11-7-2012

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

Shawn Averkamp
University of Iowa

© 2012 Shawn Averkamp

Creative Commons License
This work is licensed under a Creative Commons Attribution-Share Alike 3.0 License.

Hosted by Iowa Research Online. For more information please contact: lib-ir@uiowa.edu.
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
DIY History
Help build the historical record by doing it yourself!

Make digitized artifacts easier to find, use, and share.

Libraries lack the resources to describe these items in-depth, so we're appealing to the public for help:

- Transcribe diaries, letters, and cookbooks
  - Type entries for handwritten pages to make them full-text searchable, or review other contributors' work

- Correct printed text
  - Improve accuracy by fixing errors in machine-generated transcriptions

- Add tags and comments to historic photographs
  - A picture is worth a thousand words, so add a few to improve accessibility

Select a task below to start adding value to our shared cultural heritage!

---

**Transcribe**
Enter text for handwritten pages

**Correct**
[coming soon]
Edit transcriptions for typewritten pages

**Tag & Comment**
Add info to historic photos

---

**About**
- About the project
- FAQ
- Transcription tips
- Contact us

**Contribute**
- 17,020 pages transcribed
- Discussion Forum

**News**
- UI Libraries launches new crowdsourcing site with manuscript cookbooks and more
- Crowdsourcing sneak preview
- Civil War letters back in Iowa and ready for transcription

**Tweets**
- DIYHMW: 1863: This is the quiet Sabbath (as quiet with us as though war raged not in our beloved land) http://t.co/z9yr8CEN
- DIYHMW: 1864: skirmish with the Rebs the other day they didn't hold out long thought it best to run while they had the chance http://t.co

---

diyhistory.lib.uiowa.edu
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.
To make White Mungolet

[Note: The text is partially legible. The transcription includes corrections and annotations for clarity.]
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, 1851-1856
- 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 on the morning. We heard and I once met girls running to go out and tell their loved ones. The news was some as we and almost 10:30 and he did not know when that we could get to reach and when the whole - him and it continued on the ledge that we felt like asking them to leave us. I shall not forget the smile with tears in which they met us. We asked our the girls to let us have an arm. I asked the girls to come and saw them in...
Crowdsourcing transcription workflow

web → email → staff → CDM → web

(people-ware)

Credit: Jen Wolfe
[–] Dude_man79  32 points 4 months ago

Anyone else getting a really slow connection to this site?
permalink

[–] CommentsInHaiku  234 points 4 months ago

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>1862-05-18 Page 04</td>
<td>13:22:54 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters ... and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-18 Page 02</td>
<td>13:18:15 Oct 11, 2012</td>
<td>Clara Vinson Weaver’s letters ... and James B. Weaver, 1862-1864</td>
</tr>
<tr>
<td>1862-05-18 Page 01</td>
<td>13:18:13 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>
</tbody>
</table>
### Your Watchlist

<table>
<thead>
<tr>
<th>Changes</th>
<th>Document Page Name</th>
<th>Changed on</th>
<th>Changed</th>
<th>Changed by</th>
<th>Document Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edited (hist)</td>
<td>Page 6</td>
<td>16:24:26 Nov 02, 2012</td>
<td>+2</td>
<td>128.255.22.153</td>
<td>Chicago cookbook, 1930s?</td>
</tr>
<tr>
<td>Created (hist)</td>
<td>Page 6</td>
<td>18:17:54 Oct 27, 2012</td>
<td>+575</td>
<td>128.255.22.153</td>
<td>Chicago cookbook, 1930s?</td>
</tr>
</tbody>
</table>

### About
- About the project
- FAQ
- Transcription tips
- Contact us

### Contribute
- 17,276 pages transcribed
- Discussion Forum

### News
- Welcome, MetaFillarians!
- UI Libraries launches new crowdsourcing site with manuscript cookbooks and more
- Crowdsourcing sneak preview

### Tweets
- DIYHCivilWar: Today in 1864: your Mother & I have husked a thousand bushels of corn...we expect to elect Lincoln by a large majority
- DIYHCivilWar: Today in 1864: Election is drawing near...I am afraid it is the calm before the storm I hope it is not http://t.co/M5F9x6nX
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 Brown's book of recipes and prescriptions, 1849

Page 8

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

线1：
- Cough Mixture No. 2
- Gum Opii 2 dr
- Antimonial Wine 1 oz
- Liqorice 2 dr
- Gum Arabic 2 dr

线7：
- Cough Mixture No. 3
  - Babson Telm 2 oz
  - Emulsion Gum Amoneur 1 oz
  - Pareyoric 1/2 do
  - Dose a ten spoonfull 2 or 3 times a day

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

线1：
- Cough Mixture No. 2
- Gum Opii 2 dr
+ Antimonial Wine 1 oz
- Liqorice 2 dr
- Gum Arabic 2 dr

线7：
- Cough Mixture No. 3
  + Babson [Tolu?] 2 oz
  + Emulsion Gum [Ammonia?] 1 oz
  + Pareyoric 1/2 do
  + Dose a ten spoonfull 2 or 3 times a Day

- Corn Pluister No. 1
+ 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
well, then add some Lemon Peel, a little Pepper, Made and Salt, take the yolks of two Eggs beat them well, mix them with a little Cream, and grated Apple, 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 pan, gently stirring them over a slow fire to prevent the brains curdling. Once it sets for the table, have the other half boiled, to line it with, before you beat it, so over 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 in little cubios, browned, baked Mushrooms, Babkins, or any Pickle you like best, with sliced Lemon, and the brains in Little made of South or Beef flour.
well, then add some Lemmon Peel, a little Ketchup, Mix and Salt, take the yolks of two Eggs beat them with, mix them with a little Cream, and grated Grating, a little Parsley, boiled and finely shred, a little piece of Lemmon, and a little Flower, mix them all together, and beat them well, then put them into your Kasha, gently stirring them over a slow Fire to prevent the Kasha.  "Cooking the Kasha.

When it sets for the Table, leave the other half boiled, to let it cool.  Before you boil it, so over it the stock of an Egg, Bread Crumbs, Parsley, and a little butter.  When nicely broil, lay it in the middle of the Dish, with the Kasha.  For a little Cabbage, boil it brown, pickled Mushrooms, Pickles, or any Pickle you like best, with a slice Lemon atop the Kasha in little pieces of yang at Egg flower.
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.

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

And "side roast is evenly shaped, for it so easier to carve,"
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
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
pycdm

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

# get all the metadata for a Civil War letter and put it into a variable called ‘letter’

>>> letter = pycdm.item('cwd', '4331')
```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')

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

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

'Frank Malcolm letters, 1864'
```
pycdm

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

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

```python
>>> letter.info['findin']
'http://lib.uiowa.edu/collguides/?MSC0026'
```
# view the page labels for all pages in letter
>>> 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
>>> for p in letter.pages:
    p.fileurl
```
pycdm

# view the file URLs for each page in letter

>>> 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

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