Low Cost Metadata Management
Tools and Tips for Catalogers

Lisa Furubotten
Anna J. Dabrowski
Joseph Olivarez
Eric Pennington
Introduction

Objectives

Preliminaries
Objectives

- Demonstrate the thought process for deciding which metadata transformation strategies might be most appropriate for different projects
- Demonstrate creative ways to use familiar tools to efficiently produce metadata for digital projects
- Elicit from attendees other tools and strategies they can contribute to our list
- Inspire collaboration with/within institutions to efficiently produce metadata for digital projects
Tools

- Microsoft Office
  - Access
  - Sharepoint
  - Excel
- Google
  - Sheets
  - Forms
- MARCEdit
- OpenRefine
- OCLC Connexion
- Browser Add-on
Why cataloging staff?

- Cataloging staff are experienced in creating/reusing metadata according to international standards.
- Too many projects and not enough help.
- Don’t want to use expensive development resources for unique or simple projects.
- Cataloging staff are agile, used to change, and can easily master new tools for working with digital collection metadata.
Digital project preliminary questions

What?
- What metadata is available that could be reused?
- What form is that metadata in?

Where?
- Where will the final product be going?
  - What format should the metadata take?
  - Do you need the same metadata in multiple formats?
  - Are you going to need records for either OCLC or the local catalog, and to what standards?
Digital project preliminary questions

How?

■ How large is the project?
■ How are you going to implement quality control/data cleanup and ensure accountability?
■ How are the metadata creators going to be trained?

Who?

■ Who is available and what tools do they have expertise in?
■ Who is going to produce the metadata? A single person or a group?
Where is the metadata going?

How should the metadata be formatted for the system in which it will reside?

Two common cases seem to be:

1. MARCXML for HathiTrust Projects
   Bibliographic Metadata Specifications at: https://www.hathitrust.org/bib_specifications

2. Data organized in spreadsheets, usually with Dublin Core elements or a similar schema.
Partial record:

```xml
<marc:controlfield tag="008">160728s1955 txua s000 0 eng d</marc:controlfield>
<marc:datafield tag="245" ind2="0" ind1="1">
    <marc:subfield code="a">Fattening lambs</marc:subfield>
</marc:datafield>
<marc:datafield tag="955" ind2=" " ind1=" ">
    <marc:subfield code="b">taeb158300</marc:subfield>
</marc:datafield>
```

Where “taeb158300” is the name of the file containing the digital object “Fattening lambs.”
<table>
<thead>
<tr>
<th>filename</th>
<th>dc.title</th>
<th>dc.date</th>
<th>dc.creator</th>
</tr>
</thead>
<tbody>
<tr>
<td>04_Adachi.pdf</td>
<td>Torsional Instability Of Cooling Tower Fan During Induction Motor Startup</td>
<td>2016</td>
<td>Adachi, Akira</td>
</tr>
<tr>
<td>14_Vengurlekar_ProcessGas.pdf</td>
<td>Investigation of process gass compressor shaft vibration phenomena</td>
<td>2016</td>
<td>Vengurlekar, Ashutosh</td>
</tr>
</tbody>
</table>
Two sample projects

**Flood Insurance Studies.** Destination: HathiTrust, OCLC, and local catalog

**Turbomachinery Laboratory Proceedings.** Destination: DSpace Repository
Repository single submission[1]

Item submission

Describe Item

Authors:
Last name, e.g. Smith

First name(s) + "Jr", e.g. Donald Jr

Enter the names of the authors of this item.

Title: *

Enter the main title of the item.

Other Titles:

If the item has any alternative titles, please enter them here.

Date of Issue: *

Year
Month
Day

2017
Repository form[2]

Publisher:

Enter the name of the publisher of the previously issued instance of this item.

Citation:

Enter the standard citation for the previously issued instance of this item.

Series/Report No.:

Series Name

Report or paper No.

Enter the series and number assigned to this item by your community.

Identifiers:

DOI

If the item has any identification numbers or codes associated with it, please enter the types and the actual numbers or codes.

Type:

Animation
Article
Book
Book chapter
Dataset
Learning Object

Select the type(s) of content of the item. To select more than one value in the list, you may have to hold down the "CTRL" or "Shift" key.
Repository form

Type:
- Animation
- Article
- Book
- Book chapter
- Dataset
- Learning Object

Select the type(s) of content of the item. To select more than one value in the list, you may have to hold down the "CTRL" or "Shift" key.

Language:
- N/A

Select the language of the main content of the item. If the language does not appear in the list, please select 'Other'. If the content does not really have a language (for example, if it is a dataset or an image) please select 'N/A'.

Department:
- Accounting
- Aerospace Engineering
- Agricultural Economics
- Agricultural Leadership, Education, and Communications
- Anesthesiology
- Animal Science

Select your academic department. If your department does not appear in the list, please select 'Other'. To select more than one value in the list, you may have to hold down the "CTRL" or "Shift" key.
Repository form

Describe Item

Subject Keywords:

Enter appropriate subject keywords or phrases.

Subject Categories

Abstract:

Enter the abstract of the item.

Sponsors:

Enter the names of any sponsors and/or funding codes in the box.
Repository form

Upload File(s)

File: *

Browse...

Please enter the full path of the file on your computer corresponding to your item. If you click "Browse...", a new window will allow you to select the file from your computer.

File Description:

Optionally, provide a brief description of the file, for example "Main article", or "Experiment data readings".

Embargo until specific date:

The first day from which access is allowed. Accepted format: yyyy, yyyy-mm, yyyy-mm-dd

Embargo reason:
Repository constant data[1]

License:

Provenance:

Upload new logo:

Item template: Create...
Create your own form
Google Docs

**Advantages**
- Free
- Concurrent online access for multiple users
- Google Forms are simple to make and use
- Forms can add responses to Google Sheets or export to comma separated value (CSV) files.
- Google Sheets export to multiple formats (XLSX)

**Disadvantages**
- Requires Google account
- Need internet access
- Files stored on external servers
Google Docs

- Live Demo
Microsoft Excel [1]

- Advantages
  - Comes installed on most computers
  - A familiar tool and interface
  - Export to many different formats

- Disadvantages
  - Only one person can work on at a time
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Author</td>
<td>Publisher</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>Pride and Prejudice</td>
<td>Austen, Jane</td>
<td>Oxford</td>
<td>1970</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Author</td>
<td>Publisher</td>
<td>Date</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------</td>
<td>-------------------------------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Pride and Prejudice</td>
<td>Austen, Jane</td>
<td>Oxford</td>
<td>1970</td>
<td></td>
</tr>
</tbody>
</table>
Microsoft Access [1]

- Advantages
  - Ability to create forms and have concurrent users
  - Export in multiple formats
  - Dropdown values, Yes/No, default values, etc.
  - Autofill
  - Files can be stored locally or externally

- Disadvantages
  - Not built to easily manipulate data
  - Fields with multiple values can be tricky
  - Difficult to change form (formatting issues)
SharePoint [1]

- **Sharepoint: Advantages** - Can create a form that pushes to a spreadsheet, and is accessed by multiple users concurrently via a web interface. Can download spreadsheets to local computer.

- Also allows for users to work within an Excel spreadsheet concurrently in the web interface.

- The software can be hosted on local servers.

- **Disadvantages** - Expensive software, web interface is accessible only with internet connection, needs to be hosted on local servers.
<table>
<thead>
<tr>
<th>Printer: Transcribed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcázar, Juan Blanco de, +d 1620-1657</td>
</tr>
<tr>
<td>Arévalo, Mariano</td>
</tr>
<tr>
<td>Balli, Pedro, +d active 1574-1600</td>
</tr>
<tr>
<td>Cromberger, Juan, +d -1540</td>
</tr>
<tr>
<td>Cumplido, Ignacio, +d 1811-1887</td>
</tr>
<tr>
<td>Espinosa, Antonio de, +d -1578</td>
</tr>
<tr>
<td>Fernández de Jáuregui, María, +d -1815</td>
</tr>
<tr>
<td>Guillena Carrascoso, Juan José</td>
</tr>
<tr>
<td>Hogal, José Antonio de, +d 1731-1802?</td>
</tr>
<tr>
<td>Hogal, José Bernardo de, +d -1741</td>
</tr>
<tr>
<td>Jáuregui, José de, +d -1778</td>
</tr>
<tr>
<td>Ocharte, Melchior, +d 1571-1616</td>
</tr>
<tr>
<td>Ocharte, Pedro, +d approximately 1532-1592</td>
</tr>
<tr>
<td>Pablos, Juan, +d -1561?</td>
</tr>
<tr>
<td>Pineda Ibarra, José de, +d 1629-1680</td>
</tr>
<tr>
<td>Ricardo, Antonio, +d 1540?-1606</td>
</tr>
<tr>
<td>Ribera, María de, +d active 1732-1754</td>
</tr>
<tr>
<td>Rivera Calderón, Francisco de</td>
</tr>
<tr>
<td>Robledo, Francisco, +d active 1640-1647</td>
</tr>
<tr>
<td>Rodríguez Lupercio, Francisco</td>
</tr>
<tr>
<td>Rosa, Pedro de la</td>
</tr>
<tr>
<td>Sansorí, María de, +d active 1572-1597</td>
</tr>
<tr>
<td>Viuda de don Joseph Bernardo de Hogal (100 0#)</td>
</tr>
<tr>
<td>Viuda de Francisco Rodríguez Lupercio (100 0#)</td>
</tr>
<tr>
<td>Ximeno, José</td>
</tr>
<tr>
<td>Zúñiga y Ontiveros, Felipe de, +d 1717?-1793</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Printer: Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fernández de León, Diego, +d active 1683-1710</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Printer: Corporate Body</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
However ....

If you are going to need catalog records anyway.....

- If your project is one that will require MARC records either for OCLC or the local catalog, or MARCXML records for HathiTrust, *and*

- There are no MARC records available, *then*

- It’s easier to begin your workflow in OCLC or in the local catalog and create your original metadata in MARC, *then*

- Those records can be converted to other uses such as Dublin Core, etc.
Reusing Non-MARC Data
Data from indexes[1]
1. Alteration of activity variables relative to clinical diagnosis of bovine respiratory disease in newly received feedlot cattle.

Our objective was to explore change in activity variables relative to clinical identification of bovine respiratory disease (BRD) in cattle maintained in a commercial feedlot. Four blocks (n=364) of high-risk, crossbred beef calves were affixed with an accelerometer device upon feedlot arrival that ...

**Author(s):** Pillen, J. L.; Pinedo, P. J.; Ives, S. E.; Covey, T. L.; Naikare, H. K.; Richeson, J. T.

**Publisher:** American Association of Bovine Practitioners, Stillwater, USA

**Journal article; Conference paper:** Bovine Practitioner 2016 Vol.50 No.1 pp.1-8 ref.31

2. Efficacy and safety of a novel DNA immunostimulant in cattle.

Four studies were performed to evaluate efficacy (Studies 1 and 2) and safety (Studies of a novel DNA immunostimulant (Zelnote®; Bayer Healthcare) in cattle. Zelnote® (ZF...
Data from Indexes

<table>
<thead>
<tr>
<th>PAN</th>
<th>Item Types</th>
<th>Title</th>
<th>Sub Title</th>
<th>Foreign Title</th>
<th>Authors</th>
<th>Editors</th>
<th>Corporate Auth</th>
<th>Document till ISBN</th>
<th>ISSN</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0153E+10</td>
<td>Journal articl</td>
<td>Comparing respiratory disease treatment</td>
<td>Wells, M. R.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Journal issue</td>
<td>45th Annual Conference of the American Association of Bovine Practitioners, M Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Journal issue</td>
<td>46th Annual Conference of the American Association of Bovine Practitioners, M Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Journal articl</td>
<td>Protection against bovine herpesvirus</td>
<td>Zimmerman, A. D.; Klein, A. L.; Buterbaugh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Journal articl</td>
<td>A beef producer's view of the role of bovine Rhea</td>
<td>J. P. Smith, R. A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Conference</td>
<td>Proceedings of the 46th Annual Conference of the Association of Bovine Practitioners</td>
<td>Smith, R. A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0143E+10</td>
<td>Conference</td>
<td>Improving client profitability by increased Bowen, J. S. Smith, R. A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Alteration of activity variables relative to milk protein and pI, Pilla, J. J.; Pinedo, P. J.; Ives, S. E.; Cove Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Efficacy and safety of a novel DNA immunoassay Nickell, J. S.; Keil, D. J.; Settje, T. L.; Lech Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Serological effect of two concurrent IBR Stokka, G. L.; Neville, B.; Seeger, J. T.; C Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Emergence of bovine viral diarrhea virus</td>
<td>Ferro, P. J.; Fearnhead, M. G.; Calvillo Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Comparison of ceftiofur crystalline free Hawley, J.; Powell, J. G.; Kegley, E. B. Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>A survey to describe beef producer opinion Lee, T. L.; Thomson, D. U. Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>A stochastic model designed to estimate Nickell, J. S. Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0163E+10</td>
<td>Journal articl</td>
<td>Description and justification of a consensus Dewell, R. D.; Dewell, G. A.; Bear, D. A. Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0153E+10</td>
<td>Journal articl</td>
<td>Diverse outcomes of bovine viral diarrh</td>
<td>Fulton, R. W.; Rezabek, G. B.; Grant, R. Bovine Practitioner</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data from website

- Grabbing Metadata
  - Copy and Paste
  - Excel’s Get External Data from Web
  - Source Code

- Grabbing Files
  - Firefox Add-on *DownloadThemAll*

TurboMachinery Proceedings at:
http://turbolab.tamu.edu/proc/
44th Turbomachinery Symposium (2015)

Lectures

Lecture T01: Coupled Torsional and Lateral Analysis for the Determination of the Damping of the First Torsional Mode of Synchronous Motor Driven Compressor Trains
Authors: Jean-Claude Pradetto, Urs Baumann (MAN Diesel & Turbo Schweiz AG)

Lecture T02: Practical Guidelines For Oil & Gas Plant Design Against Sub-Synchronous Torsional Interaction Phenomena
Authors: Federico Sveti, Paola Rotondo, Daniele Sgro, Francesco Meucci (CE Oil & Gas)

Lecture T03: Development and Testing of Multi-Stage Internally Cooled Centrifugal Compressor
Authors: Jason Kerth, Jorge Pacheco (Dresser-Rand)

Lecture T04: Measured Effects Of Liquid Distribution
Authors: Grant Musgrove, Griffin Beck (SWRI)

Lecture T05: The Effect Of Starvation On The Dynamic Response Of Centrifugal Compressors
Authors: John Whalen (John Crane), Yadav Chudakar

Lecture T06: High Performance Bearing Components
Authors: Josh Ronan, Thomas Sheup, Gwendolyn C. Dailey

Lecture T07: Operational modal analysis applied on gas seal
Authors: Alberto Guglielmo, Leonardo Baldassarre, Timbo Silva (Petrobras)

Lecture T08: A Novel Method for Measurement of the Load Applied to the Seal
Authors: Staffan Lundholm, Stefano Morosi

Lecture T09: Measuring and Modeling Vertical Vibration Tests
Authors: Brian Howard, Charles Hatch (CE Oil & Gas)

Lecture T10: Numerical Prediction And Experimental Validation of New Natural Frequencies
Authors: Daniele Panara, Leonardo Baldassarre

Lecture T11: Axial Thrust In High Pressure Compressors
Authors: Michele Fontana, Leonardo Baldassarre

Lecture T12: Applying CFD to Solve a Vibrating Blade Problem
Authors: Roberto Firmento de Noreinha, Diogo Peralta Muniz, Moreira (ESS)

Lecture T13: "Equation 1a," with the Velocity of Vibration of Machinery & Structures When Changing Material & Size
Author: S.T. Myrick, Jr (S.T. Myrick, Jr, PhD, LLC)

Tutorials

Lecture T14: "Equation 1b," with the "Velocity of Vibration of Machinery & Structures When Changing Material & Size"
Author: S.T. Myrick, Jr (S.T. Myrick, Jr, PhD, LLC)
Metadata: *Excel’s external from web* [1]
Lecture T01: Coupled Torsional and Lateral Analysis for the Determination of the Damping of the First Torsional Mode of Synchronous Motor Driven Compressors
Authors: Jean-Claude Pradetto, Urs Baumann (MAN Diesel & Turbo Schweiz AG)

Lecture T02: Practical Guidelines for Oil & Gas Plant Design Against Sub-Synchronous Torsional Interaction Phenomena
Authors: Federico Svetti, Paola Rotondo, Daniele Sgro, Francesco Meucci (GE Oil & Gas)

Lecture T03: Development and Testing of Multi-Stage Internally Cooled Centrifugal Compressor
Authors: Jason Kerth, Jorge Pacheco (Dresser-Rand), Jeffrey Moore, Tim Allison, Neal Evans (SWRI)

Lecture T04: Measured Effects of Liquid Distribution on Compressor Performance During Wet Gas Ingestion
Authors: Grant Musgrove, Griffin Beck (SWRI), Michael Matheidas and Stan Upton (ExxonMobil)

Lecture T05: The Effect of Starvation on the Dynamic Properties of Tilting Pad Journal Bearings
Authors: John Whalen (John Crane), Vaclav Cerny, Vaclav Polreich (Doosan Skoda Power), Minhui He (BRG Machinery Consulting)

Lecture T06: High Performance Bearing Comparison
Authors: Josh Ronan, Thomas Shoup, Gwenael Perney, Chad Robertson (Lufkin), Bernard Quoix, Alain Gelin (Total E&P)

Lecture T07: Operational modal analyses application for the measure of logarithm decrement in centrifugal compressor
Authors: Alberto Guglielmo, Leonardo Baldassarre and Michael Catanzaro (GE), Leandro de Oliveira Zague, Leonardo Ishimoto, Marcelo Accorsi Miranda, Silva (Petrobras)

Lecture T08: A Novel Method for Measurement and Evaluation of Torsional Damping for Geared Turbomachinery During Operation Used to Mitigate Lateral Vibrations
Authors: Staffan Lundholm, Stefano Morosi, Niklas Sehlstedt, Anders Crone and Petter Pehrson (Lloyd's Register Consulting)
| A                  | B                        | C          | D          | E          | F          | G          | H          | I          | J          | K          |
|-------------------|--------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| `turbolab/`       | `proceedings`            | `cds`      | `turbo 44` |            |            |            |            |            |            |            |            |
| Texas A&M University |                           |            |            |            |            |            |            |            |            |            |            |
| Turbolab Website  |                           |            |            |            |            |            |            |            |            |            |            |
| Symposium Technical Papers |                   |            |            |            |            |            |            |            |            |            |            |
| Search            |                           |            |            |            |            |            |            |            |            |            |            |
| 44th Turbomachinery Symposium (2015) |               |            |            |            |            |            |            |            |            |            |            |
| Lectures | Tutorials | Discussion Groups | Case Studies | Short Courses |            |            |            |            |            |            |            |
| Lectures          |                           |            |            |            |            |            |            |            |            |            |            |
| Lecture T01: Coupled Torsional and Lateral Analysis for the Determination of the Damping of the First Torsional Mode | Authors: Jean-Claude Pradetto, Urs Baumann (MAN Diesel & Turbo Schweiz AG) |            |            |            |            |            |            |            |            |            |            |
| Lecture T02: Practical Guidelines For Oil & Gas Plant Design Against Sub-Synchronous Torsional Interaction Phenomenon | Authors: Federico Svetti, Paola Rotondo, Daniele Sgro, Francesco Meucci (GE Oil & Gas) |            |            |            |            |            |            |            |            |            |            |
| Lecture T03: Development and Testing of Multi-Stage Internally Cooled Centrifugal Compressor | Authors: Jason Kerth, Jorge Pacheco (Dresser-Rand), Jeffrey Moore, Tim Allison, Neal Evans (SWRI) |            |            |            |            |            |            |            |            |            |            |
| Lecture T04: Measured Effects Of Liquid Distribution On Compressor Performance During Wet Gas Ingestion | Authors: Grant Musgrove, Griffin Beck (SWRI), Michael Matheidas and Stan Uptigrove (ExxonMobil) |            |            |            |            |            |            |            |            |            |            |
| Lecture T05: The Effect Of Starvation On The Dynamic Properties Of Tilting Pad Journal Bearings | Authors: John Whalen (John Crane), Vaclav Cerny, Vaclav Polreich (Doosan Skoda Power), Minhui He (BRG Machine Tools) |            |            |            |            |            |            |            |            |            |            |
| Lecture T06: High Performance Bearing Comparison | Authors: Josh Ronan, Thomas Shoup, Gwenael Perney, Chad Robertson (Lufkin), Bernard Quoix, Alain Gelin (Total) |            |            |            |            |            |            |            |            |            |            |
Metadata: Source code[1]
Files: DownloadThemAll[2]
Files: DownloadThemAll[3]
Files: DownloadThemAll[4]

turbolab.tamu.edu/proc/turboproc/T44/L01.pdf
Files: DownloadThemAll

Lecture T02: Practical Gui...  Default
Lecture T03: Development...  Default
Lecture T04: Measured Eff...  Default
Lecture T05: The Effect Of ... Default

Mask:
44thTurbo|name*.ext*

Use mask only this time

Audio (mp3, wav, ...)
Images (jpg, png, ...)
AntiContainer

Start! Add paused Cancel

Preferences Sequences
Files: Grabbing file names

- One at a Time
- Get Them in a Batch
Files: *Grabbing file names*[2]

- A short script file placed in desired folder and ran to generate a list of files that can then be used to enter into spreadsheet
- `dir /b` gets a list of all the contents of the folder
- `findstr /v /i "\.bat\$"` excludes the script file
- `findstr /v /i "\ListOfFileNames.txt"` excludes file about to create
- `> "\ListOfFileNames.txt"` Text File created from the above results
Files: Grabbing file names

The publisher could not be verified. Are you sure you want to run this software?

Name: ...\Downloads\GenerateFileNames.bat
Publisher: Unknown Publisher
Type: Windows Batch File
From: C:\Users\\Downloads\GenerateFileNames.bat

Always ask before opening this file

This file does not have a valid digital signature that verifies its publisher. You should only run software from publishers you trust.

How can I decide what software to run?
Files: *Grabbing file names*
Cleaning or Manipulating Spreadsheet Data

MS Excel
OpenRefine
MS Excel[1]

- Advantages
  - Compatible to work with many file types
  - Wide array of tools and functions
  - Tooltips when hovered over or clicked
  - Able to utilize Macros
  - Many resources available online for help

- *Disadvantages
  - Only be edited by one person at a time
Live Demo and Exercise
- Data clean up and manipulation
OpenRefine[1]

Advantages
- Powerful browser-based tool for cleaning and manipulating data
- Many online resources
- Open source community for help and add-ons

Disadvantages
- Features can take a while to learn
OpenRefine

- General Refine Expression Language (GREL)
  - Write commands to manipulate data.
  - Comparable to formulas in Excel.

- Regular Expressions
  - “A regular expression is a string that describes a text pattern occurring in other strings.”
  - Use them to identify parts of strings in a more general way than a simple search.

- Live Demo and Exercise
### OpenRefine: Reconciliation

#### Web Slice Gallery
- Course: Introduction to Me...
- Page: 1
- Safety: Off
- Tool: Permalink

#### Data Table

<table>
<thead>
<tr>
<th>All</th>
<th>dc.contributor</th>
<th>dc.description</th>
<th>Author 1</th>
<th>Author 2</th>
<th>Column 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>International Pump Symposium (1st: 1984)</td>
<td>Lecture</td>
<td>Facet</td>
<td>rei, arno</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>International Pump Symposium (1st: 1984)</td>
<td>Lecture</td>
<td>Sort...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reconciliation[2]

Pick a Service or Extension on Left

Add Standard Reconciliation Service

Enter the service's URL:

http://lc-reconcile.cmh2166.webfactional.com/

Add Service Cancel

http://lc-reconcile.cmh2166.webfactional.com/
<table>
<thead>
<tr>
<th>Date</th>
<th>Series</th>
<th>Page</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>Proceedings of the 1st Turbomachinery Symposium</td>
<td>Pg. 79-89.</td>
<td>Boyce, Meherwan P. Choose new match</td>
</tr>
</tbody>
</table>
OpenRefine: **Reconciliation**[4]

<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Pages</th>
<th>Authors</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>Proceedings of the 1st Turbomachinery Symposium</td>
<td>Pg. 56-61</td>
<td>Herbage, B. S.</td>
<td>Create new topic</td>
</tr>
<tr>
<td>1972</td>
<td>Proceedings of the 1st Turbomachinery</td>
<td>Pg. 75-78</td>
<td>Housmnn, J. G.</td>
<td>Create new topic</td>
</tr>
</tbody>
</table>
Using MARC Records
MARC records to what?

- MARCXML [HathiTrust]
- Dublin Core, etc.
- Spread Sheet
  - Microsoft Access
  - MARCEdit
- From Spreadsheet/table to MARC
MARC to spreadsheet: Access

Advantages
- Run queries on data
- Query local catalog
- Combine metadata in different tables together

Disadvantages
- Training needed to build queries
- Not built to easily manipulate data
Getting MARC records

- Ask ILS manager to retrieve records
- OCLC Export File
  Declare export file destination on Export tab under Tools -- Options Menu.
- From Local Catalog
  Example: Voyager. Go to Session Defaults and Preferences—Folders/Files. Set local save file to where you would like to collect the records. Also set Character Set as desired.
MARC to MARCXML: OCLC

Options

Destination

E

File: (Prompt for filename)

File:

Record Characteristics

Bibliographic Records

Record Standard:

MARC 21

Authority Records

Record Standard:

MARC 21

Character Set:

UTF-8 Unicode
Flood insurance study.
City of Abilene, Texas, Taylor and Jones Counties
Abilene, Texas
Taylor County, Texas
Jones County, Texas
=245 00$aFlood insurance study.$pCity of Abilene, Texas, Taylor and Jones Counties.
=246 1$aCity of Abilene, Texas, Taylor and Jones Counties
=246 3$aAbilene, Texas
=246 3$aTaylor County, Texas
=246 3$aJones County, Texas
=300 $aiii, 33 pages, 18 folded leaves of plates :$billustrations, maps ;$c28 cm
=500 $a"January 1978."
=504 $aincludes bibliographical references (pages 32-33).
=500 $aLocation map on cover.
=500 $aTitle from cover.
=650 0$aFlood forecasting$zTexas$zAbilene.
=650 0$aFlood forecasting$zTexas$zTaylor County.
Editing records: MARCEdit[1]
=LDR 01890cam a2200481 i 4500
=001 5432554
=005 20170425103949.0
=008 170424s1978\\dcuabf\b\f000\0\eng\c
=035 \$a(OCoLC)ocn983796457
=040 \$aTXA\$beng\$serd\$cTXA\$dTXA
=035 \$a(OCoLC)983796457
=042 \$apcc
=050 \$aGB1399.2
=086 0\$aHH 10.9:Ab 5
=049 \$aTXAM
=245 00\$aFlood insurance study.$pCity of Abilene, Texas, Taylor and Jones Counties.
=246 1\$aCity of Abilene, Texas, Taylor and Jones Counties
=246 30\$aAbilene, Texas
=246 31\$aTaylor County, Texas
=246 31\$aJones County, Texas
=300 \$a33 pages, 18 folded leaves of plates :$billustrations, maps ;$c28 cm
=500 \$a"January 1978."
=504 \$aIncludes bibliographical references (pages 32-33).
=500 \$aLocation map on cover.
=500 \$aTitle from cover.
=650 \$0aFlood forecasting$zTexas$zAbilene.
=650 \$0aFlood forecasting$zTexas$zTaylor County.
=650 \$0aFlood forecasting$zTexas$zJones County.
Editing records: MARCEdit[3]
MARC to spreadsheet: MARCEdit

Utilities Dialog

Batch Process
MARC Records

Export Tab Delimited Records

Step 1: Set File Paths

Select Field Delimiter

In field delimiter

Contextual delimiter

Next >>

Close
Questions
Tools: Questions

- What do you already have in-house?
- What are the financial considerations?
- How flexible does your tool need to be in letting you define elements?
- To what schema do you want to migrate the data?
- How easily does any tool permit you to move the data into that schema?
- Who will be inputting the data? How much training will they need? Do you have an in-house expert?
- How many people will be inputting data? From where?
- How much assistance do you need from IT?