Creating context for the experiment record: user-defined metadata


16 March 2014
Automatic capture of metadata

Flexibility could mean a lack of context.
LabTrove and User defined metadata
Metadata Survey

blogs@ChemTools
The thoughts of Chemists

ALTC Enotebook

UNSW
The University of New South Wales
Sydney • Australia
enotebook

blogs@BioLab
Open output from the BioLab

our experiment
alpha

blogs@xray
The thoughts of the ORC Xray Group

104 blogs
A big proportion of our users are not really adding metadata!
Number of authors

- Mean metadata count
  - Avg sections
  - Avg Keys
  - Avg Values
Most metadata is high-level not specific

Specific
- Malaria
- SHG
- Porphyrin chemistry
- PS4
- CCD Positions
- Surface tension
- Model compounds
- Perl

High-level
- Samples
- Papers
- Experiments
- Reagents and materials
- Meetings
- Data analysis
- Conference
- Procedures
- Safety
- Planning
Sections as labels
Keys as labels
Values

- Materials: 21%
- Activities: 13%
- Codes: 14%
- Other: 36%
- Date and Values: 7%
- Equipment: 2%
- Location: 3%
- Adjectives: 3%
- Labels: 2%

Metadata anxiety and other problems

- ‘Blank canvas’ effect
- Consistency
- Duplication
- Mistakes
- Inappropriate metadata
A preference for ‘things’ over ‘activities’?

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Adjectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sections</td>
<td>513</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>Keys</td>
<td>263</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Values</td>
<td>893</td>
<td>40</td>
<td>26</td>
</tr>
</tbody>
</table>

Nominalization:

making a noun from a verb or adjective

“zombie nouns”
Categorizing ‘activity’ nouns

I’ll come back to classifying activities later..
Is LabTrove metadata different?

- Flickr
- NASA blogs
- Chemistry-related blogs
- myExperiment
Editing 1 photo:

**On the bike**
Add a description

Add tags
Separate tags with a space

Add people

Add to sets

Add to groups

Owner settings
   - None (All rights reserved)
   - Visible to everyone

**Tags**
- sunset ×
- landscape ×
- doodle ×
- art ×
- acrylics ×
- totoro ×

**Tags**
click a tag: 2012 animals archaeology autumn berlin bw cats cleveland dig dinnington doodle eb eb2008 eclipse england expo extremeblue f1 florida flower flowers g7 gardens green hampshire hilliers history honeymoon hursley hursley50 iphone japan july keukenhof kimahri landscape legoland macro marwell may me moscow nature niigata ohio orlando panto paris phil picnic pink portrait red roman sc7 silverstone spring sunset sydney team tree ut utpicnic villa water wedding winchester windsor xmas yellow

get more of your tags | get all of your tags

Separate each tag with a space: cameraphone urban moblog. Or to join 2 words together in one tag, use double quotes: "daily commute".
Flickr results

1381 tags

Top 500 most interesting photos tagged with ‘chemistry’ and ‘experiment’
Flickr results

- Nouns: 79%
- Adjectives: 14%
- Verbs: 7%

- Other: 35%
- Material: 8%
- Equipment: 11%
- Location: 11%
- Activity: 10%
- People: 9%
- Event: 2%
- Other: 14%
when we were on the plane and were doing the set-up and startup, I was really “uber-focussed” on the payload and not on myself for the first few cycles. When things started to get into a rhythm around Parabola #5 I had no idea we were 1/5th of the way done. Wow.

“That was short. That was very short.” My comments after the very first parabola, which was a Martian (0.33 G) scenario. This image shows our team’s positions in between parabola 1 & 2. We did not have space enough to fully lay down so we reclined against the side of the aircraft. Left to right is Con Tsang, myself (monitoring a payload via a table), Cathy Olkin, and Alan Stern (face not visible). The photo is taken via Go-Pro camera on the head of Dan Durda who was across the way. Eric Schindhelm, who rounded out our team, was next to Dan and not in this view.
NASA results

53 NASA blogs – 1226 ‘tags’
Nasa results
Chemistry blogs results

chemistry, biology, nmr, mathematics, research, drug discovery, technology, medicine, science, medicine, health space, open access, academia, dna, education, protein, software, evolution, msnc, ecology, china, malaria, xml, pharmacy, pharmacology, medicinal chemistry, polymer, conferences, review, books, agriculture, art, safety, carbon dioxide, recycling, travel, nanotechnology, polymer, medicine, bone, bone, uncategorized, biology, chemistry, organic chemistry, cancer, physics, fund, design, music, patents, culture, fish, periodic table, business, computing, cheminformatics, quantum mechanics, chemspider, engineering, merck, google, food, spirituality, synthesis, competition, aroma, religion, solar power, radiation, coffee, diabetes, cdk, proteomics, green, patent, pubchem, molecules, video, chemicals, spectroscopy, history, ecology, china, malaria, uncategorized, biology, chemistry, organic chemistry, cancer, physics, fund, design, music, patents, culture, fish, periodic table, business, computing, cheminformatics, quantum mechanics, chemspider, engineering, merck, google, food, spirituality, synthesis, competition, aroma, religion, solar power, radiation, coffee, diabetes, cdk, proteomics, green, patent, pubchem, molecules, video, chemicals, spectroscopy, history, ecology, china, malaria
Chemistry Blog results
Survey of 10% of 2390 flows

Number of tags per flow:
- 1-5: 50%
- 6-10: 38%
- 11-15: 5%
- 16-20: 1%
- >20: 2%
- 0: 4%

Tags (7):
- cme
- hec
- heliophysics
- lasco
- propagation
- solar
- wsdl

Add Tags:
(use commas to separate different tags)

Add

This Workflow has no tags!
myExperiment results

- Nouns: 90%
- Verb: 7%
- Adjective: 3%

- Other: 83%
- Activity: 9%
- Adjective: 3%
- Event: 0%
- Location: 1%
- Material: 1%
- Person: 3%

2349 tags
Classifying activities

- Potential to differentiate between action, activity, process, and technique
- But need to understand who is doing the activity or to what the activity is being done: person, material, or data?
- Intent is difficult to determine from ‘activity-related’ metadata
Summary

• Metadata has to be present to be useful!

• Common mistakes and “sense-making”

• Tells us about the interests and needs of the communities

• Users will use predefined values if they are available

• A number of possible strategies to encourage better use (and user experience) of metadata in ELNs
Futures

• Our findings have already had an influence on our ELN developments

• Considerations for our ELN interface designs in the future

• Links to our current research on how we can use cues to influence what students and researchers record.
Metadata Study

http://sites.google.com/site/cmetastudy/