

NCI Protege 5 History Processing

The publishable history files will be located at /priv-file-repo/protege/archive/Production-History. The source history files will be located in /priv-file-repo /PROD/NCIT/<hash>/<datetime>. The operations team will be responsible for updating the publishable history directory with the appropriate source files and generating a history file that can be loaded into LexEVS.

The protege editors are responsible for generating the source history files. Editors do a review and "squash" of edits weekly, which will generate a squash folder with two temporary history files. Those are not used for producing the publishable history. At end of month the workflow manager will do a full export of history. This will generate the datetime folder and the 3 history files we are interested in.

- ncitevs_history - this is all the raw history edits since the last export
- cumulative_ncitevs_history - this is all the raw history edits since history was first recorded, including the current month
- nцитconcept_history - this is a scrubbed history with editor names and duplicate modify records removed.

```
/priv-file-repo/protege/archive
|_PROD
|   |_NCIT
|       |_hash1
|           |_datetime (monthly history)
|               |_cumulative_ncitevs_history
|               |_ncitconcept_history
|               |_ncitevs_history
|           |_squash-datetime (weekly history)
|               |_history
|               |_history-snapshot
|           |...
|           |_hash2
|           |...
|_Production-History
|   |_cumulative_history.txt
|   |_cumulative_ncitconcept_history
|   |_cumulative_ncitevs_history
```



EVS History

The cumulative_ncitevs_history is the detailed history written for internal purposes. It includes editor names and the exact date of any and all modifications to the concepts. Protege creates a full, cumulative EVS history with every history export. This file needs to merely be copied from the source directory to the Production-History, overwriting the file that is there.

Concept History

The ncitconcept_history is a scrubbed version of history that removes editor names and eliminates duplicate modifies within the same month. While internal review might find it important that a concept has been edited 5 times in a month, external users are only informed that there have been modifications since the last release. The monthly history process examines the raw ncitevs_history and throws out all of the extra, uninformative modify records. The resulting ncitconcept_history only shows records since the last history export. It needs to be manually appended to the cumulative_ncitconcept_history in the Production_History folder.

```
cp priv-file-repo/protege/archive/PROD/NCIT/<hashversion>/<datetime>/* /priv-file-repo/protege/archive/Production-History
cp cumulative_ncitconcept_history cumulative_ncitconcept_history.bak
cat ncitconcept_history >> cumulative_ncitconcept_history
```

The cumulative_ncitconcept_history must be reformatted in order to be loadable into the LexEVS terminology server. This is done using Format_History on the processing server. When formatting is complete, the output file "concept_history.txt" should be copied back to the Production-History folder as a backup.