**Hanvey Names Workflow**

**Background**

The authoritative Hanvey Excel sheets created by AnneMarie and others often contain names. The names as they appear on the spreadsheet will appear in the metadata as the mods:title element.

The names that appear in the title will be repeated a second time in a mods:name element. (The mods:name element could be on its own or, more commonly, inside mods:subject.) Why repeat the name in the mods:name element if it already appears in the title?

The version of the name repeated in the mods:name element is the *controlled* (or *authorized*) form of the name. In a database, it is important that every person have *authorized* form of their name that is unique to that person. The person Ted Smith could appear in title as Ted Smith, or Teddy Smith, or Mr. Smith, or Theodore Smith. However, in the mods:name field, the same form of the name will always be used. This helps with searching – the searcher only has to enter one version of the name to get all records for Ted Smith. It also helps with building companion web sites. When the name is always entered in exactly the same way, you can run a script on the database to create a web page that groups together all the Ted Smith items and separately group together the items by all the other people. Authority control has become even more important now that xml metadata records can be used to create derivative products such as collection web sites.

The other thing authority control tries to do is to make sure that no two people in a database will have the same form of the name. So, if Ted Smith is in our database as *Smith, Ted* and another Ted Smith comes along, something will be added to the second Ted Smith to distinguish him from the first Ted Smith. The distinguishing additions are usually birth and/or death dates, but could be profession or something else. So the second Ted Smith would have the authorized form: *Smith, Ted, 1944-* or *Smith, Ted, musician.*

In libraries, the authorized forms of the names and the information used to tell which name goes with which person are kept track of in an authorities database. The Boston College Libraries use the OCLC authorities database. OCLC brings us WorldCat which is a catalog of books and other materials. They also bring us a separate authorities database.

You can access the database by logging onto Worldcat via connexion: <http://www.oclc.org/us/en/connexion/interface/browser/default.htm>

Log on at <http://connexion.oclc.org/>

Documentation about what all the different MARC tags, etc mean is at <http://www.oclc.org/us/en/support/documentation/worldcat/authorities/authformat/default.htm>

**Step 1:**

Meg is in charge of authority control for Hanvey names. When Rick finishes scanning a Hanvey part, Meg reviews the authoritative Excel file for that part. She looks at each name and tries to find the established form in the OCLC authority file. Sometimes this involves researching the names on the web to make sure there is a match between the Hanvey collection and the name in the OCLC authority file. When she finds the correct form for each name she adds at the a Hanvey Names Database. This is an ACCESS database and is at [\\dennis2\BHanvey\hanvey.mdb](file:///\\dennis2\BHanvey\hanvey.mdb).

Meg fills in all the relevant fields in the database except for ID which is assigned by ACCESS and code which is assigned later.

Often times with a collection like Hanvey, the person’s name doesn’t appear in the OCLC authority file. In cases like this, Meg follows a detailed set of rules (Chapter 22 of the Anglo American Cataloging rules) to “establish” a form of the name which is unique.

**Step 2:**

We generate our Hanvey METS records by running a PERL script on AnneMarie’s authoritative spreadsheets. So, somehow the authority information Meg records in ACCESS has to get connected to the Hanvey PERL script and also to the authoritative spreadsheet. This is done by use of “name codes.”

After Meg enters a bunch of new names, Betsy gives each name a unique code in the *code* column of the database. The code is usually the person’s last name, but when there are many people with the same last name, a first initial is added. The codes are all lower case and contain no spaces or special characters. The *codes* are guaranteed to be unique because the ACCESS database is set up so that it won’t allow duplicates in this column/field.

**Step 3:**

Once the codes are assigned in the ACCESS database, the codes are also added to column R of the EXCEL spreadsheet: Each title is reviewed and if a name is contained in the title it’s code is added to column S. If there are multiple codes to be added, they are separated by a **comma.**

**Step 4:**

Before generating the METS for each part, any new names have to be added to the Hanvey perl script. The ACCESS database is sorted by ID so new records appear at the end. The External Data tab is selected and the database is exported to EXCEL. In EXCEL, all rows except for those containing new names are deleted.

A PERL script is run on the export of the ACCESS database. This is a special Names PERL script that formats the EXCEL export into the mods:names format.

**Step 5:**

The new names are pasted into the names section of the Hanvey PERL script.

**Step 6:**

The Hanvey PERL script is further updated to handle any subject headings that Meg indicates should be used for every record .

**Step 7:**

The updated Hanvey PERL script is run on the authoritative xls sheet to create a METS record for each Hanvey item.

**Step 8:**

Each METS record is opened in Oxygen and reviewed for corrupt characters and potential problems with the script.