Transformation and Enrichment: Activating archival descriptions as Linked Data

Jeff Mixter
Software Engineer, OCLC Research

Bruce Washburn
Consulting Software Engineer, OCLC Research
What we’ll be talking about …

• How OCLC Research generates Linked Data
• How we’re experimenting with Linked Data from archival collection records in an experimental web application
• Findings and next steps
Understanding “RDF Triples”

A triple is a statement that relates one thing to another, specifying a Subject, Predicate, and Object. RDF triples use URIs for those three elements.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barack Obama</td>
<td>Was born in</td>
<td>Honolulu, Hawaii</td>
</tr>
</tbody>
</table>
Creating Knowledge Triples from record-oriented data

MARC Records

- FRBR Clustering
- String matching with controlled vocabularies
- Addition of standard identifiers
Creating Knowledge Triples from record-oriented data

MARC Records

- Enhanced WorldCat MARC Record
  - FRBR Clustering
  - String matching with controlled vocabularies
  - Addition of standard identifiers

RDF Entities

- Persons
- Organizations
- Places
- Concepts
- Events
- Works
Creating Knowledge Triples from record-oriented data

<table>
<thead>
<tr>
<th>MARC Records</th>
<th>RDF Entities</th>
<th>Triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced WorldCat MARC Record</td>
<td>Persons</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Organizations</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Places</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Concepts</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Events</td>
<td>Subject</td>
</tr>
<tr>
<td></td>
<td>Works</td>
<td>Subject</td>
</tr>
</tbody>
</table>

- FRBR Clustering
- String matching with controlled vocabularies
- Addition of standard identifiers
MARC Records for archival materials, and linked data challenges

• A single record may contain many subjects: people, groups, and places
• The relationships between these subjects isn’t always clear
• These entities are not always “notable” … they may lack identifiers in library authority systems (if not affiliated with published works) and lack identifiers elsewhere (if not notable enough to warrant a Wikipedia article)
Creating a Library Knowledge Vault

• Triples in a library knowledge vault provide opportunities for applications supporting discovery, editing, visualization, and more

• OCLC Research is investigating what it’s like to assemble and work with this kind of data in an experimental discovery system we call “EntityJS”
The EntityJS Research Project

Get some real-life experience with using Linked Data, test entity refinement and editing, and push triples back to the knowledge vault.
Search across entities

**Person**
- Abraham Lincoln, 1797 - 1865
- Mary Todd Lincoln, 1810 - 1878
- Robert Todd Lincoln, 1843 - 1926
- Lincoln Family
- Tad Lincoln, 1853 - 1865
- Nancy Hanks Lincoln, 1814 - 1818
- Lincoln Steffens, 1866 - 1936
- Levi Lincoln, 1782 - 1868
- George Lincoln Burr, 1857 - 1938

**Organization**
- William E. Barton Collection of Lincolniana (University of Chicago)
- Abraham Lincoln Presidential Library and Museum.
- Chamber Music Society of Lincoln Center.
- Jazz at Lincoln Center (Organization)
- United States President (1861-1865; Lincoln).
- Lincoln Center Jazz Orchestra.
- McLellan Lincoln Collection (Brown University)
- Lincoln Memorial University
- Lincoln University (Pa.)
- Ford Motor Company, Lincoln-Mercury Division.

**Concept**
- Abraham Lincoln's political career before 1861
- Lincoln automobile
- Lincoln automobile
- Funeral journey of Abraham Lincoln to Springfield
- Lincoln Memorial (Washington, D.C.)
- Lincoln-Douglas Debates, Ill., 1858
- Lincoln Continental automobile
- Lincoln Continental automobile
- Lincoln Day
- Lincoln (Bomber)

**Place**
- Nebraska—Lincoln
- Massachusetts—Lincoln
- Maine—Lincolnville
- Illinois—Lincoln
- England—Lincolnshire

**Event**
- Lincoln-Douglas Debates (Illinois : 1858)
- Lincoln Center Festival.
- Lincoln Center Festival.
- Lincoln Center Festival.
- Lincoln Center Festival.

**Work**
- Gettysburg address (Lincoln, Abraham)
- Emancipation Proclamation (United States. President (1861-1865 : Lincoln))
- Lincoln Truck Rental photographs, 1947 Feb. 17-1958 April 19 : 73 Gilbert Street
- Fort Lincoln : history and construction
Show related entities

Lincoln, Robert Todd, 1843-1926

About
Born: Aug 1, 1843
Died: Jul 26, 1926

Robert Todd Lincoln (August 1, 1843 – July 26, 1926) was an American lawyer and Secretary of War, and the first son of President Abraham Lincoln and Mary Todd Lincoln. Born in Springfield, Illinois, United States, he was one of two of Lincoln's four sons to live to adulthood, and the only one to survive into the 1900s.

Father:
Abraham Lincoln

Mother:
Mary Todd Lincoln

Spouse:
Mary Eunice Harlan

Related Works
As a Creator

[Letter] 1911 April 28 [to] Mr. Starrett?


Autograph letter signed from Robert Todd Lincoln, Chicago, to Miss Post 1901 January 29.

Autograph signature of Robert Lincoln to a typewritten letter : Chicago, to Mr. Morgan, 1908 Jan. 20.

Letter: Chicago, to Belle F. Keyes, Boston, 1885 Dec. 16.

Letter: Chicago, to Clinton L. Conkling.

Related People
Abraham Lincoln
Mary Todd Lincoln
William Eleazar Barton
Mrs. James G. Blaine

Related Organizations
William E. Barton Collection of Lincolniana (University of Chicago)
Republican Party (U.S. : 1854- )
United States. Supreme Court.
United States. War Department

Related Places
United States
Illinois.
Illinois–Chicago.
New York (State)
Show related entities

<table>
<thead>
<tr>
<th>Abraham Lincoln</th>
<th>William E. Barton Collection of Lincolniana (University of Chicago)</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary Todd Lincoln</td>
<td>Republican Party (U.S. : 1854-)</td>
<td>Illinois</td>
</tr>
<tr>
<td>Mrs. James G. Blaine</td>
<td>United States. War Department</td>
<td>New York (State)</td>
</tr>
<tr>
<td>William Tecumseh Sherman</td>
<td>United States. Congress</td>
<td>Kentucky</td>
</tr>
<tr>
<td>William McKinley</td>
<td>League to Enforce Peace (U.S.)</td>
<td>Kentucky</td>
</tr>
<tr>
<td>Thomas Lincoln</td>
<td>United States. Department of Justice. Office of the Attorney General</td>
<td>Europe</td>
</tr>
</tbody>
</table>

**Related Concepts**
- Families
- Clippings (Books, newspapers, etc.)
- Political science
- Presidents' spouses
- Mental health
- Homes
- Lawyers
- Diplomatic relations
- Manners and customs
- Travel

**Related Events**
- American Civil War (1861-1865)
- World War (1914-1918)
- Reconstruction (United States : 1865-1877)
- World's Columbian Exposition.
- Spanish-American War (1898)
- Indian Wars (Dakota : 1862-1865)
- Appomattox Campaign (1865)
- World War (1914-1918)
- Paris Peace Conference (1919-1920)
- Cowpens, Battle of (South Carolina : 1781)

**Also Known As ...**
- Robert Todd Lincoln
- ロバート・トッド・リンカーン
- 龍伯特·托德·林肯
Show related entities
User-contributed “same as” relationships
User-contributed “same as” relationships

Casablanca Conference

About
The Casablanca Conference (codenamed SYMBOL) was held at the Anfa Hotel in Casablanca, French Morocco from January 14 to 24, 1943, to plan the Allied European strategy for the next phase of World War II. In attendance were United States President Franklin D. Roosevelt, British Prime Minister Winston Churchill, and representing the Free French forces, Generals Charles de Gaulle, and Henri Giraud. Premier Joseph Stalin had declined to attend, citing the ongoing conflict in Stalingrad required his presence in the Soviet Union. The conference agenda addressed the specifics of tactical procedure, allocation of resources and the broader issues of diplomatic policy. The debate and negotiations produced what was known as the “Casablanca Declaration,” and what is, perhaps, its most historically provocative statement of purpose, “unconditional surrender.” The doctrine of “unconditional surrender” came to...

Related Works
As a Subject
[Address by Franklin D. Roosevelt at the dinner of the White House Correspondents’ Association, Washington, D.C. on February 12, 1943. Part 1]
[Address by Franklin D. Roosevelt at the dinner of the White House Correspondents’ Association, Washington, D.C. on February 12, 1943. Part 2]
[Address by Franklin D. Roosevelt at the dinner of the White House Correspondents’ Association, Washington, D.C. on February 12, 1943. Part 3]
[Address by Franklin D. Roosevelt at the dinner of the White House Correspondents’ Association, Washington, D.C. on February 12, 1943. Part 4]
[Address by Franklin D. Roosevelt at the dinner of the White House Correspondents’ Association, Washington, D.C. on February 12, 1943. Part 5]
[Gabriel Heatter news. 1943-02-23, part 1].

Related Media
Source: Wikimedia Commons

Related People
J David Goldin
Douglas Cornell
Joseph Stalin
Winston Churchill
Viscount Bernard Law Montgomery Montgomerey

Related Organizations
Marr Sound Archives.
Andrew W. Mellon Foundation.
White House Correspondents’ Association.
United States. Department of State.
United States. Armv. Armv. Third

Related Places
United States
Soviet Union
Italy
China
United States
Testing with a subset of Knowledge
Just the “ArchiveGrid” WorldCat MARC records

- WorldCat
- ArchiveGrid
- EntityJS
- VIAF
- FAST
- DBPedia
- Wikidata

Extractor

Knowledge Triples

Fusers

Scored Triples

Vault Services

Application Triples

Knowledge Vault
Continued Experimentation

- Build a way to assign confidence levels to data contributed by EntityJS
- Use confidence levels as input to a Fusion process to create Scored Triples
- Extend the EntityJS application to incorporate additional Linked Data resources and support further entity relationship refining and editing
Looking inside the Library Knowledge Vault

Contact us

Jeff Mixter  
Software Engineer, OCLC Research  
mixterj@oclc.org

Bruce Washburn  
Consulting Software Engineer, OCLC Research  
bruce_washburn@oclc.org

Because what is known must be shared.℠