DIYHistory: Scaling Up With the Crowd at the University of Iowa

Shawn Averkamp

University of Iowa
DIYHistory: Scaling up with the crowd at the University of Iowa

Shawn Averkamp
Digital Research & Publishing
University of Iowa Libraries

shawn-averkamp@uiowa.edu
github.com/saverkamp
@saverkamp
Help improve access to historic documents by transcribing handwritten pages, reviewing transcriptions (look for Needs Review pages), and correcting typewritten text. Select an item below to start transcribing or create an account to enjoy additional features.

Szathmary Culinary Manuscripts and Cookbooks (browse all)

Iowa Byington Reed Diaries (browse all)

Nile Kinnick Collection (browse all)

Civil War Diaries and Letters (browse all)
Szathmary Culinary Manuscripts and Cookbooks

Handwritten cookbooks, ca. 1600s-1960s, documenting culinary history in America and Europe and how tastes have changed over the years.

- Abigail Wellington Townsend cookbook, circa 1840
  - 95% transcribed

- Alice Electa Pickard recipe book, 1868
  - 27% transcribed

- American cookbook, 1759
  - 100% transcribed

- American cookbook, 1838-1851
  - 100% transcribed

- American cookbook, 1850-1870
  - 100% transcribed

- American cookbook, 1886
  - 18% transcribed

- American cookbook, 1900
  - 26% transcribed

- American cookbook, 1900-1920
  - 100% transcribed

- American cookbook, 1900-1920
  - 59% transcribed
To make White Marmalade of Quince

Take the quinces, wash them, put on a spit above a fire, and when they are soft, take them off the fire, drain them and cut them into small pieces, and boil them in a little water with the juice of a lemon, put your spice in the quinces and let it cook until it is nearly done, and then turn it into a cloth bag to strain it, and to be used as filling for pie.
Select an Item to Begin Transcribing:

- Clara Barton letters to cousin Len, 1863-1864
- Bean family letters, 1845-1860
- Bean family letters, 1862-1863
- Bean family letters, 1868-1898
- Miles Beatty letters, 1862-1863
- Miles Beatty letters, 1863
- Oliver Boardman letters, 1861-1863
- John C. Brown diary, 1864
- Anson R. Butler letters, 1861-1900
- Letters to Hattie Smith Cree, 1863-1865
- Andrew F. Davis letters to wife and daughters, 1861
- Andrew F. Davis letters to wife and daughters, 1862
- Andrew F. Davis letters to wife and daughters, 1863
- Laura Gibson letters to mother, 1863
- Iowa Cavalry, 1st Regiment, letters, 1861-1862
- Frank Malcolm letters, 1864
- Mann family letters, 1851-1956
- Byron McClain letters, 1863-1864
- McDowell family letters, 1857-1863
- Mead family letters, 1861
- Mead family letters, 1862
- Mead family letters, 1863
- Mead family letters, 1863-1864
- Mead family letters, Jan.-Mar. 1864
- Mead family letters, Mar.-May 1864
- Mead family letters, June-July 1864
- Mead family letters, Aug.-Sept. 1864

Rediscovering Voices

Thanks to the development of “crowdsourcing” or collaborative transcription of manuscript materials, libraries are now able to use the knowledge and interest of the general public to meet goals that they would never have the time, financial, and staff resources to achieve on their own. Please help us improve access by transcribing the hand-written pages in this collection.

Simply select an item and enter the text as it appears on the digitized image.

May 2011
Saturday April 23rd

This has been a day of some excitement. The surgeon in charge and Executive both called early in the morning Miss Hargis and I were sent with them. It was so near to Kem and almost nothing. And he did not know when we could get to Kem and after the whole medical staff had come in. It was almost then that we felt like asking them to leave us. I shall not forget the little piece of paper which they put in the envelope to let us know we had come in. The girls walked to town and raised the alarm.
Crowdsourcing transcription workflow

Credit: Jen Wolfe
Anyone else getting a really slow connection to this site?

Reddit the giant
Wants to pet the small website
Squishes it instead

http://www.reddit.com/r/todayilearned/comments/humy3/til_how_to_participate_in_history_while_sitting/
not sustainable
Requirements for new site

• Preserve functionality and ease of use
• Expand scope of content and allow for easy addition of new collections
• Streamline workflows to reduce staff time
• Allow users to view their contributions immediately
• Enable the crowd to review and improve transcriptions
A free, open source tool enabling community transcriptions of document and multimedia files

Scripto brings the power of MediaWiki to your collections. Designed to allow members of the public to transcribe a range of different kinds of files, Scripto will increase your content’s findability while building your user community through active engagement.

Download  Learn More
Scripto

• Open-source software
• Developed by Center for History and New Media (CHNM) at George Mason University
• Built on MediaWiki
• Works as a plugin with a CMS: Omeka, Drupal, Wordpress
• http://scripto.org
Serious Web Publishing

Create complex narratives and share rich collections, adhering to Dublin Core standards with Omeka on your server, designed for scholars, museums, libraries, archives, and enthusiasts. Learn More

Download Omeka
Linux, Apache, MySQL5, PHP5

System Requirements
No server? Try Omeka.net!
Omeka

• Open-source software
• Developed by CHNM
• Built on the Zend MVC framework
• Uses the LAMP stack (Linux, Apache, MySQL, PHP)
benefits

• Immediate results for transcribers
• The crowd can edit and improve their work
• Deputized users can proofread and finalize transcriptions (saving staff time)
• Users can track their contributions and bookmark pages with a watchlist
<table>
<thead>
<tr>
<th>Document Page Name</th>
<th>Most Recent Contribution</th>
<th>Document Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page 2</td>
<td>14:49:01 Oct 18, 2012</td>
<td>Iowa Byington Reed diary, January 1, 1920-December 31, 1922</td>
</tr>
<tr>
<td>Front cover</td>
<td>18:55:45 Oct 15, 2012</td>
<td>English cookbook, 1700</td>
</tr>
<tr>
<td>1864-02-24 -- 1864-02-26</td>
<td>16:46:19 Oct 11, 2012</td>
<td>Mary Shelton Diary, 1864</td>
</tr>
<tr>
<td>1864-07-23 Page 05</td>
<td>16:45:41 Oct 11, 2012</td>
<td>Lt. Shelton Diary, 1864</td>
</tr>
<tr>
<td>1862-05-20 Page 07</td>
<td>16:14:40 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters … and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-18 Page 07</td>
<td>16:14:40 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters … and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-18 Page 06</td>
<td>16:14:40 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters … and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-18 Page 05</td>
<td>15:20:55 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters … and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-16 Page 06</td>
<td>13:18:00 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters … and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>Front recto flyleaf</td>
<td>13:17:57 Oct 11, 2012</td>
<td>Wendell D. Wiltsie diary, 1865</td>
</tr>
<tr>
<td>Changes</td>
<td>Document Page Name</td>
<td>Changed on</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Edited</td>
<td>Page 6</td>
<td>16:24:26 Nov 02, 2012</td>
</tr>
<tr>
<td>Created</td>
<td>Page 4</td>
<td>14:14:35 Nov 01, 2012</td>
</tr>
<tr>
<td>Created</td>
<td>Page 6</td>
<td>18:17:54 Oct 27, 2012</td>
</tr>
<tr>
<td>Created</td>
<td>Page 9</td>
<td>17:17:44 Oct 27, 2012</td>
</tr>
</tbody>
</table>
benefits

• Immediate results for transcribers
• The crowd can edit and improve their work
• Deputized users can proofread and finalize transcriptions (saving staff time)
• Users can track their contributions and bookmark pages with a watchlist
• On admin side, easier to monitor activity
Dr. Artemus Browns book of recipes and prescriptions, 1849

Page 8

**Revision as of 21:11:23, Oct 09, 2012**
**Created by Ncronin**

<table>
<thead>
<tr>
<th>Line 1:</th>
<th>Line 1:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough Mixture No. 2</td>
<td>Cough Mixture No. 2</td>
</tr>
<tr>
<td>Gum Opii 2 dr</td>
<td>Gum Opii 2 dr</td>
</tr>
<tr>
<td><strong>Antimonial Wine 1 oz</strong></td>
<td><strong>Antimonial Wine 1 oz</strong></td>
</tr>
<tr>
<td>Liqorice 2 dr</td>
<td>Liqorice 2 dr</td>
</tr>
<tr>
<td>Gum Arabic 2 dr</td>
<td>Gum Arabic 2 dr</td>
</tr>
</tbody>
</table>

**Revision as of 11:14:05, Oct 18, 2012**
**Edited by Ctade**

<table>
<thead>
<tr>
<th>Line 4:</th>
<th>Line 4:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cough Mixture No. 3</td>
<td>Cough Mixture No. 3</td>
</tr>
<tr>
<td>Babson Tolf 2 oz</td>
<td>Babson [Tolu?] 2 oz</td>
</tr>
<tr>
<td>Emulsion Gum Amoneur 1 oz</td>
<td>Emulsion Gum [Ammonia?] 1 oz</td>
</tr>
<tr>
<td>Ppareyorin 1/2 do</td>
<td>Pparegoric 1/2 do</td>
</tr>
<tr>
<td>Dose a ten spoonfull 2 or 3 times a da</td>
<td>Dose a ten spoonfull 2 or 3 times a Day</td>
</tr>
</tbody>
</table>

**Line 7:**

- Corn Pluister No. 1

**Line 8:**

+ Corn Plaister No. 1
drawbacks

• Omeka ingest means duplication of content
• Needs significant customization in order to scale for library production:
  – Theme
  – Functionality
  – Ingest process
  – Data export for completed items
theme

• Simple, uncluttered interface
• Fewer clicks to transcribe
• More images, less text for an intuitive flow
functionality

• Page status and item progress
Szathmary Culinary Manuscripts and Cookbooks

Handwritten cookbooks, ca. 1600s-1960s, documenting culinary history in America and Europe and how tastes have changed over the years.
functionality

• Page status and item progress
• Hide certain features to manage quality, incentivize participation
will, then add some Lemon Peel, a little Pepper, Mace, and Salt, take the yolks of two Eggs beat them well, mix them with a little Cream and grated Onion, a little Parsley, boiled and finely shred, a little piece of Lemon, and a little Flour, mix them all together, and beat them well, then put them into your Hash, gently stirring them over a slow fire, to prevent the Hash curdling. Dish it up for the Table, have the other half boiled, to be served it, before you eat it, so over of the yolk of an Egg, bread crumbs, Parsley and a little butter, when nicely broil, lay it in the middle of the dish, with the Brains just in little Cabbots, briskly Browned, pickled Mushrooms, Pickleaways, or any Pickle you like best, with sliced Lemon, dip the Brains in little made of yeast at basic Flour.
well, then add some Lemon Peel, a little Pippin, Mace, and Salt, take the yolks of two Eggs beat them well, mix them with a little Cream, and grated Nutmeg, a little Parsley, boiled and finely shred, a little piece of Lemon, and a little Flour, mix them all together, and beat them well, then put them into your MIash, gently stirring them once a slow Fire to prevent the Hash, burning the dish. Disht up for the Table, have the other half boiled, to lay upon it, before you serve it, so over the half of an Egg, Bread Crumbs, Parsley and a little butter, when nicely broil, lay it in the middle of the Dish, with the brains, jellied in little Lobes, built a Brown, Pickled Mushrooms, Bamboccio, or any Pickle you like best, with cutted Lemon, and the Brains in horse made of yeast, at Savoy Flour.
import/export

• Batch import page-level metadata with Omeka’s CSV Import plugin
CSV Import

Step 1: Select File and Item Settings

Upload CSV File  *   Choose File  No file chosen
Maximum file size is 8 MB.

Record type
- Item
- File

Use an export from Omeka CSV Report
Selecting this will override the options below.

Select Item Type

Select Collection

Make All Items Public?
**Step 2: Map Columns To Elements, Tags, or Files**

<table>
<thead>
<tr>
<th>Column</th>
<th>Example from CSV File</th>
<th>Map To Element</th>
<th>Use HTML?</th>
<th>Tags?</th>
<th>File?</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>&quot;Iowa Byington Reed diary, January 1, 1878- May 11, 1879&quot;</td>
<td>Title</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>identifier</td>
<td>&quot;byington_999&quot;</td>
<td>Identifier</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>source</td>
<td>&quot;<a href="http://digital.lib.uiowa.edu/u/?byington,999">http://digital.lib.uiowa.edu/u/?byington,999</a>&quot;</td>
<td>Source</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Step 2: Map Columns To Elements, Tags, or Files

<table>
<thead>
<tr>
<th>Column</th>
<th>Example from CSV File</th>
<th>Map To Element</th>
<th>Use HTML?</th>
<th>Filename?</th>
</tr>
</thead>
<tbody>
<tr>
<td>filename</td>
<td>&quot;<a href="http://s-lib017.lib.uiowa.edu/convert/byington/814.jpg">http://s-lib017.lib.uiowa.edu/convert/byington/814.jpg</a>&quot;</td>
<td>Select Below</td>
<td></td>
<td></td>
</tr>
<tr>
<td>title</td>
<td>&quot;Front cover&quot;</td>
<td>Title</td>
<td></td>
<td></td>
</tr>
<tr>
<td>identifier</td>
<td>&quot;byington_999_813&quot;</td>
<td>Identifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>source</td>
<td>&quot;<a href="http://digital.lib.uiowa.edu/u/?byington,999">http://digital.lib.uiowa.edu/u/?byington,999</a>&quot;</td>
<td>Source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>status</td>
<td>&quot;Not Started&quot;</td>
<td>Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>transcription</td>
<td>&quot;&quot;</td>
<td>Transcription</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Import CSV File]
import/export

• Batch import page-level metadata with Omeka’s CSV Import plugin

• Pull in CDM full text (existing transcriptions, dirty ocr) to Scripto database
Rib Roast Choicest of Cuts, Whether Rolled or Standing

By MARTHA HARMAN

A rib roast of beef is one of the choicest cuts of meat, whether it is a standing rib roast or whether the bones have been removed and the roast is rolled.

It is largely a matter of your own preference whether or not you have the rib roast rolled. Experiments have shown that the presence or absence of bones makes no difference in the flavor of the meat.

They do make a difference in the cooking time, however. Bones are a good conductor of heat, and the rolled rib roast requires from 10 to 15 minutes more per pound than unboned roasts. Keep that in mind.

EASIER TO CARVE

A rolled roast is evenly shaped. heat is recommended for roasting. At this temperature, there is no danger of burning the meat nor the drippings: so you do not need water in the pan.

The searing, which was supposed to be necessary in order to "seal in" the juices, doesn't do that at all. A roast which is not seared is juicier than one which is seared.

A roast cooked at a constant low oven temperature, even if for a longer period than was formerly given at higher temperatures, is preferred.

ROLLED RIB ROAST

Select a two or three rib roast, boned and rolled. Wipe with a damp cloth and rub with salt and pepper.
import/export

• Batch import page-level metadata with Omeka’s CSV Import plugin
• Pull in CDM full text (existing transcriptions, dirty ocr) to Scripto database
• Export data in CSV with Omeka’s Reports plugin
Setting our content free

Cadet at camp getting a blanket toss, The University of Iowa, 1912
digital.lib.uiowa.edu/cdm/ref/collection/ictcs/id/1550
Student at typesetting machine in Close Hall, The University of Iowa, 1920s
Student at typesetting machine in Close Hall, The University of Iowa, 1920s

Creator:
Kent, Frederick W. (Frederick Wallace), 1894-1984

Subjects:
United States -- Iowa -- Iowa City
University of Iowa
University of Iowa. School of Journalism and Mass Communication
Close Hall (Iowa City, Iowa)
Universities & colleges
Typesetting
Students
Journalism

Category:
Colleges and departments
Journalism

Collection ID: RG30.0001.01
Item ID: cd23-0023

This image is part of the University of Iowa Libraries' Iowa City Town and Campus Scenes digital collection, documenting the UI campus and its surrounding community. These photos have been added to Flickr so you can comment, tag, and share them easily. Please add words to make the images easier to find -- we welcome your contribution of names, descriptions, locations, tags, and general reactions.

Rights Information: There are no known copyright restrictions on this image. The digital file is owned by the University of Iowa Libraries, which is making it freely available with the request that the Libraries be credited as its source.

More information about this image: digital.lib.uiowa.edu/cdm/ref/collection/ictcs/id/9410
Visit the Iowa City Town and Campus Scenes collection at the Iowa Digital Library: digital.lib.uiowa.edu/ictcs
View all Iowa Digital Library collections: digital.lib.uiowa.edu
Setting our content free

- DIYHistory
- Flickr (historic photographs)
- OpenGeoPortal (historic maps)
- Transcription data
metadata challenges

• Keeping track of content
• Creating meaningful identifiers in new contexts
• Preserving context: links to CDM item, CDM collection, finding aid
• Extracting page-level metadata from complex compound objects: page label, existing full text (transcription, ocr)
identifiers

**Item:**

alias + item ID

cookbooks_4717

**Page:**

alias + item ID + page pointer

cookbooks_4717_4558
item-level metadata

• From CONTENTdm:
  – Title
  – Identifier
  – Reference URL
  – Digital collection URL
  – Finding aid URL
  – Comma-delimited list of file URLs

• Auto-generated
  – Progress (integer for % transcribed)
page-level metadata

- From CONTENTdm
  - File URL
  - Page label
  - Identifier
  - Reference URL (item-level)
  - Transcription (existing or ocr)

- Auto-generated
  - Status (‘Not Started’, ‘Needs Review’, ‘Completed’)

---

[178x459]page-level metadata

• From CONTENTdm
  – File URL
  – Page label
  – Identifier
  – Reference URL (item-level)
  – Transcription (existing or ocr)

• Auto-generated

---
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>title</td>
<td>identifier</td>
<td>source</td>
<td>ispartof</td>
<td>relation</td>
<td>files</td>
</tr>
<tr>
<td>Julie M. Clagett cookbook, June 1, 1886</td>
<td>cookbooks_7078</td>
<td><a href="http://digital.lib.ui.edu">http://digital.lib.ui.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Julie M. Clagett cookbook, 1908-1914</td>
<td>cookbooks_7250</td>
<td><a href="http://digital.lib.ui.edu">http://digital.lib.ui.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alice Electa Pickard recipe book, 1868</td>
<td>cookbooks_7489</td>
<td><a href="http://digital.lib.ui.edu">http://digital.lib.ui.edu</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
pycdm

• a simple Python library for working with CONTENTdm item and collection metadata as Python objects
• github.com/saverkamp/pycdm
import pycdm
pycdm

```python
>>> import pycdm

# get all the metadata for a Civil War letter and put it into a variable called 'letter'
>>> letter = pycdm.item('cwd', '4331')
```
>>> import pycdm

# get all the metadata for a Civil War letter and put it into a variable called ‘letter’
>>> letter = pycdm.item('cwd', '4331')

# view the title of letter
>>> letter.info['title']
>>> import pycdm

# get all the metadata for a Civil War letter and put it into a variable called ‘letter’
>>> letter = pycdm.item('cwd', '4331')

# view the title of letter
>>> letter.info['title']

'Frank Malcolm letters, 1864'
pycdm

# view the reference URL for letter

>>> letter.refurl
# view the reference URL for letter

```python
>>> letter.refurl
'http://digital.lib.uiowa.edu/cdm/ref/collection/cwd/id/4331'
```
# view the finding aid URL for letter

```python
>>> letter.info['findin']
```
```python
# view the finding aid URL for letter

>>> letter.info['findin']

'http://lib.uiowa.edu/collguides/?MSC0026'
```
# view the page labels for all pages in letter

```python
>>> for p in letter.pages:
    p.label
```
# view the page labels for all pages in letter

```python
>>> for p in letter.pages:
    p.label

'1864-03-10 Page 01'
'1864-03-10 Page 02'
'1864-03-13 Page 01'
'1864-03-13 Page 02'
```
pycdm

# view the file URLs for each page in letter

```python
define p in letter.pages:
    p.fileurl
```
# view the file URLs for each page in letter

```python
>>> for p in letter.pages:
    p.fileurl

'http://digital.lib.uiowa.edu/utils/getfile/collection/cwd/id/4276/filename/4277.jpg'
'http://digital.lib.uiowa.edu/utils/getfile/collection/cwd/id/4277/filename/4278.jpg'
'http://digital.lib.uiowa.edu/utils/getfile/collection/cwd/id/4278/filename/4279.jpg'
'http://digital.lib.uiowa.edu/utils/getfile/collection/cwd/id/4279/filename/4280.jpg'
...
pycdm

https://github.com/saverkamp/pycdm

(feedback, please!)
metadata challenges ahead

• When is a transcription ‘done’?
• Persistence vs. ephemerality – how long should we leave completed items up?
• Can we get transcriptions back into CONTENTdm more efficiently?
• How to better track our content out in the wild?
• How to automate pulling tags, comments, annotations, etc. back from third party sites?
Thank you!

Shawn Averkamp
Digital Research & Publishing
University of Iowa Libraries

shawn-averkamp@uiowa.edu
github.com/saverkamp
@saverkamp