For the selected group, all members are listed here. Select the member(s) you want to remove and click the "Remove" button.

All users are listed here. Select the user(s) you want to add to the selected group and click the "Add" button.

!Note: Holding down the Ctrl-key makes it possible to select more names one by one. If you press the Shift-key you can select a group of names (blocked).

4 Defining and removing access rights for groups

4.1 Defining access rights for groups

From the main index page, click the "define access rights for a group" link. The opened page allows you to associate read rights to a group or to all users, for all resources that can be found under the selected node. Therefore, please make sure that you are at the right node in the corpus tree. The current starting node has to be within your domain of authority. All nodes and all resources that will be affected by your operation are listed at the bottom of the web page.

To define access rights for a group or for all users, do the following:

Select the group whose access rights you want to define. The group-members are listed in the field on the right. If you want to list only the groups that you have created, check the "creator's only" box.

The current node in the tree is shown here. This node and nodes below in the tree will be affected by your operation.

Define the access rights by checking the boxes. Access rights can be defined for a specific type of media (i.e. all annotations, all audio, all images, all information files or all videos). Access rights can be allowed/denied for the selected group or allowed/denied for all users. Note that allowing/denying access rights for all users automatically overrules access rights defined for the group.

The affected corpus nodes and resources are indicated in the pull-down menus at the bottom of the page. You can navigate through the menus using the scroll wheel on your mouse, the PageUp and PageDown-keys or the arrow-keys.

Next, click the "AddRights" button to complete the operation.

4.2 Remove access rights for a group

From the main index window, click the "remove access rights from a group" link. The opened page lists all groups, with creator and the path to the corpus node. If you want to list only the groups that you have created, check the "creator's only" box. You can select how you want the list to be ordered (by group name, creator first name, creator last name or path) from the pull-down menu.
Next, click the "Del" button in front of the group you want to remove. You can only remove groups that you have created.

5   List access rights

This feature will be added soon.

Appendix D

IMDI CV-Editor, version 1.0

Manual

This manual was last updated: 14 Jul 2003
The latest version of the manual can be downloaded from the following webpage: www.mpi.nl/tools
Author: Birgit Hellwig

Introduction

The IMDI CV-Editor (ISLE Metadata Initiative – Controlled Vocabulary Editor) was developed at the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. The CV-Editor is used to create lists of controlled vocabulary items. As such, it complements the other IMDI tools, notably the IMDI Editor (also known as the IMDI Metadata Editor).

At the moment, the following IMDI tools exist (tools and manuals can be downloaded from www.mpi.nl/IMDI/tools):

- IMDI (Metadata) Editor: a tool for creating metadata descriptions;

- IMDI Browser: a tool for displaying and accessing data and metadata;

- IMDI Metadata Search Tool: a tool for searching through metadata information.

This manual explains and exemplifies the features of the IMDI CV-Editor. It is organized around the following three chapters:

4. Basic information

5. Creating controlled vocabularies with the IMDI CV-Editor

6. Working with controlled vocabularies in the IMDI Metadata Editor

Notation Conventions

The following notation conventions are used:

- Menu items, icons and screen displays are written in the font MS Sans Serif.

- (SHORTCUT) KEYS ARE WRITTEN IN SMALL CAPS.

- Information on troubleshooting starts as follows: **!**

Table of Contents

8 Basic information

All the IMDI tools are concerned with metadata information. As such, they make use of so-called "controlled vocabularies". Each vocabulary consists of a metadata category (e.g., a category labeled "languages") and of a list of predefined values to be entered into this category (e.g., all the names of the world's languages).

The IMDI CV-Editor allows you to create such controlled vocabularies, i.e., to define a metadata catgeory and its values. For example, in the illustration below, the category "MPI-Languages" is specified, and its values are defined as consisting of the names of all those languages that researchers at the MPI for Psycholinguistics in Nijmegen work on (i.e., a subset of the world's languages).
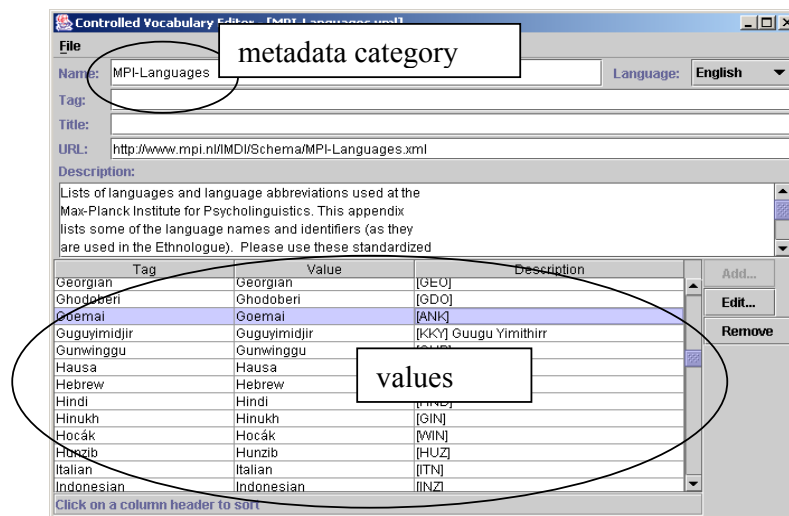
In other IMDI tools (e.g., in the IMDI Metadata Editor) such values are then displayed in form of pull-down menus. For example, in the illustration below, all values of the metadata category "MPI-Languages" are displayed.
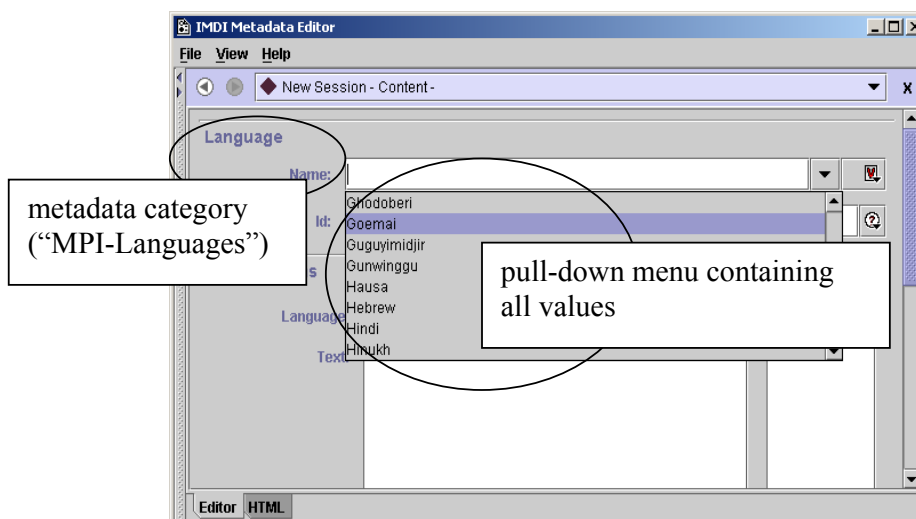
Controlled vocabularies thereby (a) facilitate the entering of metadata information (as the values need not be typed, but are selected from a pull-down menu) and (b) ensure consistency (as the values are selected from a predefined list).

9 Creating controlled vocabularies with the IMDI CV-Editor

Starting the IMDI CV-Editor opens up the Controlled Vocabulary Editor window. This window is structured as follows:
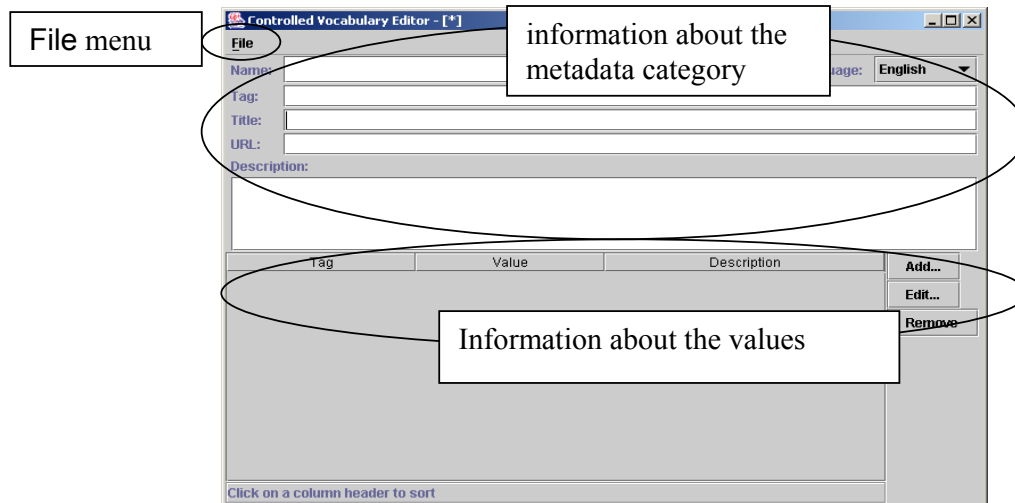
The File menu of the Controlled Vocabulary Editor window allows you to create, open, save and print controlled vocabularies file (see section 9.1). Each file contains information about the metadata category (see section 9.2) and its values (see section 9.3).

9.1 Managing controlled vocabularies: the File menu

The File menu is used for managing the controlled vocabularies files. It contains the following menu items:
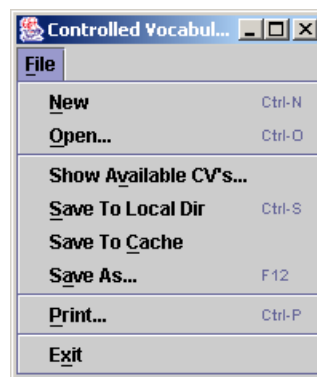
To access any menu item, do one of the following:

(g) Click on that item.

(h) Or use the shortcut key ALT plus the underlined letter(s) (e.g., to access the item File – New, press first the keys ALT+F and then N).

(i) Or use the shortcut key that is displayed to the right of the corresponding menu item (e.g., to access the item New, press the keys CTRL+N) (only available for some items).

The following menu items are available:

New	Click on New to create a new controlled vocabulary.
Open …	Click on Open … to open an existing controlled vocabulary.

! Note: The IMDI tools make use of a number of predefined controlled vocabularies. These vocabularies are stored on the Web, and the links to them are stored in the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE". Any such vocabulary can be opened in the IMDI CV-Editor, and can then be taken as the basis for creating a new controlled vocabulary in a language other than English (see section 9.1). However, the predefined vocabularies themselves cannot be modified.

Show Available CV's …	Click on Show Available CV's … to view a list of all controlled vocabularies that are stored in the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE". The Vocabs overview window opens and displays information on the Name, the Tag, the Link To Canonical CV, the Description and the Language of a controlled vocabulary (see section 9.1 for details), e.g.:
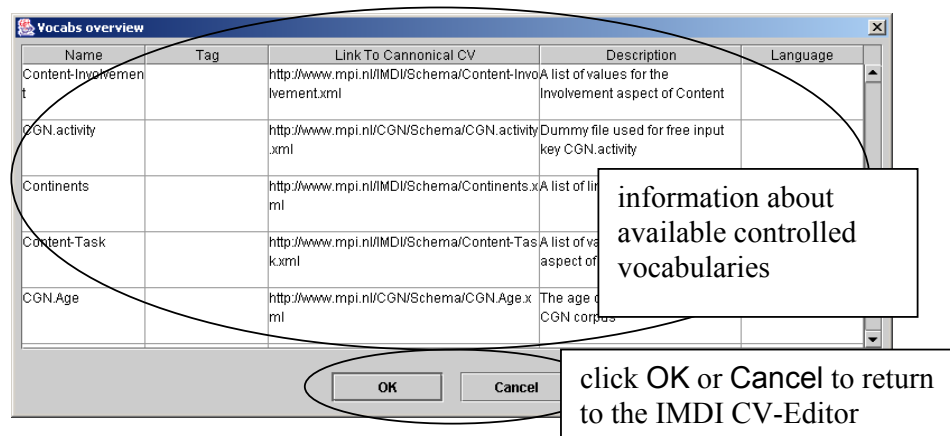
! Note: You can open any of the available controlled vocabularies from within the **Vocabs overview** window. Do the following:

[Screenshot: Vocabs overview window]

Name	Tag	Link To Cannonical CV	Description	Language
Content-Involvement		http://www.mpi.nl/IMDI/Schema/Content-Ivement.xml		
CGN.activity		http://www.mpi.nl/CGN/Schema/CGN.ac.xml		
Continents		http://www.mpi.nl/IMDI/Schema/Contine...ml		
Content-Task		http://www.mpi.nl/IMDI/Schema/Content-Task.xml	A list of values for the Task aspect of Content	
CGN.Age		http://www.mpi.nl/CGN/Schema/CGN.Age.xml	The age categories used by the CGN corpus	

1. In the window, click on the controlled vocabulary to be opened.

It will be highlighted in blue color.

OK

2. Click on **OK**.

[Screenshot: Controlled Vocabulary Editor - [Continents.xml]]

File

Name: Continents
Tag:
Title:
URL: http://www.mpi.nl/IMDI/Schema/Continents.xml
Description:
A list of linguistic continents

Tag	Value	Description
Unknown	Unknown	
Unspecified	Unspecified	
Africa	Africa	
Asia	Asia	
Europe	Europe	
Australia	Australia	
Oceania	Oceania	
North-America	North-America	
Middle-America	Middle-America	
South-America	South-America	

Add...
Edit...
Remove

Click on a column header to sort

3. The highlighted vocabulary will be displayed in the IMDI CV-Editor.

Sav   ...Local Dir to save the controlled
                                 ...he local directory. The file is
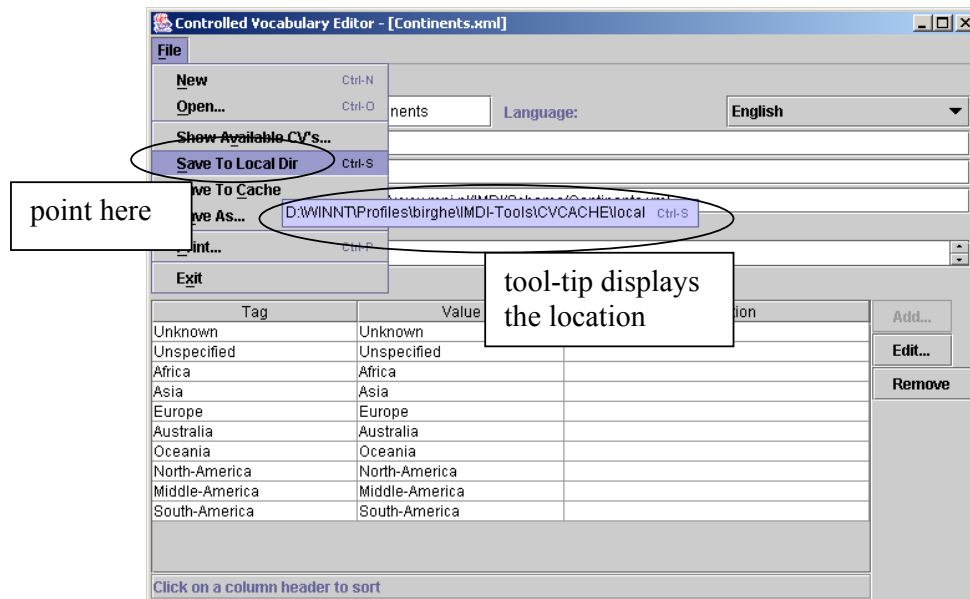   ...user-directory \ IMDI-TOOLS \

Save To Cache   Click on **Save To Cache** to save the controlled vocabulary file to the cache. The file is saved to the folder "user-directory \ IMDI-TOOLS \ CVCACHE \ www.***.***".

Save As …   Click on **Save As** … to save the controlled vocabulary file anywhere on the computer.

! Note: The files will be saved as XML files.

! Note: If you are in doubt where the local directory or the cache is located on your computer, point with the mouse to Save To Local Dir or Save To Cache. The tool-tip is activated and displays the corresponding location, e.g.:

! Note: The IMDI tools can only make use of controlled vocabularies that are saved to one of the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE". Files that are saved anywhere else on the computer cannot be accessed by the IMDI tools.

! Note: It is recommended that all controlled vocabularies are stored on the Web and are saved in one of the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE". In this case, the IMDI tools will be able to check for newer versions of the corresponding controlled vocabularies and to update the cache accordingly.

Print …   Click on Print … to print the controlled vocabulary.

Exit   Click on Exit to exit the IMDI CV-Editor.

9.2 The metadata category

The IMDI CV-Editor asks you to specify the following kind of information about the metadata category:

1. Name, Tag and Title

Name contains the name of the underlying metadata category, Tag contains the form that is displayed in the user-interface of the IMDI tools, and Title contains the spelled-out version of the Tag.

- The Name is obligatory. It is automatically the same as the name of the controlled vocabulary file (plus a language abbreviation, if applicable) (see below).

- The Title is optional. It contains the information that is displayed in the IMDI tools whenever the tool-tip is activated. That is, if you point with the mouse to a metadata category, a blue box containing the Title will be displayed.

- The Tag is optional. It is only necessary if you have a single underlying controlled vocabulary (a so-called "canonical" vocabulary) that corresponds to vocabularies in different languages. For example, in the above illustration, the canonical vocabulary is "Continents". The abbreviation "ger" in the Name informs us that this particular vocabulary is the German language version of the canonical "Continents". The Tag is "Kontinente" (i.e., the German word for "Continents"). That is, the IMDI tools will display the canonical "Continents" as "Kontinente" in the user-interface.

  ! Note: Currently, the IMDI tools display the English language version of all predefined controlled vocabularies. However, in a future version, it will be possible for the user to select a different language version, e.g., German. That is, all vocabularies will then be displayed in the appropriate language.
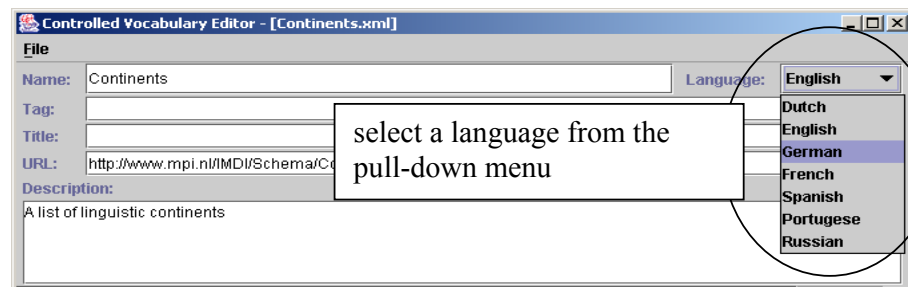
2. Language

Select the language of the controlled vocabulary from the pull-down menu, e.g., "German" in the illustration above.

Note that you can open an already existing controlled vocabulary file and create a different language version of it (see above). Do the following:
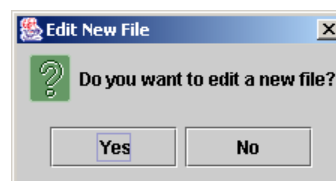
1. Open the file that contains the underlying controlled vocabulary (i.e., the so-called "canonical" vocabulary), e.g., the canonical "Continents" file:
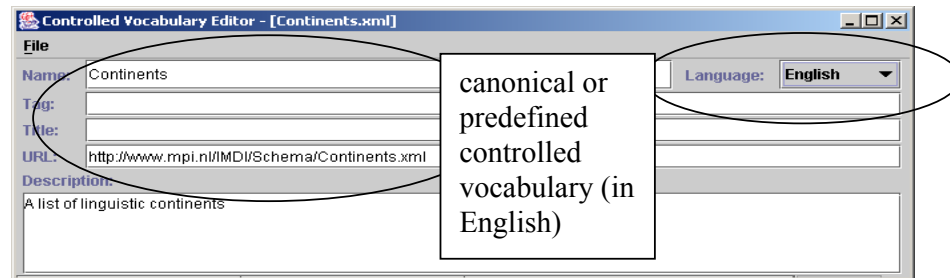
2. Go to Language, and select the appropriate language from the pull-down menu, e.g., "German":


3. The following window appears:

Click on Yes to create a new file for the new language.

Click on No to overwrite the existing file with the new language.

Click on the cross icon in the upper right corner to return to the IMDI CV-Editor without making any changes.

If you have selected either **Yes** or **No** the, IMDI CV-Editor will display the underlying category plus the new language, e.g.:
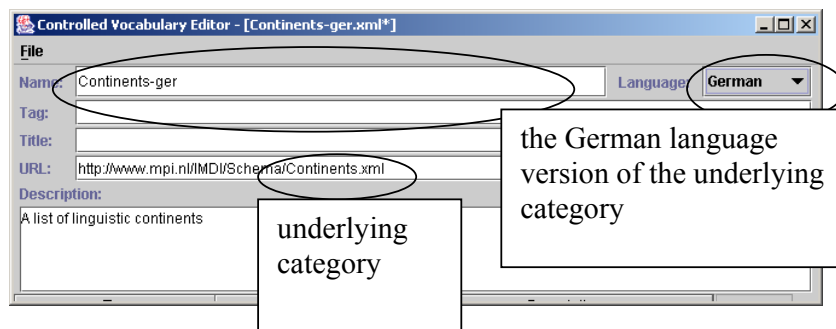

You can now edit the file, e.g., give it a German **Tag**, **Title** and **Description** (see above), and enter the appropriate German translation of the values (see section 9.3).

! Note: Please do not change the **Name** and the file name (as it appears in the **URL**):

- The **Name** contains the name of the underlying metadata category plus the appropriate language extension (e.g., "Continents-ger")

- The file name (in the **URL**) only contains the underlying metadata category (e.g., "Continents").

- The file will then automatically be saved under the name of the category plus the appropriate language extensions (e.g., "Continents-ger")

This procedure is necessary for the IMDI tools to understand that the different language versions go back to the same canonical vocabulary. The relationships between different language versions and canonical vocabularies are visualized in the **Vocabs Overview** window as follows (see section 9.1):

Specify the URL of the controlled vocabulary file
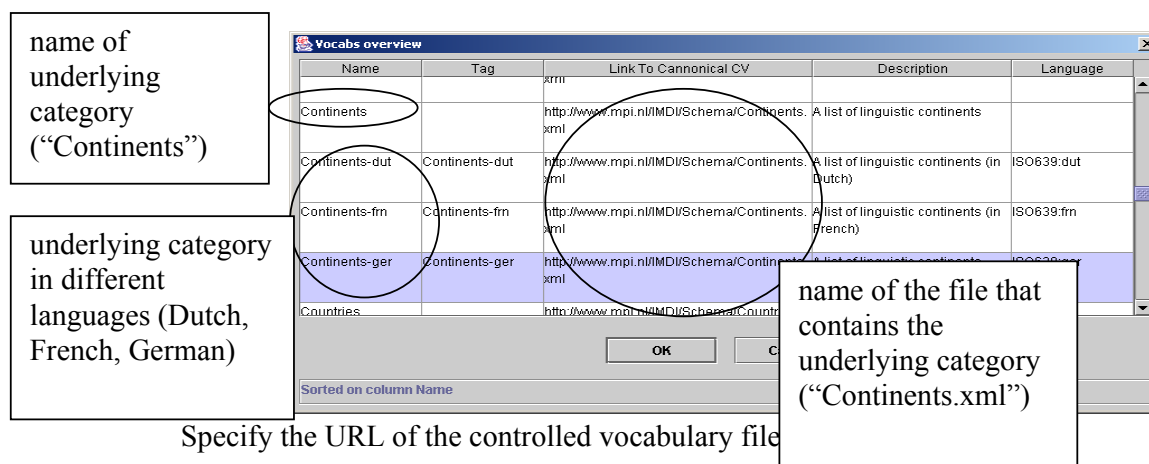
! Note: It is recommended to store all such files on the Web.

The controlled vocabulary file is an XML file, and the file name is identical to the name entered into the category **Name** (minus the language abbreviation) (see above).

4. Description

The Description is optional. It contains a prose description of the category.

9.3 The values

The IMDI CV-Editor asks you to specify the following kind of information about the values:

To add, modify or delete a value to or from the list, make use of the buttons Add …, Edit … and Remove:

- To add a value, click on Add …. The Controlled Vocabulary Entry window will be displayed.

- To modify a value, click on it to select it, then click on Edit … to modify it. The Controlled Vocabulary Entry window will be displayed.

- To delete a value, click on it to select it, then click on Remove to remove it. It will be removed without further warning.
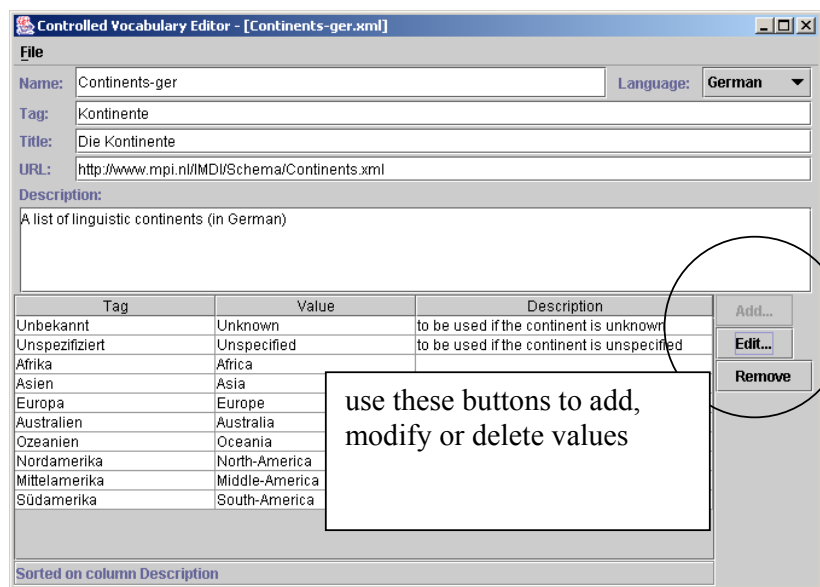
! Note: If you work with a controlled vocabulary that is predefined by the IMDI tools, the buttons Add … and Remove … will be disabled. This is also true for different language versions of predefined controlled vocabularies. That is, in such cases, you are only allowed to modify existing entries.

When the Controlled Vocabulary Entry window opens up, you are asked to enter the following information:

Value contains the underlying value, Tag contains the form that is displayed in the user-interface of the IMDI tools, and Description contains a prose description of the Tag.

- The Value is obligatory.

- The Tag is optional. It is only necessary if you have a single underlying controlled vocabulary (a so-called "canonical" vocabulary) that corresponds to vocabularies in different languages. For example, in the above illustration, the underlying value is "Africa" (= the Value), and the Tag is "Afrika" (i.e., the German word for "Africa"). That is, the IMDI tools will display the underlying value "Africa" as "Afrika" in the user-interface.

- The Description is optional. It contains descriptive information that is displayed in the IMDI tools whenever the tool-tip is activated. That is, if you point with the mouse to a value, a blue box containing the Description will be displayed.

10 Working with controlled vocabularies in the IMDI Metadata Editor

The IMDI Metadata Editor is a tool that allows you to create metadata information. As such, it makes frequent use of controlled vocabularies – both predefined vocabularies and user-defined vocabularies.

This section of the manual describes how the IMDI Metadata Editor uses the controlled vocabularies that were created in the IMDI CV-Editor. A more comprehensive manual of the IMDI Metadata Editor can be downloaded from

www.mpi.nl/IMDI/tools.

10.1   Viewing the available controlled vocabularies

The IMDI Metadata Editor can display a list of all available controlled vocabularies. To view this list, click on the menu item View, and then on Vocabularies …. The Vocabs overview window opens and displays information on the Name, the Tag, the Link To Canonical CV, the Description and the Language of a controlled vocabulary (see section 9.2 for details). Furthermore, it informs you whether the file is stored locally (an "X" symbol is entered into the column Local Only) or on the Web (the column Local Only remains blank). For example:


information about available controlled vocabularies

options for managing controlled vocabularies
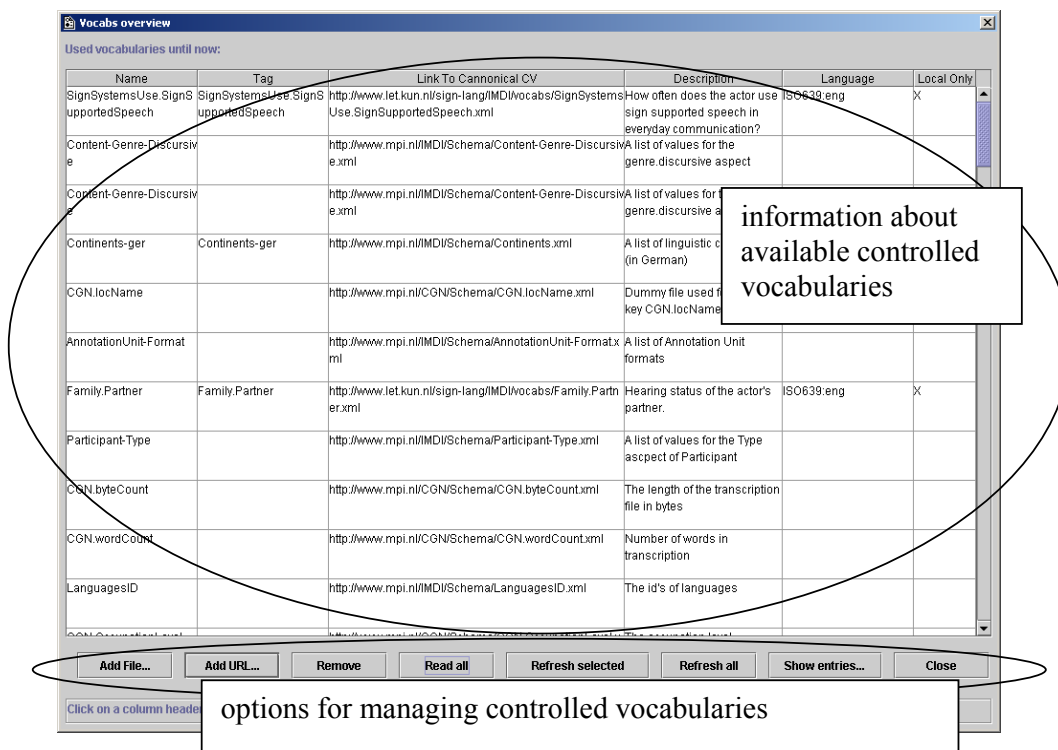
The Vocabs overview window offers you the following options for managing the controlled vocabularies:

Add File …	Click on Add File … to add a controlled vocabulary file that is stored locally. The Open dialog window appears, and asks you to navigate to the controlled vocabulary file.
Add URL …	Click on Add URL … to add a controlled vocabulary file that is stored on the Web. The Vocab add window appears: enter the URL of the file, click on Apply, then click on OK.
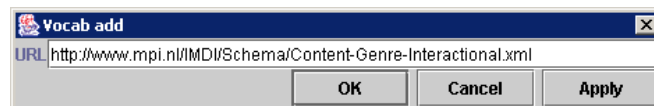
|	
Remove	Click on a controlled vocabulary in the Vocabs overview window to select it. It will be highlighted in blue color. Then click on Remove to remove this vocabulary from the window. It will be removed without further warning.
Read all	Click on Read all to view all available controlled vocabularies that are stored in the cache (i.e., all links that are stored in the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE).
Refresh selected	Click on a controlled vocabulary in the Vocabs overview window to select it. It will be highlighted in blue color. Then click on Refresh selected to update the selected vocabulary. The IMDI tools will check the corresponding URL for newer versions of this vocabulary.
Refresh all	Click on Refresh all to update all vocabularies. The IMDI tools will check the corresponding URLs for newer versions of the vocabularies.

Show entries …   Click on a controlled vocabulary in the Vocabs overview window to select it. It will be highlighted in blue color. Then click on Show entries … to view its content, e.g.:
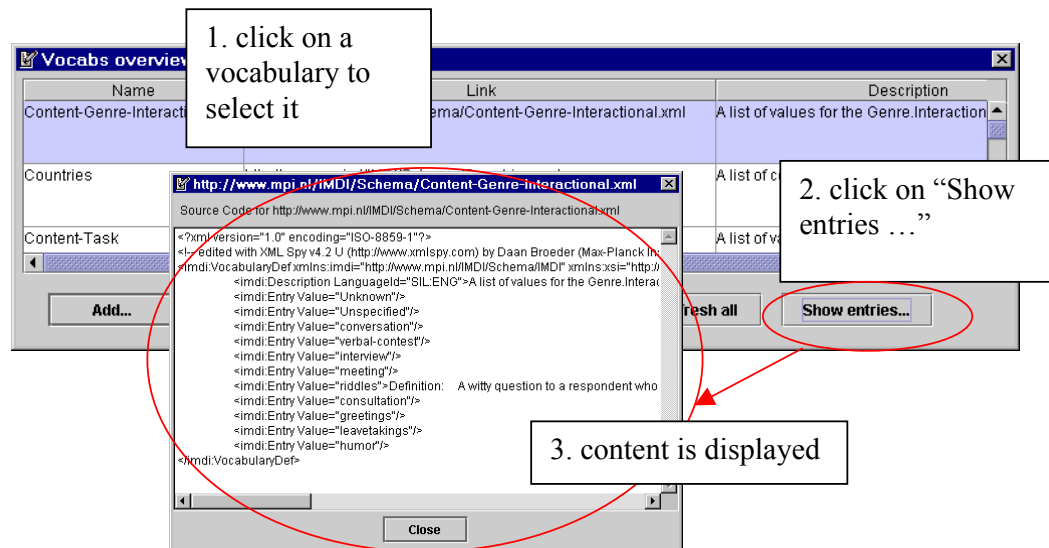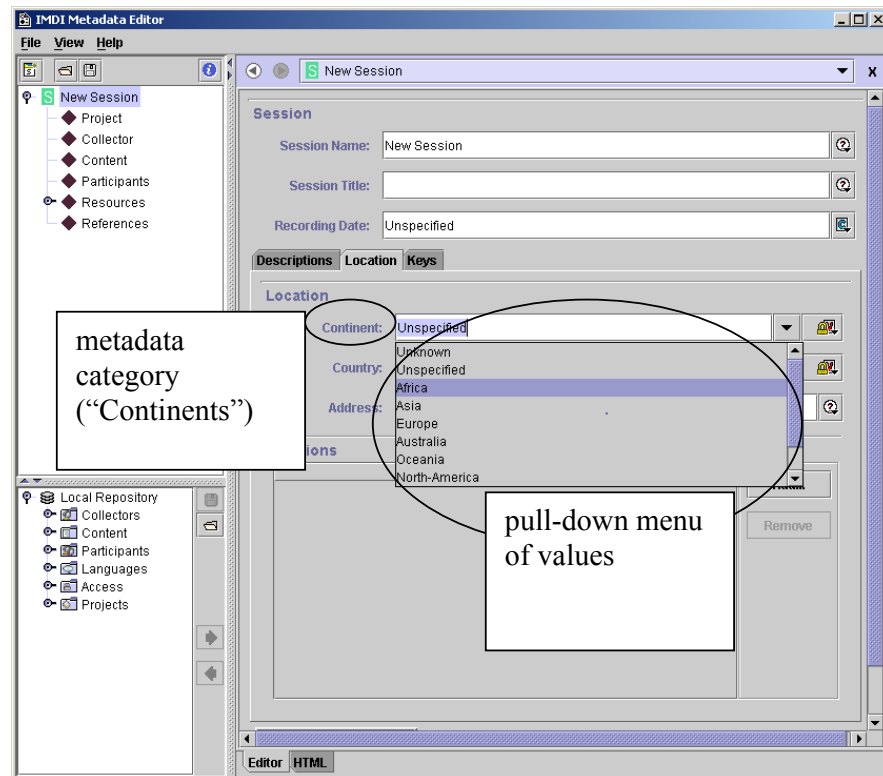

Close   Click on Close to close the Vocabs overview window.

10.2   Working with predefined controlled vocabularies

The IMDI Metadata Editor makes use of a number of predefined controlled vocabularies. These vocabularies are stored on the Web, and the links to them are stored in the subfolders contained under "user-directory \ IMDI-TOOLS \ CVCACHE".

In the IMDI Metadata Editor, these predefined vocabularies are displayed in form of a metadata category and a pull-down menu of its values, e.g.:

! Note: Currently, the IMDI Metadata Editor display the English language version of all predefined metadata categories. However, in a future version, it will be possible for the user to select a different language version, e.g., German. That is, all metadata categories and controlled vocabularies will then be displayed in the appropriate language.

10.3   Working with user-defined controlled vocabularies

The IMDI Metadata Editor makes use of a number Keys fields. These Keys contain user-defined metadata information:

- you can either type in a metadata category and a value for that category;

- or you can use a user-defined controlled vocabulary.

For this manual, only the second option is of interest. To add a user-defined controlled vocabulary, do the following:

1. Create a controlled vocabulary by means of the IMDI CV-Editor (see section 9).

2. Click on the Add Vocab Link … button, e.g.:
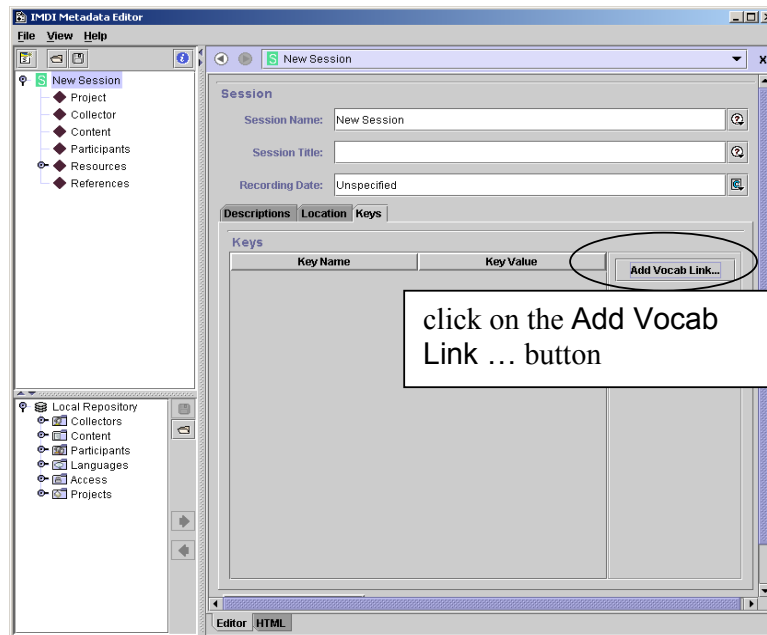
The Vocabs overview window appears and displays the available controlled vocabularies.

3. Click on the controlled vocabulary that you want to add. It will be highlighted in blue color, e.g.:
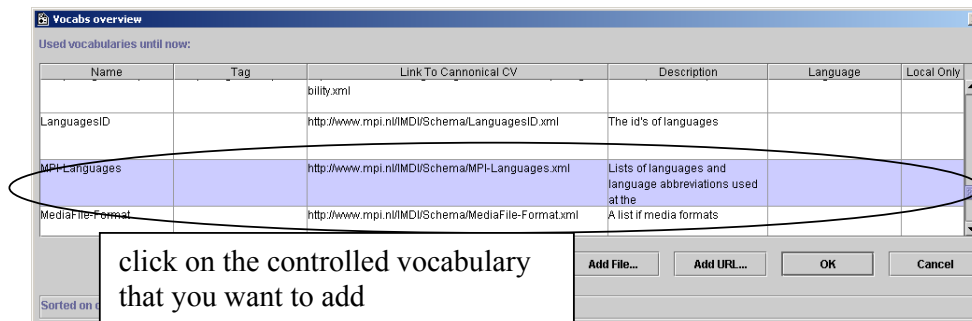
! Note: If the required controlled vocabulary is not displayed, click on Add File … or Add URL … to add it to the Vocabs overview window.

4. Click on **OK**. The **Specify Vocabulary Type** window appears, e.g.:
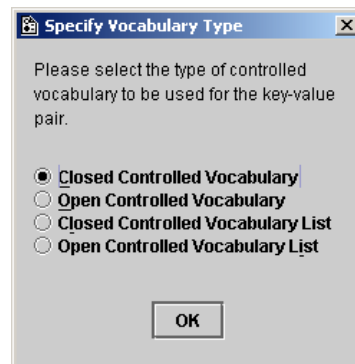
5. Select the vocabulary type by clicking in the box to its left:

Closed Controlled Vocabulary   only values defined in the controlled
                                       vocabulary will be accepted by the
                                       IMDI Metadata Editor

Open Controlled Vocabulary   values <u>not</u> defined in the controlled
                                       vocabulary can be typed in and will be
                                       accepted by the IMDI Metadata Editor

Closed Controlled Vocabulary List

Open Controlled Vocabulary List

6. Click on **OK**. The metadata category and its list of values are added to the **Keys** schema, e.g.:

The controlled vocabulary is added to the **Keys** schema