Connections:

Piloting linked data to connect library and archive resources to the new world of data, and staff to new skills

Laura Akerman  
Metadata Librarian  
Robert W. Woodruff Library  
Emory University

Zheng (John) Wang  
AUL, Digital Access, Resources, and IT  
Hesburgh Library  
Notre Dame University
Who has presented most frequently at CNI?
Current Model: Search and Discover
Who presented most frequently at CNI

Membership - Coalition for Networked Information
www.cni.org \ About CNI
Apr 24, 2012 – More information is available from Key Benefits of CNI Membership and the ... Higher education institutions most often appoint their heads of ...

CNI Shared Shelf presentation (James Shulman).pptx - ARTstor
www.artstor.org/shared-shelf/5-pdf/CNI-Shared-Shelf-ARTstor.pdf
File Format: PDF/Adobe Acrobat - Quick View
CNI. December 15, 2009. • Genesis. • Our partners. • Why Shared Shelves? • ... Most requested enhancements to hosting service: • more frequent updates ...

CNI News
www.cni.org/news/
4 days ago – CNI Mtg Program/Schedule of Events Now Available ... The document includes a schedule of project briefings to be presented at the meeting, as well as ... For more information, contact Angela Murillo, Project Manager, ...

California Numismatic Investments - What Others Have Said About ...
www.golddealer.com/testimonials.html
He gave me the tracking number and said that he would check with your shipper. .... the website must be bringing eyeballs to the site on a more frequent basis.

From the Guest Editor - CNI Review Medical Journal - Online Library ...
www.theeni.org/reviews/10-2-editor.htm
Everyone who contributed to this journal did so in a most gracious manner, ...
Alzheimer's disease is a topic that is often presented in the lay media, which ...

Epilepsy - CNI Review Medical Journal - Online Library - Colorado
www.theeni.org/reviews/14-0-abstracts.htm
To view these articles: This issue of the CNI REVIEW is provided in Adobe Acrobat ... (PNES) are perhaps most often encountered as something of a hidden syndrome. ...
The medical information presented on this website is meant for general ...

Latest CNI News - Colorado Neurological Institute
www.theeni.org/Public/News-Events/LatestCNINews/index.cfm
About CNI - Find a Physician - Patient Programs - Services - Rehab Therapy .... balance and other symptoms, Parkinson's disease most often affects people over the
Metadata Published as Documents
Require Human to Decipher
Linked Data Model: Find
Semantic Graph Model
Machine Understands Semantics
RDF Triple

Subject — Predicate — Object
RDF Triple

Laura \(\rightarrow\) Lecture \(\rightarrow\) Connections
Relevant to What We Do

Reuse, Authority Control, Knowledging Linking...
Connections Pilot

To Interlink EAD, Catalog, and Other External Resources
Connections: Context

Little Time to Learn Additional New Things
Hands-on learning
Ingredients

• Leader/teacher/evangelist
• Learning group – open to all
  o 2 "classes" a month, 5 months.
• Pilot: 3 months
  o Brainstorming a pilot project
  o Start small
    o Team: programmer, subject liaison, metadata specialists, archivist, digital curator, fellow.
  o 1-3 hrs/week for all but leader
  o A sandbox running Linux
The Pilot: Grand Ambitions
Integrate linked data into discovery layer (catalog)?

Our Own Triplestore

RDF from EAD

DBPedia

id.loc.gov

RDF from TEI

Civil War

RDF from MARCXML (and MARC)

SPARQL

User interface Navigation

Timelines

Maps

Crowdsourcing

Faculty project

Redesign metadata creation as RDF

Data from other archives

CW150

Other data

National Park Service Data

Rosters
3 months later...
Sampling little bites of the meal:

- id.loc.gov URIs for LC subjects and names (scripted)
- DBPedia/subjects (by hand)
- EAD (starting from ArchiveHub stylesheet)
- MARCXML (starting from LC DC stylesheet)
- Make some RDF metadata
- Sesame triplestore
- Visualization – Simile Welkin
HTTP:OurResourceURL

HasSubject

"Mobley, Thomas"
hasSubject
HTTP:OurPersonMobleyT1

memberOf
Confederate States of America. Army. Georgia Infantry Regiment, 48th
Confederate miscellany collection, 1860-1865

isPartOf

heldBy

MARBL
Manuscript, Archives, & Rare Book Library

EMORY LIBRARIES
EmoryFindingAids
We learned:

Selecting material that will “link up” without SPARQL, is too hard!

Even when items are in a unified “discovery layer”, the types of search are limited.

Get it into triples, then find out!
We learned:

There are many ways of modeling data
ArchivesHub handles subjects:

<associatedWith>

<!--About the Concept (Person)-->

<skos:Concept
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
rdf:about="http://duchamp.library.emory.edu/resource/id/concept/person/lcnaf/gearyjohnwhite1819-1873">


  <skos:inScheme>
    <skos:ConceptScheme rdf:about="http://duchamp.library.emory.edu/resource/id/conceptscheme/lcnaf">
      <rdfs:label xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xml:lang="en">lcnaf</rdfs:label>
    </skos:ConceptScheme>
  </skos:inScheme>

  <foaf:focus xmlns:foaf="http://xmlns.com/foaf/0.1/">
    <!--About the Person-->
    <foaf:Person
      rdf:about="http://duchamp.library.emory.edu/resource/id/person/lcnaf/gearyjohnwhite1819-1873">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
      <rdf:type rdf:resource="http://purl.org/dc/terms/Agent"/>
    </foaf:Person>
  </foaf:focus>

</skos:Concept>
</associatedWith>
dc:subject "Geary, John White, 1819-1873."
Simile MARC to MODS to RDF:

<modsrdf:subject rdf:resource="http://simile.mit.edu/2006/01/Entity#Geary_John_White_18191873"/>

<rdf:Description rdf:about="http://simile.mit.edu/2006/01/Entity#Geary_John_White_18191873">
  <rdf:type rdf:resource="http://simile.mit.edu/2006/01/ontologies/mods3#Person"/>
  <modsrdf:fullName>Geary, John White</modsrdf:fullName>
  <modsrdf:dates>1819-1873</modsrdf:dates>
</rdf:Description>
We learned:

Linked data is HUGE
It’s coming at us FAST
It’s not “cooked” yet
More learnings

• We learned more by doing than by "class".

• Making DBPedia mappings or links by hand is very time consuming! We need better tools.

• We need to spend a lot more time learning about OWL, and linked data modeling.
Challenges

• Easily available tools are not ideal!
• Skills we needed more of: HTML5, CSS, Javascript
• Time!
• Visualization/killer app not there yet.
• Can't do things without the data! No timeline if no dates!
What we got out of it

Test triplestore for training and more development
Better ideas on what to pilot next
Convinced some doubters
"Gut knowledge“ about triples, SPARQL, scale
Beginning to realize how this can be so much more than a better way to provide "search"
Outside our reach for now

Transform ILS system to use triple store instead of MARC
Create hub of all data our researchers might want
Make a bank of shared transformations for EAD, MARC, etc.
Shared vocabulary mappings
Social/networking aspect (e.g. Vivo, OpenSocial...) - need a culture shift?
Next? Maybe...

Build user navigation?
More Civil War triples including other local institutions’ stuff?
Publishing plan?
Integrate ILS with DBPedia links?
Suite of “portal tools” for scholars?
Use linked data for crowdsourcing metadata?
More classes?
Connect with others at Emory around linked data
Recommendation: Individual Institutions

• Focus on unique digital content
• Publish unique triples
• Reuse existing linked data
Recommendation: Community

- Create standards or best practices
- Grow our skills
- Test and evaluate tools
- Develop tools
Recommendation:
Librarians’ Role?

- Interdisciplinary linking?
- Metadata librarians - Linking association and normalization
Acknowledgements

Connections group sponsors: Lars Meyer, John Ellinger

Connections Pilot team: Laura Akerman (leader), Tim Bryson, Kim Durante, Kyle Fenton, Bernardo Gomez, Elizabeth Roke, John Wang

Fellows who joined us: Jong Hwan Lee, Bethany Nash

Our website: https://scholarblogs.emory.edu/connections/

Laura Akerman, liblna@emory.edu

John Wang, Zheng.Wang.257@nd.edu
Thanks

Q&A