## Database properties

There is a file with properties used by Kettle where you can set the values of the database properties.

(Unix/Linux/OSX)

$HOME/.kettle

(Windows)

C:\Documents and Settings\<username>\.kettle\ (Windows)

Mine looks like this

# This file was generated by Pentaho Data Integration version 3.2.0.

#

# Here are a few examples of variables to set:

#

# PRODUCTION\_SERVER = hercules

# TEST\_SERVER = zeus

# DEVELOPMENT\_SERVER = thor

#

# Note: lines like these with a # in front of it are comments

#

catissue.db.host = localhost

catissue.db.port = 3306

catissue.db.name = catissuecore12a

catissue.db.username = catissuedba

catissue.db.password = secret

cbm.db.host = localhost

cbm.db.port = 3306

cbm.db.name = CBM

cbm.db.username = cbmdba

cbm.db.password = verysecret

## Organization

This now runs as a Kettle job rather than a single transformation. The job runs a number of transformations in sequence. Some of the tables need to be loaded before other so that data required by foreign keys is present. In an transformation the steps run in parallel so there's no guarantee of the sequence in which each table will be loaded.

The job file is Organization.kjb

It runs

Organization.ktr

Organization1.ktr

Organization2.ktr

The tables loaded by each transformation can be seen by looking at the transformation.

As previously the database parameters are set as variables in the .kettle file.

## Collection protocol

This also now runs as a job collectionProtocol data.kjb. This job runs three transformations.

Profiles.ktr – loads Annotation Availability profiles and Specimen Availability profiles.

The two collection protocol steps create first the

CollectionProtocol.ktr

Creates the collection protocols

CollectionProtocol2.ktr

Maps collection protocols to institutions.

### Profiles

This transformation currently creates profiles with all possible combinations of options. Not all such combinations may be necessary and sites may wish to create only those that are needed.

### Assigning profiles to Collection Protocols

caTissue does not contain data from which it is universally possible to extrapolate what the Annotation Availability profile or Specimen Availability profile should be for any particular collection protocol. The collection protocol transformation here (CollectionProtocol.ktr) assigns to all collection protocols a single Annotation Availability profile and a single Specimen Availability profile. Sites will have to determine which is the appropriate single protocol in each case, or the appropriate logic to assign different profiles to each collection protocol.

Note: In the case of Annotation Availability profiles it may be possible to derive an appropriate profile from the annotation that is used on patients and specimens registered to the Collection Protocol.

In some cases Specimen Availability profiles could be derived from the consent data in caTissue, perhaps also using the security data in caTissue.

It is likely, however, to be simpler to consider each protocol and assign appropriate profiles in each case.

## Lookup tables

In this revision of the CBM the lists of values for various attributes have been defined as part of the model. For convenience of loading data into the CBM database schema these lists have been loaded into tables in the database. Most of these tables are not included in the object model. The tables used by the object model are, however, linked by foreign keys to the list of values tables.

The state of mapping caTissue permissible values to the lists used in the CBM is addressed as follows

## Diagnosis

Diagnoses in caTissue map well to the CBM values.

## Anatomic source

The following discussion is based on the Tissue Site.xls file which comes from the caTissue Compatibility Review Package. It was helpful to load this table into the database to query the data more effectively and to join the data to the caTissue. It is also useful to use it in a transformation script.

This and two other tables useful for mapping purposes are created by mappings.sql. A transformation to load them is available as mappings.ktr.

In the CBM we have an anatomic source of "Lip" which has been mapped to an NCI code C12220. This is the code for "Lip (NOS)' i.e. Not Otherwise Specified.

In caTissue there are more fine detailed anatomic sources for Lip. In CBM we would want to use the more general term to map all of them. In the case of this group of values the appropriate mapping could be done by mapping the parent column of caTissue Site.xls to the truncated ICD code of the "Lip (NOS)” item i.e. the one that is mapped to the CBM list.

However, this approach would not work in all cases. Some of the