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
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 typeset 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
DIYHCMWar: 1863: This is the quiet Sabbath (as quiet with us as though war raged not in our beloved land) http://t.co/z9yr8CE
DIYHCMWar: 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
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 (59% transcribed)
- American cookbook, 1900-1920 (100% transcribed)
To make White Marmalade of Quince

Take the Quinces, allow for Spots, sit on a Plate of Salt, make it Boil, and put to good Quinces &c. appearing Large, and remove cut out, make it Boil again, let it Boil a while, then take in a Sieve, press them & set good pulp from the Core of Quince, put your Salt, the Quince being salted, let it be Small, then ketch it and so a Boon of Quince.
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 Hogan and I were sent girls to go out, so Prue had to talk with them. Hogan said the fight was near us and almost within. Ish worded no word, and we had to make plans. We did not know when we could get to the hospital. The whole situation was very difficult and almost overwhelming. We felt like asking them to leave us. I shall not forget the mile filled with words that went into the ladies to let us have our own place. The girls walked to town and raised the money.
Crowdsourcing transcription workflow

web → email → staff → CDM → web

(people-ware)

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.
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
## Your Contributions

<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>
</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>
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

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

Line 7:
- Cough Mixture No. 3
- Babson Tolu 2 oz
- Emulsion Gum Ammonia 1 oz
- Paregoric 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

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

Line 7:
- Cough Mixture No. 3
- Babson [Tolu?] 2 oz
- Emulsion Gum [Ammonia?] 1 oz
- Paregoric 1/2 do
- Dose a ten spoonfull 2 or 3 times a day
- 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.

- 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
  - 59% transcribed

- American cookbook, 1900-1920
  - 100% transcribed
functionality

• Page status and item progress
• Hide certain features to manage quality, incentivize participation
American cookbook, ca. 1850

Page 36
image 37 of 117
view in Iowa Digital Library

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 Juniper, a little Parsley, boiled and finely shred, a little piece of Lemon, and a little Flower, mix them all together, and beat them well, then put them into your Hash, gently stirring them over a slow flame to prevent the Hash, curdling, thickening.

Dish it up for the Table, have the other half boiled, to live when it, before you eat it, so even at the yolks of an Egg, bread crumbs, Parsley and a little butter, when nicely broil, lay it in the middle of the Dish, with the Brains, junk in little Lard, breast cream, pickled Mushrooms, Blackbird, or any Pickle you like best, with slice Lemons, and the Brains in a little Made of stale at each flour.

Enter your transcription below:
- Copy the text as is, including misspellings and abbreviations.
- No need to account for formatting (e.g. spacing, line breaks, alignment); the goal is to provide text for searching.
- If you can't make out a word, enter "[illegible]"; if uncertain, indicate with square brackets, e.g. "[town?]

Save edits
well, then add some Lemon Peel, a little Pernice, Wine 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 price of Lemon, and a little Flower, mix them all together, and beat them well, then put them into your Mould, gently stir them once a slow Fire to prevent the Heash, twisting the Dish it up for the Table, leave the other half boiled, to cut when it, before you seel it, so over all theught of an Egg, Bread Crumbs, Parsley and a little butter, when nicely broil, lay it in the middle of the Dish, with the Brain in little Cubes, boil in Brown, Pickled Mushrooms, Brine, or any Pickle you like best, with a little Benz in the Brain in Brine made of salt and 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

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>
### CSV Import

#### 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 and is easily carved.

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

<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>
</tbody>
</table>

```
half a cupful of

RUTTI JAR
```
pycdm

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

>>> import 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

`>>> letter.refurl`

'http://digital.lib.uiowa.edu/cdm/ref/collection/cwd/id/4331'
pycdm

# view the finding aid URL for letter

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

```python
>>> letter.info['findin']
'http://lib.uiowa.edu/collguides/?MSC0026'
```
pycdm

# 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
>>> 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' ...
# 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

```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

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