May 11, 2012

Staff Sourcing for Digitization

Rose Fortier, Marquette University

Available at: https://works.bepress.com/rose_fortier/13/
Staff Sourcing for Digitization
What does it even mean?

- **outsource** – To send out (work, for example) to an outside provider or manufacturer in order to cut costs. (*American Heritage College Dictionary, 4th ed.*)

- **staff source** – To send out work to staff members outside of Digital Projects to increase productivity. (Rose Fortier, 1st ed.)
How it works @ MPL

- It works better for the digitization of some types of materials than for others

- Digitization steps
  - Scanning
  - Image editing
  - Text creation (depending on the project)
  - Metadata creation
  - Import to content management software (we use CONTENTdm)
How it works @ MPL

• We staff source primarily for metadata creation
• Can also do it for text creation if not using Optical Character Recognition (OCR)
  • We use staff to proofread text created using OCR
Setting it up

• Need a (virtual) space to set up what staff will be working with
  • We use a common drive on the local network that is accessible by Neighborhood Library and Central Library staff

• Need the components that staff will work on
  • We have a few basic components:
    • Image
    • Text (if necessary for collection)
    • Database
    • Controlled vocabularies
Pineapple and Rhubarb Marmalade

1 cup water
1 box powdered pectin
5 cups sugar
3 tablespoons vanilla

Wash rhubarb but do not peel. Cut into thin slices. Peel pineapple, chop very fine and place in a five to six quart saucepan along with sliced rhubarb. Add water. Cover and simmer until soft about one minute. Add powdered pectin. Bring to a full rolling bone. Boil one minute at one minute and add for several minutes to prevent fruit from floating. Ladle into hot sterilized jars, leaving one-half inch space at top of jar. Seal at once. Makes about seven medium glasses.
Image editing

- Images need to be prepared before assigned to staff
- We use Photoshop but generally any basic image editor will do for minor edits
Any OCR conversion needs to be done beforehand.

We use ABBYY Finereader.

CONTENTdm has an OCR component which works well for collections that need very little proofreading.

We save files as .txt to strip out unnecessary formatting.

Some files require much more proofreading than others.

The older it is, the more work it is likely to require.

---

Before:

Recipe:

Pineapple and Rhubarb Marmalade

1 pound rhubarb
1 medium pineapple (fully ripe)
1/2 cup water
1 box powdered pectin
5 cups sugar
3 tablespoons vanilla

Wash rhubarb but do not peel. Cut into thin slices. Peel pineapple, chop very fine and place in a five to six quart saucepan along with sliced rhubarb. Add water. Cover and simmer until soft about one minute. Add powdered pectin. Bring to a full rolling boil and boil hard one minute. Add sugar all at one time and boil hard one minute. Remove from heat and add vanilla. Stir and skim for seven minutes to cool slightly to prevent fruit from floating. Ladle into hot sterilized jars, leaving one-half inch space at top of jar. Seal at once. Makes about seven medium glasses.

After:

Recipe:

Pineapple and Rhubarb Marmalade

1 pound rhubarb
1 medium pineapple (fully ripe)
1/2 cup water
1 box powdered pectin
5 cups sugar
3 tablespoons vanilla

Wash rhubarb but do not peel. Cut into thin slices. Peel pineapple, chop very fine and place in a five to six quart saucepan along with sliced rhubarb. Add water. Cover and simmer until soft about one minute. Add powdered pectin. Bring to a full rolling boil and boil hard one minute. Add sugar all at one time and boil hard one minute. Remove from heat and add vanilla. Stir and skim for seven minutes to cool slightly to prevent fruit from floating. Ladle into hot sterilized jars, leaving one-half inch space at top of jar. Seal at once. Makes about seven medium glasses.
Database

- We use Microsoft Access for metadata entry
- There are advantages and disadvantages to using Access
• Create drop down menus for controlled vocabulary lists
• Can’t enter multiple values in a field from a dropdown list
Controlled Vocabularies

- Controlled vocabularies allow you to limit the terms that are used in certain fields.
  - Important for consistency when multiple people are working on the same project
- We use pre-existing controlled vocabularies i.e. LCSH, and create our own for local or specific subjects that don’t already have one
Selected LCSH headings

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcoholic Beverages</td>
<td>Amish cooking</td>
<td>Angelica</td>
<td>Appetizers</td>
<td>Bagels</td>
<td>Baking</td>
<td>Baking - Ireland</td>
<td>Banana breads</td>
</tr>
<tr>
<td>B1</td>
<td>B2</td>
<td>B3</td>
<td>B4</td>
<td>B5</td>
<td>B6</td>
<td>B7</td>
<td>B8</td>
</tr>
<tr>
<td>Barbecuing</td>
<td>Beans</td>
<td>Beverages</td>
<td>Biscuits</td>
<td>Blintzes</td>
<td>Bread</td>
<td>Broccoli</td>
<td>Butter</td>
</tr>
<tr>
<td>B9</td>
<td>B10</td>
<td>B11</td>
<td>B12</td>
<td>B13</td>
<td>B14</td>
<td>B15</td>
<td>B16</td>
</tr>
<tr>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
<td>C5</td>
<td>C6</td>
<td>C7</td>
<td>C8</td>
</tr>
<tr>
<td>Cake</td>
<td>Candy</td>
<td>Casserole Cooking</td>
<td>Cheese</td>
<td>Cheese</td>
<td>Cheese</td>
<td>Cheese</td>
<td>Cheese</td>
</tr>
<tr>
<td>C9</td>
<td>C10</td>
<td>C11</td>
<td>C12</td>
<td>C13</td>
<td>C14</td>
<td>C15</td>
<td>C16</td>
</tr>
<tr>
<td>D1</td>
<td>D2</td>
<td>D3</td>
<td>D4</td>
<td>D5</td>
<td>D6</td>
<td>D7</td>
<td>D8</td>
</tr>
<tr>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
<td>Drink</td>
</tr>
</tbody>
</table>

Controlled vocabulary created in-house

- Adriatic Sea Theater
- Aegean Sea Theater
- Aisne-Chemin des Dames
- Aisne-Marne
- Albert
- Amiens
- Andevanne use Meuse-Argonne, bois d'Andevanne
- Andon Creek USE Meuse-Argonne, Andon Creek
- Anzio
- Angers
- Archangel USE Russian Front - Archangel
- Ardenne Canal USE Meuse-Argonne, Ardenne Canal
- Ardennes
- Ardeuil
- Argeonne use Meuse-Argonne, Argeonne Forest
- Argonne Forest USE Meuse-Argonne, Argeonne Forest
- Argonne Forest Sector
Bringing it all together

- Once all the pieces are together create a place where staff can go to access them
- We organize the pieces into a main folder and then subfolders by staff name
Staff assigned to project

Unassigned text for recipes

Database for metadata creation

Controlled vocabulary list

Unassigned recipe images
Final steps

• Once the metadata is complete, the Access databases are harvested and converted to tab-delimited Excel files

• Images and metadata are parsed together while being imported into CONTENTdm

• Final quality control happens at this stage

• http://content.mpl.org/cdm4/browse.php?CISOROOT=/histrecipe
Some things to consider

• Create a project guide that takes staff through the project step by step
  • A good reference for them to consult if you’re not available

• Consider creating a local style guide for your metadata fields
  • Increases consistency of entries in fields
  • Also good to consult if you’re unavailable
What do I do if I don’t have...

• There are other options beyond those gone over here, many which are much cheaper

• Determine delivery of images and accompanying materials to staff
  • Can be as simple as a flash drive with images loaded if no network is available

• Determine way of entering metadata
  • Google Docs has a spreadsheet with an option to create forms
  • [https://docs.google.com/spreadsheet/ccc?key=0ArkFKY8gDS4wdDh6ZjZJQ3p1c2hleG9iYURINIR6VUE&pli=1#gid=0](https://docs.google.com/spreadsheet/ccc?key=0ArkFKY8gDS4wdDh6ZjZJQ3p1c2hleG9iYURINIR6VUE&pli=1#gid=0)