IMDI Editor, version 1.5

Manual

This manual was last updated: 27 Nov 2002
The latest version of the manual can be downloaded from the following webpage: www.mpi.nl/tools
Author: Birgit Hellwig

Introduction

The IMDI (ISLE Metadata Initiative) Editor was developed at the Max Planck Institute for Psycholinguistics, Nijmegen. It is concerned with metadata information, i.e., with information about a piece of data, such as the date when the data was collected, the location where it was collected, the larger project, the participants involved, the content, the original audio or video tapes, the corresponding annotation and digitized media files, etc.

The IMDI Editor provides a standardized format that makes it possible to enter metadata information with a minimum of effort. It supports the following features to facilitate the creation of metadata information:

- templates (i.e., recurring information needs to be entered only once and is re-loaded when needed);
- pull-down menus of controlled vocabulary items.

The IMDI Editor serves a second purpose: to prepare the standardized display of metadata information and to thereby facilitate the search process and the access to media and annotation files. The corresponding IMDI Browser displays all information that is entered into the IMDI Editor: it visualizes links between media, annotation and metadata files, and it allows simultaneous access to all data through a single interface, i.e., through the desktop computer.

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

I. The IMDI Editor and the IMDI Browser

This part introduces the connection between the IMDI Editor and the IMDI Browser. It illustrates how files that have been created with the Editor are displayed and accessed in the Browser.

II. Getting started

This part introduces (a) the key concepts, structure and terminology of the IMDI Editor, and (b) those features of the Editor that make it possible to enter recurring information.

III. The User's Guide to the IMDI Session Editor

This part gives detailed information and examples of how to fill in the different fields of the IMDI Session Editor. It is organized on a screen by screen basis.

IV. The User's Guide to the IMDI Corpus Editor

This part gives step-by-step instructions on how to manage a corpus with the help of the IMDI Corpus Editor.

V. Appendices

Notation Conventions

The following notation conventions are used:

- Main screens of the Editor start with a caption on a gray background.
- 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

PART I: THE IMDI EDITOR AND THE IMDI BROWSER

The information that is entered into the IMDI Editor is visualized by means of the IMDI Browser. Part I of this manual introduces the connection between these two tools. It illustrates the following three points:

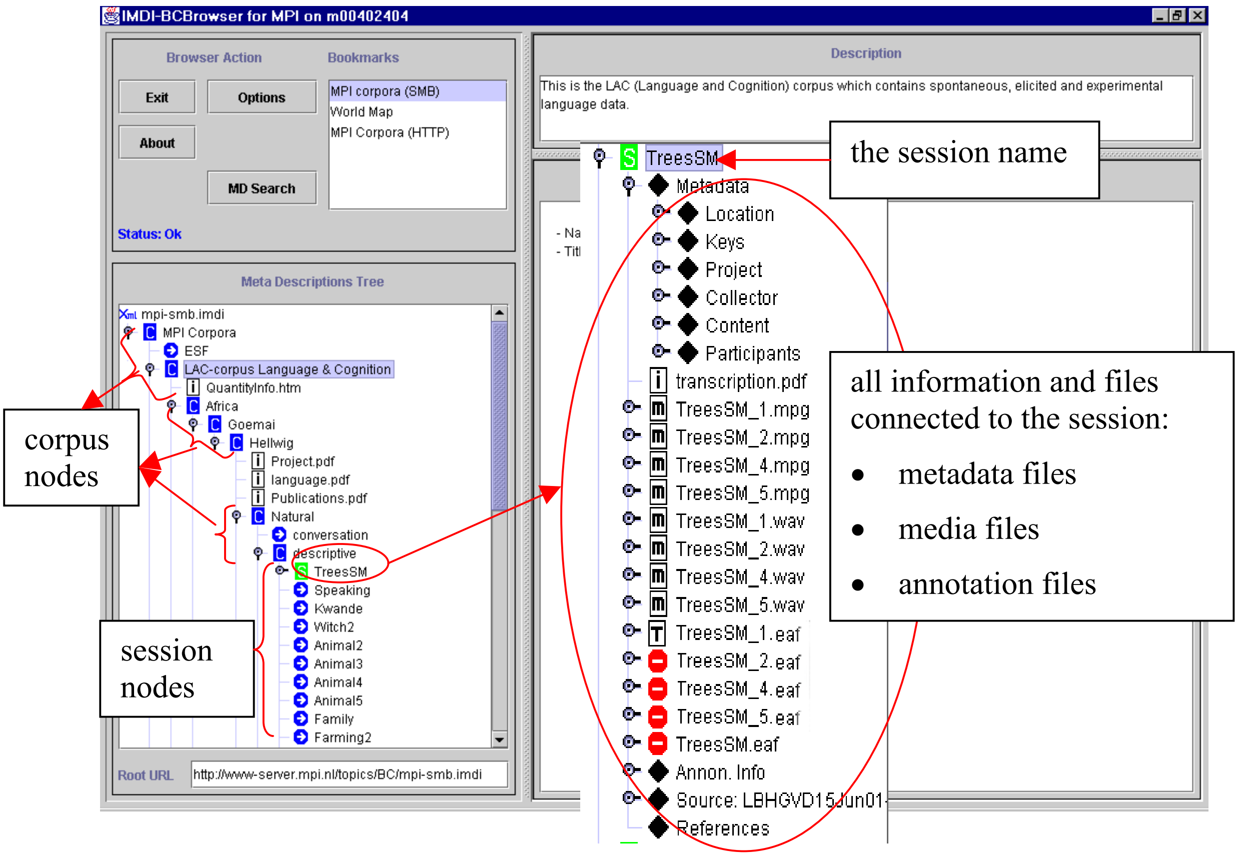
1. how the organization of your data may look like;
2. how the information that you have entered into the IMDI Editor will be displayed in the IMDI Browser;
3. and how you can access the information that you have entered into the IMDI Editor.

Note that this part only gives a brief illustration of the available possibilities. See the separate manual "IMDI Browser" for details.

1 The organization of your data

The IMDI Browser visualizes the organization of your data in form of a corpus hierarchy.

The corpus is organized in form of a tree structure that allows you to navigate through to the actual data. For example, the corpus node "MPI Corpora" in the illustration below contains the subcorpus nodes "ESF" (i.e., the European Science Foundation Second Language Acquisition Database) and "LAC" (i.e., the corpora of the Language and Cognition group), whereby each subcorpus contains further subcorpora. The lowest level of the hierarchy is the session, which contains the actual data.

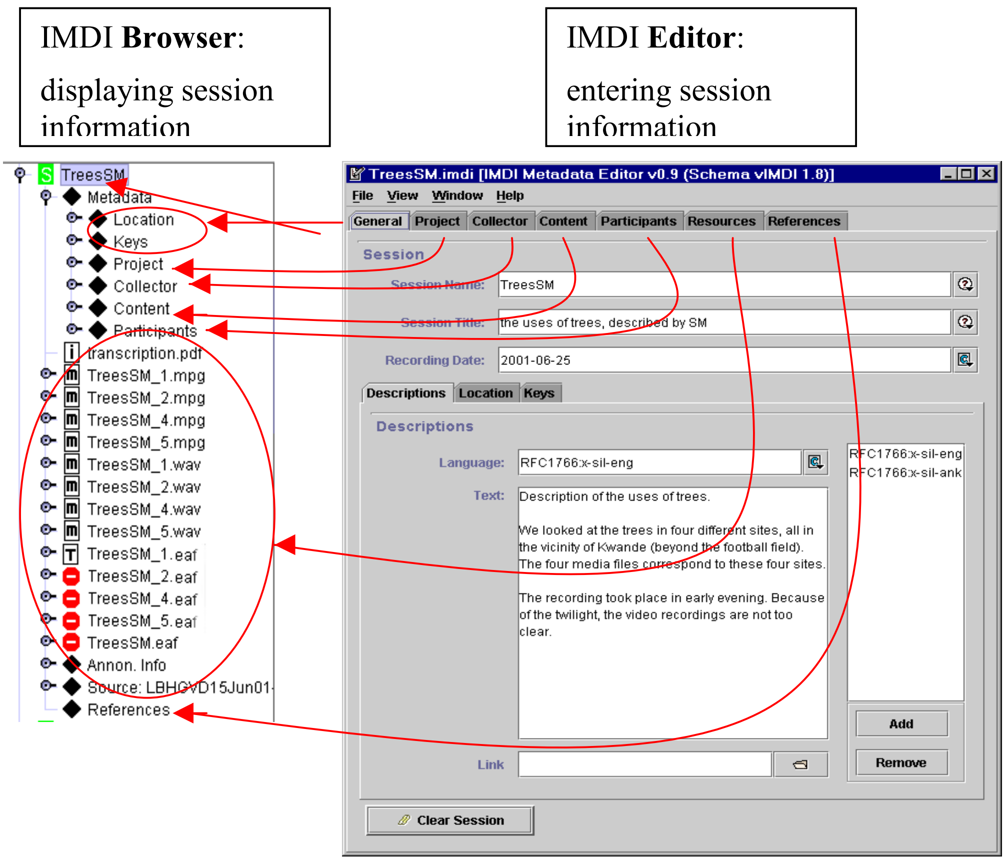


A session is the unit that is described by one IMDI Editor Session file. Such a unit can contain, e.g., a particular descriptive text told by a particular speaker at a particular date and time (e.g., the text named “TreesSM” in the illustration above). The session node displays all the information relevant to this unit. It contains the following information:

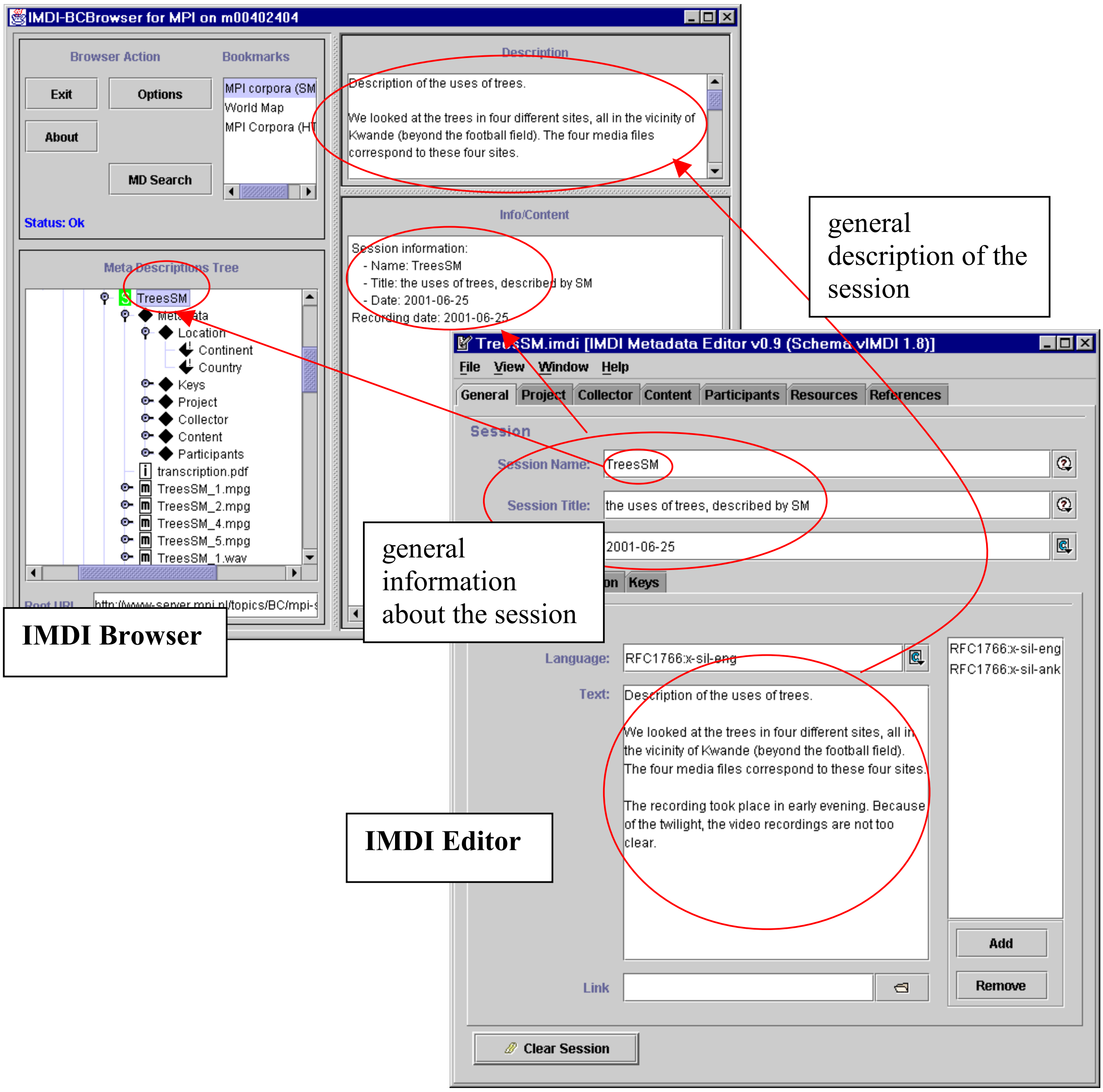
- link to the metadata file of the session (i.e., the IMDI file).
- link to the digitized media file(s), containing the audio/video data of the session.
- link to the transcription and annotation file(s) of the session.
- link to the info file(s) providing general background information.
- link not available (either because the file does not exist yet, or because access is denied).

2 Display of metadata information

At the session level, the IMDI Browser displays metadata information. This metadata information corresponds to the information that you have entered into the IMDI Session Editor: the participants of a session, the session’s location and date, its content, its corresponding audio and/or video tapes, etc. In the following illustration, you see how the structure of the Editor corresponds to the nodes displayed in the Browser.



The following illustration examplifies how a piece of metadata information is displayed in the IMDI Browser:



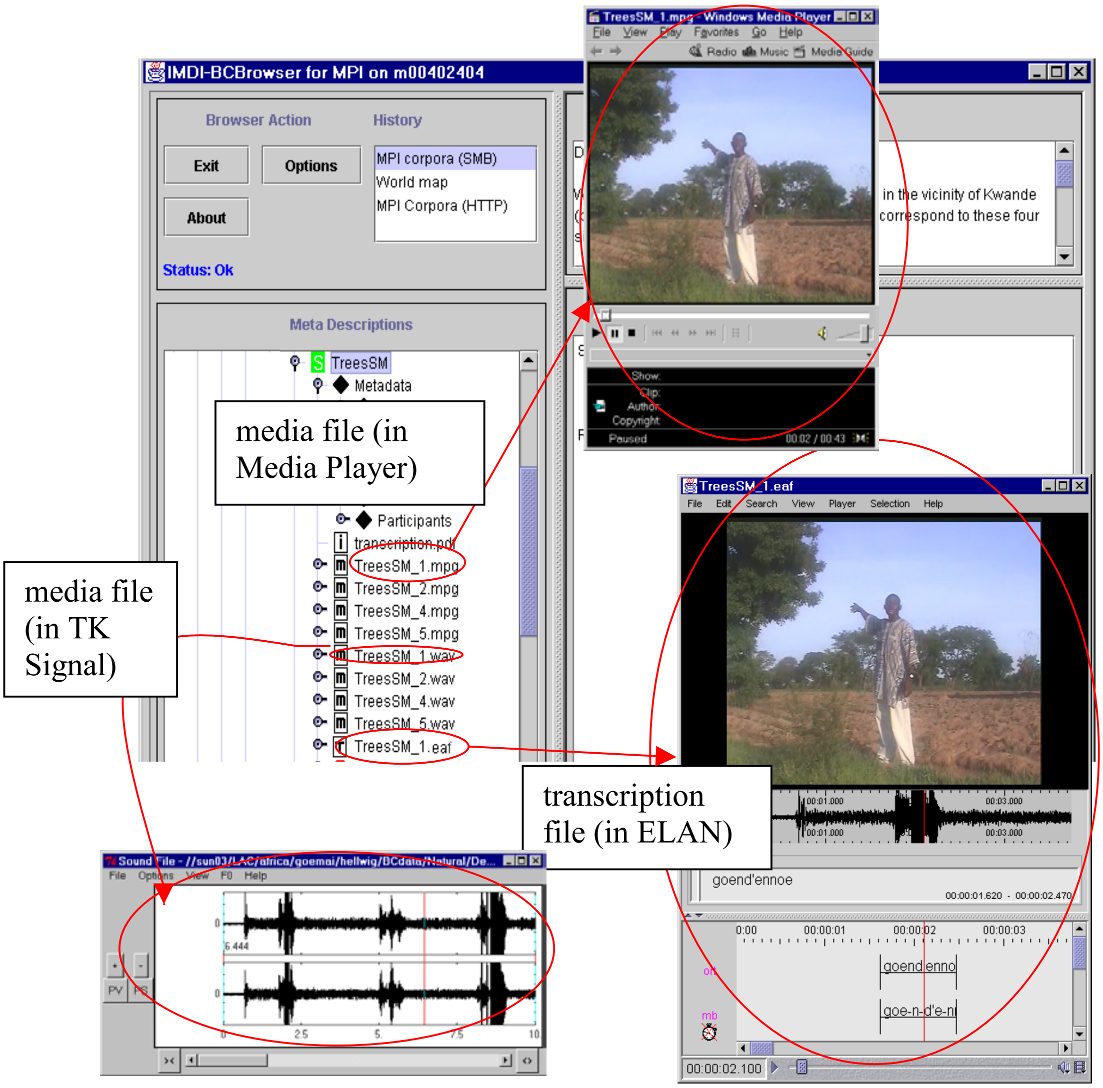
3 Accessing information

The IMDI Browser gives access to all types of data. It supports the following four features:

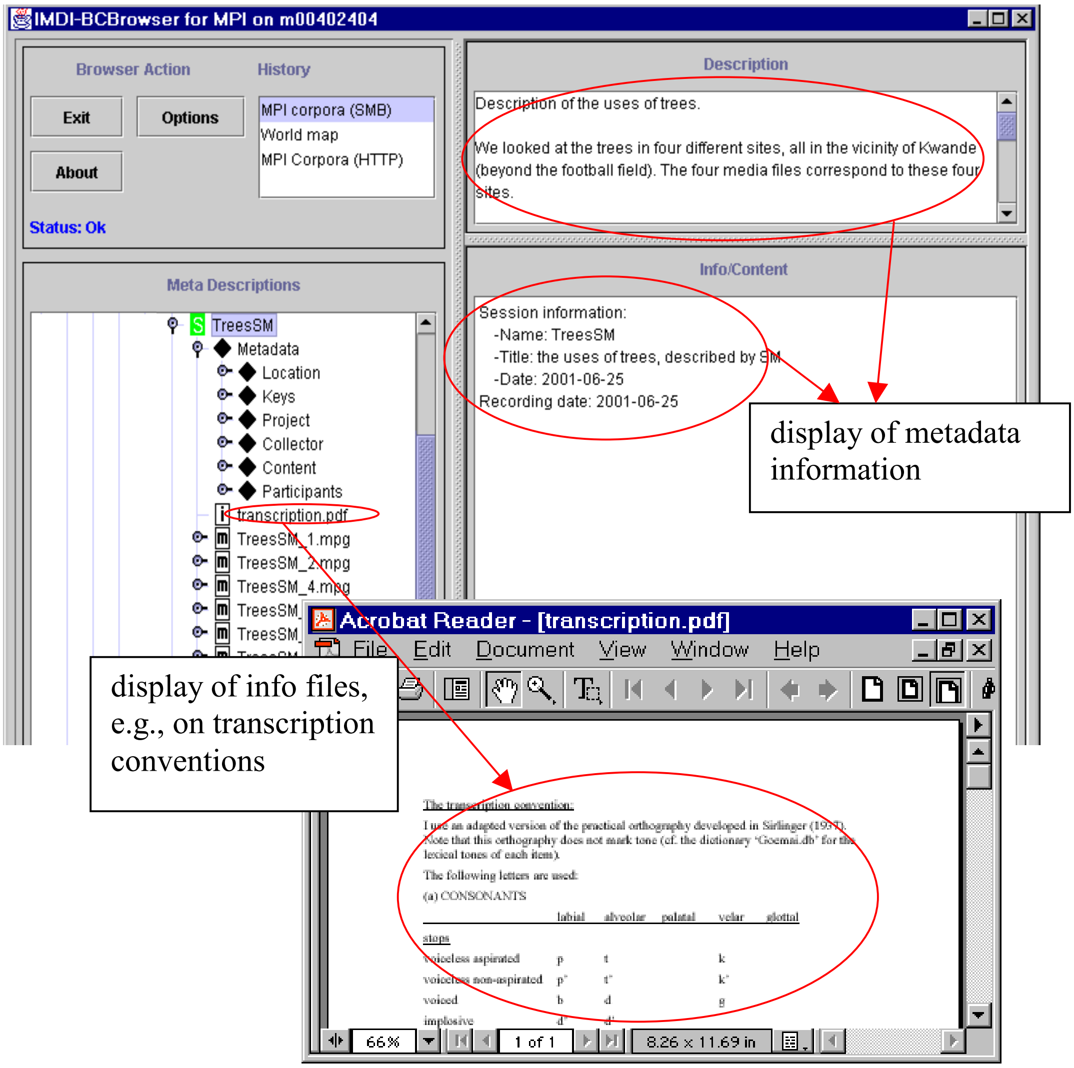
(1) A hierarchy of nodes and subnodes that allows you to structure your data in a systematic way. For example, you can navigate via the corpus nodes “Africa”, “Goemai” and “Natural data” to the subnode “Descriptive texts”. This subnode

then displays all its session nodes, i.e., all sessions that contain descriptive natural text, collected for the African language Goemai (see section 1).

(2) Each session node displays media and annotation files that can be accessed immediately. Click on an icon to open the corresponding file, e.g.:



(3) Each session node displays metadata information, e.g.:



(4) All metadata information can be searched. For example, as illustrated below, you can search for all texts by female speakers above 60 years of age. The IMDI Browser then displays the search results, and allows you to directly access the corresponding session data.

The IMDI Browser is a very flexible tool. A hierarchy of nodes will be created to suit your individual needs, and you can even create your own temporary hierarchies (see part IV of this manual). There are no restrictions on the format of your annotation files, and only very few on the format of your digitized media files. Moreover, you can constantly update all annotation and metadata files (provided you do not change their file names) to reflect changes or to enter additional information.

The IMDI Browser thus facilitates the access to and the work with your data. The corpus manager (e.g., corpus.manager@mpi.nl) helps you to integrate your data into the Browser: they develop a hierarchy of nodes in consultation with you, they link metadata information to media and annotation files, and they digitize video/audio tapes.

PART II: GETTING STARTED

This part of the manual introduces the underlying structure of the IMDI Editor. It addresses the following two topics:

- basic information about the key concepts, structure and terminology of the IMDI Editor;
- features of the Editor that allow you to enter recurring information.

Note that this part of the manual is not yet concerned with the actual content of the metadata files – this is the topic of part III.

4 Basic Information

4.1 The access policy

The metadata information entered into the IMDI Editor is meant, in principle, to be visible to the outside world via the Internet. Note that this does **not** include the audio, video and annotation files, as these are visible to you alone. And note also that the IMDI Editor itself allows for the possibility to suppress certain types of information.

Two considerations lie behind the decision to make IMDI files available: (a) the documentation of little known languages, and (b) a general policy of the Max Planck society and of many scientific journals relating to the transparency of the database.

Please be aware that other people will have access to the IMDI files. Keep this in mind when you enter information into the Editor, and make sure that you exclude all sensitive information.

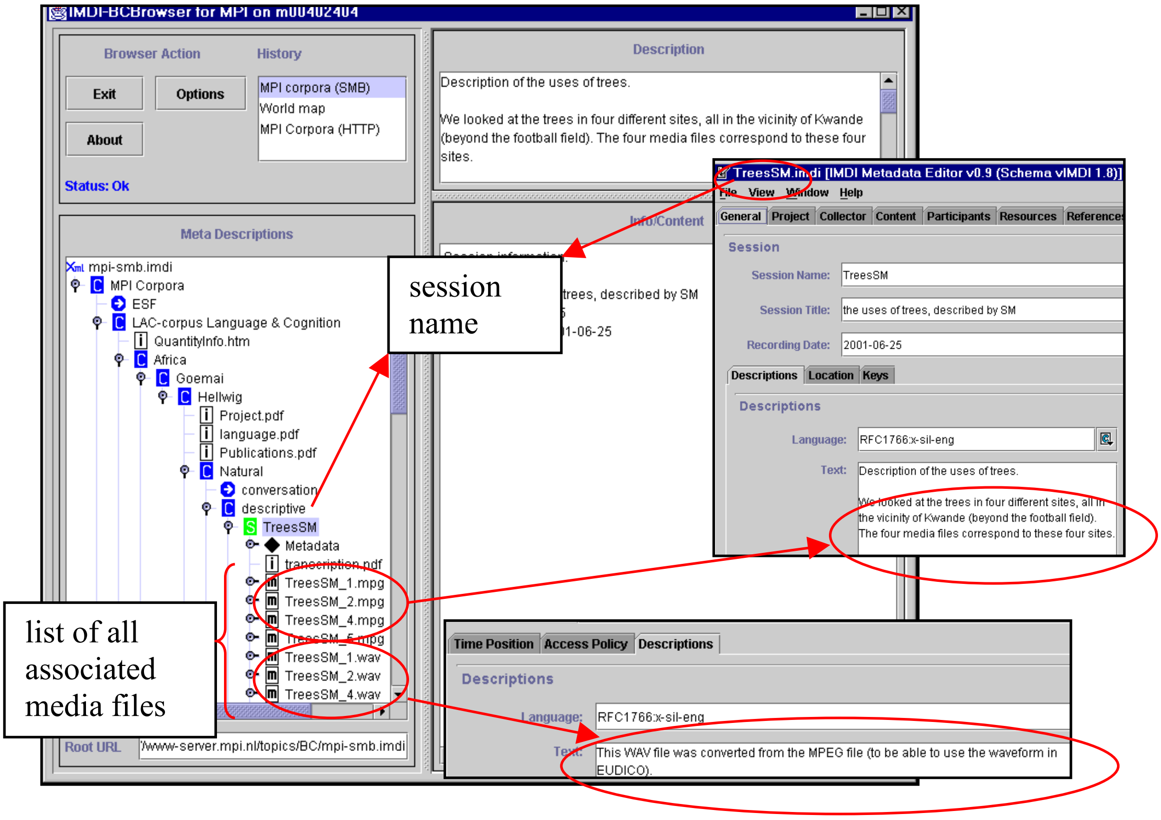
> **!** Note: We are aware that there exist old IMDI files that still contain sensitive information. If this is the case, please contact the corpus manager (e.g., corpus.manager@mpi.nl), who will propose a solution.

4.2 The session

The session is the unit that is described by one IMDI Session file (i.e., the file created with the IMDI Session Editor) (see part III). It corresponds to one unit 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’.

A session consists of media files, annotation files, info files and/or metadata (i.e., IMDI Session) files, which are linked to each other. The information that is necessary for linking these files (i.e., file names, directory location) is entered into the corresponding IMDI Session file (see section 7.6).

Note that a session does not need to correspond to a single media file and/or tape and/or annotation file. For example, in the following illustration the session "TreesSM" contains a number of different media files. These files correspond to different subparts of one session – they are included in one session because they have the same topic, the same speaker, the same location and the same date. As such, the one IMDI Session file **TreesSM.imdi** contains information about all subparts.



Conversely, one media file and/or tape may contain different sessions. For example, a speaker first explained the uses of trees; following that, another speaker talked about marriage customs. In this case, you would have to divide the tape into two sessions (one containing the 'trees' session, and the other one containing the 'marriage' session). Each of the two sessions corresponds to a single IMDI file.

4.3 The corpus

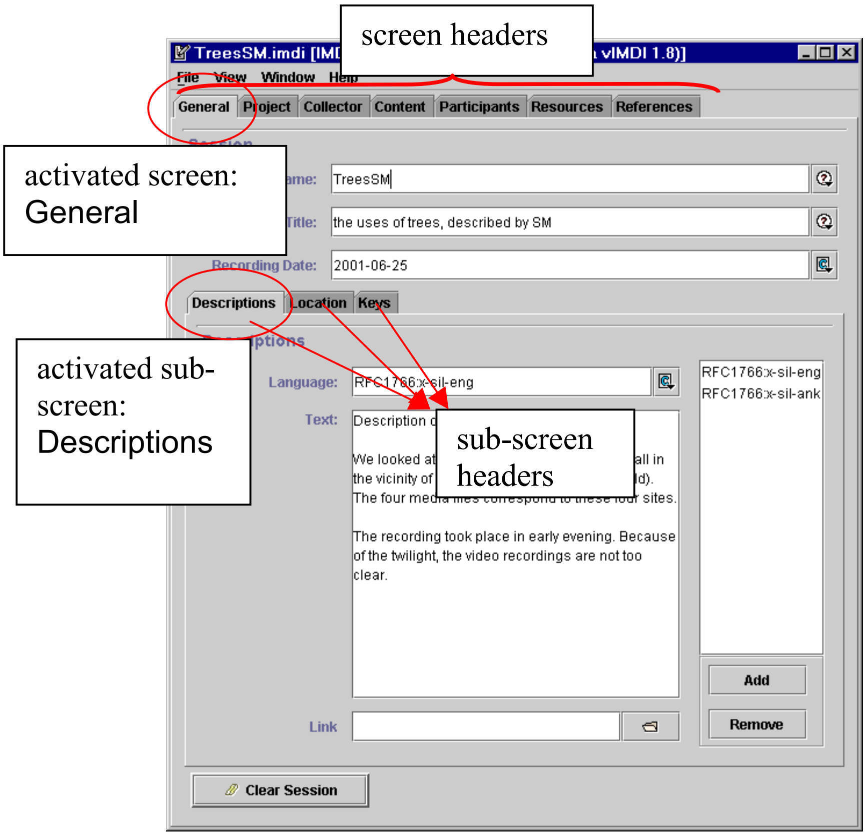
All sessions belong to a corpus, e.g., to the corpus of the language XY. Generally, the structure of the corpus is devised by the corpus manager in consultation with you. In the case of the MPI corpora, you need to contact corpus.manager@mpi.nl. In the case of other corpora, a different corpus management may be responsible.

Alternatively, you can manage your own corpus with the help of the IMDI Corpus Editor (see part IV).

4.4 The structure of the IMDI Editor

4.4.1 Screens

The IMDI Editor is organized around a number of screens that are displayed like filing cards, e.g.:

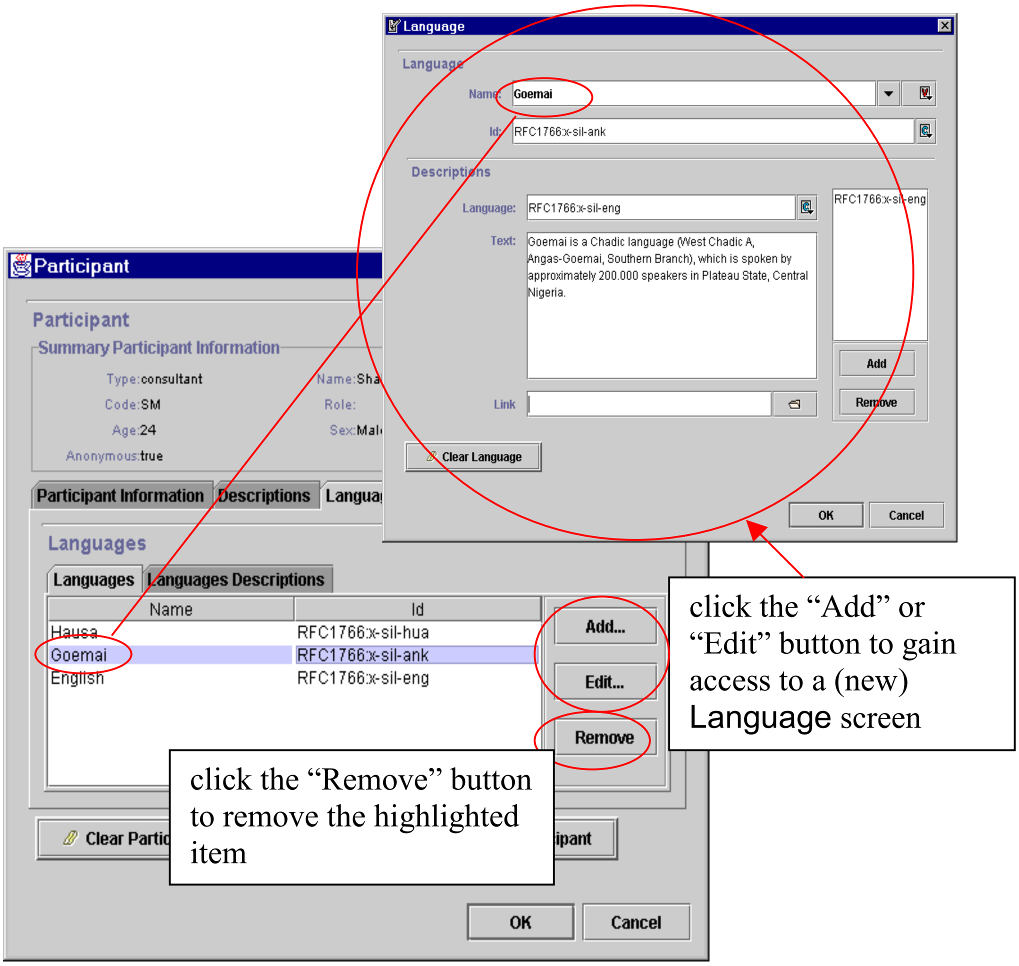


Click on a screen header to activate the corresponding screen. The following main screens are available:

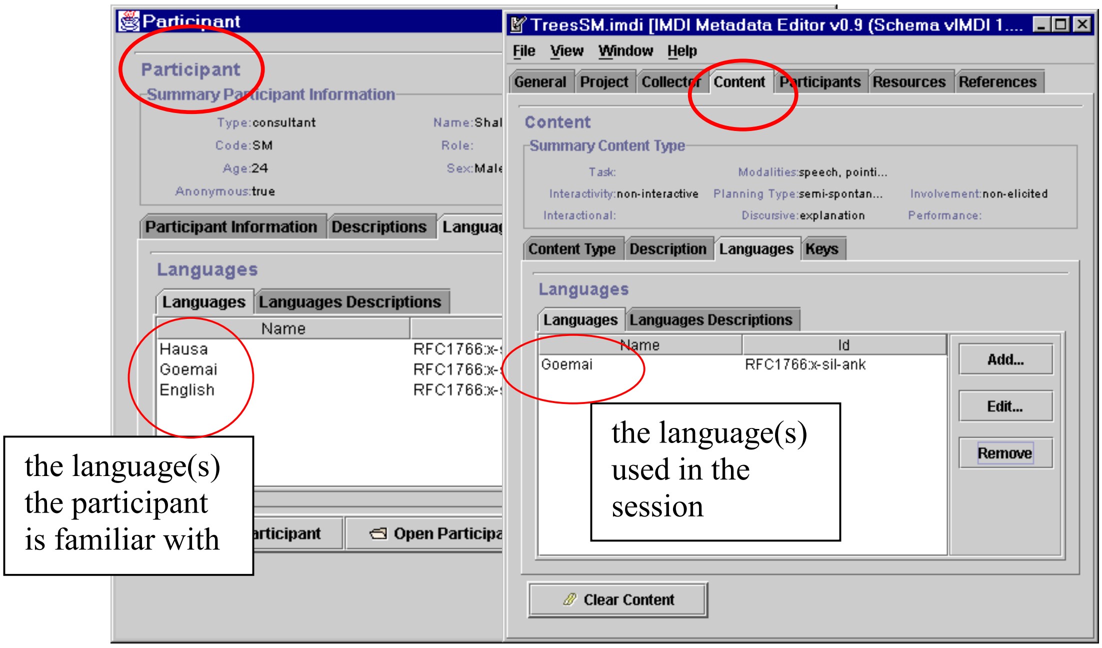
1. General: general information about (a) the session (name, date, location) and (b) the IMDI file.
2. Project: information about the larger project within which the session data was collected.
3. Collector: information about the person who collected the session data, and about the person who is responsible for collecting the session data.
4. Content: information about the content of the session.
5. Participants: information about (a) each participant of the session and (b) the interrelations among different participants of the session.
6. Resources: information about (a) access rights, (b) the annotation and digitized media files, and (c) the original media tapes.
7. References: cross-references to other sessions, fieldnotes, or publications, which are relevant to the content of the session.

These seven main screens bundle all information available on an overall topic. In addition, some of the main screens give access to additional screens. The following two types are available:

- Sub-screens that give access to additional fields and schemata relevant to the main screen. They are activated by clicking on the sub-screen header (see the screenshot above).
- “Add” buttons (and their corresponding “Remove” and “Edit” buttons) that allow you to enter multiple types of information of the same kind, e.g., one screen added for each language spoken by the participant, e.g.:

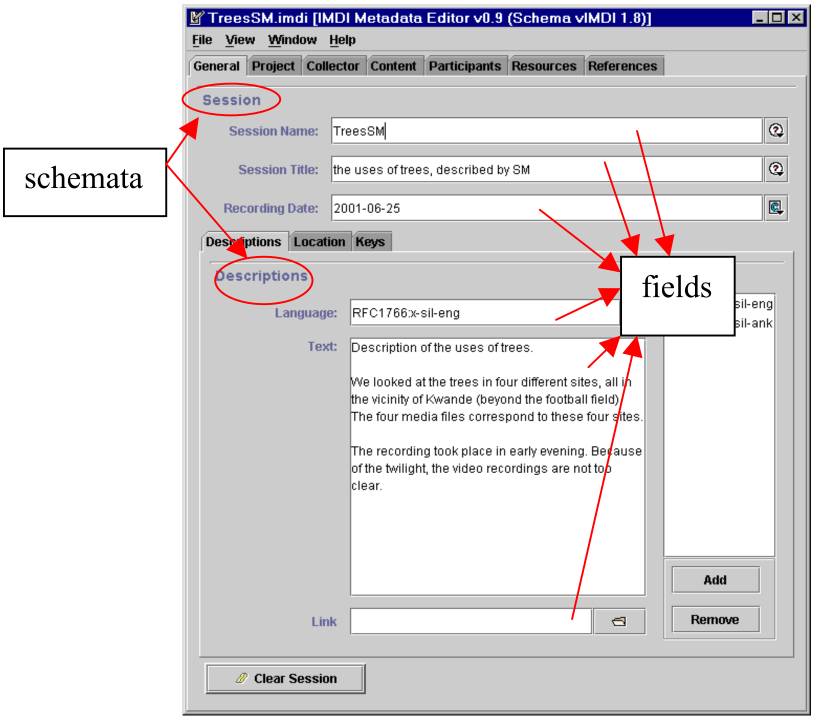


It is important to remember that all information is relevant to the main screen only. An example may illustrate this point: you will find that you are asked to provide information about ‘languages’ both on the Participant screen and on the Content screen. But the information that you enter differs: in the first case, you enter the language(s) that participant X is familiar with; in the second case, you enter the language(s) used during the session.



4.4.2 Schemata and fields

Screens are made up of a number of units that are labeled ‘schemata’. The schemata contain a number of boxes into which the actual information is entered. These are labeled ‘fields’, e.g.:

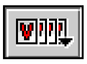


Most of the fields are more or less standardized (see “Appendix 3: Lists of recommended vocabularies”). The following options are available:

‘open vocabulary’

There exist recommendations for the content (which can be selected from a pull-down menu), but since the list is potentially endless, you will often have to type in a different content.

‘open vocabulary lists’



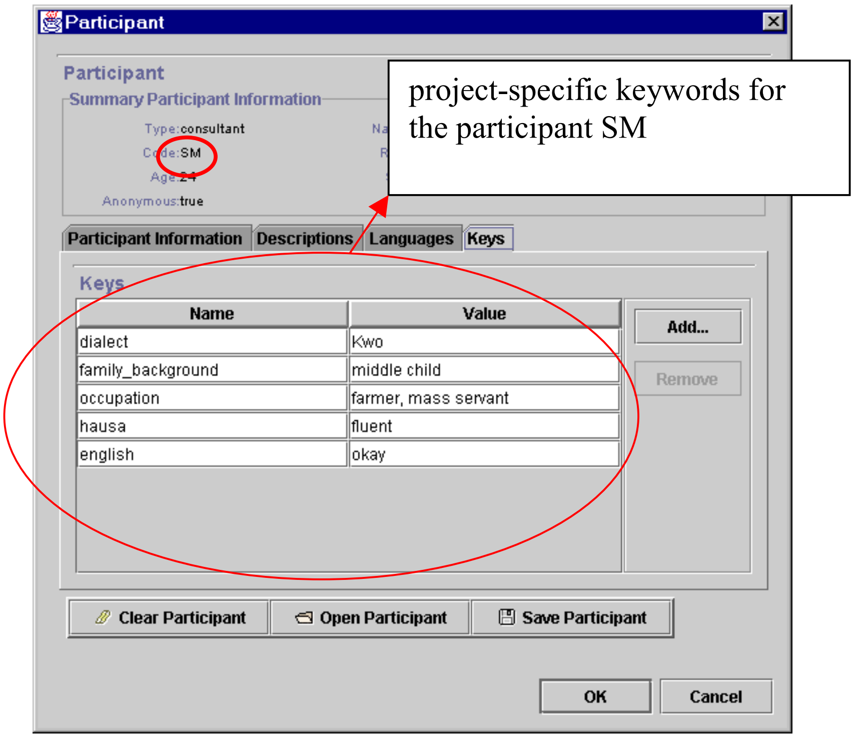
There exist recommendations for the content (which can be selected from a pull-down menu). This list is nearly exhaustive, and you will only occasionally have to type in a different content.

‘closed controlled vocabularies’

The content must be selected from the pull-down menu.

In all three cases, you can choose the item 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 options.

In addition to these standardized fields, the IMDI Editor allows for the possibility to enter additional, project-specific information. Such information is entered into a Keys schema (see “Appendix 2: Recurring schemata” on instructions of how to fill in a Keys schema), e.g.:

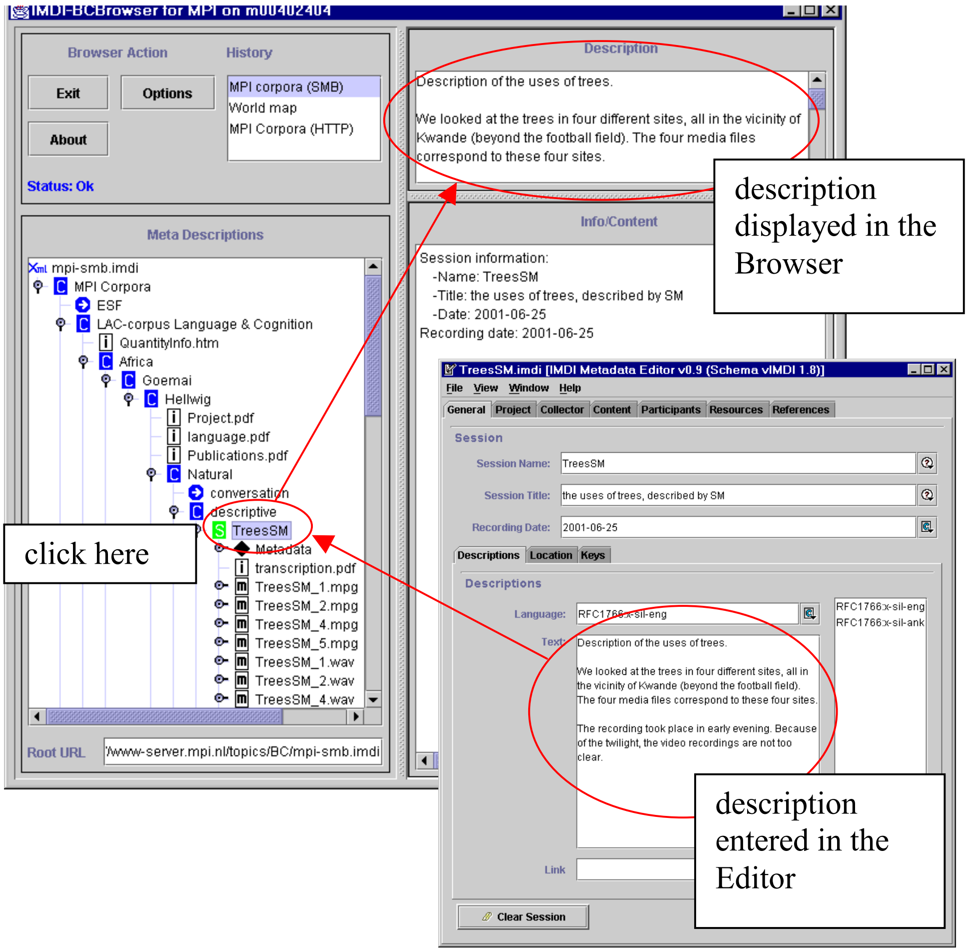


Like all other fields, keyword fields are searchable.

We recommend to enter the following kind of information into a Keys schema: project-specific information that is (a) not taken care of in the standardized fields (provided by the IMDI Editor), and is (b) meant to be searched. If such additional information is irrelevant for your purposes, feel free to leave a Keys schema blank.

In a future version of the IMDI Editor, it will be possible to standardize such project-specific keywords by compiling project-specific controlled vocabulary lists.

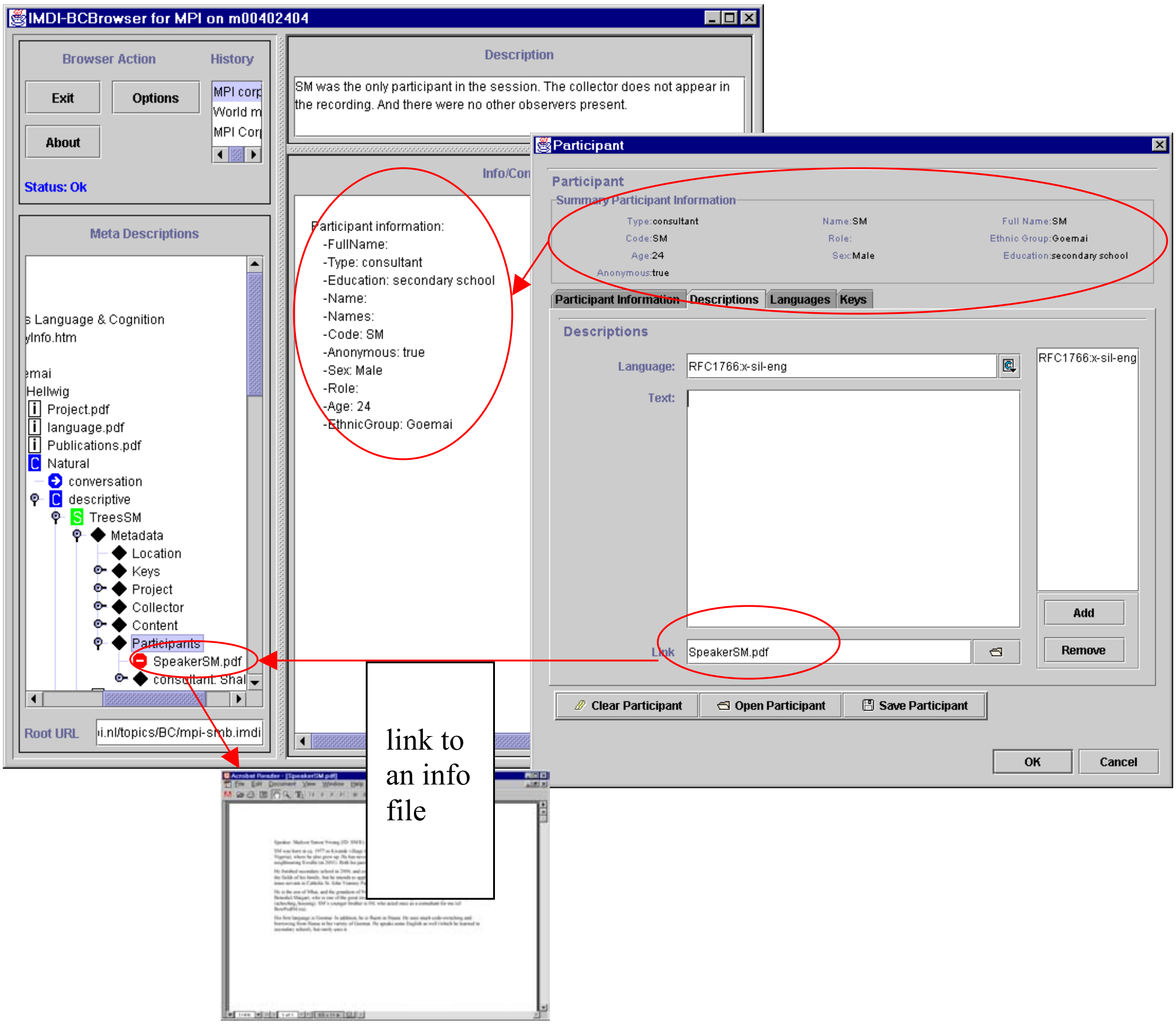
In addition to the information entered into standardized fields and keyword fields, you can enter descriptive information. Descriptive information is entered into a Descriptions schema (see “Appendix 2: Recurring schemata” on instructions of how to fill in a Descriptions schema). A Descriptions schema should contain a prose description that serves as a reminder of the circumstances of data collection. It is displayed whenever you access the corresponding session in the IMDI Browser, e.g.:



! Note: Every main screen of the IMDI Editor contains a Descriptions schema. Recall that the information entered in a schema is only relevant to the main page, e.g., the Descriptions schema on the page Participants should contain only information about the participants (see part III for details).

4.4.3 Info files

In addition to the information that is entered directly into the IMDI file, it is possible to create links to so-called info files. The links are specified in the IMDI Session file, and the info files can be accessed through the IMDI Browser, e.g.:



We suggest that you use info files for storing the following kinds of information:

(1) Any protocols that are connected to the session.

(2) Any digitized photos that are connected to the session (in ***.html** format). Note that it is currently not possible to integrate photos in any other way.

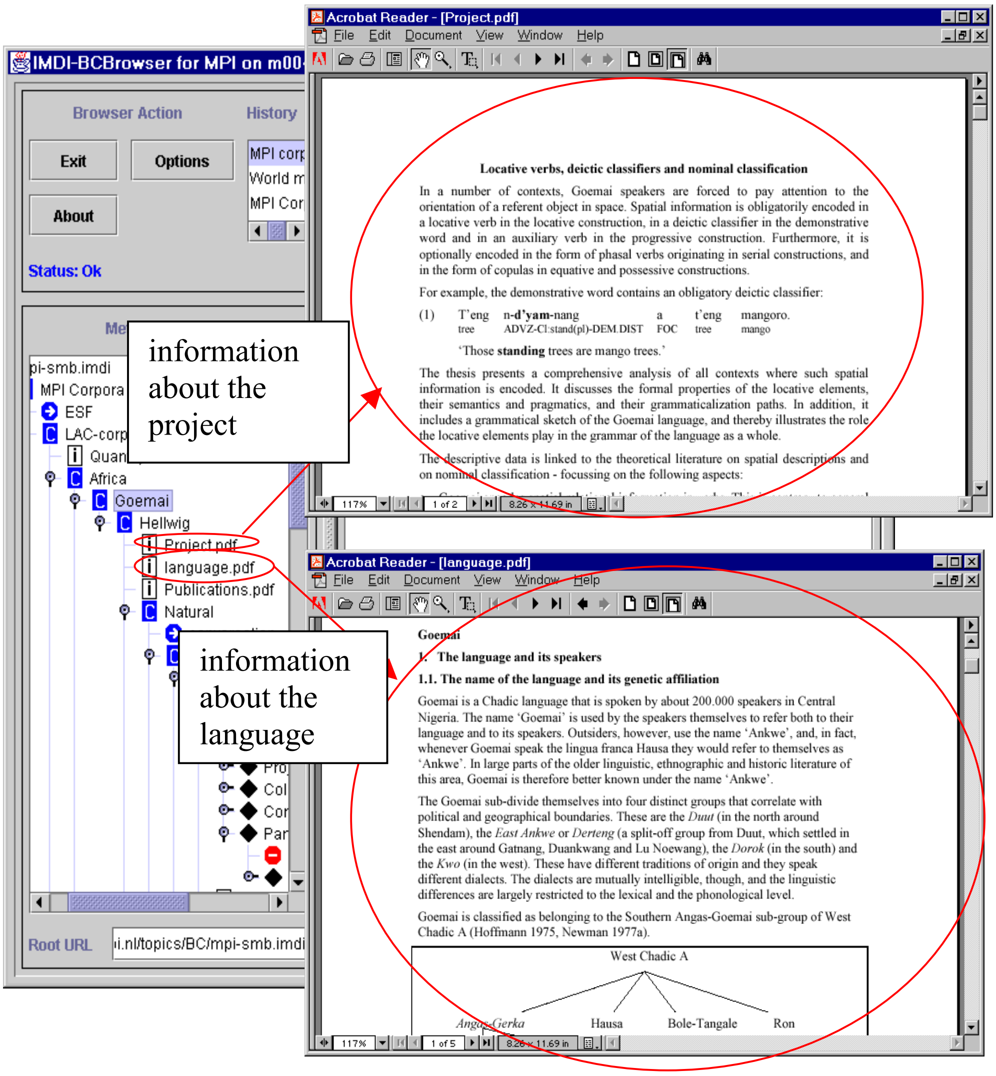
(3) Any kind of sensitive information that is connected to the session. By default, info files are accessible to the outside world, but you can request that the content of an info file should be hidden. In this case, please contact the corpus manager (e.g., corpus.manager@mpi.nl).

! Note: The info file has to be saved as either a ***.txt** file, a ***.pdf** file or an ***.html** file. The IMDI Browser cannot display any other formats. If you need help with converting files into a ***.txt**, ***.pdf** or ***.html** format, please contact the corpus manager (e.g., corpus.manager@mpi.nl).

! Note: If you specify a link to an info file in your IMDI Session file, please specify only the name – do **not** make use of the "open folder" icon to the right of the Link field (unless you want to manage your own corpus, see part IV).

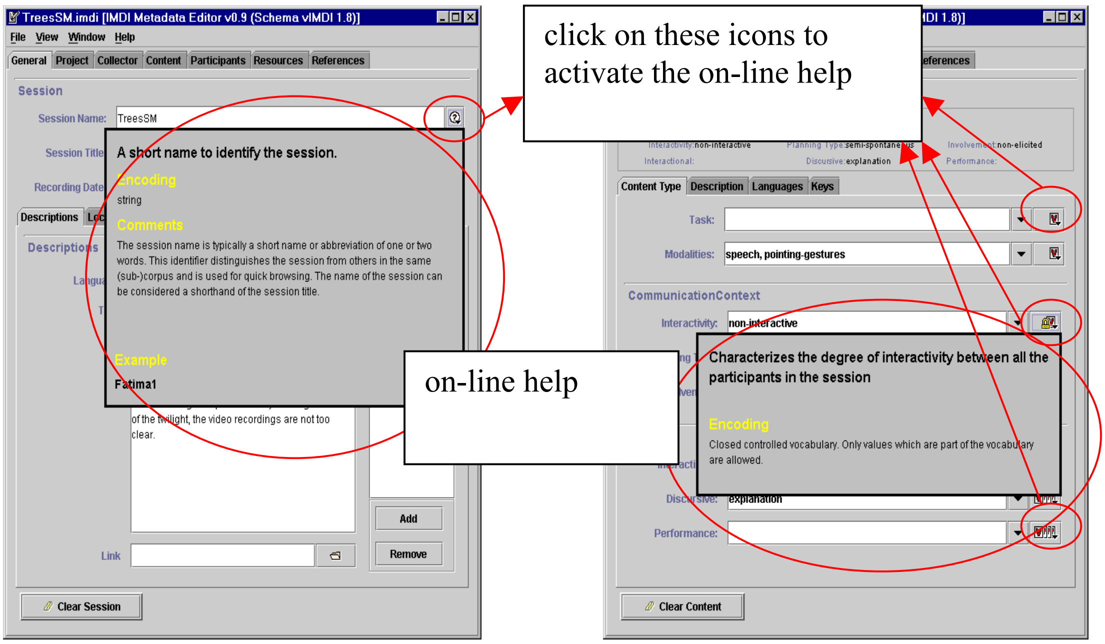
! Note: Please use such info files for session-specific information only. It is, of course, possible to create additional info files that are not relevant to the individual session, but to the corpus as a whole. Such information is very important, but it should not appear under the session node as this will multiply information unnecessarily. We ask you **not** to include links to such info files in the IMDI Session file. Instead, contact the corpus manager (e.g., corpus.manager@mpi.nl), who will advise you about the available possibilities.

In the following screenshot, background information relevant to the whole corpus (about the project and about the language) is stored as ***.pdf** info files under the first node in the corpus hierarchy.



4.4.4 On-line help

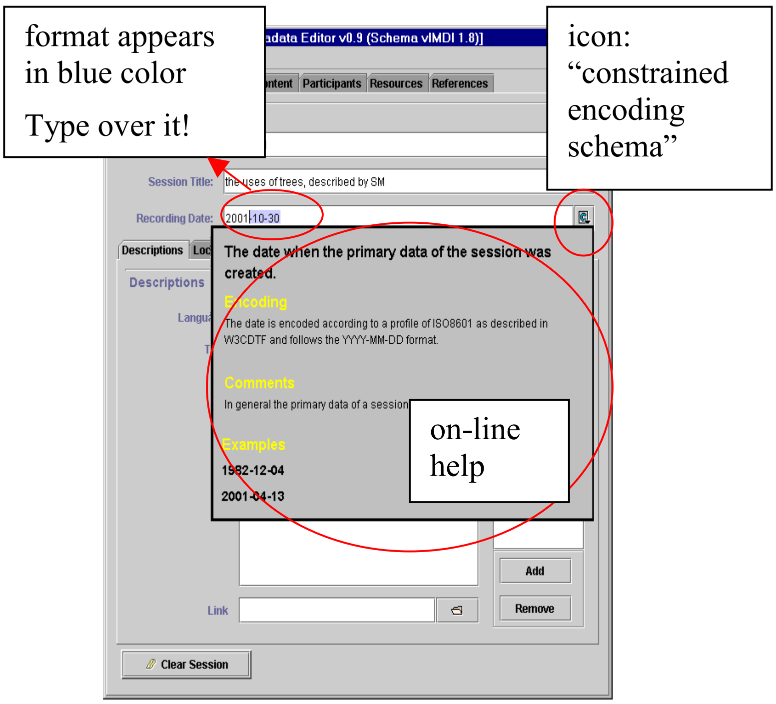
Most fields display an “information” icon to their right. Whenever you click on this icon, a textbox appears that gives information about how to fill in the corresponding field, e.g.:



! Note: The textbox remains open until you click into the corresponding field.

In some cases, the data has to be entered in a standardized format (e.g., date, language (identifier), e-mail address, (time) position). These fields display the icon “constrained encoding schema” to their right. Click on this icon to activate the on-line help that informs you about the format.

As soon as you start typing, the standardized format is displayed in the field (highlighted in blue color). Please type over this format.



5 Preliminaries: Preparing an IMDI Session file

It is useful to make the following preparations before you create an IMDI Session file:

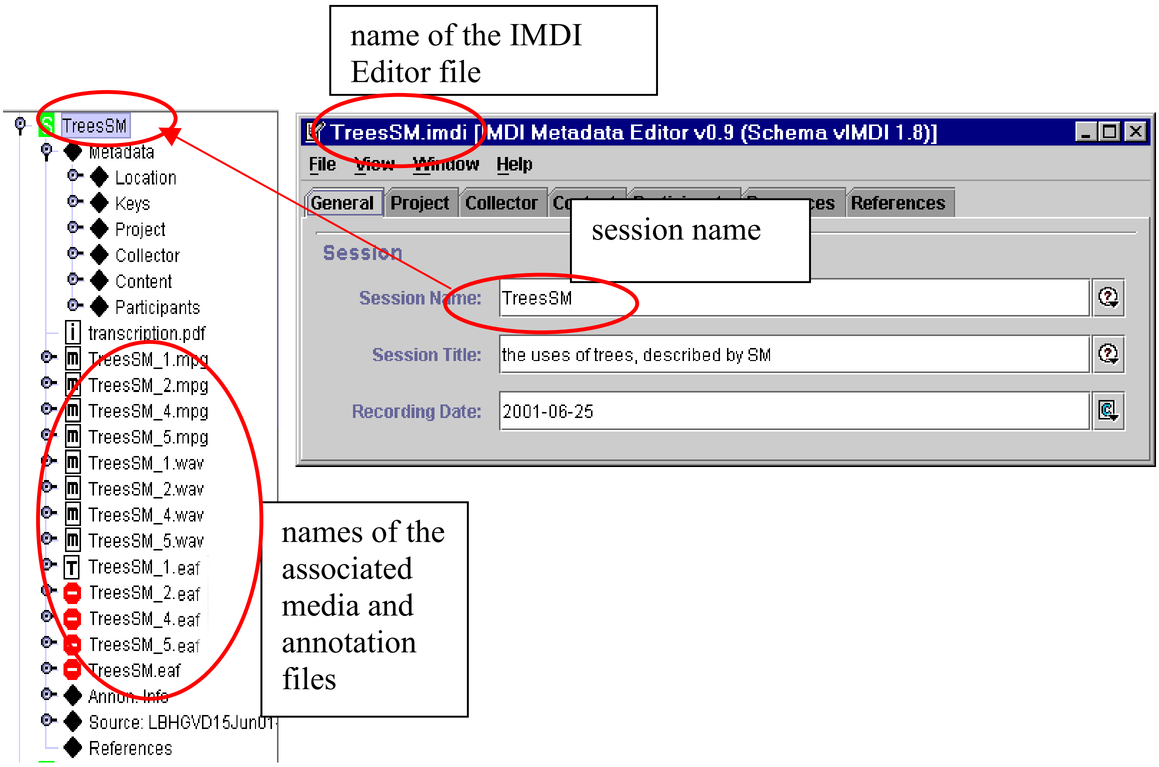
1. Decide on the unit that should be described by the IMDI file.
2. Think about the information that is likely to occur repeatedly.
3. Create master IMDI files and templates for recurring information.

These three strategies will help you to create IMDI files with a minimum of time and effort.

5.1 Decide on the unit that should be described by the IMDI file

The unit to be described is a ‘session’, i.e., a piece of data with a single overall topic (e.g., a folktale, an elicitation session, a matching game), collected at a specific date at a specific location and with a specific set of participants (see section 4.2).

When you have identified the session, assign a name to it. Please note that this name should be used for (a) the IMDI Editor Session file, (b) the corresponding annotation and media file(s), and (c) the field Session Name on the General screen of the Editor, e.g.:



! Note: If necessary, use cardinal numbers to differentiate between different annotation or media files belonging to the same session.

! Note: 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.

5.2 Think about the information that is likely to occur repeatedly

Much of the metadata information remains the same for different sessions. The following types of information are likely to occur repeatedly:

Type of information	**Screen/ Sub-screen(s)**
The location where the data was collected	General/ Location
The larger project within which the data was collected	Project
The person who collected the data, and the person who is responsible for the collection of the data	Collector
The language(s)	Content/ Languages/ Language Participants/ Participant/ Languages/ Language
The participant(s)	Participants/ Participant
The access policy by which you grant/deny access to different parts of your data	Resources/ Media Files/ Media File/ Access Policy Resources/ Annotation Units/ Annotation Unit/ Access Policy Resources/ Sources/ Source/ Access Policy

Table (1): Recurring information

Some of the information illustrated in the table above will not change at all (possibly the larger project), or only within a certain limit (e.g., the participant is likely to come from a set of recurring participants). It thus makes sense to supply this information only once, to save it and to then re-use it for other sessions.

You have two options for entering recurring information:

(1) A master IMDI file: This is used for information that is relevant to all/most sessions (e.g., the project).

(2) A template: This is used for information that recurs in many sessions, but that (a) does not occur in all sessions or that (b) does not occur in the same combinations in all sessions (e.g., different constellations of participants).

The following section shows you how to create and use master files and templates.

5.3 Create master IMDI files and templates

5.3.1 Master IMDI files

Master IMDI files are used for information that is relevant to all/most sessions. To create such a file, do the following:

1. Open the IMDI Editor.
2. Fill in all fields that will recur in all/most sessions.
3. Save the file under a dummy name, e.g., **master.imdi**

Every time you create an IMDI Session file, do the following:

1. Open the IMDI Editor.
2. Open the master IMDI file, e.g., **master.imdi**.
3. In the "File" menu, choose the option "Save As". Specify a name for the new IMDI Session file.

   ! Note: Please choose the session name as the name for the IMDI Session file (see section 5.1 above).

   All information that was contained in the master IMDI file is now also contained in the IMDI Session file. You only need to add the session-specific information.

! Note: You can always delete information from an IMDI file, either manually, or by making use of the option Clear … (located at the bottom of all screens – it automatically deletes all information on that screen). It is thus possible to create master IMDI files even when some part of the information changes occasionally. For example, you might usually collect your data at the same location, and only occasionally at a different location. In this case, it is still useful to enter the usual location into the master IMDI file, and to change it manually for those sessions where you collected the data elsewhere.

5.3.2 Templates

Templates are used for information that recurs in many sessions, but that (a) does not occur in all sessions or that (b) does not occur in the same combinations in all sessions.

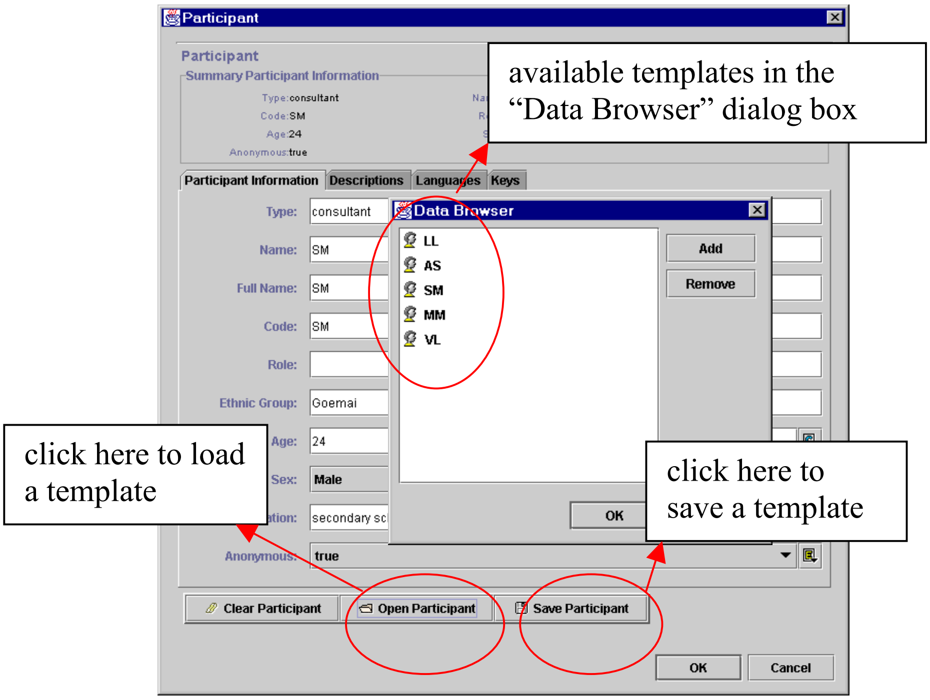
Templates are available for the following (sub-) screens:

- Project
- Collector
- Participant (accessed via the screen Participants)
- (Access) Policy (accessed via the four sub-screens on the screen Resources)

To create a template, do the following:

1. Open the IMDI Editor.
2. Fill in the screen for which you want to create a template.
3. Click on the button "Save …". The "Save" dialog box appears.
4. Browse to the folder where you want to save the template.
5. Save the template under a descriptive name, e.g. **SpeakerSM.imdi**.

! Note: 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.



Every time you want to re-use a template (e.g., whenever the saved participant takes part in a session), do the following:

1. Create/open the IMDI Session file.
2. On the relevant screen, click on the button "Open …". The "Data Browser" dialog box appears.
   (a) When the template is displayed in the dialog box, click on it to highlight it, and then click "OK".
   (b) When the template is not displayed, click the "Add" button, browse to the folder that contains the template and open it. The template will be displayed in the "Data Browser" dialog box. Click on it to highlight it, and then click "OK".

   The IMDI Editor automatically fills in all fields.

! Note: Sometimes the “Data Browser” dialog box erroneously displays a template that has already been deleted or moved to a different folder. If you attempt to load such a template, nothing will happen. Please remove it manually by using the button “Remove”.

! Note: When you load a template, it may be necessary to change some of the information. For example, when you load the template **speakerXY.imdi**, it may be that his/her role in the current session differs (e.g., she might occur in the role of ‘mother’ in one session, but in the role of ‘wife’ in another). In this case, change the information manually.

! Note: You can create templates in retrospect. For example, a participant who only appeared once in the beginning may occur more often later on. In this case, you can open the old IMDI Session file that contains the information about the participant, and save him/her as a template.

! Note: You can always change the information contained in a template. Do the following:

1. Click on “Open …” to load the template.
2. Change it manually.
3. Save it (using the “Save …” button) under the same name as the old template. The old information will be overwritten.

5.3.3 Session-specific information

The two strategies outlined in sections 5.3.1 and 5.3.2 above will save you much time and effort when creating IMDI files. In the end, only little session-specific information has to be entered manually. Session-specific information is likely to be of the following kind:

- Information on the General screen, which contains brief session and IMDI file information. Some information may be included in a master IMDI file, though (e.g., information on the location).
- Information on the Content and the References screens. Since both screens are concerned with the session-specific content, it is only to a very limited extent possible to create master files or templates (e.g., the main language could be included in a master IMDI file).
- Information on the Participants screen. This information concerns the session-specific constellation of participants. Information on individual participants, by contrast, can be loaded from templates (with some manual changes relating to the type, role and age of the participant).
- Information on the Resources screen. This information needs to be entered manually for each session. Although many fields have to be filled in, much of this process is semi-automatic (in form of pull-down menus and lists of controlled vocabularies). The information is needed for (a) linking the different files (media,

annotation, and IMDI) to each other and (b) digitizing the data from the original tape.

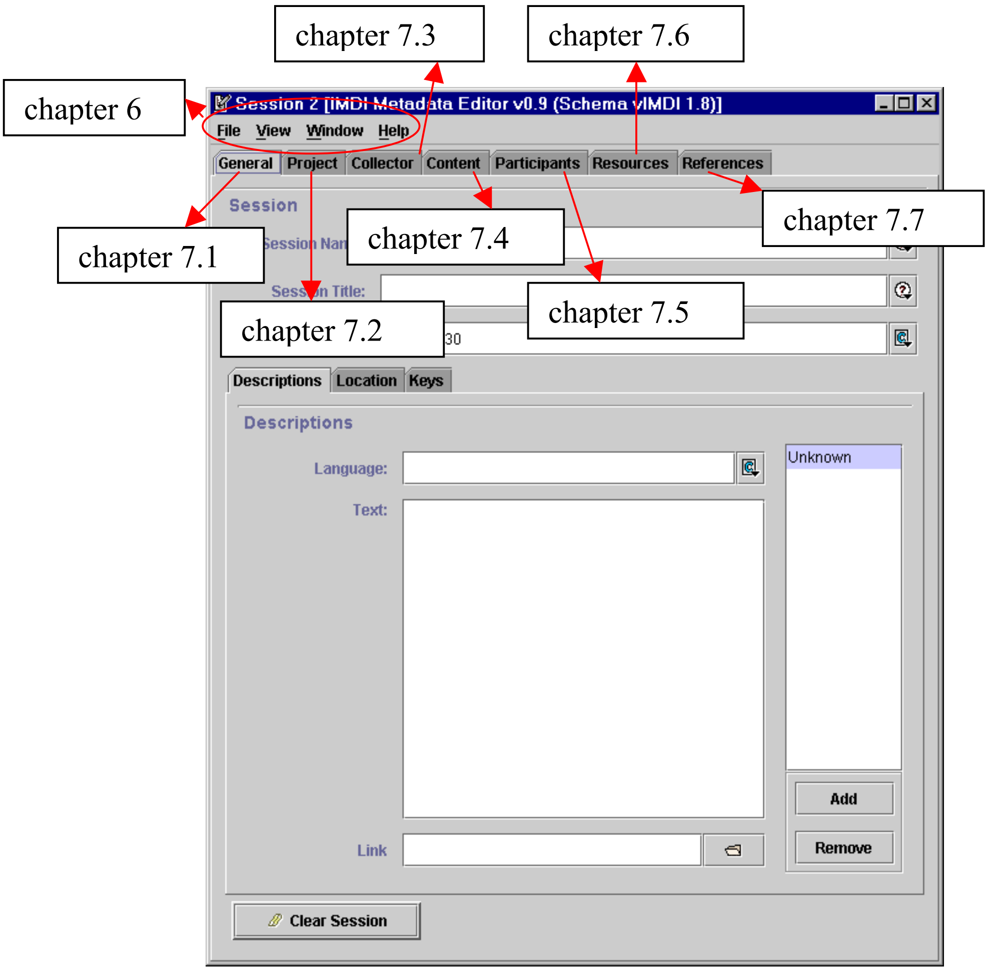
Please keep in mind that elaborate descriptions can be entered into separate info files, e.g., background information about the participants, the language, the project, the transcription conventions, etc. Such information need not – and should not – be repeated in every session (see section 4.4.3 for further information).

PART III: THE USER'S GUIDE TO THE IMDI SESSION EDITOR

This part of the manual guides you through using the IMDI Session Editor (see part IV for the IMDI Corpus Editor). It explains the purpose of each screen, schema and field, and it illustrates the kind of information to be entered.

The IMDI Editor creates an IMDI Session file (***.imdi**), or IMDI file for short.

When you start the IMDI Editor, the following window appears:



! Note: After you have worked with the IMDI Editor for some time, it will become very slow, and it is possible that files will not be saved anymore. Currently, there is no way around this problem other than to (a) close all unused programs while working with the Editor and (b) periodically close the Editor and restart it again.

6 Menu Items

The menu items are located at the top of the IMDI Editor window. The following items are available:

"File": use this menu to create, open, save, or exit an IMDI file.

"View": use this menu to set preferences.

"Window": use this menu to jump between open IMDI files.

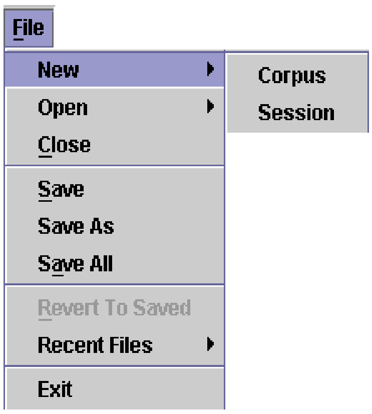
"Options":

"Help": use this menu to read about the copyright and version information.

Some of the menu items are displayed with an underlined letter. Such items can be accessed through the shortcut key ALT plus the underlined letter, e.g., the "File" menu can be accessed through pressing ALT+F.

"File" Menu

The "File" menu is used to create, open, save, or exit an IMDI file. It includes the following options:



"New"

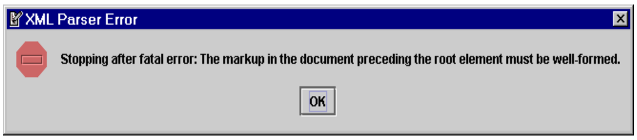
Click on "New" and then on "Corpus" to open an IMDI Corpus Editor window (see part IV).

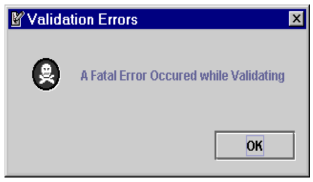
Click on "New" and then on "Session" to open an IMDI Session Editor window.

“Open”

Click on “Open” and then on “Corpus” or “Session” to open an existing IMDI Corpus or Session file. Browse to the folder that contains the IMDI file and double-click on it to open it. A second IMDI Editor window containing this file appears.

> **!** Note: You can only open files of the IMDI Editor format. If you try to open a file in a different format, the following two warnings appear:

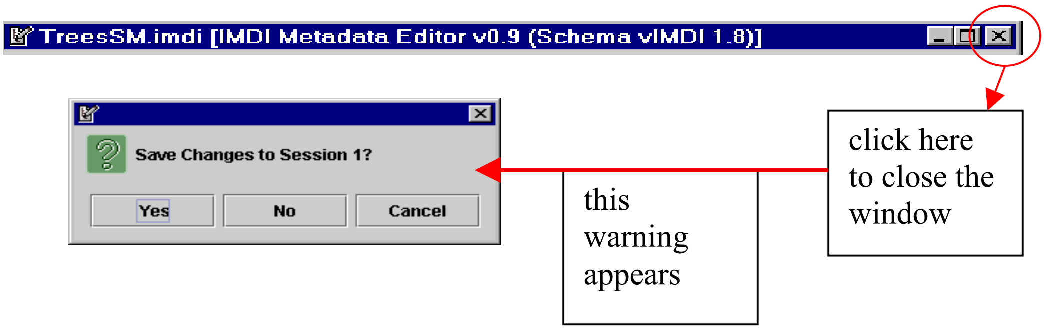




“Close”

Click on “Close” to close he current IMDI file. Note that all other currently open IMDI files will remain open.

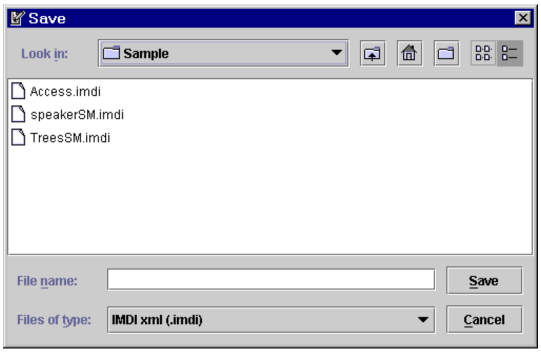
> **!** Note: You could also close an IMDI Editor window through clicking on the cross icon in the upper right corner. In this case, the following warning is displayed:



If you now click “Cancel” to return to the IMDI Editor window, the Editor will close without a warning, i.e., you will loose all unsaved changes. Please do not make use of this option.

“Save”

Click on “Save” to save the current IMDI file. When you save a file for the first time, the “Save” window appears, e.g.:



Browse to the folder where you want to save the IMDI file. Assign a name to the file, and click on “Save”.

> **!** Note: Please remember that the name of the IMDI file should be the same as the name for the corresponding annotation and media file(s). And it should be the name that is entered in the field Name on the General screen of the Editor (see section 4.2).

> **!** Note: 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.

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

“Save As”

Click on “Save As” to save the current IMDI file under a different file name. You are then asked to assign a new name to that file (see the explanation under “Save” above).

> **!** Note: This option is especially relevant when you work with master IMDI files (see section 5.3.1).

“Save All”

Click on “Save All” to save all open IMDI files.

“Revert to Saved”

Click on “Revert to Saved” to return to the last saved version of the current IMDI file.

> **!** Note: This option will automatically delete all changes that you have made since last saving the file. The following warning will appear:

“Recent Files”

Go to “Recent Files” to access an IMDI file that you have recently worked with. A submenu with the names of recently save files appears. Click on the file that you want to open. A second IMDI Editor window containing this file appears.

“Exit”

Click on “Exit” to exit the IMDI Editor window. All open IMDI files will be automatically closed.

“View” Menu

The “View” menu is used to set preferences. Click on “View” / “Preferences …” to access the “Preferences” dialog window.

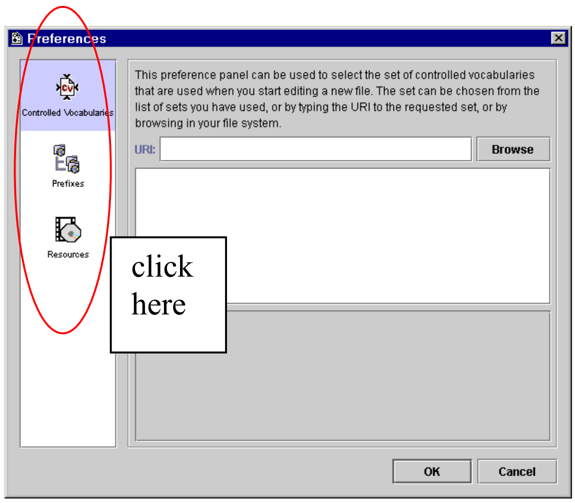
In the “Preferences” dialog window, the following three options are available:

(1) “Controlled Vocabularies”

(2) “Prefixes”

(3) “Resources”

Click on a heading in the left panel to activate the corresponding option.



> **!** Note: Generally, you do **not** need the options “Prefixes” and “Resources” because the relevant information is specified by the corpus manager (e.g.,

corpus.manager@mpi.nl). If you manage your own corpus, you may want to make use of these options, though (see part IV for details).

"Controlled Vocabularies"

Please ignore this button for the moment.

"Prefixes"

See part IV for details.

"Resources"

See part IV for details.

"Window" Menu

The "Window" menu is used for jumping between open IMDI files. Go to "Window" to see the names of all files that are currently open. Click on the name of the file that you want to jump to. This file will become the active file.

"Options" Menu

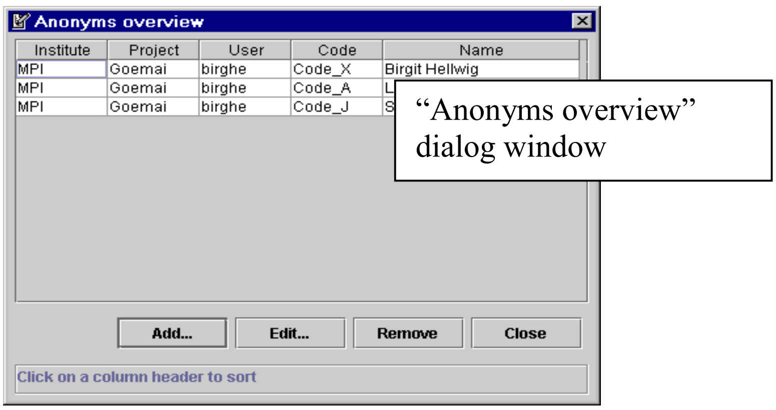
"Anonyms …"

The option "Anonyms …" can be used if you have chosen to make the name of a participant anonymous, i.e., if you have chosen to enter a code (instead of the full name) into the field Full Name on the screen Participant (see section 7.5.1). If you have done this, the option "Anonyms …" allows you to create a file that maps this code to a full name. You (and everybody whom you have granted access to this file) will be able to see the full name, while everybody else will see only the code.

To create such a file, do the following:

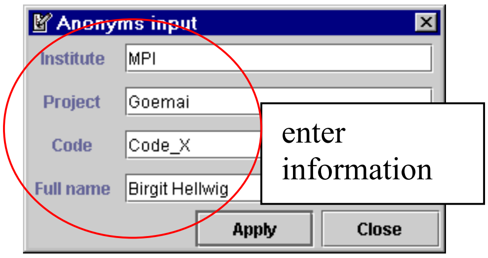
1. Click on "Options".
2. Click on "Anonyms …".

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



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 participant and his/her full name.



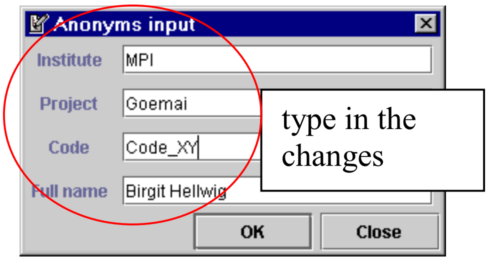
To save the mapping, click first on "Apply" and then on "Close".

The new mapping is 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 screen Participant – if the codes are identical, the IMDI Editor will display the full name in the field Code on the Participant screen – not the intended code (see section 7.5.1).

! 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.



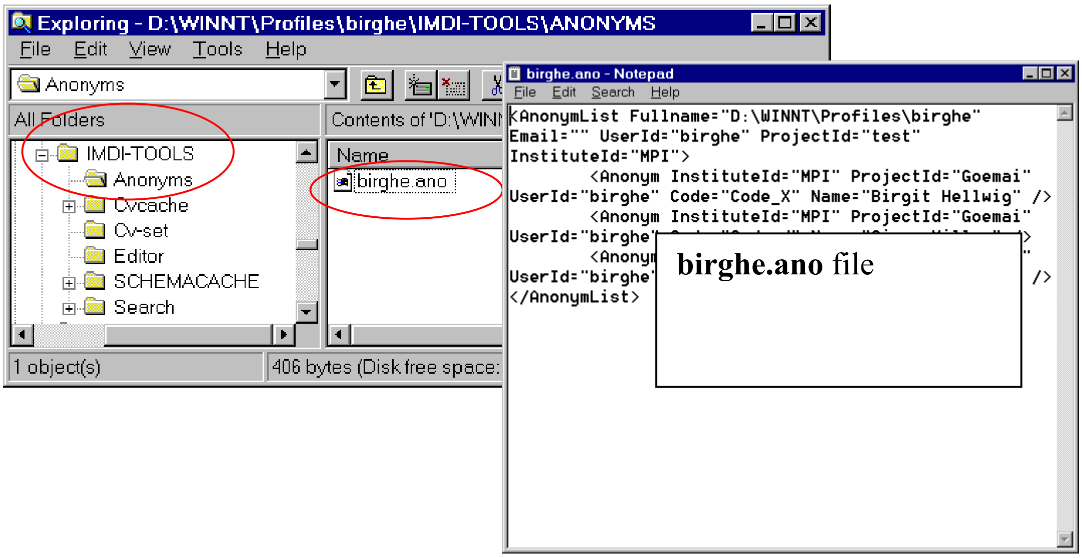
Click on "OK" to save the changes.

Or click on "Close" to exit the window without saving 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.

   The 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, give them a copy of the file **user-name.ano**. They should copy it into their folder user-directory \ IMDI-TOOLS \ Anonyms.

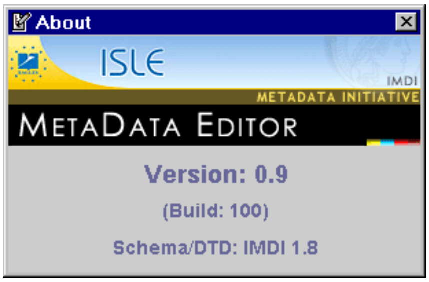
   ! Note: If you are in doubt where the folder user-directory \ IMDI-TOOLS \ Anonyms is located on your computer, please contact your corpus manager (e.g., corpus.manager@mpi.nl).

“Vocabs …”

Please ignore this option for the moment.

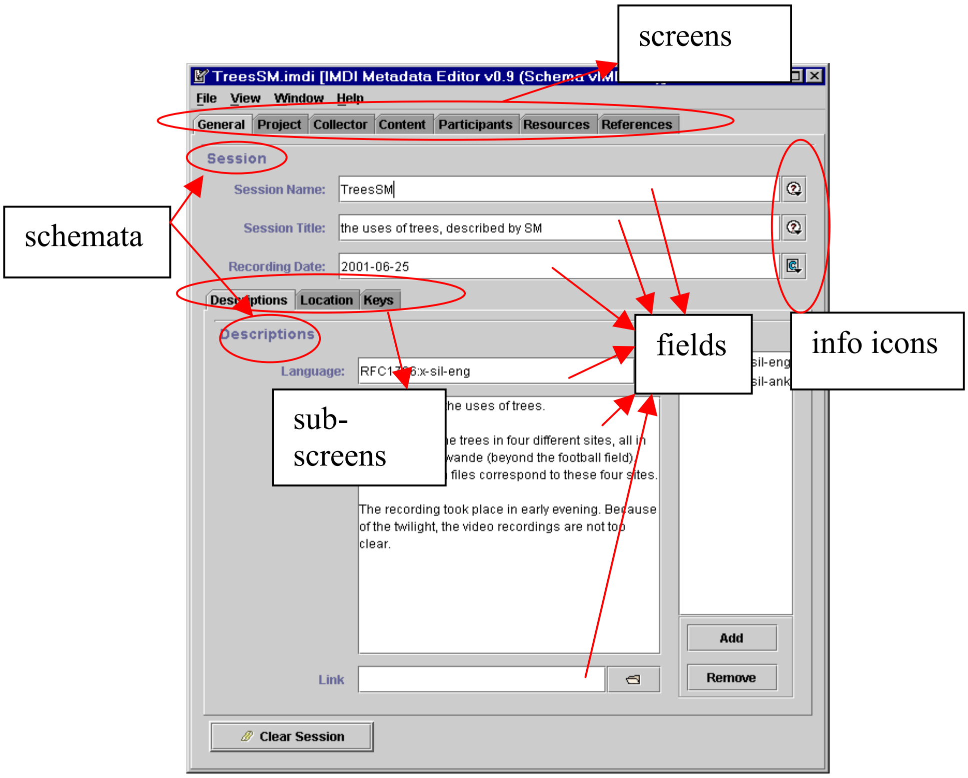
“Help” Menu

The “Help” menu displays the copyright and version information. Go to “Help”, and then click on “About …” to view this information:



7 Screens

The IMDI Editor contains a number of screens, schemata and fields (see section 4.4), e.g.:



There are two options for navigating through the IMDI Editor window:

(1) You can use the mouse.

- Click with the mouse button somewhere in a field to position the cursor in this field.
- Click with the mouse button on a (sub-) screen header to activate the corresponding (sub-) screen.
- Click with the mouse button on any info icon to display its corresponding textbox.

(2) You can use the following shortcut keys.

- Press TAB to move to the next field, info icon or sub-screen header.
- Press SHIFT+TAB to move to the previous field, info icon or (sub-) screen header.
- When an info icon is highlighted, press SPACE to display its corresponding textbox.
- Press the left and right arrow key to move from screen to screen header, or from sub-screen to sub-screen header.

The following sections give instructions and examples of how to fill in the various screens, schemata and fields. In addition, it is shown how the information entered into the IMDI Editor will be displayed in the IMDI Browser.

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

(1) You do not have to fill in every field.

The IMDI Editor was created to be as comprehensive as possible – as a result, it includes a lot of features that may be irrelevant for your individual purposes. In this case, feel free to leave fields blank.

Keep also the following in mind: obviously, it would be nice to have elaborate descriptions and extensive keyword lists available for each session. But you have to weigh this advantage against the time and effort it will take you to enter the relevant information. Please take the following considerations into account:

- What kind of information would you want to search for? Make sure that this information is entered into standardized fields or Key schemata.
- How much descriptive information do you need in order to remind yourself about the circumstances of the session? This will determine how extensive the descriptions in a Descriptions schema should be.

(2) All information entered on a screen is only relevant to that screen (see section 4.4.1).

This point is especially important for a recurring schema 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. So, when you are asked to enter another description, you can safely assume that a different type of information is required this time.

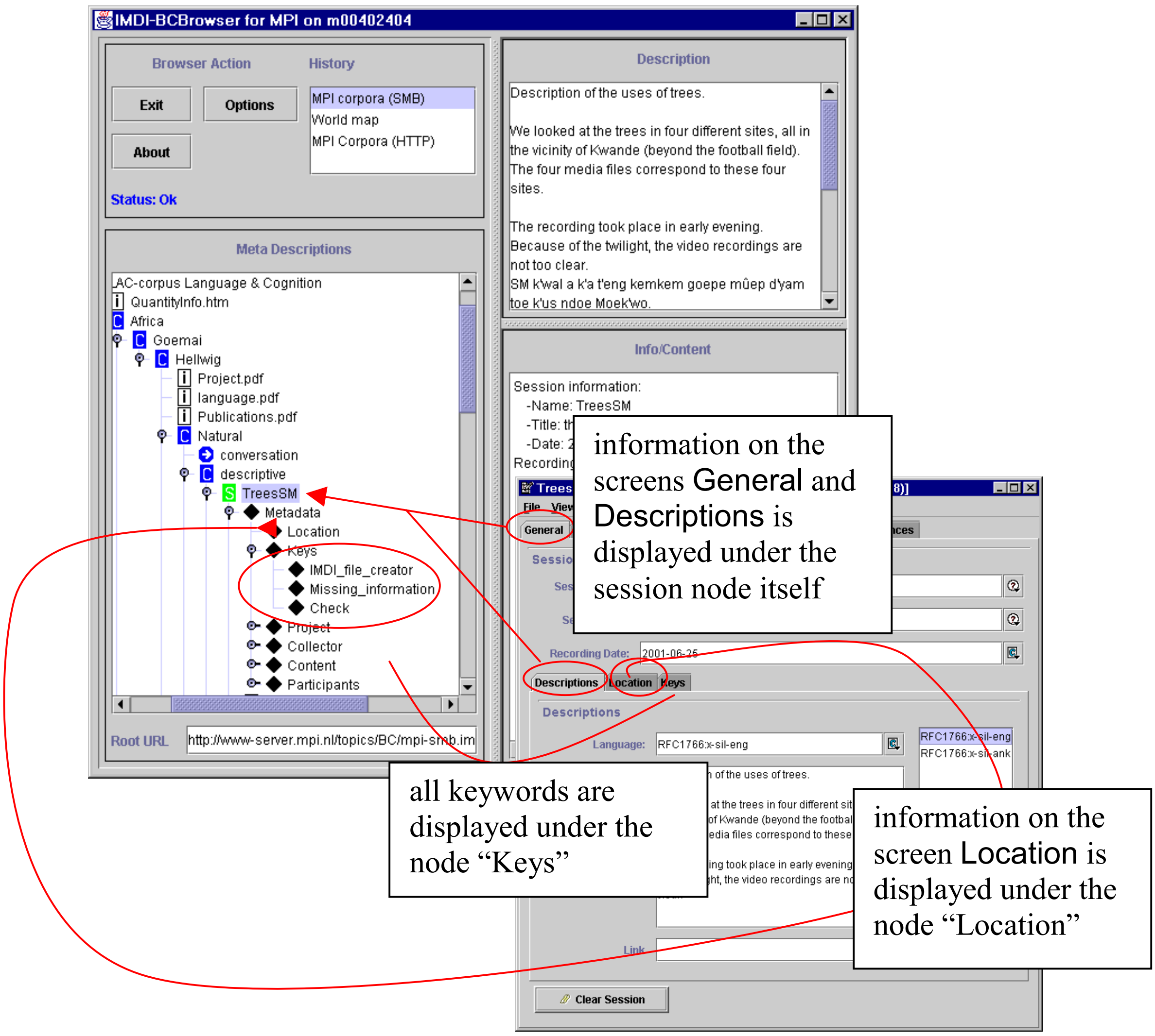
(3) Remember to make use of master IMDI files and templates to enter all recurring information (see section 5.3).

7.1 General

This screen contains general information about (a) the session (name, date, location) and (b) the IMDI file. The following schemata and sub-screens are available:

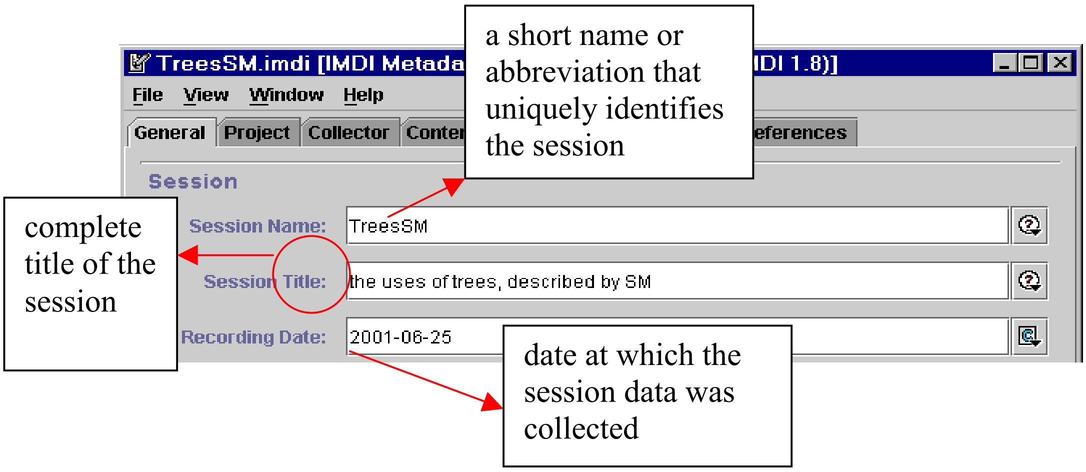
Session schema	Name, title and recording date of the session.
Descriptions sub-screen	Description of the circumstances under which the session data was collected.
Location sub-screen	Information about the location at which the session data was collected.
Keys sub-screen	Keywords that are relevant to (a) the collection of the data or (b) the creation of the IMDI file.

The information that you enter on these four screens is displayed in the IMDI Browser as follows:



7.1.1 General: Session

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



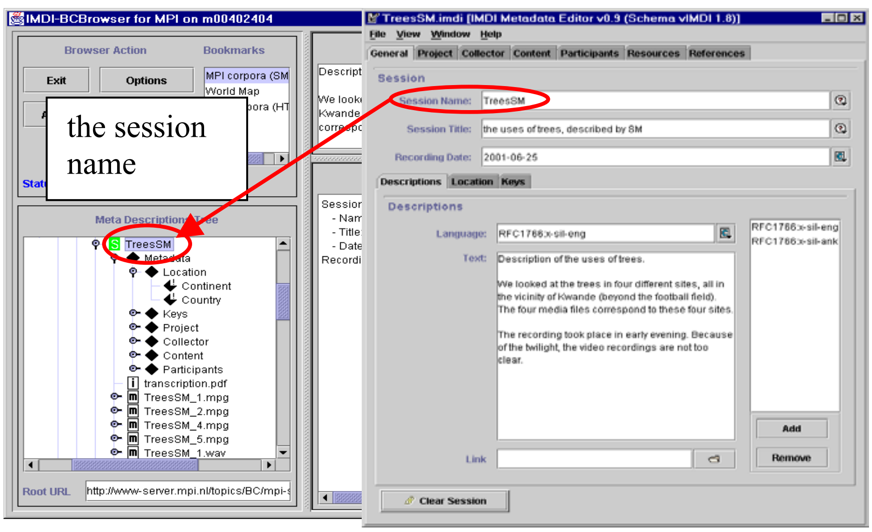
Name

A short name or abbreviation that uniquely identifies the session. It should be identical to the name of the IMDI file, the annotation file(s) and the media file(s) (see section 5.1).

! Note: This field is obligatory.

! Note: The name has to be Unix compatible: do not use names longer than 14 characters, do not use non-letter or non-number characters (except for the underscore: _), and do not use blank spaces.

! Note: The Name is displayed as the session name in the IMDI Browser. It should therefore be descriptive enough to remind you about the session's content.



Title

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

Recording Date

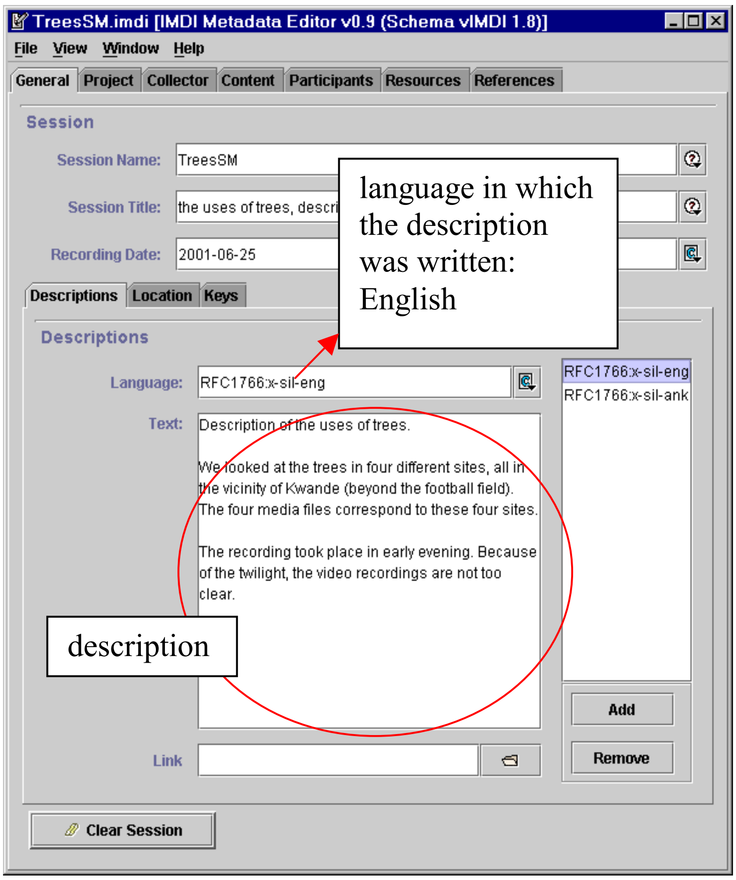
The date at which the session data was collected. Once you start typing a date, the IMDI Editor will automatically display the format (highlighted in blue color). Type over this format to enter the date.

> ! Note: Please enter the date in the following format: **YYYY-MM-DD**, e.g. 2000-12-30.

> ! Note: You have to enter a complete date. If you do not know the exact date, please enter an approximate date. You can then make use of the Descriptions schema to remind you that it is only an approximation.

7.1.2 Descriptions

A description of the circumstances under which the session data was collected. It should **not** contain an elaborate description of the content (reserve this information for the Content screen), but it can contain a brief reminder of the content, e.g.:

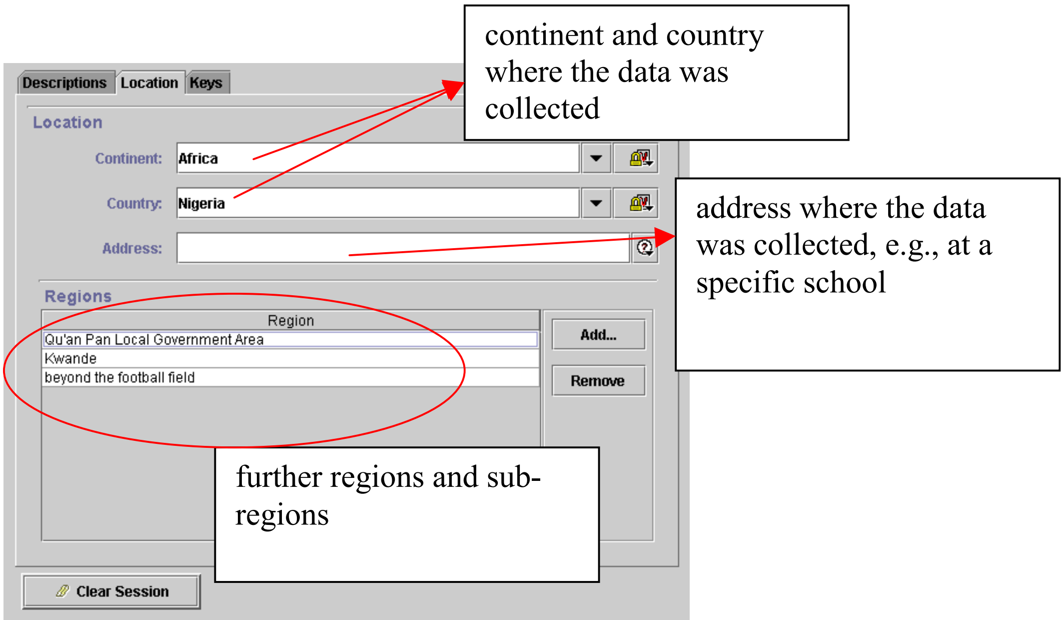


! Note: This Descriptions schema is the first schema that is displayed when you click on the session node in the IMDI Browser. It should therefore contain all the information that you need in order to quickly remind you about this session.

See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.1.3 Location

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



Continent

The continent where the session data was collected. Choose the continent from the pull-down menu, or start typing (in this case, the pull-down menu will automatically open to display the available options).

Country

The country where the session data was collected. Choose the country from the pull-down menu, or start typing (in this case, the pull-down menu will automatically open to display the available options).

Address

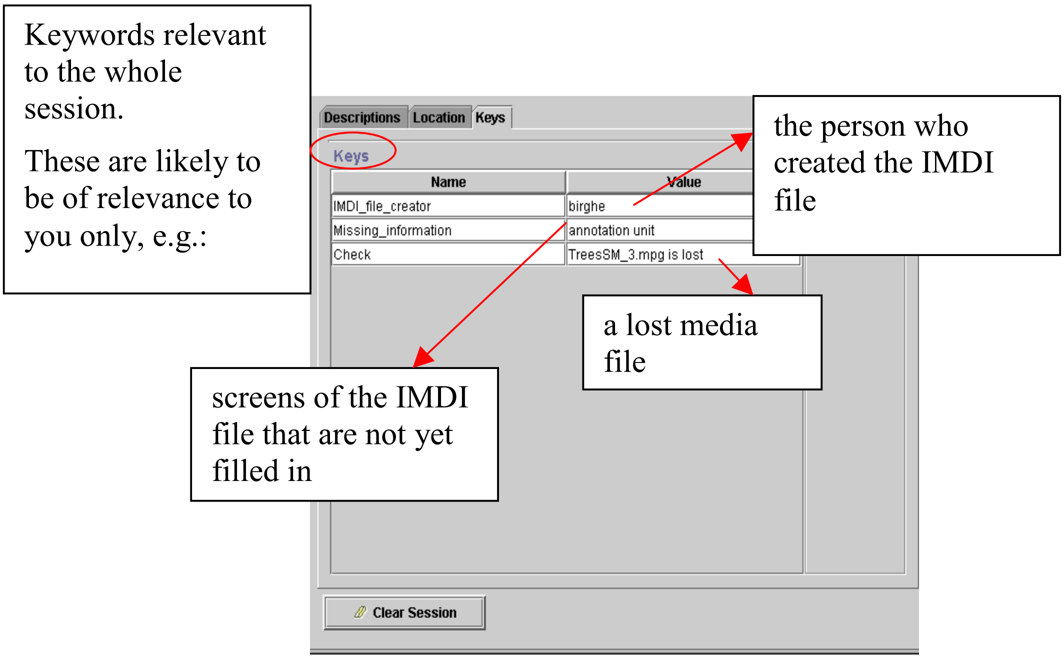
The address where the session data was collected (e.g., at a specific school).

Region

The region (province, town, suburb, etc.) where the session data was collected. Use the “Add” button to add another region, and use the “Remove” button to remove the highlighted region.

7.1.4 Keys

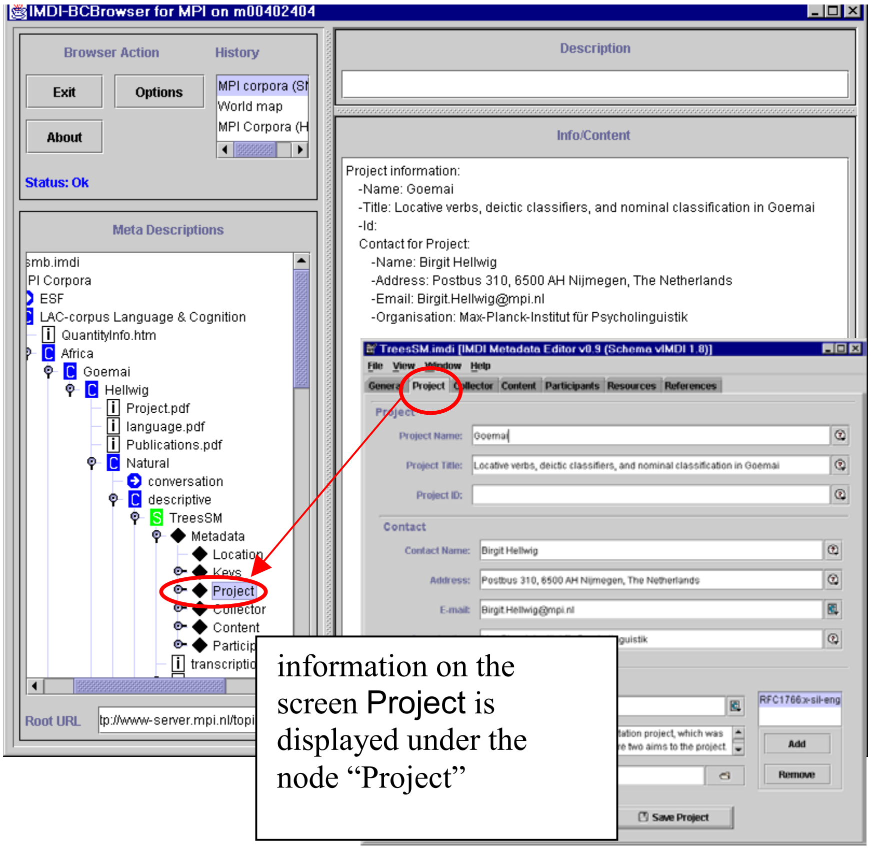
This sub-screen contains keywords that are relevant to (a) the collection of the data or (b) the creation of the IMDI file. For example:



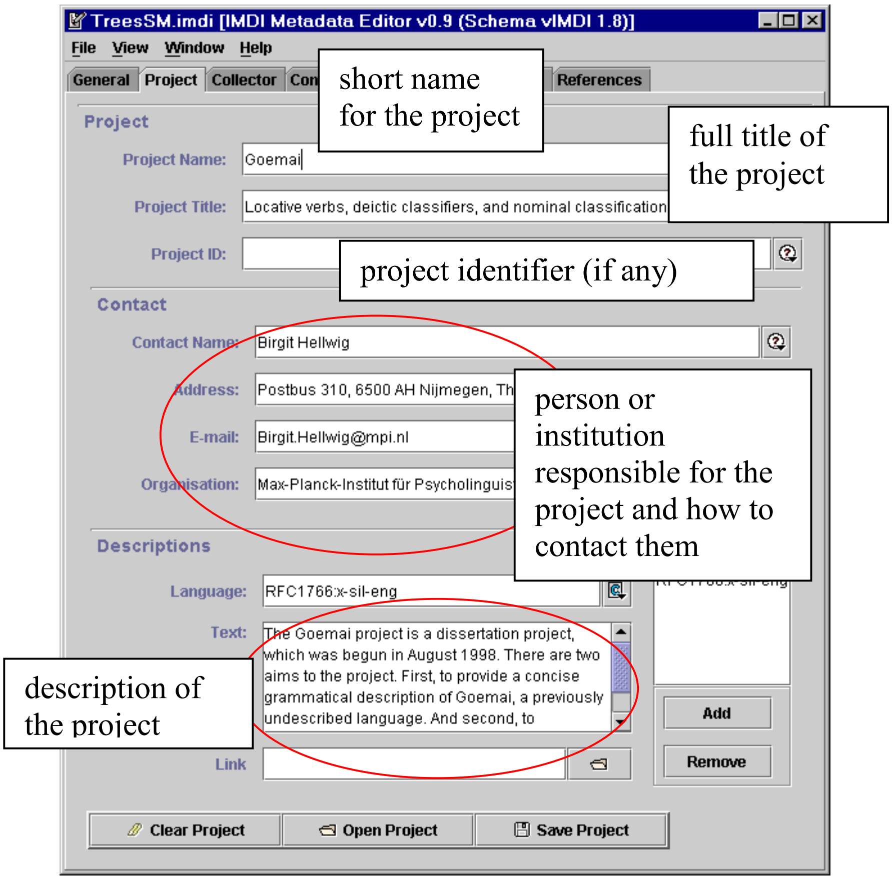
See “Appendix 2: Recurring schemata” for instructions on how to fill in a Keys schema.

7.2 Project

This screen contains information about the project within which the session data was collected. The information that you enter on this screen is displayed in the IMDI Browser as follows:



Enter the information into the IMDI Editor as follows:



Project

- **Name**
  A short name or abbreviation that uniquely identifies the project.

- **Project Title**
  The full title of the project.

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

Contact

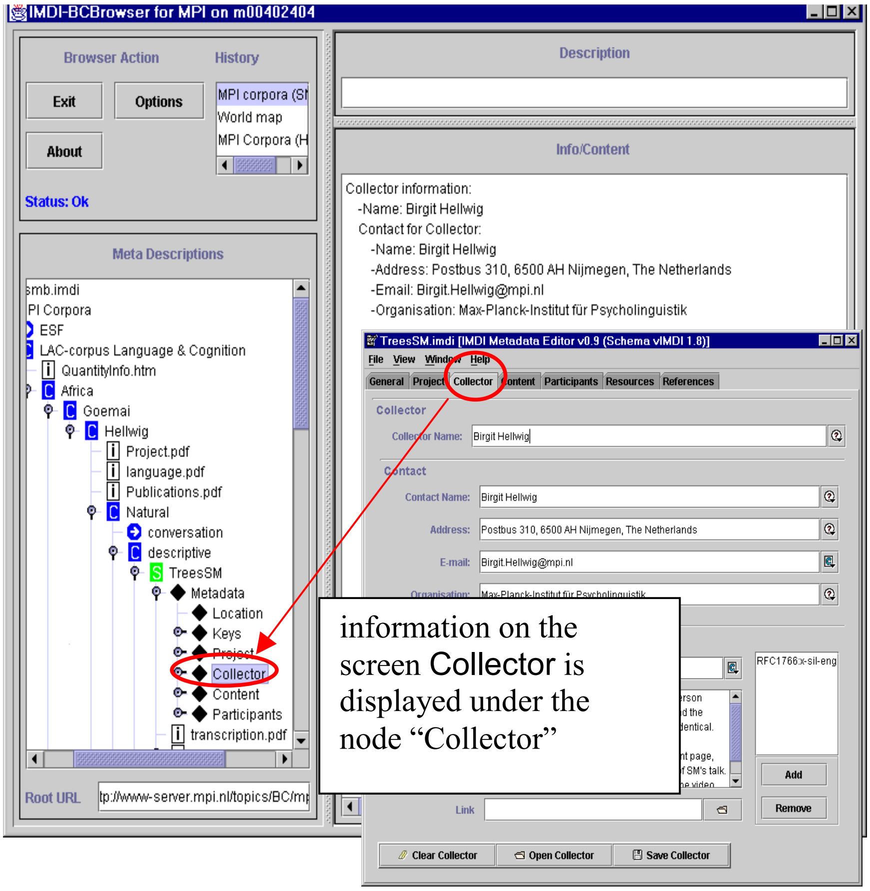
The contact information about the person or institution responsible for the project (the name, the address, the e-mail address and the organization he/she belongs to).

Descriptions

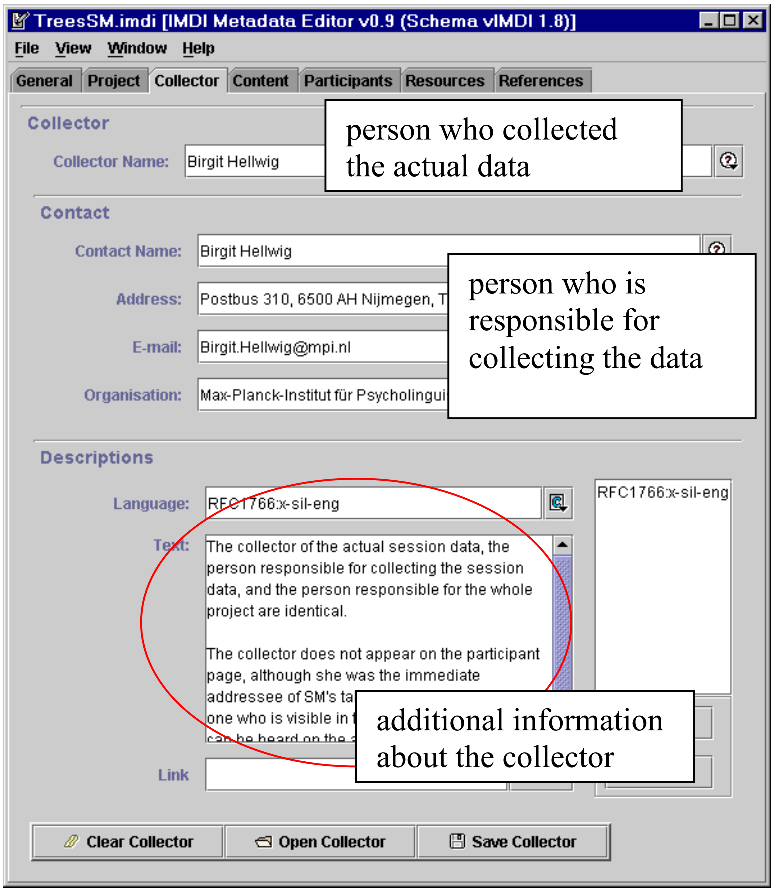
A description of the scope and goals of the project. See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.3 Collector

This screen contains information about the person who collected the session data and the person who is responsible for collecting the session data. Note that this may be the same person. The information that you enter on this screen is displayed in the IMDI Browser as follows:



Enter the information into the IMDI Editor as follows:



Collector Name

The name (or ID code) of the person who collected the session data.

Contact

The contact information about the person responsible for collecting the session data (the name, the address, the e-mail address and the organization he/she belongs to).

Descriptions

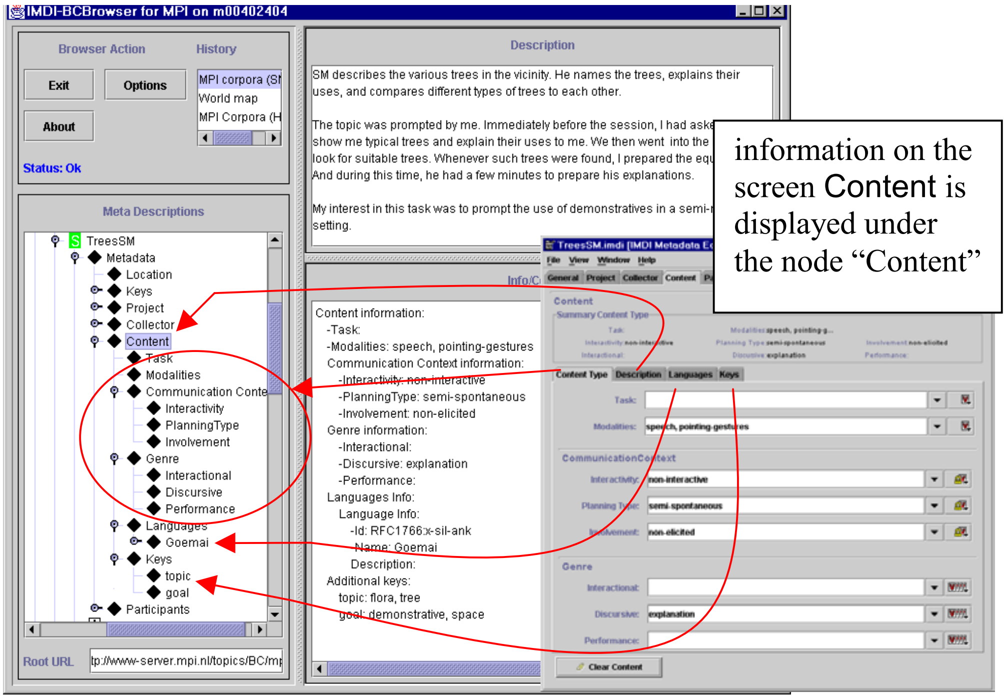
Additional information about the person who collected the session data and the person who is responsible for collecting the session data. See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.4 Content

This screen contains information about the content of the session. The following sub-screens are available:

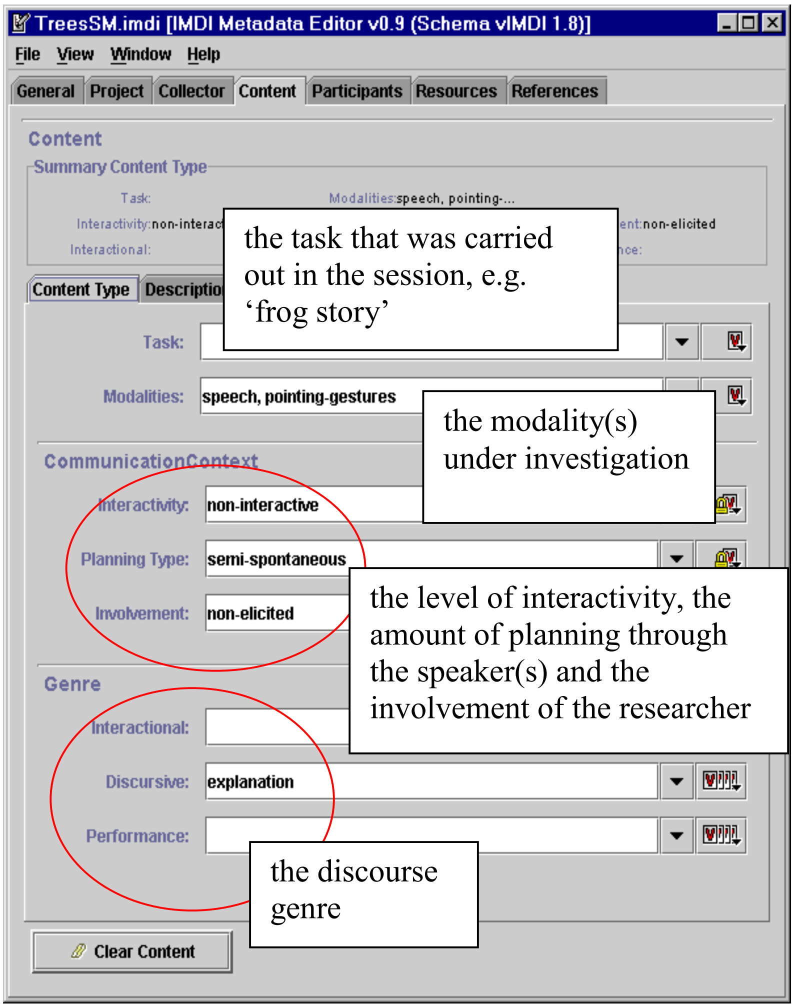
Content Type sub-screen	Information about the task, the modalities, the communication context, and the genre of the session.
Description sub-screen	Description of the content of the session.
Languages sub-screen	Information about the language(s) used in the session.
Keys sub-screen	Keywords that are relevant to the content of the session.

The information that you enter on these four screens is displayed in the IMDI Browser as follows:



7.4.1 Content Type

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



The information that is entered into these fields is standardized to a certain extent (see "Appendix 3: Lists of recommended vocabularies"). Although there is room for variation, we recommend that you keep as closely as possible to the standardized lists as this will facilitate the search process.

Pull-down menus display the suggested vocabularies. Either choose an item from the pull-down menu, or start typing (in this case the pull-down menu will automatically open to display the available options).

You can enter more than one item in each field. In this case, please separate the items with a comma.

Task

The name of the (experimental) task that was carried out. Choose one of the following items or type in another item:

"info-kiosk"
"wizard-of-oz"
"travel-planning"
"room reservation"
"frog story"

Modalities

The modalities under investigation. Choose one of the following items or type in another item:

"speech"
"writing"
"gestures"
"pointing-gestures"
"signs"
"eye-gaze"
"facial-expressions"
"emotional-state"
"haptic"

CommunicationContext

Information about the communication context, i.e., levels of participant interaction, the degree of planning through the speaker(s) and the involvement of the researcher. In all cases, an item from the pull-down menu must be chosen – other items will not be accepted. Enter the following information:

- Interactivity

  The level of participant interaction. Choose from the following items:

name	comments
"interactive"	a verbal 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 • procedural texts
"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. Choose from the following items:

name	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. Choose from the following items:

name	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) 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)

Genre

Information about the genre of the session. Again, wherever possible, the genre should be chosen from the list of recommended vocabularies. Enter the following information:

- **Interactional**

  If the session contains an interactional genre, choose one of the following items, or type in another item:

name	comments
"conversation"	
"verbal-contest"	this also includes debates.
"interview"	
"meeting"	
"riddles"	
"consultation"	this does not refer to an interview with the investigator, but rather to, e.g. a visit to a shaman
"greetings"	
"leavetakings"	
"humor"	

- **Discursive**

  If the session contains a discursive genre, choose one of the following items, or type in another item:

name	comments
"procedure"	a directive description of the procedures involved in the preparation or production of something, e.g., 'how to make tortillas'
"explanation"	practical, theoretical, or historical reality statements, e.g., 'how the monkey got its tail'

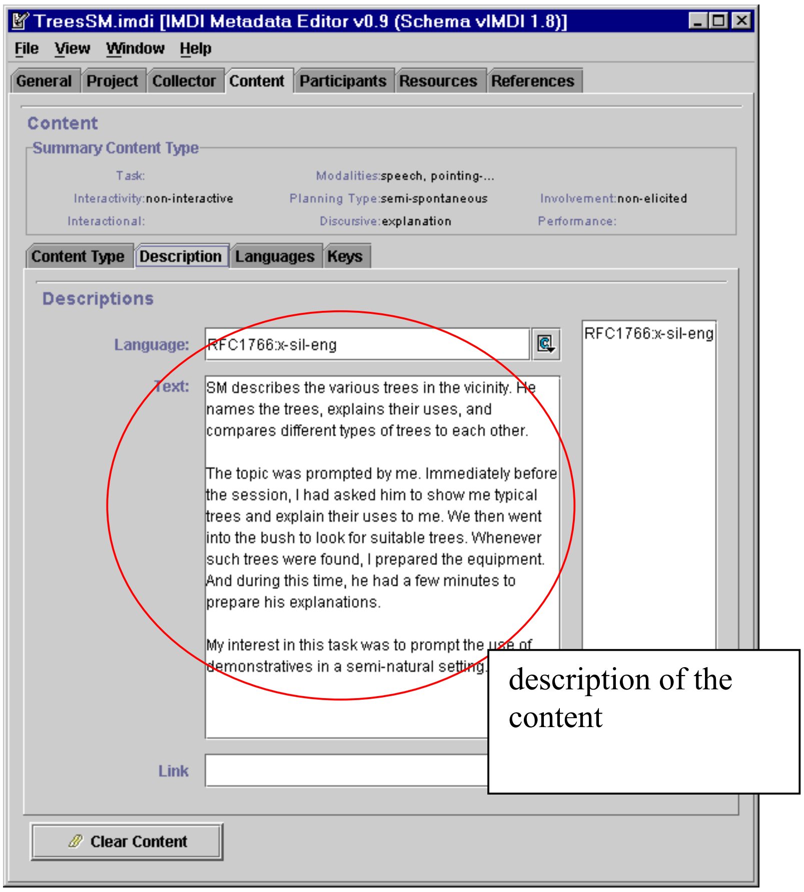
- Performance

  If the session contains a performative genre, choose one of the following items, or type in another item:

name	comments
"oratory"	using speech effectively in a conventionalized format to address an audience within political, legal, ceremonial, or religious settings
"oral-history"	an account of firsthand experience, recalled retrospectively and communicated to an interviewer for historical purposes
"historical-narrative"	a secondhand account of the experience of historical figures and events, which may be partly or wholly fictional, communicated to both locals and outsiders for both historical purposes and entertainment
"narrative"	a recounting of one or more fictional events by one or more narrators to an audience of at least one
"oral-poetry"	spoken or sung, in a relatively structured form (in prosody and syntax), often with distinctive language, e.g. oral epics, narrative poetry, ballads (shorter, lyrical narratives), and panegyric odes
"song"	a tune with recognizably structured lyrics, e.g., popular and love songs, lullabies
"proverb"	a summary of the wisdom of collective experience, often one line long; formulaic
"lament"	
"insult"	an insolent verbal act creating animosity

7.4.2 Description

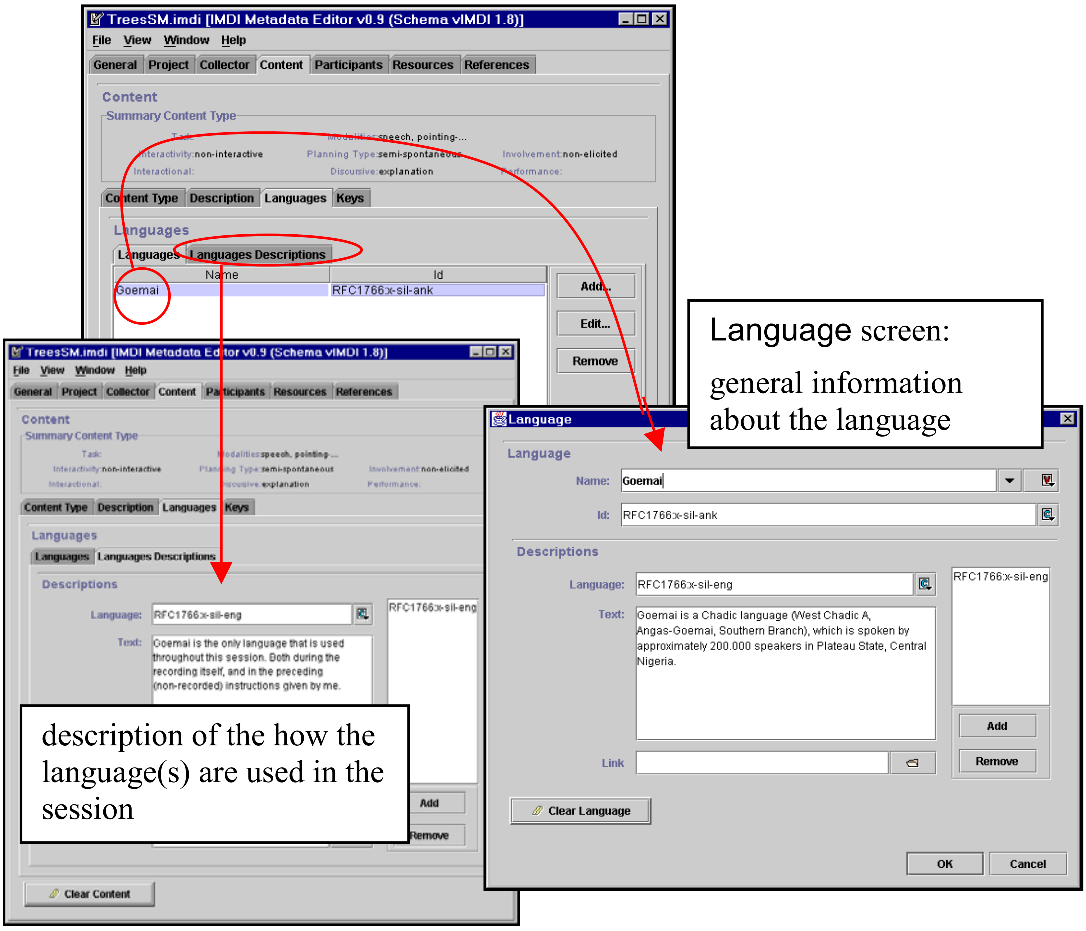
A description of the content of the session. For example:



See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.4.3 Languages: Languages and Languages Descriptions

Information about the language(s) **used** in the session.



Languages

Enter all languages that are used in the session. The first language to be entered should be the main language of the session.

Use the “Add” button to add another language.

If you want to remove a language from the list, click on it to highlight it and then click the “Remove” button.

If you want to modify a language on the list, click on it to highlight it and then click the “Edit” button.

The “Add” and “Edit” buttons give you access to the Language screen that contains general information about the language (independent of its use in the session).

Enter the following information:

- Name

  The name of the language.

  > **!** Note: The name of the language is standardized (see "Appendix 4: List of languages and language abbreviations"). Choose a name from the pull-down menu, or start typing (in this case, the pull-down menu will automatically open to display the available options). You may have to wait a few moments before the Editor will be able to process the entered information.

  > **!** Note: You have to enter the name in capital letters.

- ID

  The identifier of the language. It has to be entered in a standard format. Do the following:

  1. Click with the mouse button into the ID field and press the key SHIFT+R (or the key SHIFT+I). A dummy code "RFC1766:x-sil-aaa" (or "ISO639:aaa") is displayed.
  2. Replace the last three characters "aaa" with the proper language identifier (see "Appendix 4: List of languages and language abbreviations" for a list of language identifiers).

- 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 below).

  See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

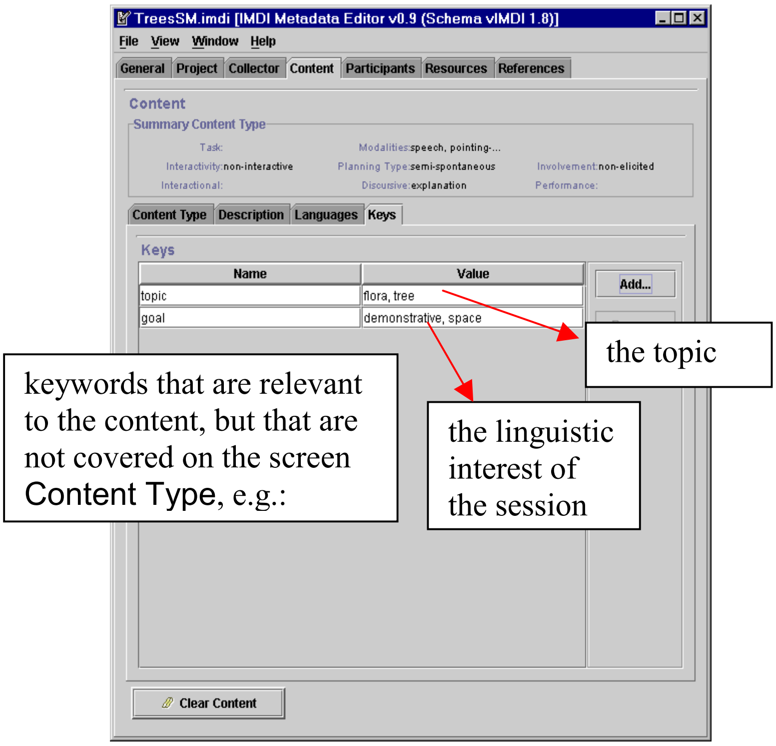
Languages Descriptions

Enter a description of (a) the set of languages used in the session and (b) 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/ Descriptions schema above).

See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.4.4 Keys

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



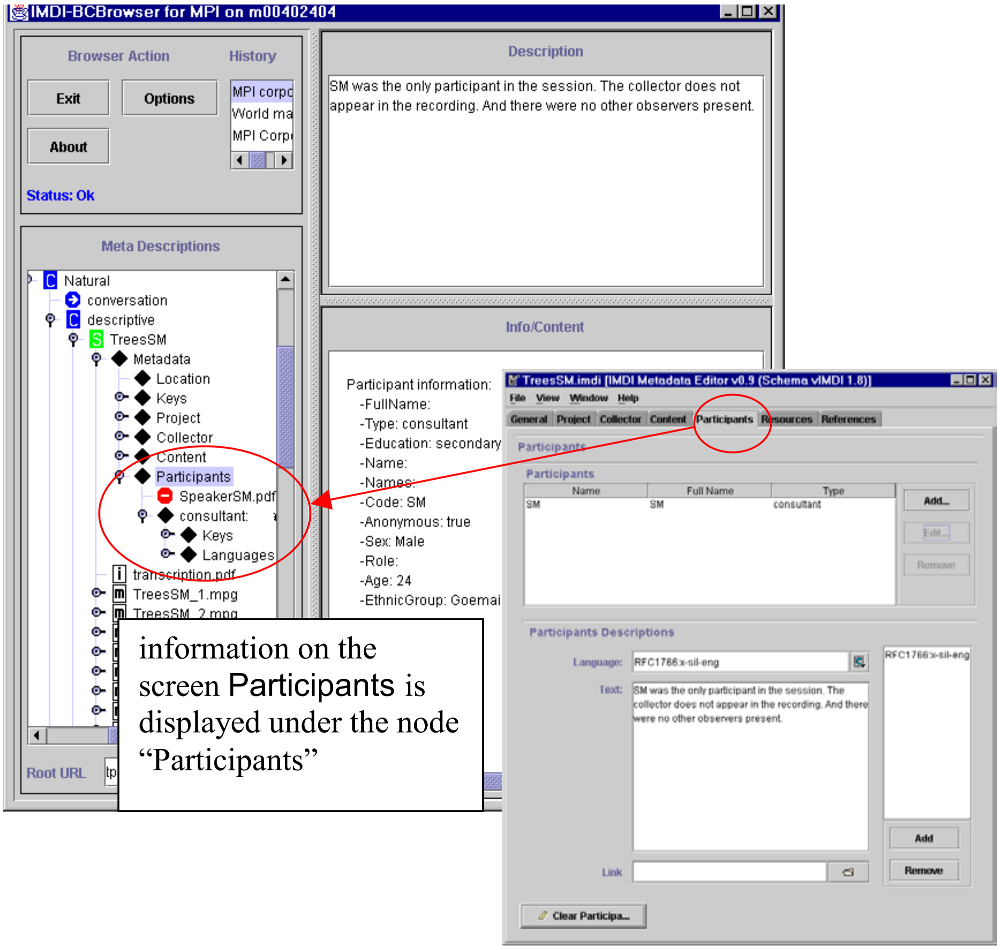
See "Appendix 2: Recurring schemata" for instructions on how to fill in a Keys schema.

7.5 Participants

This screen contains information about (a) each participant of the session and (b) the interrelations among different participants of the session. The following schemata are available:

Participants schema	Information about each participant of the session.
Participants Descriptions schema	Description of the interactions and interrelations among different participants of the session.

The information that you enter on this screen is displayed in the IMDI Browser as follows:



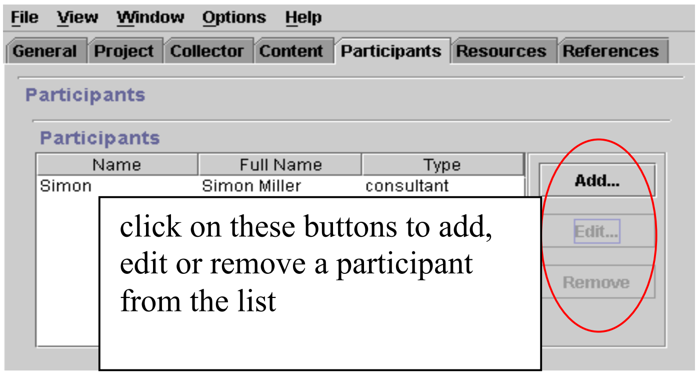
7.5.1 Participants: Participant Information, Descriptions, Languages and Keys

The **Participants** schema gives you access to the **Participant** screen and its four sub-screens, containing information about each individual participant of the session.

Use the "Add" button to add another participant.

If you want to remove a participant from the list, click on it to highlight it and then click the "Remove" button.

If you want to modify a participant on the list, click on it to highlight it and then click the "Edit" button.



The "Add" and "Edit" buttons give you access to the **Participant** screen. On the **Participant** screen, the following four sub-screens are available:

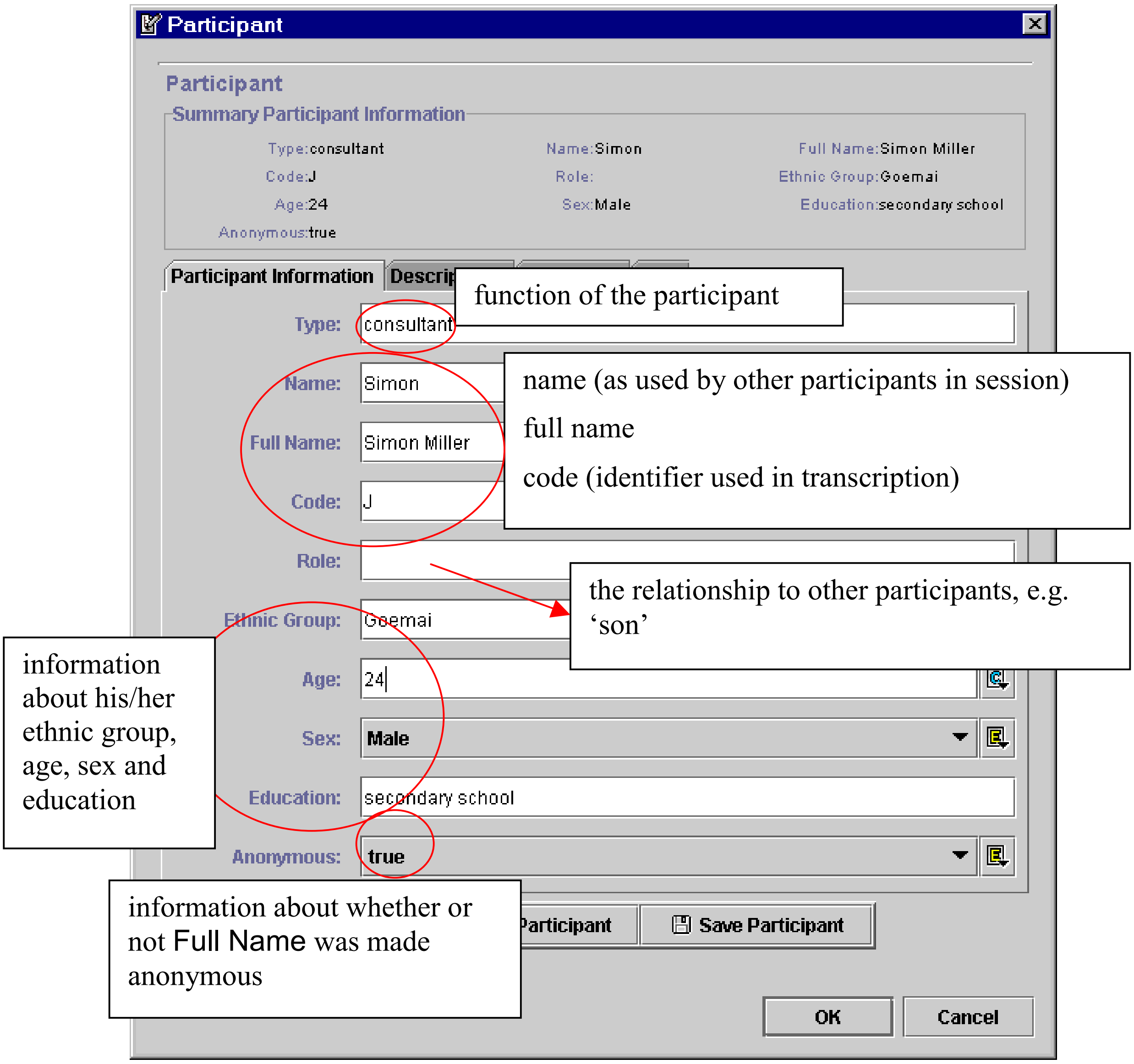
Participant Information sub-screen	Information about the individual participant.
Descriptions sub-screen	General description of the individual participant (independent of his/her role in the session).
Languages sub-screen	Information about all the languages that the participant is familiar with (independent of whether or not the participant uses them in the session).
Keys sub-screen	Keywords that are relevant to the individual participant (independent of his/her role in the session).

Note that this schema concerns the individual participant – information about the interrelations among different participants of the session should be entered into the **Participants Descriptions** schema (see section 7.5.2).

! Note: The collector should only appear on this screen if he/she takes an active part in the session.

Participant Information

Information about the individual participant. For example:



- Type

  The function of the participant of the session. For example:

  "interviewer"
  "consultant"
  "contributor"
  "computer"

- Name
  The name of the participant, i.e., the name that other participants in the session use to identify him/her. It is usually not the same as his/her full name.

- Full Name
  The full name of the participant.

- Code
  The code that is used in transcriptions and annotations to identify parts that were spoken by this participant.

- Role
  The role of the participant, i.e., his/her relationship to the other participants of the session. For example:

  Family relations, e.g. mother, father, child, husband, wife, etc.
  Work relations, e.g. boss, partner, student, teacher, etc.

- Ethnic Group
  The ethnic group of the participant.

- Age
  The age of the participant. Enter the age in the following format: **YY;MM.DD**

  ! Note: If the exact age is not known, we would ask you to enter the approximate age in years (you do not need to enter the months and days). This will allow you to later conduct searches on all participants who are in the age group between, e.g., 20 and 30 years of age.

  ! Note: The options "unknown" (for an unknown age) and "undefined" (for an artificial participant such as a computer) are available, too.

- Sex
  The sex of the participant. Choose it from the pull-down menu.

  ! Note: The option "undefined" refers to an artificial participant such as a computer.

- Education
  The education or literacy level of the participant. For example:

  primary school, secondary school, etc.
  literate, illiterate, etc.

- Anonymous

  Specify whether or not you have made the participant's name anonymous.

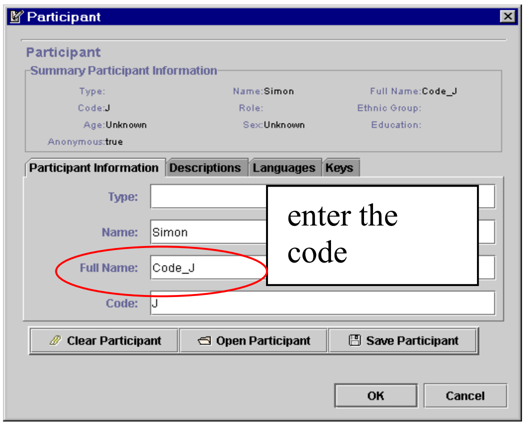
  Choose "true" if you have entered a code in the field Full Name.

  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 a full name. Make use of the menu item "Options" / "Anonyms …" (see section 6 "Options" Menu). If you have created such a file, the participant information is entered, saved and displayed as follows:

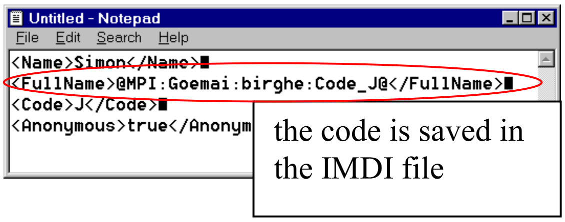
1. Enter the code into the field Full Name (instead of the full name).

   ! Note: The code entered into the field Full Name should **<u>not</u>** be the same as the code entered into the field Code.

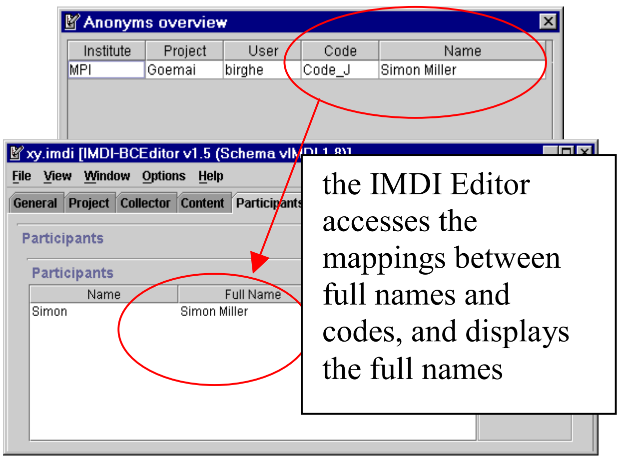


2. The code that you have entered into the field Full Name is saved in the IMDI file as follows:

   @Institute: Project: User: Code@



3. The IMDI Editor accesses the mappings that you have defined in the "Anonyms overview" dialog window (see section 6 "Options" Menu), and displays the full name.

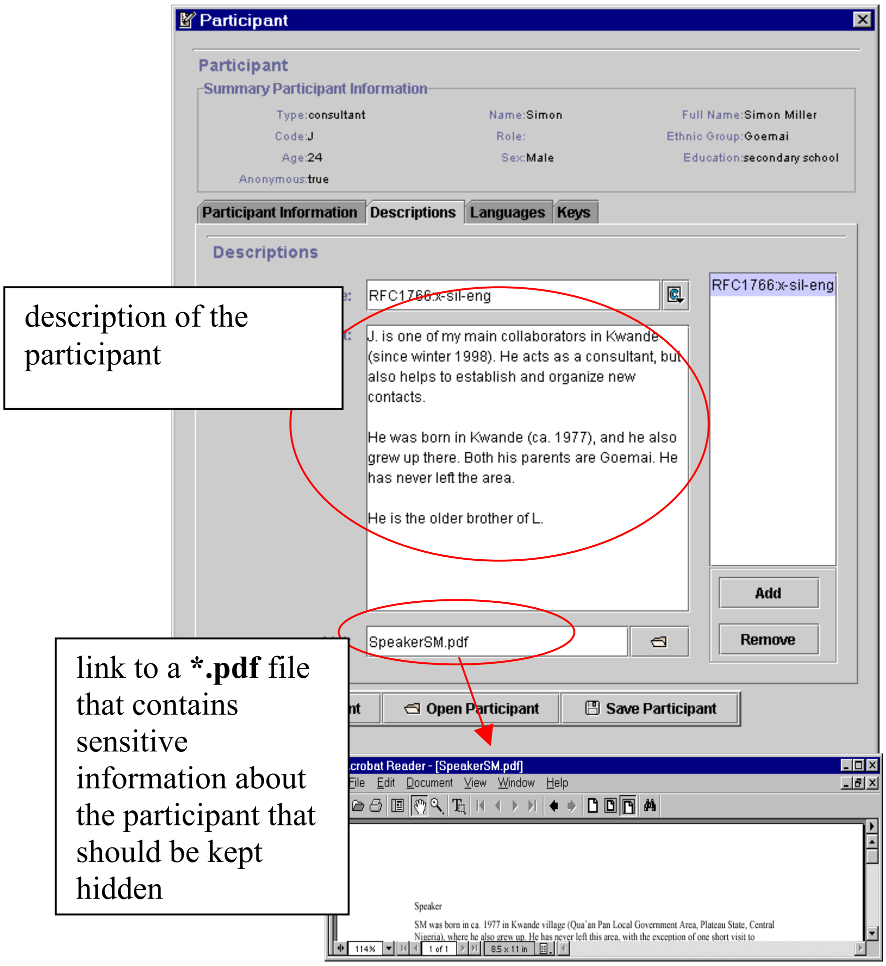


People who have access to the file that specifies the mappings will be able to see the full names, all others will only see the code.

! 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 (e.g., corpus.manager@mpi.nl).

Descriptions

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

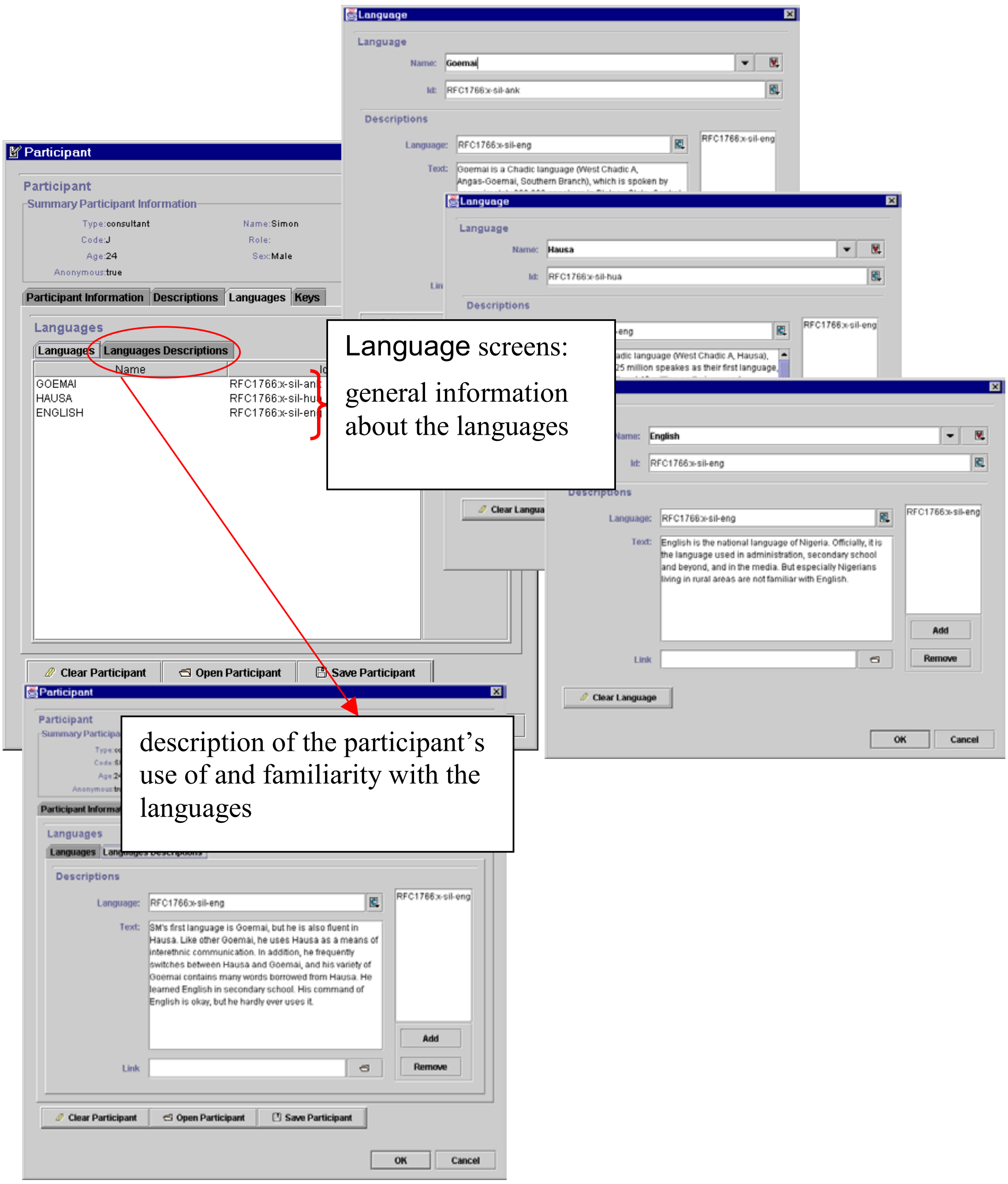


! Note: Recall that you can specify links to info files (see section 4.4.3). It is possible to store sensitive information about the participant in such files, i.e., information that is accessible to you alone. If you want to make use of this option, please contact the corpus manager (e.g., corpus.manager@mpi.nl).

See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

Languages

Information about all the languages that **<u>the participant is familiar with</u>** (independent of whether or not the participant uses them in the session).



- ### Languages

  Enter all the languages that the participant is familiar with.

  Use the "Add" button to add another language.

  If you want to remove a language from the list, click on it to highlight it and then click the "Remove" button.

If you want to modify a language on the list, click on it to highlight it and then click the “Edit” button.

The “Add” and “Edit” buttons give you access to the Language screen that contains general information about the language (independent of the participant’s familiarity with it). Enter the following information:

Name

The name of the language.

! Note: The name of the language is standardized (see “Appendix 4: List of languages and language abbreviations”). Choose a name from the pull-down menu, or start typing (in this case, the pull-down menu will automatically open to display the available options). You may have to wait a few moments before the Editor will be able to process the entered information.

! Note: You have to enter the name in capital letters.

ID

The identifier of the language. It has to be entered in a standard format. Do the following:

1. Click with the mouse button into the ID field and press the key SHIFT+R (or the key SHIFT+I). A dummy code “RFC1766:x-sil-aaa” (or “ISO639:aaa”) is displayed.
2. Replace the last three characters “aaa” with the proper language identifier (see “Appendix 4: List of languages and language abbreviations” for a list of language identifiers).

Descriptions

A description that gives background information about the language in general. Note that the description is independent of the participant’s familiarity with it (reserve such information for the Languages Descriptions sub-screen below).

See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

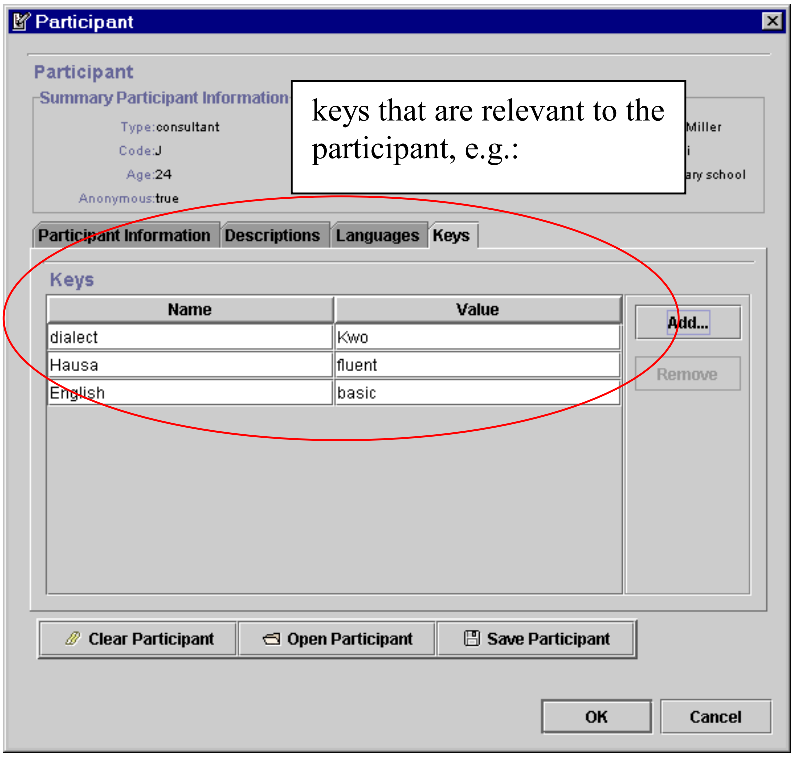
- Languages Descriptions

  A description of the set of languages that the participant is familiar with. Note that the description does **not** contain background information about the language in general (reserve this for the Language/ Descriptions schema above).

  See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

Keys

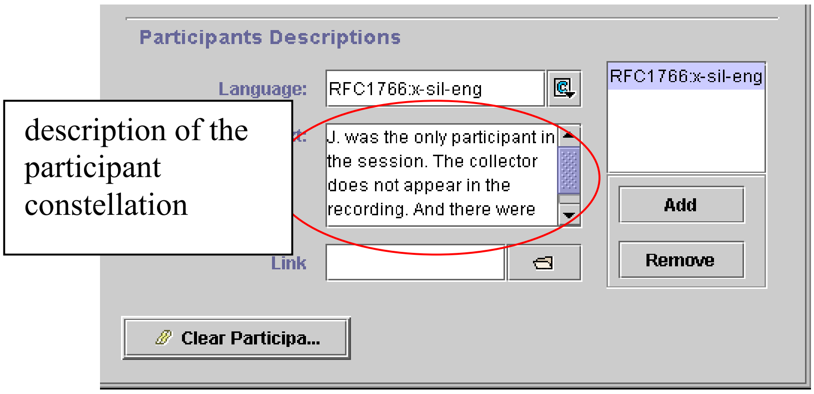
This sub-screen contains keywords that are relevant to the individual participant (independent of his/her role in the session).



See "Appendix 2: Recurring schemata" for instructions on how to fill in a Keys schema.

7.5.2 Participants Description

A description of the interactions and interrelations among different participants of the session. Note that this description concerns the set of participants as a whole - information about specific participants should be described with the Participants/ Participant schema (see section 7.5.1 above).



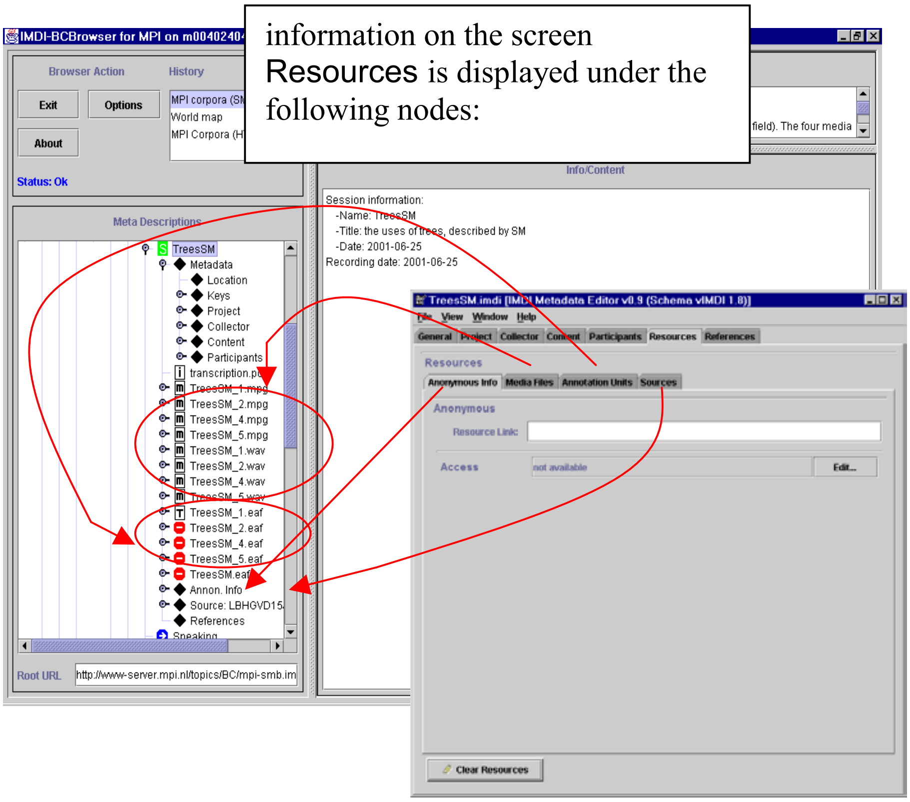
See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.6 Resources

This screen contains information about (a) access rights, (b) the annotation and digitized media files, and (c) the original media tapes. The following sub-screens are available:

Anonymous Info sub-screen	Please ignore this screen for the moment.
Media Files sub-screen	Information about the digitized media file(s), i.e. audio, video or image files.
Annotation Units sub-screen	Information about the annotation units (i.e., the annotation tiers).
Sources sub-screen	Information about the sources, i.e., the original tape(s) of the session.

The information that you enter on these four screens is displayed in the IMDI Browser as follows:



7.6.1 Anonymous Info: Anonymous and Access

Please ignore this screen for the moment.

7.6.2 Media Files: Media File, Time Position, Access Policy and Descriptions

Information about the digitized media file(s), i.e. audio, video or image files. Enter all media files that are connected with the session.

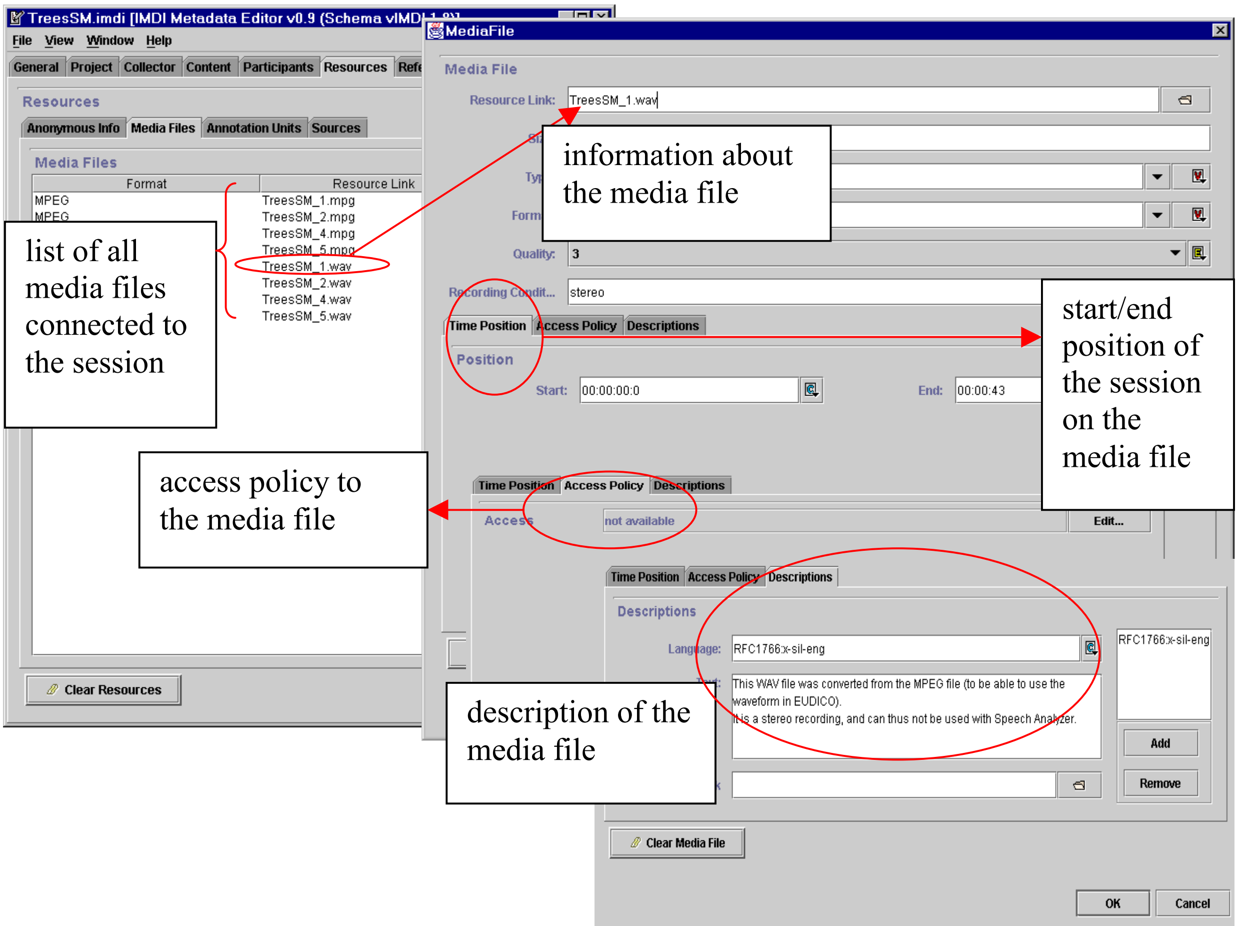
Use the “Add” button to add another media file.

If you want to remove a media file from the list, click on it to highlight it and then click the “Remove” button.

If you want to modify a media file on the list, click on it to highlight it and then click the “Edit” button.

The “Add” and “Edit” buttons give you access to the Media File screen that contains the following schemata and sub-screens:

Media File schema	General information about the media file.
Time Position sub-screen	The start and end position of the session within the media file.
Access Policy sub-screen	Access rights to the media file.
Descriptions sub-screen	Description of the media file.



Media File

General information about the media file. Enter the following information:

- ### Resource Link

  The link to the corresponding media file. Enter the name of the media file. Please do **not** make use of the "open folder" icon to the right of that field (unless you want to manage your own corpus, see part IV).

- ### Size

  The size of the media file in bytes. You do not have to fill in this field.

- Type

  The type of the media file. Choose an option from the pull-down menu, or specify a different type, e.g.:

  "unknown"
  "unspecified"
  "audio"
  "video"
  "image" (i.e., for photo image)
  "drawing"

- Format

  The format of the media file. Choose an option from the pull-down menu, or specify a different format, e.g.:

  "unknown"
  "unspecified"
  "MPEG1"
  "MPEG2"
  "MPEG4"
  "WAV"

- Quality

  The quality of the media file. 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, i.e., the equipment used in the recording (e.g., microphone type, amplifier type, mono/stereo recording, etc.).

Time Position

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

This information is necessary for the corpus management to correctly segment your data (see "Appendix 5: Digitization Policy").

Access Policy

Information about the access rights to the **media file**. See "Appendix 2: Recurring schemata" for instructions on how to fill in an Access schema.

Descriptions

A description of the media file. See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.6.3 Annotation Units: Annotation Unit, Information, Access Policy and Descriptions

Information about the annotation units (i.e., the annotation tiers). Enter all annotation units that are connected with the session.

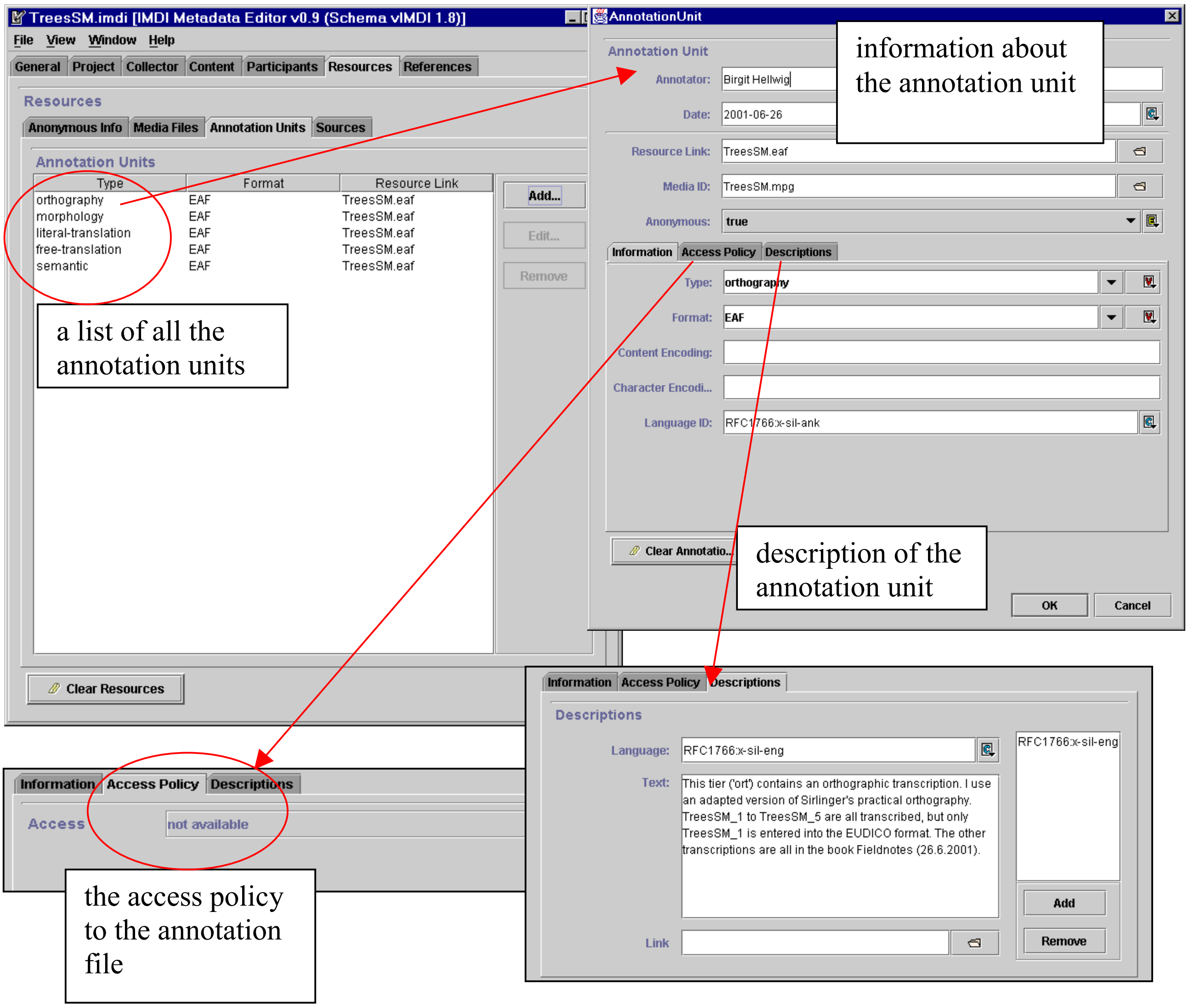
Use the “Add” button to add another annotation unit.

If you want to remove an annotation unit from the list, click on it to highlight it and then click the “Remove” button.

If you want to modify an annotation unit on the list, click on it to highlight it and then click the “Edit” button.

The “Add” and “Edit” buttons give you access to the Annotation Unit screen that contains the following schemata and sub-screens:

Annotation Unit schema	General information about the annotation unit and its corresponding annotation file.
Information sub-screen	Information about the annotation unit.
Access Policy sub-screen	Access rights to the annotation file.
Descriptions sub-screen	Description of the annotation unit.



! Note: The label 'annotation unit' does **not** refer to the annotation file itself, but to one annotation tier (e.g., the free translation, or the phonetic transcription).

! Note: You do not have to enter all annotation units, but it is necessary that you list at least one annotation unit per annotation file. This is needed for technical reasons in order to create a link to the annotation file(s).

Annotation Unit

General information about the annotation unit and its corresponding annotation file. Enter the following information:

- Annotator

  The name of the person who did the annotation/transcription. List all annotators in this field (separated by commas).

- Date

  The date when the annotation unit 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 annotation unit. Enter the name of the annotation file. Please do **not** make use of the "open folder" icon to the right of that field (unless you want to manage your own corpus, see part IV).

- Media ID

  The link to the media file from which the annotation unit originated. Enter the name of the media file. Please do **not** make use of the "open folder" icon to the right of that field (unless you want to manage your own corpus, see part IV).

- Anonymous

  Specify whether or not you have made the participant's name anonymous in the annotation file.

  Choose "true" if you have used a code.

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

Information

Information about the annotation unit. Enter the following information:

- Type

  The type of annotation unit. Choose an option from the pull-down menu, or specify a different type, e.g.:

  “orthography”
  “phonemic”
  “phonetic”
  “morphology”
  “morphosyntax”
  “syntax”
  “free-translation”
  “literal-translation”
  “semantic”

  ! Note: You can enter all annotation units into this one field. Separate them with a comma.

- Format

  The file format of the file that contains the annotation unit. Choose an option from the pull-down menu, or type in a different format, e.g.:

  “CHAT”
  “Shoebox”
  “RDBMS”
  “TRS”
  “EAF”
  “AIF”
  “BAS”

- Content Encoding

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

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

- Character Encoding

  The name of the character encoding used for creating the annotation unit.

- Language ID

  The identifier of the language that is used for creating the annotation unit, e.g. ‘English’ for an English translation. It has to be entered in a standard format. Do the following:

  1. Click with the mouse button into the ID field and press the key SHIFT+R (or the key SHIFT+I). A dummy code “RFC1766:x-sil-aaa” (or “ISO639:aaa”) is displayed.
  2. Replace the last three characters “aaa” with the proper language identifier (see “Appendix 4: List of languages and language abbreviations” for a list of language identifiers).

Access Policy

Information about the access rights to the **annotation file**. See “Appendix 2: Recurring schemata” for instructions on how to fill in an Access schema.

Descriptions

A description of the annotation unit. See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.6.4 Sources: Source, Position, Access Policy and Descriptions

Information about the sources, i.e., the original tape(s) of the session. Enter all the sources that are connected with the session.

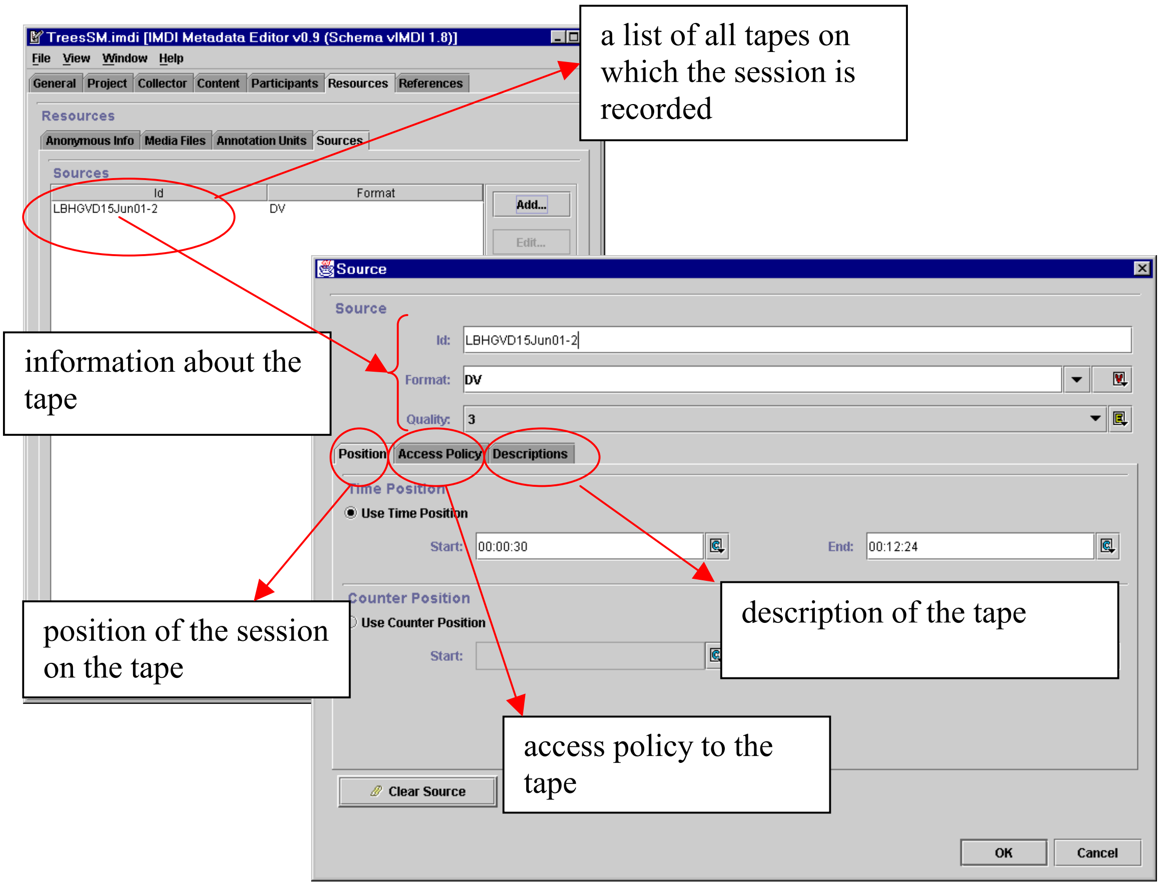
Use the “Add” button to add another source.

If you want to remove a source from the list, click on it to highlight it and then click the “Remove” button.

If you want to modify a source on the list, click on it to highlight it and then click the “Edit” button.

The “Add” and “Edit” buttons give you access to the Source screen that contains the following schemata and sub-screens:

Source schema	General information about the source.
Position sub-screen	The start and end position of the session on the source.
Access Policy sub-screen	Access rights to the source.
Descriptions sub-screen	Description of the source.



Source

General information about the source. Enter the following information:

- **ID**

  A short code to identify the source. Please make use of the tape labeling conventions (see "Appendix 6: Tape labeling conventions").

- **Format**

  The physical storage format of the tape. For example:

name	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

- Quality

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

Position

The start and end position of the session on the corresponding tape. Please enter the start/end position in the following formats:

- In the case of digital audio tapes use the field Time Position: **hh:mm:ss** (i.e., hours:minutes:seconds).
- In the case of digital video tapes use the field Time Position: **hh:mm:ss:f** (i.e., hours:minutes:seconds:frames).
- In the case of non-digital tapes, use the field Counter Position: enter any number of digits to represent the start/end position.

Access Policy

Information about the access rights to the **source**. See “Appendix 2: Recurring schemata” for instructions on how to fill in an Access schema.

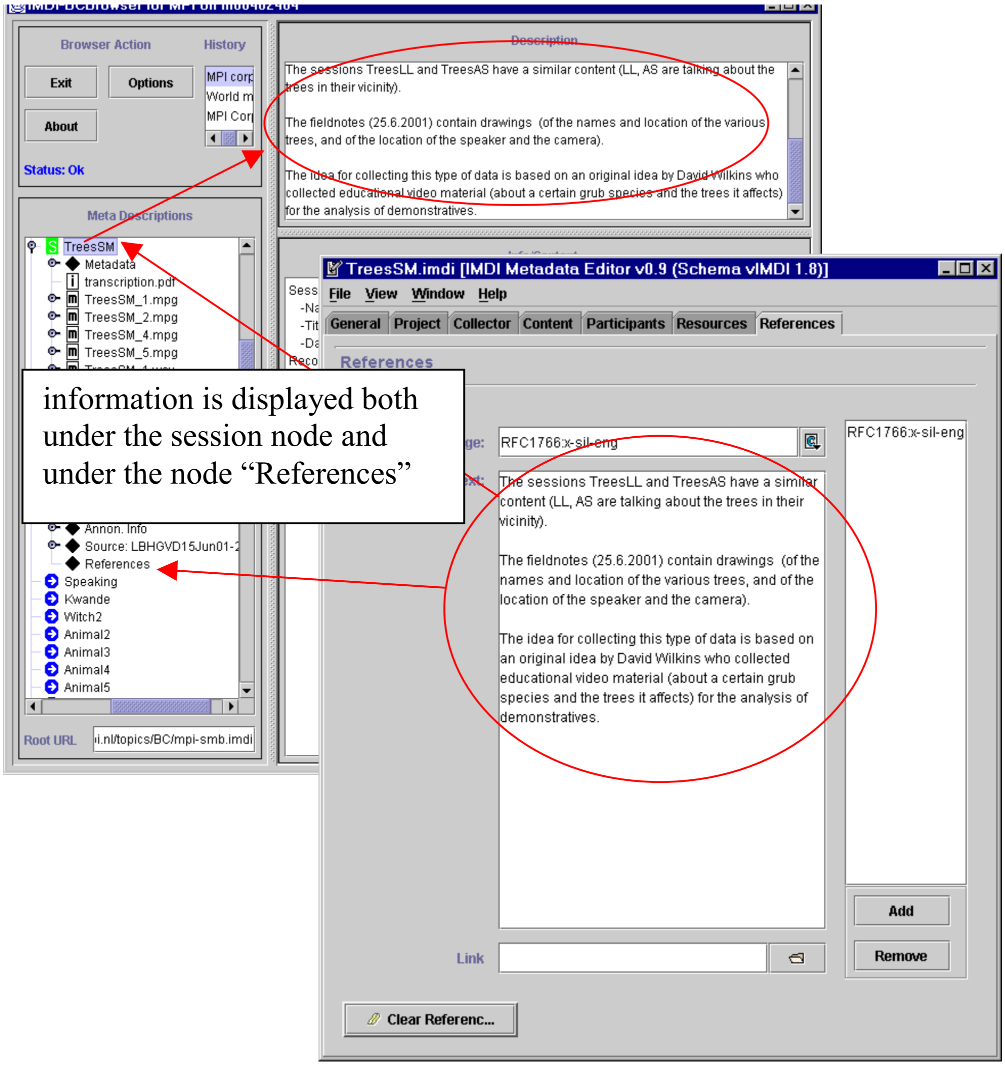
Descriptions

A description of the source. See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

7.7 References

Descriptions

This schema contains cross-references to other sessions, fieldnotes, or publications that are relevant to the content of the session. For example:



See “Appendix 2: Recurring schemata” for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

PART IV: THE USER'S GUIDE TO THE IMDI CORPUS EDITOR

8 The IMDI Corpus Editor

This part of the manual guides you through using the IMDI Corpus Editor to create a corpus hierarchy (see part III for the IMDI Session Editor).

> **!** Note: In many cases, the corpus manager (e.g., corpus.manager@mpi.nl) 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.

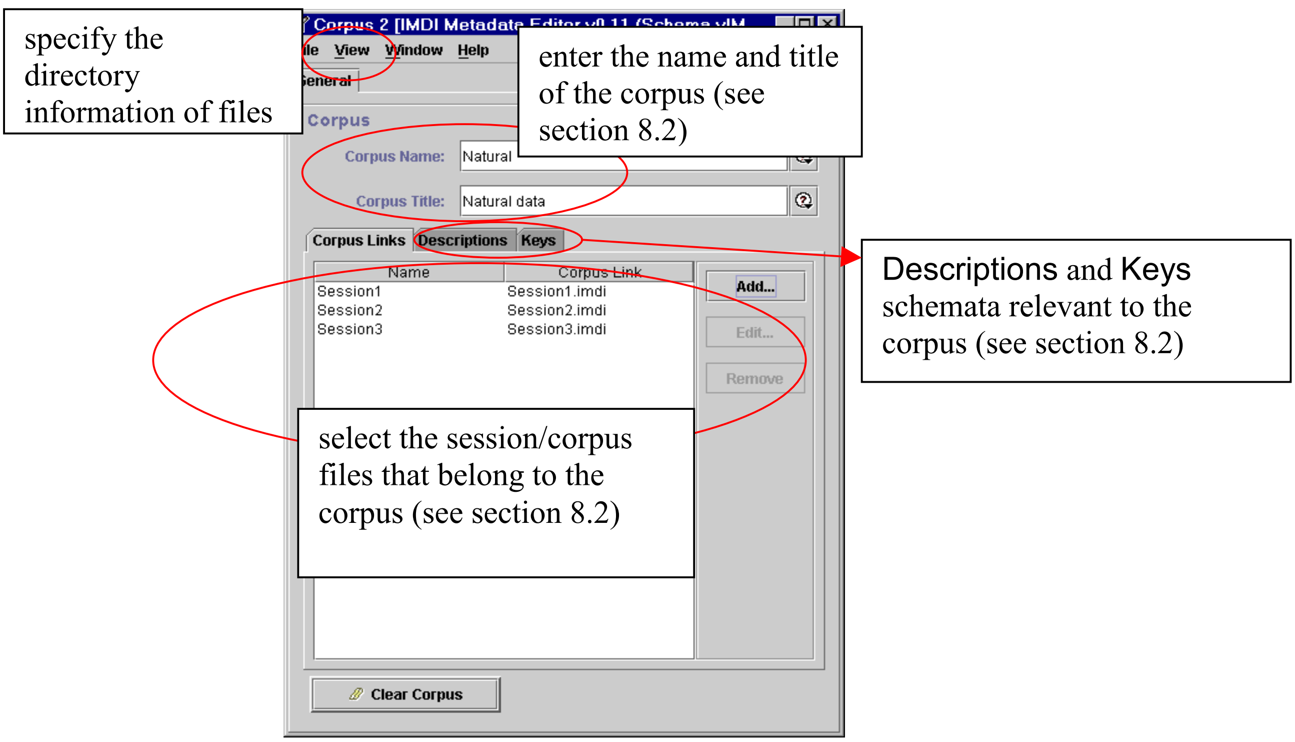
> **!** Note for researcher working at the MPI for Psycholinguistics: If you plan on using the IMDI Corpus Editor, please contact corpus.manager@mpi.nl first.

The IMDI Corpus Editor creates an IMDI Corpus file (***.imdi**). Note that this file has the same extension as the IMDI Session file created with the IMDI Session Editor (see part III). Furthermore, the menu bars of both the IMDI Corpus Editor and the IMDI Session Editor are almost identical (see section 6 for the available menu options).

The IMDI Corpus Editor is accessed as follows:

1. In the IMDI Session or Corpus Editor, click on "File" menu.
2. Click on "New".
3. Click on "Corpus".

   The IMDI Corpus Editor opens, e.g.:



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

1. Specify the directory information of all info, media and annotation files that belong to a session, and of all session and corpus files that belong to a corpus.
2. Select the sessions that belong to a subcorpus. Select the subcorpora that belong to a corpus.
3. Create a corpus node.

These steps are explained in the following three sections.

8.1 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:

- In the IMDI Session Editor, Link fields exist for info files (see section 4.4.3), media files (see section 7.6.2), and annotation files (see section 7.6.3).
- In the IMDI Corpus Editor, Link fields exist for IMDI Session and Corpus files, i.e., files created with the IMDI Session or Corpus Editor (see section 8.2).

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

If your corpus is managed by a corpus manager (e.g., corpus.manager@mpi.nl), you only need to enter the file name. In this case, type the file name directly into the Link field of the IMDI Session 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 8.1.2).

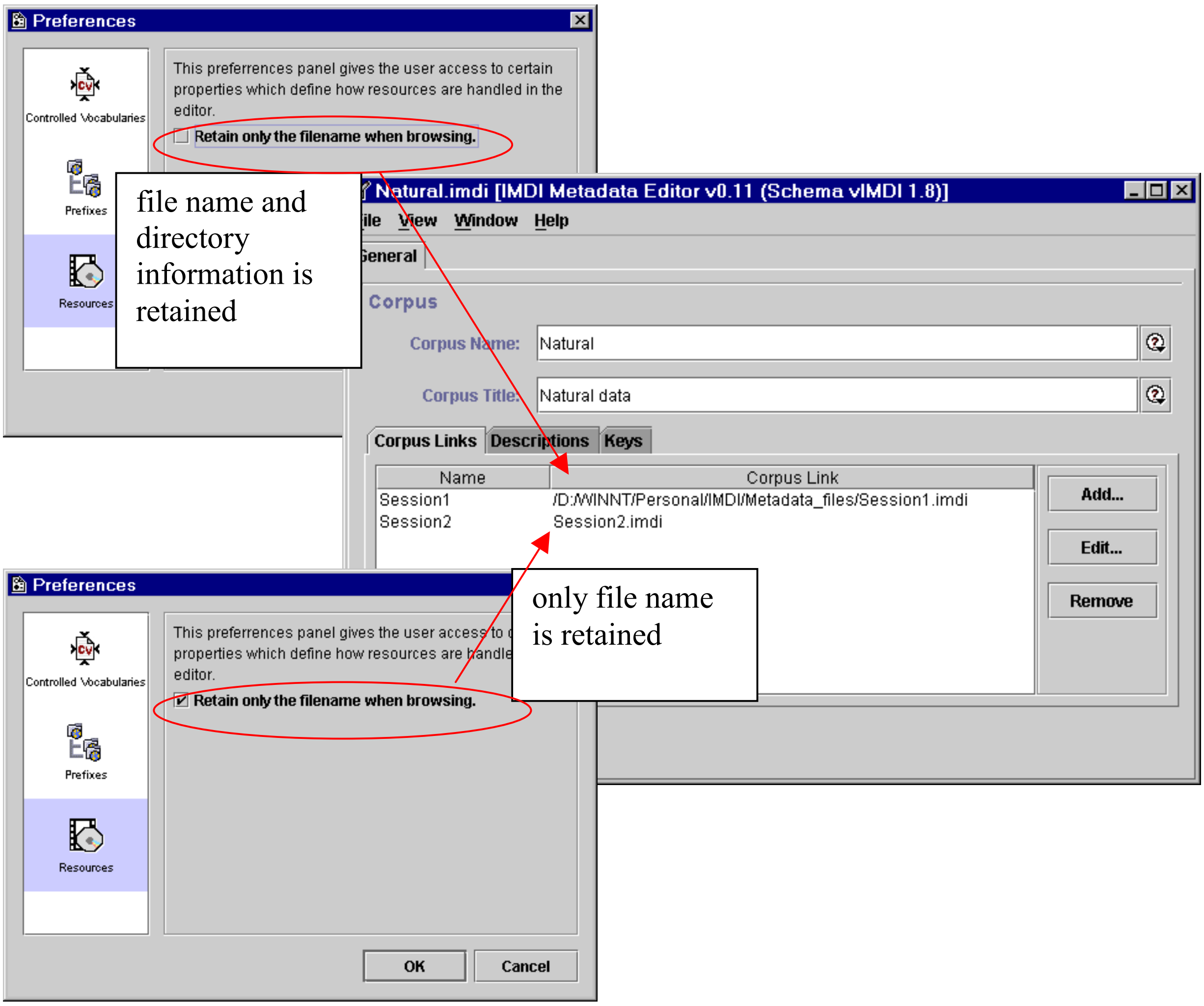
If you manage your own corpus, you need to enter both the file name and the directory location. You can enter the information either directly into the Link field (see section 8.1.1) or you can make use of prefixes (see section 8.1.2).

8.1.1 Link fields

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

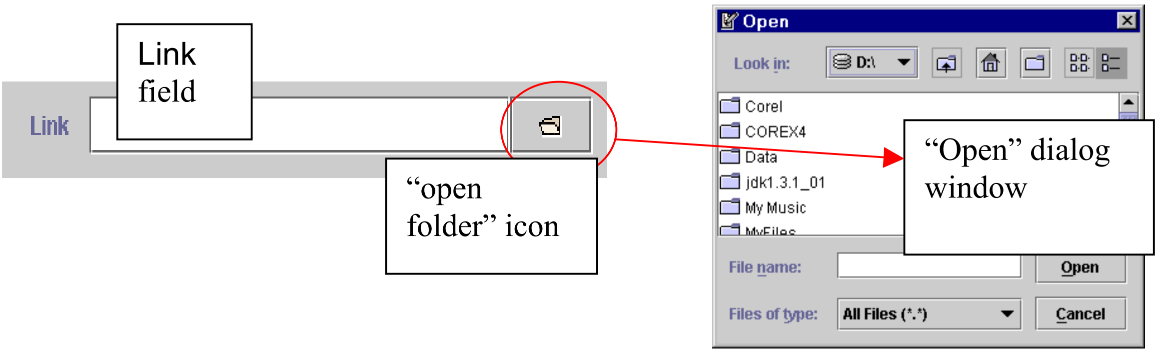
1. Make sure that the browsing option is set correctly. Do the following:
    1. In the IMDI Session or Corpus Editor, click on “View” menu.
    2. Click on “Preferences …”.
    3. In the “Preferences” dialog window, click on “Resources”.
    4. In the “Resources” dialog window, make sure that there 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 Session or Corpus Editor, click on “View” 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 8.1.2).

3. In the IMDI Session or Corpus 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 info, media, annotation, session or corpus file.
5. Double-click on the file to insert its name and directory location into the Link field.

8.1.2 Prefixes

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

1. In the IMDI Session or Corpus Editor, click on "View" menu.
2. Click on "Preferences …".
3. In the "Preferences" dialog window, click on "Prefixes". The "Prefixes" dialog window opens.

   ! Note: The "Prefixes" dialog window does not display all its information. Please enlarge the window by pointing with the mouse to its edges. The mouse cursor turns into a double-headed arrow. Click the mouse button, keep it clicked and drag the arrow to enlarge the window.

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 is already defined in the original IMDI Session or Corpus 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 defined 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 this option):

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

   Choose this option if the prefixes are defined in the “Prefixes” dialog window.

   The prefixes apply to the entire IMDI Session/Corpus file, and are thus especially useful whenever one IMDI file contains links to a number of different files that are stored in the same folder, e.g., to a number of media files, or to a number of IMDI Session files.

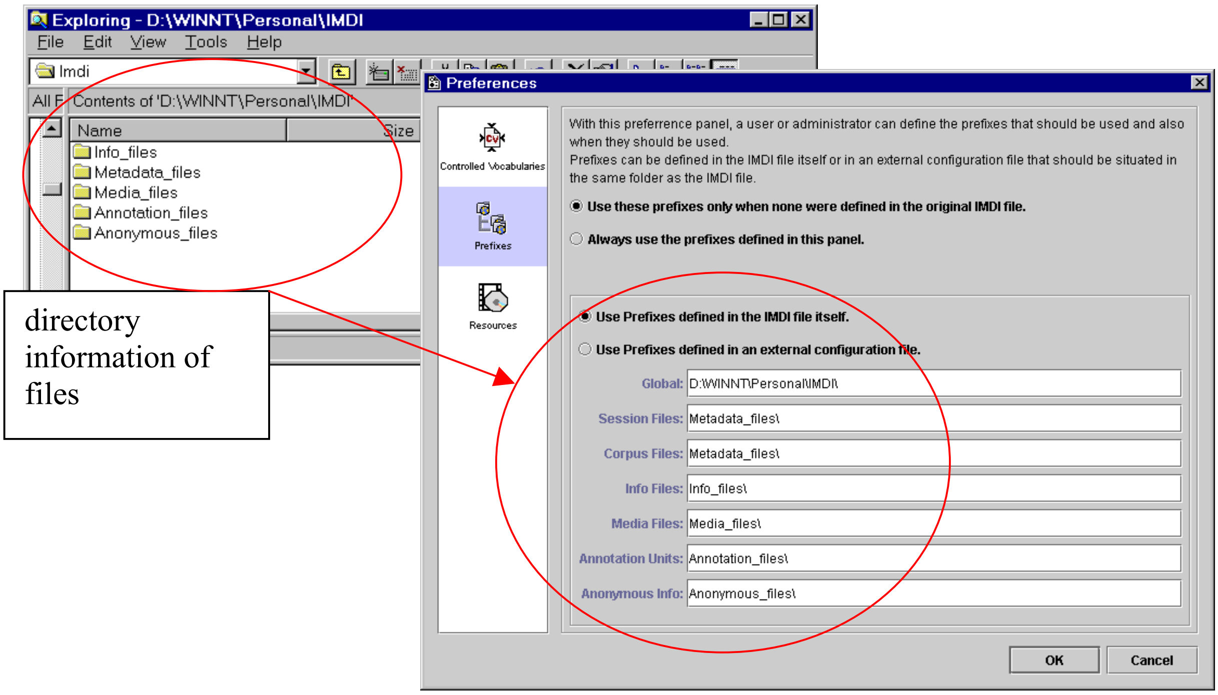
   The prefixes are saved in the IMDI Session/Corpus file.

   To define the prefixes, enter the directory 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).
   - Annotation Units: the location of your annotation files (relative to Global).

   ! Note: Please add a backslash after each folder.

   e.g.:

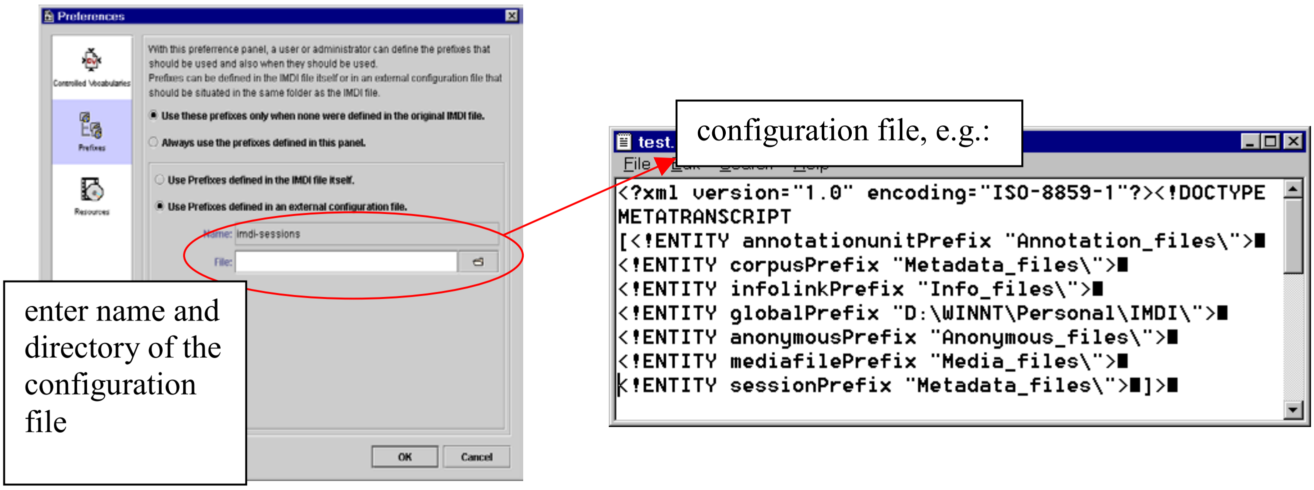


(2) Use Prefixes defined in an external configuration file.

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

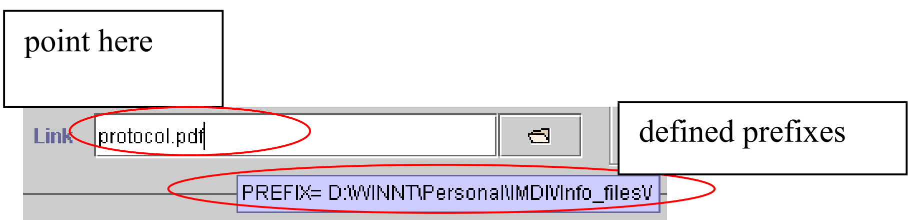
This option is more flexible than option (1) above because the prefixes are not saved in the IMDI Session/Corpus file itself. I.e., if you decide to move files to a different directory at a later stage, you need to change the prefixes only once in the external configuration file – not in each IMDI Session/Corpus 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.



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

There is an on-line help 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.:



8.2 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, and annotation 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. It should be identical to the name of the IMDI Corpus file.

> **!** Note: 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) Description:

Descriptive information about the corpus or subcorpus. See "Appendix 2: Recurring schemata" for instructions on how to fill in a Descriptions schema. Remember: The field Language refers to the language in which the description was written - **not** to the language under investigation.

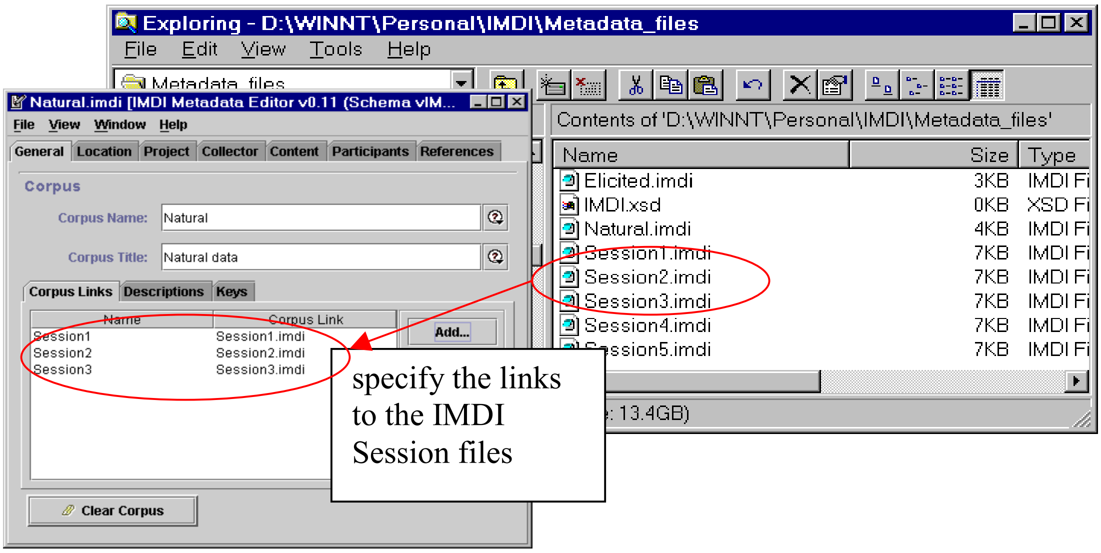
(4) Keys:

Keywords relevant to the corpus or subcorpus. See “Appendix 2: Recurring schemata” for instructions on how to fill in a Keys schema.

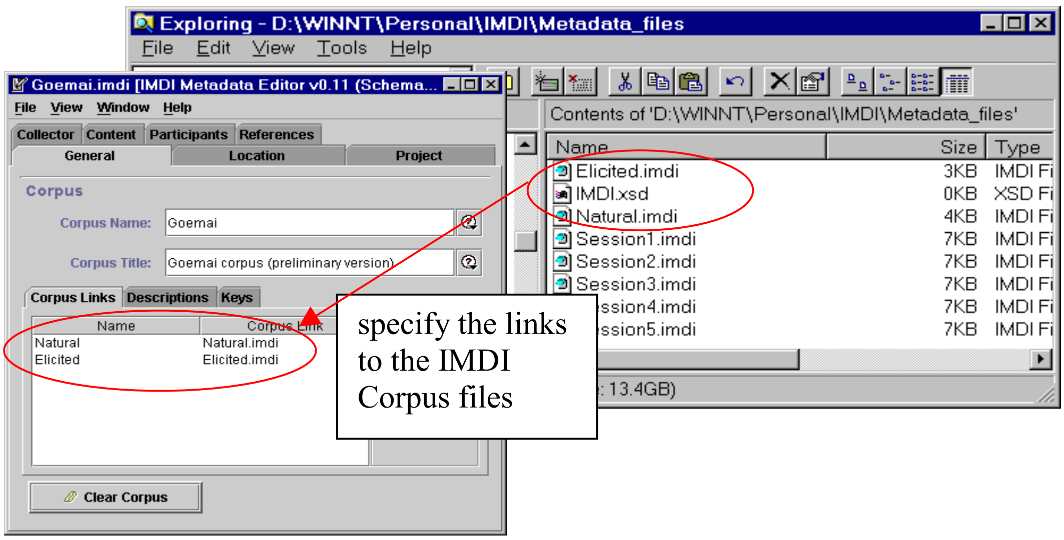
(5) Corpus Links:

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**.

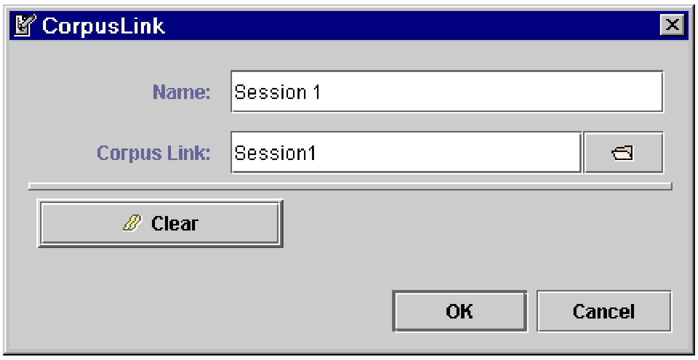


(b) Specify the IMDI Corpus files that belong to the corpus or subcorpus. E.g., in the IMDI Corpus file **Goemai.imdi** (i.e., “Goemai corpus (preliminary version)”), specify that it contains the IMDI Corpus files **Natural.imdi** (i.e., “Natural data”) and **Elicited.imdi** (i.e., “Elicited data”).



To specify a link, do the following:

1. Click on the “Add” button. The “CorpusLink” dialog window appears.



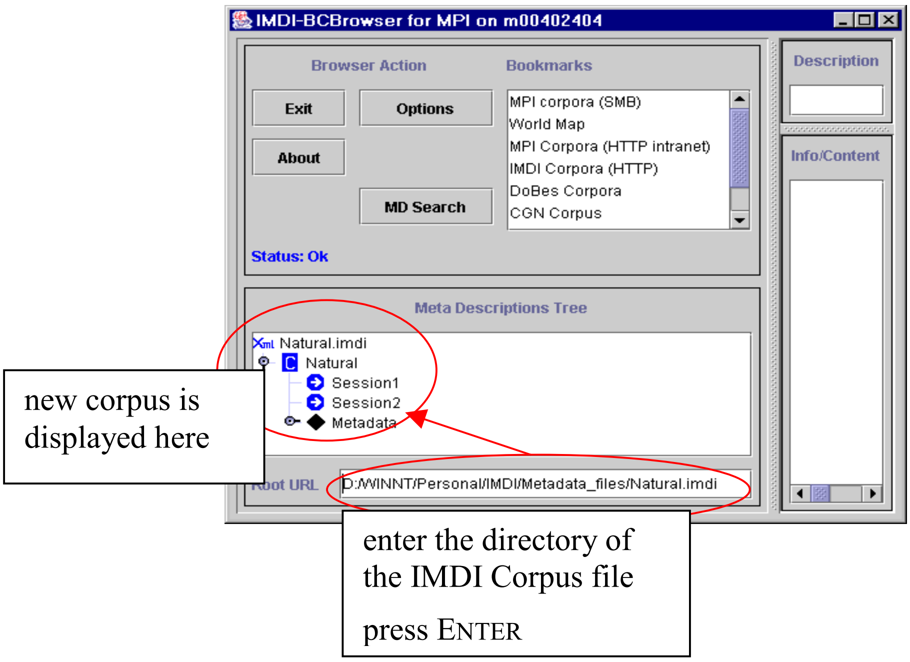
2. Enter the name of the session or corpus in the field Name.
3. Specify the directory location of the corresponding IMDI Session/Corpus file in the field Corpus Link.

8.3 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 “Meta Descriptions Tree” panel, enter the location of the IMDI Corpus file into the field labeled Root URL. Press ENTER.

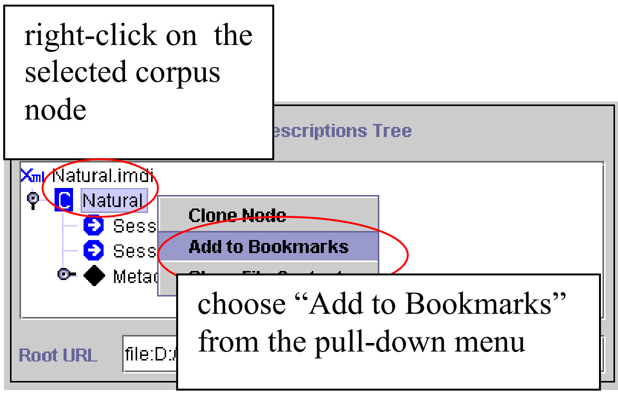
   The corpus node, together with its subcorpus nodes and sessions, is displayed in the “Meta Descriptions Tree” panel, e.g.:

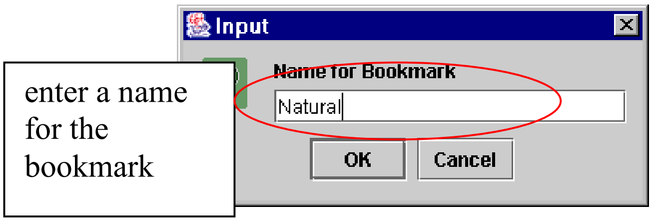


3. To save the new corpus node permanently, do the following:
    1. In the “Meta Descriptions Tree” panel, click on the corpus node to select it.
    2. Right-click on the selected corpus node.

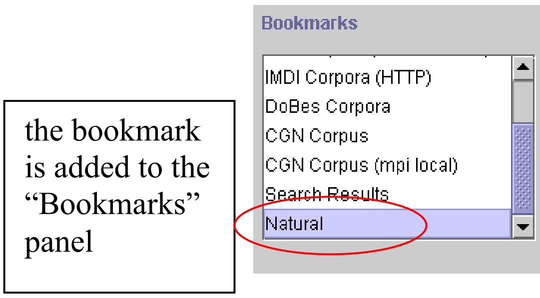
       A pull-down menu appears.
    3. Choose “Add to Bookmarks” from the pull-down menu.

       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).



PART V: APPENDICES

Appendix 1: Screens

The following table gives an overview of all (sub-) screens and their content.

General	General information about (a) the session (name, date, location) and (b) the IMDI file.
Descriptions	Description of the circumstances under which the session data was collected.
Location	Information about the location at which the session data was collected.
Keys	Keywords that are relevant to (a) the collection of the data or (b) the creation of the IMDI file.
Project	Information about the larger project within which the session data was collected.
Collector	Information about the person who collected the session data, and about the person who is responsible for collecting the session data.
Content	Information about the content of the session.
Content Type	Information about the task, the modalities, the communication context, and the genre of the session.
Description	Description of the content of the session.
Languages	Information about the language(s) used in the session.
Language	General information about the language (independent of its use in the session).
Descriptions	Description of (a) the set of languages used in the session and (b) the role of each language as it is used in the session.
Keys	Keywords that are relevant to the content of the session.

Participants	Information about (a) each participant of the session and (b) the interrelations among different participants of the session.
Participant(s)	Information about each participant of the session.
Information	Information about the individual participant.
Descriptions	General description of the individual participant (independent of his/her role in the session).
Languages	Information about all the languages that the participant is familiar with (independent of whether or not the participant uses them in the session).
Language	General background information about the language itself (independent of the participant's familiarity with it).
Descriptions	Description of the set of languages as a whole that the participant is familiar with.
Keys	Keywords that are relevant to the individual participant (independent of his/her role in the session).
Descriptions	Description of the interactions and interrelations among different participants of the session.
Resources	Information about (a) access rights, (b) the annotation and digitized media files, and (c) the original media tapes.
Anonymous Info	Please ignore this screen for the moment.
Media Files	Information about the digitized media file(s), i.e. audio, video or image files.
Media File	General information about the media file.
Time Position	The start and end position of the session on the media file.
Access Policy	Access rights to the media file.
Descriptions	Description of the media file.
Annotation Units	Information about the annotation unit(s) (i.e., the annotation tiers).
Annotation Unit	General information about the annotation unit and its corresponding annotation file.
Information	Information about the annotation unit.
Access Policy	Access rights to the annotation unit.
Descriptions	Description of the annotation unit.
Sources	Information about the source(s), i.e. the original tape(s) of the session.
Source	General information about the source.
Position	The start and end position of the session on the source.
Access Policy	Access rights to the source.
Descriptions	Description of the source.

References	Cross-references to other sessions, fieldnotes, or publications, which are relevant to the content of the session.

Appendix 2: Recurring schemata

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

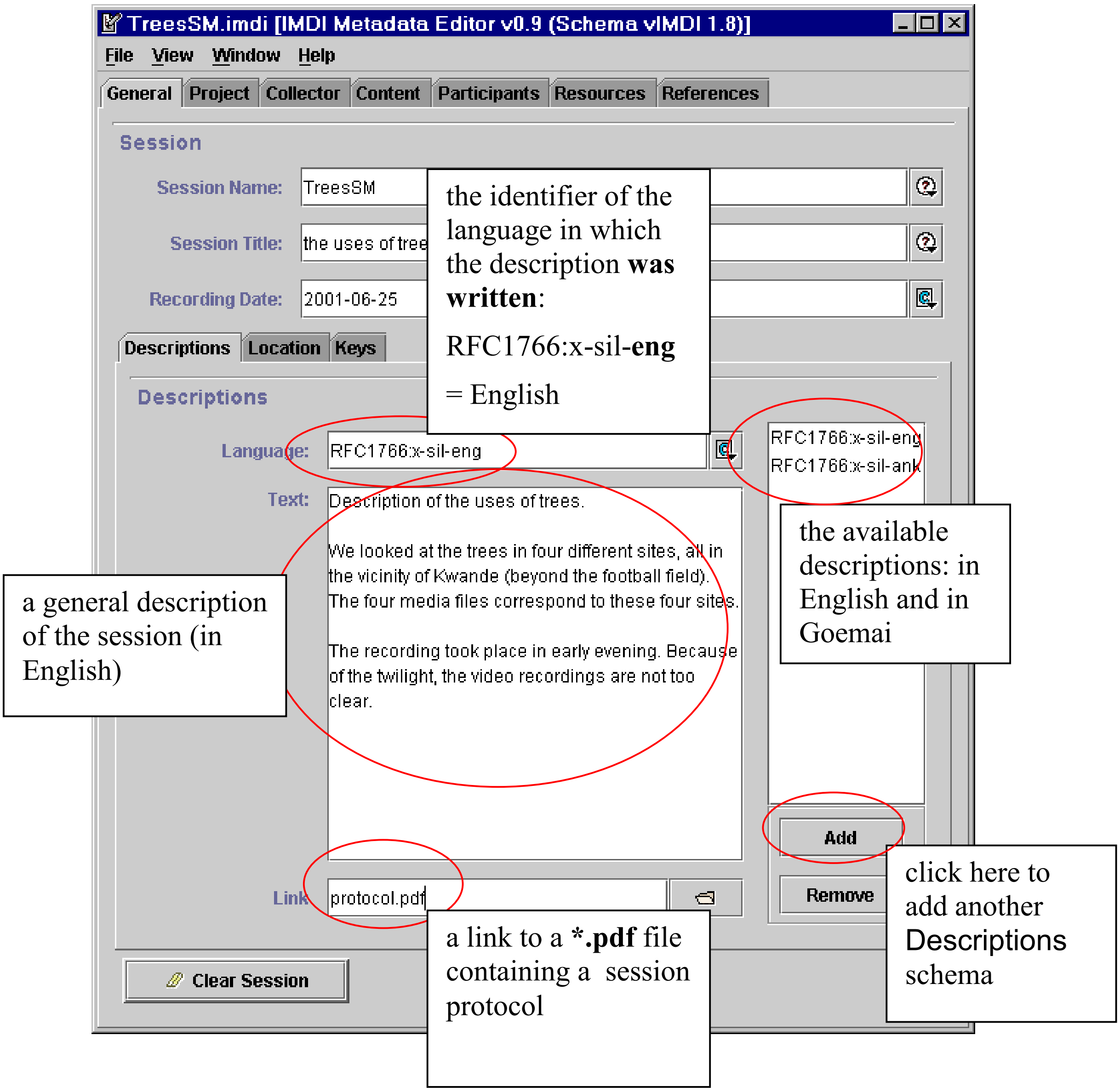
Descriptions

The following screens contain a Descriptions schema:

General	Description of the circumstances under which the session data was collected.
Project	Description of the scope and the goals of the project, within which the session data was collected.
Collector	Description of the person who collected the session data, and about the person who is responsible for collecting the session data.
Content	Description of the content of the session.
Content/ Languages	Description of (a) the set of languages used in the session and (b) the role of each language as it is used in the session.
Participants/ Participant	Description of the individual participant (independent of his/her role in the session).
Participants/ Participant/ Languages	Description of the set of languages as a whole that the participant is familiar with.
Participants	Description of the interactions and interrelations among different participants of the session.
Resources/ Media File	Description of the media file.
Resources/ Annotation Unit	Description of the annotation unit.
Resources/ Source	Description of the source.
References	Description of cross-references to other sessions, fieldnotes, or publications, which are relevant to the content of the session.
Recurring schemata on recurring screens:	
Access	Description of the applied access restrictions (to the conversion file, the media file, to annotation file, the source).
Language	Description of the language (independent of its use in the session or the participant's familiarity with it).

A Descriptions schema contains a prose description. The description serves as a reminder of the circumstances of data collection and is displayed whenever you access the relevant session in the IMDI Browser (see section 4.4.2).

It is structured as follows:



Language The language in which the **description is written**. The language has to be entered in a standard format. Do the following:

1. Click with the mouse button into the ID field and press the key SHIFT+R (or the key SHIFT+I). A dummy code “RFC1766:x-sil-aaa” (or “ISO639:aaa”) is displayed.

2. Replace the last three characters "aaa" with the proper language identifier (see "Appendix 4: List of languages and language abbreviations" for a list of language identifiers).

Text — A prose description.

> **!** Note: The description in the text field may get distorted, as new lines and tabs are sometimes automatically inserted by the XML parser. We hope to fix this problem in a future version of the IMDI Editor.

Link — A link to an info file that contains information relevant to the specific topic.

Enter the name of the info file. Please do **not** make use of the "open folder" icon to the right of that field (unless you want to manage your own corpus, see part IV).

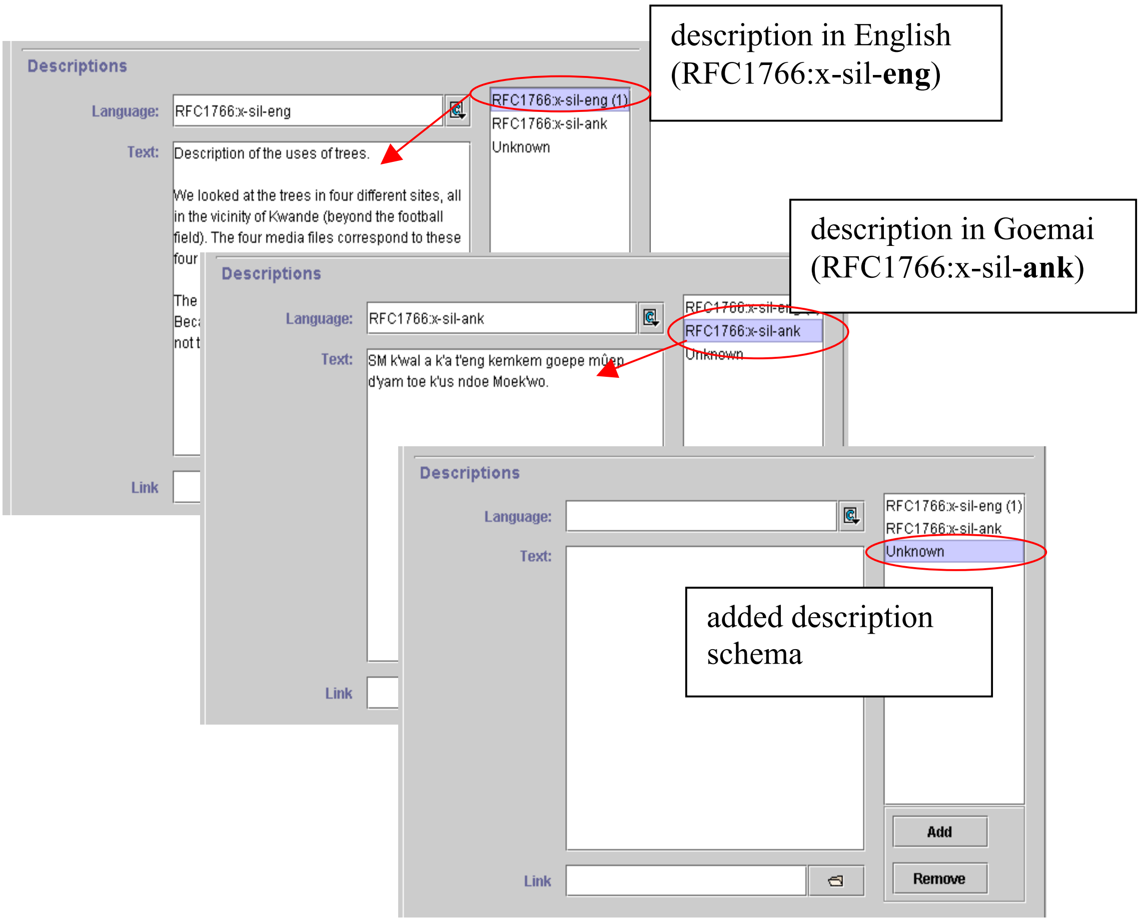
> **!** Note: Remember to include only session-specific information (e.g., session protocols, digitized photos, sensitive information) (see section 4.4.3).

> **!** Note: The info file has to be saved as either a ***.txt** file, a ***.pdf** file or an ***.html** file. The IMDI Browser cannot display any other formats. If you need help with converting files into a ***.txt**, ***.pdf** or ***.html** format, please contact the corpus manager (e.g., corpus.manager@mpi.nl).

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 identifier of the language (as defined in the Language field).
    - To access a schema, highlight its identifier 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 its identifier by clicking on it. Then click on the "Remove" button to delete it.

The screenshot below illustrates the available possibilities.



! Note: If the content of the Text field is not visible, click on the language identifier to the right of the window to highlight it.

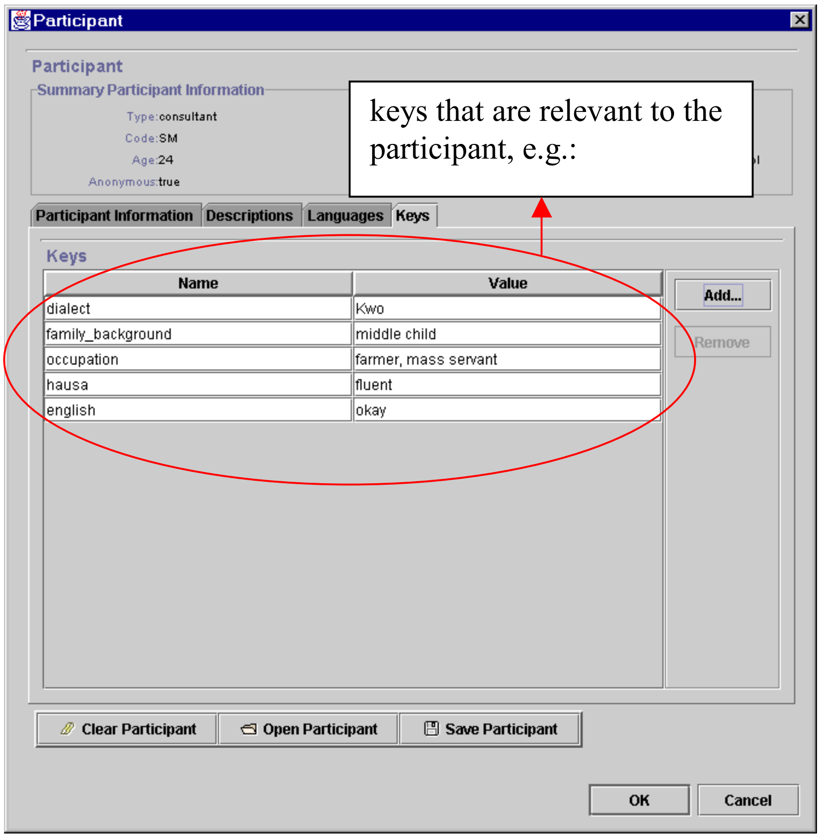
! Note: When you remove the last available language identifier from the list, the fields Language, Text and Link will disappear. To make them visible again, you have to close the Editor and re-open it.

Keys

The following screens contain a Keys schema:

General	Keywords that are relevant to (a) the collection of the data or (b) the creation of the IMDI file.
Content	Keywords that are relevant to the content of the session.
Participant	Keywords that are relevant to the individual participant (independent of his/her role in the session).

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 (see section 4.4.2). For example:



To add a key, do the following:

1. Click on the "Add" button. A new key will be added.
2. Specify a "name" for the key. For technical reasons, the name has to be one single word (e.g., 'MetaDescriptionCreator'). Do not use blank spaces.
3. Specify a "value" for the key. Multiple words are allowed (e.g., 'Student Assistant XY').

To delete a key, highlight it by clicking on it and then click on the "Remove" button.

Access

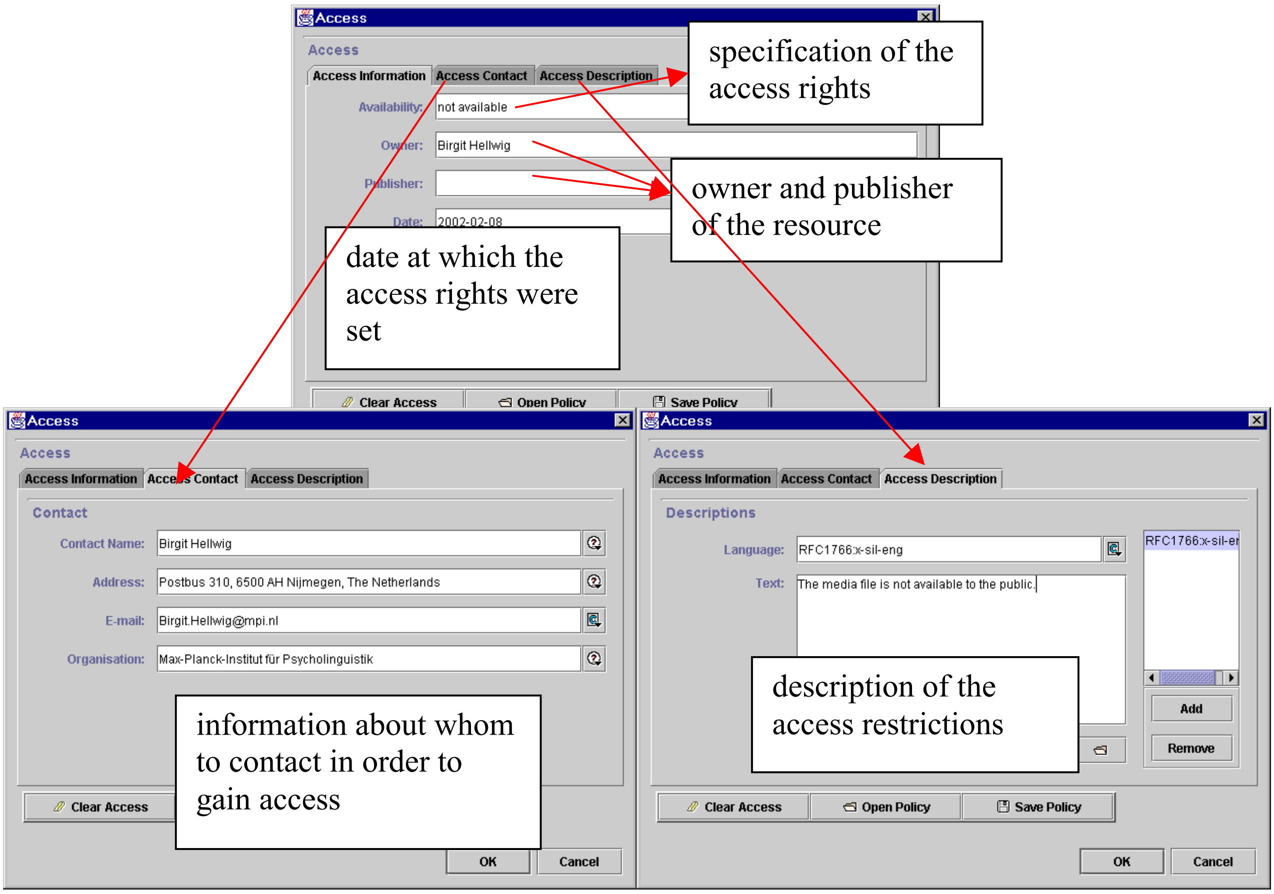
The following screens contain an Access schema:

Resources/ Anonymous	Please ignore this screen for the moment.
Resources/ Media File	Access rights to the media file(s).
Resources/ Annotation Unit	Access rights to the annotation file(s).
Resources/ Source	Access rights to the source(s).

Access schemata group information about access rights to a resource.

! Note: For the moment, all access to the actual data (media files, annotation and transcript files, sources) is automatically denied to the outside world. For future developments, we would nevertheless ask you to provide the required information.

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 who 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 institution.
Descriptions	Prose description of the access restrictions.

! Note: The information entered into an Access schema is not immediately visible after you have closed and re-opened an IMDI file. The information is not lost – it will become visible once you click on the “Edit …” button.

Appendix 3: Lists of recommended vocabularies

This appendix contains lists of recommended vocabularies. Although there is room for variation, we recommend that you keep as closely as possible to the standardized lists as this will facilitate the search process.

1. Content: Tasks

A set of names for experiments, elicitation tools and matching games that are commonly used at the Max Planck Institute for Psycholinguistics, Nijmegen (to be used on the screen Content/ Content Type/ Tasks). Note that the list is not closed, i.e., you can add other tasks.

name	comments
"info-kiosk"	
"wizard-of-oz"	
"travel-planning"	
"room reservation"	
"frog story"	

2. Content: Modalities

A set of recommended names for the modalities (to be used on the screen Content/ Content Type/ Modalities). Note that the list is not closed, i.e., you can add other modalities.

name	comments
"speech"	
"writing"	
"gestures"	
"pointing-gestures"	
"signs"	
"eye-gaze"	
"facial-expressions"	
"emotional-state"	
"haptics"	

3. Content: CommunicationContext: Interactivity

A set of names for the level of participant interaction (to be used on the screen Content/ CommunicationContext/ Interactivity). Note that this list is closed, i.e., you cannot add other items.

name	comments
"interactive"	a verbal 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 • procedural texts
"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

4. Content: CommunicationContext: Planning Type

A set of names for the degree of planning through the consultant (to be used on the screen Content/ CommunicationContext/ Planning Type). Note that this list is closed, i.e., you cannot add other items.

name	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.

5. *Content: CommunicationContext: Involvement*

A set of names for the degree of involvement of the researcher (to be used on the screen Content/ CommunicationContext/ Involvement). Note that this list is closed, i.e., you cannot add other items.

name	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) 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)

6. *Content: Interactional Genre*

A set of names for interactional genres (to be used on the screen Content/ Content Type/ Interactional). Note that the list is not closed, i.e., you can add other interactional genres.

name	comments
"conversation"	
"verbal-contest"	this also includes debates.
"interview"	
"meeting"	
"riddles"	
"consultation"	this does not refer to an interview with the investigator, but rather to, e.g. a visit to a shaman
"greetings"	
"leavetakings"	
"humor"	

7. *Content: Discursive Genre*

A set of names for discursive genres (to be used on the screen Content/ Content Type/ Discursive). Note that the list is not closed, i.e., you can add other discursive genres.

name	comments
"procedure"	a directive description of the procedures involved in the preparation or production of something, e.g., 'how to make tortillas'
"explanation"	practical, theoretical, or historical reality statements, e.g., 'how the monkey got its tail'

8. Content: Performance Genre

A set of names for performative genres (to be used on the screen Content/ Content Type/ Performance). Note that the list is not closed, i.e., you can add other performative genres.

name	comments
“oratory”	using speech effectively in a conventionalized format to address an audience within political, legal, ceremonial, or religious settings
“oral-history”	an account of firsthand experience, recalled retrospectively and communicated to an interviewer for historical purposes
“historical-narrative”	a secondhand account of the experience of historical figures and events, which may be partly or wholly fictional, communicated to both locals and outsiders for both historical purposes and entertainment
“narrative”	a recounting of one or more fictional events by one or more narrators to an audience of at least one
“oral-poetry”	spoken or sung, in a relatively structured form (in prosody and syntax), often with distinctive language, e.g. oral epics, narrative poetry, ballads (shorter, lyrical narratives), and panegyric odes
“song”	a tune with recognizably structured lyrics, e.g., popular and love songs, lullabies
“proverb”	a summary of the wisdom of collective experience, often one line long; formulaic
“lament”	
“insult”	an insolent verbal act creating animosity

Appendix 4: List of languages and language abbreviations

This appendix lists some language names and identifiers (as they are used in the Ethnologue). Please use these standardized names/identifiers in all fields of the IMDI Editor that ask for the name/identifier of a language. If a language name and identifier is not in this list, please look it up under www.ethnologue.com/web.asp.

name	identifier	comment
American Sign Language	ASE	
Arabic, Moroccan (spoken)	ARY	
Arrernte, Eastern	AER	
Basque	BSQ	
Belhariya	BYW	Belhare
Bora	BOA	Mirana
Chontal de Oaxaca - Costa	CLO	
Czech	CZC	
Dutch	DUT	
Dutch Sign Language	DSE	
English	ENG	
Ewe	EWE	
Farsi (Western)	PES	
Finnish	FIN	
French	FRE	
Fulfulde Fuuta Jalon	FUF	
Galician	GLE	
German	GER	
German Sign Language	GSG	
Goemai	ANK	
Guguyimidjir	KKY	Guugu Yimithirr
Hausa	HUA	
Hebrew	HBR	
Hindi	HND	
Italian	ITN	
Inuktitut, Greenlandic	ESG	
Inuktitut, Eastern Canadian	ESB	
Japanese	JPN	
Kgalagadi	XKV	
Katang	KGD	
Khmer (Central)	KMR	
Kilivila	KIJ	
Korean	KKN	
Kota	KFE/KOQ?	
Lao	NOL	
Longgu	LGU	
Marquesan, North	MRQ	
Maya, Yucatán	YUA	Yukatek
Maya Mopán	MOP	
Nicaraguan Sign Language	NCS	
Popoluca, Oluta	PLO	
Portuguese	POR	
Punjabi		(not in Ethnologue)
Russian	RUS	
Saliba	SBE	

Samoan	SMY	
Sekpele	LIP	Likpe
Spanish	SPN	
Swedish	SWD	
Tamil	TCV	
Tidore	TVO	
Tongan	TOV	
Totonaca, Papantla	TOP	
Turkish Sign Language	TSM	
Trumaí	TPY	
Turkish	TRK	
Tzeltal, Oxchuc	TZH	
Tzeltal, Tenejapa	TZL	(not in Ethnologue)
Tzotzil, Zinacantán	TZZ	
Yalunka	YAL	
Yele	YLE	Yélî Dnye
Zapotec, Zoogocho	ZPQ	

Appendix 5: Digitization Policy

1. Contact dobesarc@mpi.nl with your digitization wishes. A digital master file (DMF) will be created from your master tape and placed in the appropriate network folder.

2. Go through the DMF and use the Media Player program to identify the relevant units of analysis (i.e., the sessions). Make a note of where each session begins and ends (enter the information on the screen Resources/ Media Files/ Media File/ Time Position).

3. Create a metadata file for each session on your tape by using the IMDI Editor.

4. Submit the metadata file to dobesarc@mpi.nl for the digital segmentation and conversion of your data.

> **!** Note: Only data that is accompanied by metadata description files will be further processed by the technical group. If you do not plan on meta-describing your data, the DMF will be saved on a DLT tape that will then be given to you for storage.

Appendix 6: Tape labeling conventions

A label, containing the ID, written in block letters, and readable by everyone must be put on the media. The ID has to contain the following characters.

- one character to identify the team or group:

  L (Language and Cognition)
  A (Acquisition)

- two characters to identify the researcher
- one character to identify the sub-project of the researcher

  Language and Cognition Group:
  S (Space)
  G (Gesture)
  E (Event)

  Acquisition Group:
  A (Argument Structure)
  F (Finiteness)
  S (Second Language Acquisition)

- three characters to identify the type:
  - First position:

    A (Audio)
    V (Video)
    P (Photo)
    C (CD)

  - Second position:

    in case of Audio:
    C (audio cassette)
    T (reel-to-reel-tape)
    D (DAT tape)
    M (minidisk)

    in case of Video:
    D (digital video)
    8 (Hi-8)
    S ((S)VHS)
    U (U-matic)

    in case of Photo:
    D (digital)
    A (analog)

  - Third position (in case of video):

    N (NTSC)
    P (PAL)

- seven characters to specify the date of recording: 2 digits, 3 letters and 2 digits, e.g. 09Oct00
- two digits to identify the number of the tape/photo (collected on that date)
- one digit to indicate if it is a copy: “C” (otherwise omit it)
- one digit to indicate if the copy is identical to the original: “I” (otherwise omit it)

For Example, the tape label “ASECV8N09OCT0002CI” has the following meaning: