

ANNOTATOR'S WORKBENCH CONFIGURATION MANUAL

THIS PRODUCT INCLUDES SOFTWARE DEVELOPED BY THE ETHNOGRAPHIC VIDEO FOR INSTRUCTION AND ANALYSIS DIGITAL ARCHIVE PROJECT BY INDIANA UNIVERSITY AND UNIVERSITY OF MICHIGAN FUNDED BY A GRANT FROM THE ANDREW W. MELLON FOUNDATION. FOR FURTHER INFORMATION CONTACT THE INSTITUTE FOR DIGITAL ARTS AND HUMANITIES AT INDIANA UNIVERSITY, IDAH@INDIANA.EDU

CONFIGURATION FILES

There are several files that are used by the Annotator's Workbench to configure the layout and available features of the AWB. These files are all XML files and interact with each other to determine the layout of the AWB. On Windows these are found in the *conf* directory of the AWB installation and on Macs, these are found by *control-click or right-click* the AWB icon and looking in the *working/conf*.

The Main Configuration File: awbconfig.cfx

The file *awbconfig.cfx* is the main configuration file. This file is used to determine the behavior of the AWB and to tell the AWB where other configuration files are located. The directory structures shown are for a Apple Macintosh OSX 10.6 computer. The file structures for Windows would conform to the Windows standard. File locations are all relative to the directory containing the AWB executable file. There are two main areas to this file, <parameters> and <state>.

PARAMETERS SAMPLE

```
<parametersp=yt31 `Ad/
  <projectDirectory>/Users/Directory/Desktop/Demo/awb_demo</projectDirectory>
  <cvconfigFile>conf/cvconfig.xml</cvconfigFile>
  <logFilePattern>log/awb%u.log</logFilePattern>
  <toolsDirectory>tools</toolsDirectory>
  <termMappingFile>conf/cvtermmapping.props</termMappingFile>
  <categoryMappingFile>conf/cvcategorymapping.props</categoryMappingFile>
  <loggingLevel>INFO</loggingLevel>
  <volume>0.26</volume>
  <autosaveDirectory>data</autosaveDirectory>
  <cvSetFile>data/sets.sar</cvSetFile>
  <mediaLocations>
    <localDirectory>/Users/Directory/Desktop/Demo/awb_demo</localDirectory>
  </mediaLocations>
  <layoutFiles>
    <file>conf/editing.xml</file>
    <file>conf/playback.xml</file>
    <file>conf/standby.xml</file>
  </layoutFiles>
</parameters>
```

<PROJECTDIRECTORY>

This parameter contains the file location that will initially open in the file dialog box when you choose to open a file from the File menu. For PC's this tag would contain something like "c:\Users\username\demo\awb_directory".

<CVCONFIGFILE>

This is the location of the controlled vocabulary configuration file. This file determines how you connect to the database, either through a web service or using a username and directly connecting to the database. This configuration file has all the information about where the database is located and the username and password needed to connect to it. If connecting directly to the database, we recommend using a username and password that has read-only access to the database. More on the cvconfigFile later.

<LOGFILEPATTERN>

Since multiple log files will be generated, this parameter determines the naming convention to be used for the files. The % values are determined by the Java file handler and can be:

- "%t" the system temporary directory
- "%h" the value of the "user.home" system property
- "%g" the generation number to distinguish rotated logs
- "%u" a unique number to resolve conflicts
- "%%" translates to a single percent sign "%"

<TOOLSDIRECTORY>

Upon occasion there are additional .jar files needed for special conversions or temporary fixes to problems. These .jar files are kept in the directory pointed to by tools.

<TERMAPPINGFILE>

For display purposes, the literal value of the term in the Controlled Vocabulary (see below) database needs to be changed (see below for more information about Controlled Vocabulary). This file contains those transformations to make terms that may be defined differently in an authority source display better in the AWB.

<CATEGORYMAPPINGFILE>

This file does the same as the <termMappingFile> except at the category or thesaurus level. It allows you to change the displayed name of a Controlled Vocabulary category or thesaurus to a simpler or more common name.

<LOGGINGLEVEL>

The logging levels are determined by the values for the Java logger:

- SEVERE (highest value)
- WARNING
- INFO
- CONFIG
- FINE
- FINER
- FINEST (lowest value)

The default value is “INFO”.

<VOLUME>

Volume level when you start the video playback window.

<AUTOSAVEDIRECTORY>

This directory is where the automatic save file is stored. This file is generated everytime a change is made to information in the project file by the AWB. In addition this file is used for automatic recovery if necessary.

<CVSETFILE>

If the feature to create “sets” of controlled vocabulary terms is turned on, then the resulting sets are stored in this file.

<MEDIALOCATIONS>

The element <localDirectory> contains the directory names of the last several locations that were used for the video files. The project file only contains the name of the video file. When you load a project in AWB, it searches the <localDirectory> entries to try to find the video files. If none of the <localDirectory> entries contains the file, then a file dialog box opens and it asks for the location of the video file. The AWB then checks in that directory as well as the <localDirectory> entries to try to find the rest of the video files if necessary until it has found all video files.

<LAYOUTFILES>

The element <file> contains the location of three standard, default files used for the layout of windows and the available features for the AWB. There is a detailed description of each of these files later in the document. These files represent three modes.

The *standby.xml* file represents when the AWB is first loaded before a file has been created or selected. By modifying the standby.xml file, you can determine which menu items are available at startup. For instance, you may not want someone to be able to

create new files but only load existing files. By changing the *standby.xml* file, you can eliminate the option to create new files from the *File* menu.

The *editing.xml* file is the default file once a file has been created or loaded. This controls the positioning of windows, the availability of features and the items that appear in menus. By changing this file you can create a flexible interface or a more rigid interface. For instance, you may want someone to be able to add annotations to video files but not to change the start and stop positions of the video segments. That can be accomplished with this file.

The *playback.xml* file controls the application in Preview mode, a special mode to allow reviewers to comment on video segmentation and annotation without making modifications. This file controls the layout and feature availability for that mode.

The actual configuration of these files will be discussed later in this document.

STATE

The `<state>` elements store `<recentlyOpenedFile>` entries for the File menu. It keeps track of the last 5 project files that were opened.

Controlled Vocabulary Configuration Files

If you are interested in using controlled vocabularies with the Annotator's Workbench, you will need to implement the open source CV Maintenance Tool which will allow you to create hierarchical controlled vocabularies. (The CV Maintenance Tool is available separately.) There are three files that control the use and configuration of Controlled Vocabulary in the AWB: the *cvconfig.xml* file, the *cvtermmapping.props* file and the *cvcategorymapping.props* file. The *awbconfig.cfx* file determines the names and locations of these files. In general, controlled vocabularies are implemented on large institutional projects because they require some technical support. The configuration files mentioned here would allow you to link to the controlled vocabularies that you created with this tool.

THE CVCONFIG.XML FILE

This file controls how the AWB connects to the Controlled Vocabulary database, discussed in the previous paragraph, to populate the controlled vocabulary terms. The use of a controlled vocabulary is optional with the AWB and can be disabled in the *editing.xml* file. A sample *cvconfig.xml* file follows:

```
<cvconfig:CVConfiguration
xmlns:cvconfig="http://www.dlib.indiana.edu/xml/CVConfiguration/version1.0/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

```

xsi:schemaLocation="http://www.dlib.indiana.edu/xml/CVConfiguration/version1.0/CVConfiguration.xsd"
>
<categories>
  <category>
    <name>Genres</name>
    <thesaurusId>1</thesaurusId>
  </category>
  <category>
    <name>Languages</name>
    <thesaurusId>2</thesaurusId>
  </category>
</categories>
<repository>
  <identifier>AWB_CV</identifier>
  <!--
    javaClass is the MySQLThesaurusRepository class,
    jdbcUrl is the url to the CV database, username/password
    are the account credentials to connect to the database
    <javaClass>edu.indiana.shared.cv.database.MySQLDBThesaurusRepository</javaClass>
    <jdbcUrl>jdbc:mysql://localhost:3306/cv_test</jdbcUrl>
    <username>cvadmin</username>
    <password>cvadmin</password>
    <cacheDate>2007-08-06T16:03:49-04:00</cacheDate>
  -->
  <!-- web service url should point to production server -->
  <webServiceUrl>http://brie.dlib.indiana.edu:8080/cv/cv_cache.html</webServiceUrl>
  <termInclusionCondition>
    <or>
      <fieldCheck>
        <name>Status</name>
        <value>Approved</value>
      </fieldCheck>
      <fieldCheck>
        <name>Status</name>
        <value>Pending SACO</value>
      </fieldCheck>
    </or>
  </termInclusionCondition>
</repository>
<languageCategoryThesaurusId>2</languageCategoryThesaurusId>
<defaultLanguage>English Language</defaultLanguage>
<cacheFile>data/repository.cvt</cacheFile>
</cvconfig:CVConfiguration>

```

<CATEGORIES>

Within this element are the <category> elements. These elements contain the name and the specific id of the category in the Controlled Vocabulary database that is created and maintained by the Controlled Vocabulary Maintenance Tool. See the Controlled Vocabulary System Guide for details of how to use that tool. The <name> element must

contain the same spelling of the category as in the database and the <thesaurusId> must have the id value. If you want to modify the name for display purposes, use the <categoryMappingFile>. Because only the categories listed in this file will be loaded, it is possible to have a category in process but not yet complete in the Controlled Vocabulary Database that is not then loaded into the AWB. To have it loaded into the AWB when complete, just add the name and id to this tag.

<REPOSITORY>

The <identifier> tag contains the name of the CV database. This is entered at the time you create the database with the CV tool and it must be the same as the name you entered during database creation. After creation of the database, this identifier name is contained in the *identification* table in the *database_id* column.

This element contains two ways to connect to the repository. The first is by directly connecting to the database with jdbc, Java Database Connectivity. This method requires the embedding of a username and password that has at least read only privileges on the Controlled Vocabulary Database. Since the password is not encrypted in the file, we do not recommend any other privileges than read only for this user.

Since providing a password may be against the security policies of some institutions, another way to retrieve Controlled Vocabulary terms is through a web service. This technique does not require the embedding of a password in a plain text file and is the preferred method of connection. The web service is part of the available source code for the AWB and can be modified as needed.

<TERMINCLUDECONDITION>

Without any term inclusion conditions, all terms will be added to the controlled vocabulary in the AWB. The term inclusion conditions gives you the opportunity to add only, for instance, those terms that have a status of approved, as in the sample code included with the default file. This allows you to withhold specific terms which may be pending, awaiting final approval. The <termInclusionCondition> element can also contain <or> and <and> elements to allow more complex boolean expressions.

<LANGUAGECATEGORYTHESAURUSID>

This element allows you to specify where the drop down list for translation languages will come from in the transcription area. You can use any category, whether it is displayed in the AWB or not. This allows you to develop specific lists for translation languages if you want.

<DEFAULTLANGUAGE>

You can also choose a default translation language from the list of languages in the category specified in the <languageCategoryThesaurusId> element.

<CACHEFILE>

When a successful connection is made to the Controlled Vocabulary database, the CV terms are downloaded to a cached file. If you cannot make a successful connection in the future, then this file is used for CV terms in the AWB. In fact if you are unable to make a successful connection to the CV database, you can import a copy of this file from another source to update the Controlled Vocabulary in the AWB instead of connecting to the database through either jdbc or a web service.

The Layout Files

See the table at the end of this section for a summary of the layout options.

Each layout file contains information about three areas of the AWB.

COMPONENTS

1. Components are the actual elements that appear on the desktop, such as the Timeline or the Media Playback window.
2. Components are contained in layout elements that assist with the positioning of the component on the desktop. These layout elements are:
 - a. `formattingBlock` – allows you to divide the desktop into larger display areas, such a left or right and top or bottom. Components can then be placed in these `formattingBlocks`
 - b. `componentBlock` – allows you to position and size a specific component on the desktop or within a `formattingBlock`. A `componentBlock` is attached to the top, bottom, right, or left of the desktop or a `formattingBlock` and can not be freely moved. The Timeline is an example of a component Block attached to the bottom of the desktop.
 - c. `floatingComponentBlock` – allows you to define an area of the desktop as containing `floatingComponents`, i.e. components that are freely moveable within the desktop or the `floatingComponentBlock`.
 - d. `floatingComponent` – a component that can be moved freely within the desktop or a `floatingComponentBlock`
 - e. `metadataWindowArea` – the metadata window is a special component and can be independently defined with the `metadataWindowArea` component.

FEATURES

3. Features usually either appear in the menus, such as “load_project” or “save_project” or they allow certain activities within the application such as “create_segments” or “add_scenes_to_events”.
4. Features are set to “true” or “false”.

FIELDS

5. Appear in the metadata window and can be set to “read_write”, “read_only”, or “hidden”. Read_write fields can be updated; read_only fields can be viewed but not changed and hidden fields do not appear at all.
6. If all the fields on a given tab in the metadata window are set to “hidden” then that tab does not appear either.

Component	Component Layout
	within
MEDIA_PLAYBACK	componentBlock or floatingComponent
TIMELINE	componentBlock
TRANSCRIPTION_PREVIEW	componentBlock
METADATA_PREVIEW	componentBlock
COLLECTION_HIERARCHY	componentBlock or floatingComponent
CONTROLLED_VOCABULARY	componentBlock or floatingComponent
SETS	componentBlock or floatingComponent
GLOSSARY	componentBlock or floatingComponent
BIBLIOGRAPHY	componentBlock or floatingComponent
METADATA WINDOW AREA	metadataWindowArea
Feature	Feature Values
LOAD_PROJECT	TRUE
PREVIEW_PROJECT	FALSE
SAVE_PROJECT	
NEW_PROJECT	
IMPORT_LEGACY_PROJECT	
CREATE_HTML_SUMMARY	
VIEW_PROJECT_PROPERTIES	
NEW_REVIEWER_COMMENTS	
OPEN_REVIEWER_COMMENTS	
VIEW_REVIEWER_COMMENTS	
EXPORT_FOR_COPYEDITING	
APPLY_COPYEDITING	
MULTIPLE_COMMENTS_FILES	
CREATE_SEGMENTS	

DELETE_SEGMENTS	
RESIZE_SEGMENTS	
EDIT_SEGMENT_METADATA	
NOW_PLAYING_INDICATORS	
PLAYBACK_MODE_CHOICES	
TEXT_FORMATTING	
SEGMENT_LINKING	
GLOSSARY_LINKING	
CITATION_LINKING	
MULTIPLE_METADATA_WINDOWS	
ADD_SCENES_TO_EVENTS	
ADD_ACTIONS_TO_SCENES	
REMOTE_SAVE	
MODIFY_TRANSCRIPTIONS	
SPELL_CHECKER	
VARIABLE_SPEED_SLIDER	
CV_WEBSERVICE	
Field	Field Values
COLLECTION_ARCHIVE	READ_WRITE
COLLECTION_TITLE	READ_ONLY
COLLECTION_ARCHIVE_TITLE	HIDDEN
COLLECTOR_NAME	
COLLECTOR_BIO	
COLLECTION_COPYRIGHT HOLDER	
COLLECTION_MATERIALS_ACQUISITION_NOTES	
COLLECTION_CONTRIBUTOR_NOTES	
COLLECTION_RECORDING_DATES	
COLLECTION_DESCRIPTION	
COLLECTION_CV	
PARTICIPANTS_INDIVIDUAL	
PARTICIPANTS_GROUP	
TECHNICAL_PROBLEMS	
TRANSCRIPTIONS	
SUBSEGMENT_TITLE	
SUBSEGMENT_RECORDING_DATES	
SUBSEGMENT_BRIEF_DESCRIPTION	
SUBSEGMENT_DETAILED_DESCRIPTION	
SUBSEGMENT_BLOCKED_CONTENT	
SUBSEGMENT_CV	
EVENT_QUALITY_RATING	
SCENE_QUALITY_RATING	
ACTION_QUALITY_RATING	

TABLE OF COMPONENTS, FEATURES AND FIELDS

The following table shows which layout components can be contained in other layout components. The root level of layout is called <componentLayout>. These component layouts are used primarily with Editing mode and Playback mode since there is minimal layout with Standby mode. For certain layout components, the table shows some sample values to give some context for the layout possibilities.

Layout Component	Consists of one of more of the following
componentLayout	
	formattingBlock
	componentBlock
	floatingComponentBlock
	metadataWindowArea
formattingBlock <formattingBlock> <position> <direction>LEFT</direction> <percentOfParent>0.35</percentOfParent> <minimumPixels>350</minimumPixels> </position> </formattingBlock>	
	componentBlock
	floatingComponentBlock
	metadataWindowArea
componentBlock <componentBlock component="COMPONENT"> <position> <direction>BOTTOM</direction> <minimumPixels>150</minimumPixels> <maximumPixels>150</maximumPixels> </position> </componentBlock>	Component
floatingComponentBlock	
	floatingComponent
	metadataWindowArea
floatingComponent <floatingComponent component="COMPONENT"> <xOffsetPercent>0.0</xOffsetPercent> <yOffsetPercent>0.0</yOffsetPercent> <heightPercent>0.5</heightPercent> <widthPercent>0.3</widthPercent> </floatingComponent>	Component

<pre> metadataWindowArea <metadataWindowArea> <xOffsetPercent>0.32</xOffsetPercent> <yOffsetPercent>0.0</yOffsetPercent> <widthPercent>0.43</widthPercent> <heightPercent>0.98</heightPercent> </metadataWindowArea> </pre>	
---	--

LAYOUT COMPONENTS

There are certain features that cannot be used by certain modes, standby, editing and playback. Listed below are those prohibited features for each mode:

Mode	Mode Prohibited Features
EDITING	NEW_REVIEWER_COMMENTS
	OPEN_REVIEWER_COMMENTS
	VIEW_REVIEWER_COMMENTS
	MULTIPLE_COMMENTS_FILES
PLAYBACK	ADD_ACTIONS_TO_SCENES
	ADD_SCENES_TO_EVENTS
	CREATE_SEGMENTS
	DELETE_SEGMENTS
	EDIT_SEGMENT_METADATA
	RESIZE_SEGMENTS
	MODIFY_TRANSCRIPTIONS
	CITATION_LINKING
	GLOSSARY_LINKING
	SEGMENT_LINKING
	TEXT_FORMATTING
	SAVE_PROJECT
	MULTIPLE_METADATA_WINDOWS
	APPLY_COPYEDITING
REMOTE_SAVE	
SAVE_PROJECT	
STANDBY	SAVE_PROJECT
	CREATE_HTML_SUMMARY
	VIEW_PROJECT_PROPERTIES
	NEW_REVIEWER_COMMENTS
	OPEN_REVIEWER_COMMENTS
	VIEW_REVIEWER_COMMENTS
	EXPORT_FOR_COPYEDITING
	APPLY_COPYEDITING
	MULTIPLE_COMMENTS_FILES
	CREATE_SEGMENTS
DELETE_SEGMENTS	
RESIZE_SEGMENTS	

	EDIT_SEGMENT_METADATA
	NOW_PLAYING_INDICATORS
	PLAYBACK_MODE_CHOICES
	TEXT_FORMATTING
	SEGMENT_LINKING
	GLOSSARY_LINKING
	CITATION_LINKING
	MULTIPLE_METADATA_WINDOWS
	ADD_SCENES_TO_EVENTS
	ADD_ACTIONS_TO_SCENES
	MODIFY_TRANSCRIPTIONS

PROHIBITED FEATURES FOR EACH MODE

STANDBY.XML

This file is used for initially loading the desktop on startup, before any files have been loaded or created. Normally this mode would contain just enough features to load or create a file.

EDITING.XML

This file is used to configure the AWB during normal use. All components can be used, all layouts can be used and nearly all features can be used. The only restrictions on features are those that features dealing with playback mode or perhaps better called, peer-review mode.

PLAYBACK.XML

This configuration is used for playback or peer-review. Segments cannot be edited or changed but the reviewer can add comments about each segment. These comments are saved in a separate file and can then be given to the original annotator to review.