

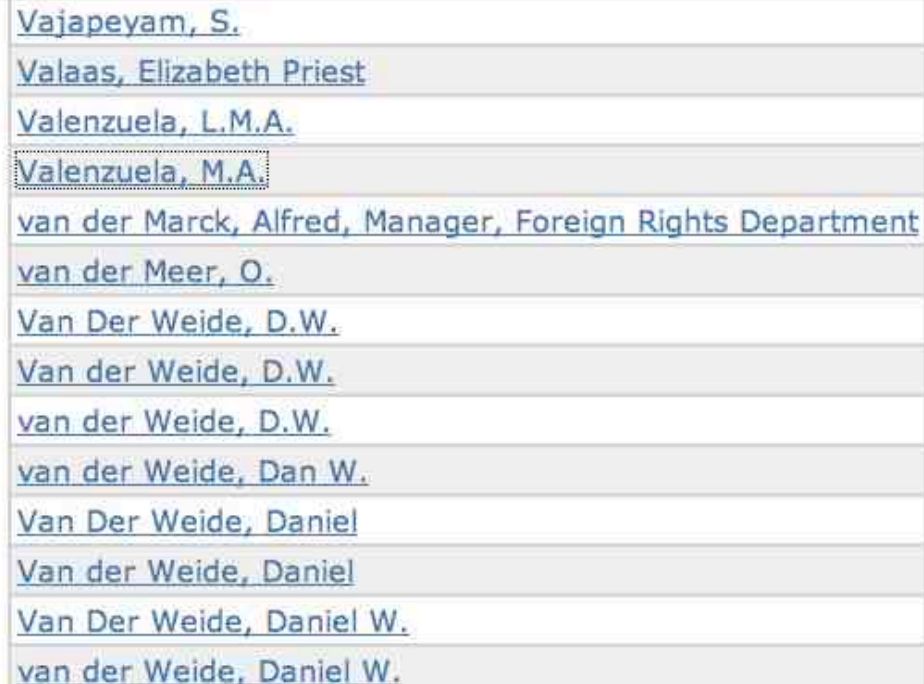
Neatening metadata

“Some metadata are born neat,
some achieve neatness,
and some have neatness thrust upon them.”

—Dorothea Salo
(with apologies to the immortal Bard)

MINDS@UW

(the seamier side of DSpace)



A screenshot of a table with 14 rows and 1 column. The text in each row is as follows:

Vajapeyam, S.
Valaas, Elizabeth Priest
Valenzuela, L.M.A.
Valenzuela, M.A.
van der Marck, Alfred, Manager, Foreign Rights Department
van der Meer, O.
Van Der Weide, D.W.
Van der Weide, D.W.
van der Weide, D.W.
van der Weide, Dan W.
Van Der Weide, Daniel
Van der Weide, Daniel
Van Der Weide, Daniel W.
van der Weide, Daniel W.

(I got an entire article
out of this one screenshot.)

Browse

All of MINDS@UW

- [Sitemap](#)
- [By date](#)
- [By author](#)
- [By title](#)
- [By subject](#)

[Exported items](#)

- [Logout](#)
- [Profile](#)
- [Submissions](#)

Administrative

Access Control

- [People](#)
- [Groups](#)
- [Authorizations](#)

Registries

- [Metadata](#)
- [Format](#)

Items

- [Withdrawn Items](#)
- [Control Panel](#)
- [Statistics](#)

About MINDS@UW

- [About us](#)
- [FAQ](#)
- [Contact us](#)

Edit Item

[Item Status](#)

[Item Bitstreams](#)

[Item Metadata](#)

[View Item](#)

Add new descriptive information

Name:

Value:
[Language](#)

[Add new descriptive information](#)

PLEASE NOTE: These changes are not validated in any way. You are responsible for entering the data correctly. If you are not sure how to do that, please do NOT make changes.

[Update](#) [Return](#)

Description

Remove	Name	Value	Language
<input type="checkbox"/>	remove dc.contributor.advisor	<input type="text" value="Oberly, James Warren, 1954-"/>	<input type="text"/>
<input type="checkbox"/>	remove dc.contributor.author	<input type="text" value="Holtan, Thomas J."/>	<input type="text"/>
<input type="checkbox"/>	remove dc.date.accessioned	<input type="text" value="2009-07-14T17:38:15Z"/>	<input type="text"/>
<input type="checkbox"/>	remove dc.date.available	<input type="text" value="2009-07-14T17:38:15Z"/>	<input type="text"/>
<input type="checkbox"/>	remove dc.date.issued	<input type="text" value="2009-07-14T17:38:15Z"/>	<input type="text"/>
<input type="checkbox"/>	remove dc.description.abstract	<input type="text" value="At the beginning of the twentieth century, White Sand Lake's shores were cutover from the logging of..."/>	<input type="text" value="en"/>

How do I fix this? I cheat.

```
Terminal — ssh — 128x28
Macintosh-3:~ dsalo$ ssh repository.library.wisc.edu
Password:
Last login: Fri Oct 30 10:43:48 2009 from 128.104.62.47
Sun Microsystems Inc. SunOS 5.10 Generic January 2005
-bash-3.00$ sudo su - mindsadm
Sun Microsystems Inc. SunOS 5.10 Generic January 2005
mindsadm@repository:~$ cd /minds/postgresql/postgresql-8.3.3/bin
mindsadm@repository:/minds/postgresql/postgresql-8.3.3/bin$ psql -d dspacedb
Welcome to psql 8.3.3, the PostgreSQL interactive terminal.

Type: \copyright for distribution terms
      \h for help with SQL commands
      \? for help with psql commands
      \g or terminate with semicolon to execute query
      \q to quit

dspacedb=# select item_id, text_value from metadatavalue where text_value like 'van der Weide%';
 item_id |
-----+-----
    3962 | van der Weide, Daniel W.
    4559 | van der Weide, Daniel W.
    4562 | van der Weide, Daniel W.
    4565 | van der Weide, Daniel W.
    4578 | van der Weide, Daniel W.
```

We're moving!

DSpace is going away.

So are most of UWDC's existing
technology platforms.

But that means a
new problem.

Moving the old metadata!

New world: FOXML

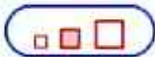
("Fedora Object XML")

- Verbose, so I won't show you a sample.
- Basic idea: mix-and-match XML vocabularies
 - MODS: for description
 - PREMIS: for provenance
 - METS: for structure
 - ... etc.

For MINDS@UW

- Export
- Munge
- Reimport
- (sounds simpler than it is, but it's very doable; I'm not worried)

A UWDCDC record on the web



Title: City visitors to the village
Photographer: Bernault, Florence
Date: 1995
Place: Congo, République du / Pool, Région du
Description: Two young women from Brazzaville stop to pay a visit to relatives in a small village in southern Congo. In the background note cassava leaves and roots for the dinner.
LCSH Subjects: Rural-urban relations
Type: Stillimage
Format: Photograph
Is Part Of: Florence Bernault Collection
Africa Focus
Terms: For information on re-use see: <http://digital.library.wisc.edu/1711.dl/Copyright>
Ownership: Original material owned by Florence Bernault.
Submitter: Bernault, Florence: University of Wisconsin--Madison. African Studies Department
Local Identifier: AfricaFocus.fb0007.bib

Metadata: CSGML

```
<resource update="2007-04-21" id="AfricaFocus.fb0007.bib">
  <dc.title type="Main">City visitors to the village</dc.title>
  <dc.creator scheme="AACR2" type="Photographer">Bernault, Florence</dc.creator>
  <dc.subject scheme="LCSH">Rural-urban relations</dc.subject>
  <dc.description>Two young women from Brazzaville stop to pay a visit to relatives in a small
    village in southern Congo. In the background note cassava leaves and roots for the
    dinner.</dc.description>
  <dc.date scheme="W3C-DTF">1995</dc.date>
  <dc.type scheme="DCT1">StillImage</dc.type>
  <dc.format>Photograph</dc.format>
  <dc.relation type="IsPartOf">Florence Bernault Collection</dc.relation>
  <dc.relation type="IsPartOf">Africa Focus</dc.relation>
  <dc.coverage type="Spatial" scheme="TGN">Congo, République du</dc.coverage>
  <dc.coverage type="Spatial" scheme="TGN">Pool, Région du</dc.coverage>
  <dc.rights type="Ownership">Original material owned by Florence Bernault.</dc.rights>
  <dc.rights type="Terms">For information on re-use see:
    http://digital.library.wisc.edu/1711.dl/Copyright</dc.rights>
  <submitter>
    <name scheme="AACR2">Bernault, Florence</name>
    <affiliation scheme="AACR2"> University of Wisconsin--Madison. African Studies
      Department</affiliation>
  </submitter>
  <media ident="AfricaFocus.fb0007" n="01">
    <basePath>/db/dlmap/AfricaFocus/htdocs/data/images/</basePath>
    <localPath>MmBib/FrenchAfrImages/</localPath>
    <objectID>fb0007</objectID>
    <numComponents>1</numComponents>
    <relativeSize>t</relativeSize>
    <relativeSize>r</relativeSize>
    <relativeSize>l</relativeSize>
    <dc.format scheme="MIME">image/jpeg</dc.format>
  </media>
</resource>
```

Getting there from here: XSLT

```
<xsl:if test="$mainnode/dc.creator">
  <xsl:for-each select="$mainnode/dc.creator">
    <mods:name>
      <mods:namePart>
        <xsl:value-of select="./text()"/>
      </mods:namePart>
      <xsl:if test="./@type">
        <mods:role>
          <mods:roleTerm type="text">
            <xsl:value-of select="./@type"/>
          </mods:roleTerm>
        </mods:role>
      </xsl:if>
    </mods:name>
  </xsl:for-each>
</xsl:if>
<xsl:if test="$mainnode/dc.subject">
  <xsl:for-each select="$mainnode/dc.subject">
    <mods:subject>
      <xsl:if test="@scheme">
        <xsl:attribute name="authority">
          <xsl:value-of select="@scheme"/>
        </xsl:attribute>
      </xsl:if>
      <mods:topic>
        <xsl:value-of select="./text()"/>
      </mods:topic>
    </mods:subject>
  </xsl:for-each>
</xsl:if>
```

Results

- Writing and testing code: 2-3 weeks (doing other things as well)
- Once the code worked, all of Africa Focus (~4000 items) moved to Fedora in less than a day!
- Code can now be reused to move other collections.

Other people's metadata

... the final frontier.

The Excel spreadsheet

ResearcherName1	DateCreated	DateEntered	SourceLocationCounty	SourceLocationCity	keyword
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Leather bindings suspended fr
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Leatehr bindings, Leather bind
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Maine snowshoe suspended fr
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Maine snowshoe suspended fr
James P. Leary	7/22/96	9/9/02	Bayfield	Washburn	James P. Leary; Bodin's Brownstone Bowl, Brownstone Bowl at Bodin's Resort
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's snowshoe workshop, Nelson poses with Cree s
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's snowshoe workshop, Top end of Cree snowsho
James P. Leary	7/23/96	3/27/02	Ashland	Ashland	James P. Leary; Tom Nelson's snowshoe workshop; rawhide fill on Cree snow
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Cree snowshoe on workbench
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop; first shoe made, First snowsho
Robert T. Teske	9/10/96		Green	Monroe	Robert T. Teske, , Historic Cheesemaking Center, Monroe Courthouse Squar
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, models of heavy equipment from Judi and Lew Tolliver Jr., Bu
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Sheets of rawhide, WIFFJL9601
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Finished snowshoes, WIFFJL96
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, finished snowshoes, WIFFJL96
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, finished snowshoes, Strips of
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, repaired snowshoes, Snowsho
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, WIFFJL960118, JL, Ashland, As
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Varnished snowshoes, WIFFJL9
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Finished snowshoes hanq fron
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary; Tom Nelson's shoe workshop, Tom Nelson's 'Snowshoes by S
James P. Leary	7/23/96		Ashland	Ashland	James P. Leary, Dee bainbridge, Ojibwa storyteller, WIFFJL960122, JL, Ashland
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, Bayfield Heritage Association, Inc., Bayfield's Mystic Knights
James P. Leary	7/24/96		Bayfield	Bayfield	James P. Leary, John Benson farm/ranch, John Benson at the reins of a team,
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, John's son Theodore 'Ted' (R) and D
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, John Benson on the wagon's buckb
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, Benson with Bigfoot and a dog in fr
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, Benson with an ax/pickaroon made
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, Horse logging gear: two chainbinds
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, John Benson with his skidding hors
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, John Benson with his skidding hors
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, Painting of Sunny and Cher by Donna Lanni, Painting of farm
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, Ted Benson puts various harness d
James P. Leary	7/24/96		Bayfield	Mason	James P. Leary, John Benson farm/ranch, Ted Benson puts various harness q

Step 1: run in circles...

- No titles!
- Abbreviations
- “Keywords” contain researcher name, item identifiers, other non-subject information
- Which photos have essentially the same descriptions? (FRBRization, after a fashion)

Procedure

(such as it is)

- Save spreadsheet as a CSV file.
- Write a program to interpret the CSV file and output the needed metadata files.
 - How easy and fast this is depends on how neat the spreadsheet is.
 - CSUMC is messy, as these things go.

My ~~knife~~ language of choice: Python

```
def main():
    folder = os.path.join("/", "users", "dsalo", "Fedora", "csumc")

    prevsubs = [] #previous subjects for comparison to see if we have an image cluster
    clusters = {} #key is title string, value is list of metadata lists

    datafile = open(os.path.join(folder, "see_all_data.csv"), 'r')

    # prepare to clean up field 16
    stopwords = []
    for entry in datafile.readlines()[2:]:
        fields = entry.split('\t')
        if fields[0].strip() not in stopwords: stopwords.append(fields[0].strip())
        if fields[2].strip() not in stopwords: stopwords.append(fields[2].strip())
        if fields[9].strip() not in stopwords: stopwords.append(fields[9].strip())
        if fields[10].strip() not in stopwords: stopwords.append(fields[10].strip())
        if fields[11].strip() not in stopwords: stopwords.append(fields[11].strip())
        if fields[12].strip() not in stopwords: stopwords.append(fields[12].strip())

    stopwords.remove("Red Lake")
    datafile = open(os.path.join(folder, "see_all_data.csv"), 'r')

    identifiers = []
    for entry in datafile.readlines()[2:]: #ignore column headers, desired markup at top
        #fix ampersands
        entry = entry.replace("& ", "& ")
        #fix other warts in the data
        entry = entry.replace("James P. Leary;", "James P. Leary,")
        entry = entry.replace("shoe workshop;", "shoe workshop,")
        entry = re.sub(r',+', ',', entry)
        entrylist = []

        fields = entry.split('\t')

        contributor1 = Name(fields[0])
```

Takeaways

- Different jobs ask for different tools. Use the tool that makes sense for the job.
- 80/20 point: sometimes the computer can't fix it all, so the optimal solution is to let it fix what it can and fix the rest by hand.
- Always something new!