The Potential of Linked Data
Stephen Rhind-Tutt, President

Singapore, August 2013
Overview

- Some background
- The landscape - challenges facing academic publishing
- What’s the role of (open) linking?
- ASP’s experience of linking between databases
- Inbound links
- Outbound links
- Are we moving fast enough?
- Where will it end?
- If we have enough time…a couple of exciting projects
Some background
• 13 years old
• 100+ people
• Based in Alexandria, Virginia
• Offices in the UK, ANZ, Kuala Lumpur, Shanghai
• ‘Born digital’ publisher
• Improve access and utility for learning and research
Our Purpose

Make silent voices heard
The Landscape… challenges facing us
By 2020 the web will contain…?

• 90% of published works prior to 1923

• Majority of works published to 2020

• > 100 billion pages of e-mail, phone logs, databases, blogs, and Web sites (currently 45 billion)

• > 1 Trillion Photos
  (As of 1/13 Facebook had 240 Bn photos. 1.1 Bn uploaded over New Years alone)

• > 100 million facsimile pages of manuscripts

• > 30 million published audio files

• > 2.4 Bn video files on YouTube (1 Bn unique users per month)
Exponentially more use…

- >1 billion unique users monthly
- >4 billion hours watched monthly
- 72 hours of video are uploaded every minute
- 70% of YouTube traffic comes from outside the US
- In 2011, YouTube had more than 1 trillion views or around 140 views for every person on Earth

What does this have to do with academic publishing?
Watch out…
We are going to need bigger boats
What happens when?

• (Almost) all content is machine readable
• (Almost) everyone can access it instantly
• (Almost) everyone uses it

• Machines must help us learn, discover, understand
  • Search
  • Publish
  • Analyze
  • Process data
  • Identify patterns
  • Draw conclusions
  • Etc…
  • To do e-science, e-humanities,
Evolution of publisher tasks

- Editorial
  - Typesetting
  - Printing
  - Print monograph
  - Print directory
  - Warehousing
  - Simple, one database search
  - Public domain reprints
  - Unified search software
  - Licensing

- Commissioning
  - Fading
  - Growing
  - Process integration
  - Workflow tools
  - Semantic indexing
  - Linking
  - Community building
  - Asset management
  - Free materials
  - Rare and unpublished material
  - Unified search software
  - Licensing
  - Quality
  - Selection
  - Editorials
  - Quality
  - Selection
  - Community building
  - Asset management
  - Free materials
  - Rare and unpublished material
  - Unified search software
  - Licensing

- Quality
  - Growing
  - Selection
  - Editorial
  - Commissioning
What role does (open) linking have to play?
It’s the next hurdle

Connectivity Era

- Mainframe Computing
  (organizational, department productivity)
- Personal Computing
  (individual productivity tools)
- Physical Connectivity
  (networked, group productivity information overload)
- Logical Connectivity
  (wireless/service oriented, semantic, event-based)
- Embedded Connectivity
  (smart, dynamic)

Computing Era


Adapted from Gartner Group research
It’s the nature of the Web

- Atomic
- Interconnected
- Interdependent
- Value of the Connection = Value of the object

- Pliable
- Constantly evolving
- Practically unlimited in size
Linking

• Speeds research and learning
  • Fewer interfaces to learn and navigate
  • Reduces duplication
  • Increases comprehensiveness
  • Allow unified annotation and

• Lowers costs
  • Libraries don’t purchase same content twice
  • Publishers don’t need to license and re-sell ubiquitous content
  • Obviates need for updates

• Maximizes usage
  • Content can be part of every search

• Increases functionality
  • Allows machine manipulation of data
  • Linkages not just between content but between apps
It’s the future of academic publishing…

• We seek knowledge and learning, not journals, books and publishers

• Publishers need to deliver the right content, in the right way to the right people.

• In a network a node is less important than the links between the nodes. No one site will have all information

• We need actionable, high quality access to content across different publishers, libraries and websites.
Where we’re headed


Who, What, When, Where?

Why?

Therefore

Who, What, When, Where?
ASP experience in linking
How to build links between lots of relevant but different files?
Began with…

100,000+ pages, including previously unpublished manuscripts
Analyzed what’s studied in the “real” world
Built a structure to capture it…
Find Battles

Use this screen to find battles by specific criteria.

**Battle Name:**

**Battle Location:**

**Battle Description Keyword:**

**Year:**

**Month Started:**

**Day Started:**

**Day of Started:**

**Duration:**

**Theater:**

**Campaign:**

**Winner:**

**Leaders:**

**Troop Size:**

**Confederate Troop Size:**

**Union Troop Size:**

**Overall Losses:**

**Confederate Losses:**

**Union Losses:**

**Confederate Wounded:**

**Union Wounded:**

**Confederate Missing in Action:**

**Union Missing in Action:**

[SEARCH] [CLEAR]
Launched the Civil War Research Database

Death by Disease Analysis

<table>
<thead>
<tr>
<th>Regiment Short Name</th>
<th>Total Men</th>
<th>Combat Pop</th>
<th>Enlist/Comm</th>
<th>Draft</th>
<th>Trans In</th>
<th>Killed/Mortal Wound</th>
<th>Died POW</th>
<th>Died Disease</th>
<th>Disabled</th>
<th>Desert</th>
<th>Disch</th>
<th>Muster Out</th>
<th>Transf Out</th>
<th>???</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>118,391</td>
<td>116,721</td>
<td>110,949</td>
<td>2,584</td>
<td>4,609</td>
<td>5,993</td>
<td>1,692</td>
<td>5,642</td>
<td>15,876</td>
<td>8,726</td>
<td>8,186</td>
<td>61,058</td>
<td>8,373</td>
<td>2,278</td>
</tr>
<tr>
<td>Weighted Averages</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 12th MA Light Artillery</td>
<td>291</td>
<td>290</td>
<td>99.65</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>8.58</td>
<td>5.16</td>
<td>29.90</td>
<td>1.38</td>
<td>51.55</td>
<td>0.00</td>
<td>1.71</td>
</tr>
<tr>
<td>2 13th MA Light Artillery</td>
<td>341</td>
<td>341</td>
<td>87.68</td>
<td>0.00</td>
<td>12.32</td>
<td>0.00</td>
<td>0.00</td>
<td>5.27</td>
<td>5.28</td>
<td>28.74</td>
<td>7.91</td>
<td>44.28</td>
<td>5.57</td>
<td>0.29</td>
</tr>
<tr>
<td>3 2nd MA Cavalry</td>
<td>2,588</td>
<td>2,578</td>
<td>99.34</td>
<td>0.04</td>
<td>0.23</td>
<td>3.10</td>
<td>2.16</td>
<td>2.55</td>
<td>6.53</td>
<td>24.15</td>
<td>8.06</td>
<td>38.60</td>
<td>3.35</td>
<td>10.77</td>
</tr>
<tr>
<td>4 15th MA Light Artillery</td>
<td>360</td>
<td>360</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7.50</td>
<td>3.60</td>
<td>23.89</td>
<td>3.60</td>
<td>41.93</td>
<td>17.21</td>
<td>0.28</td>
</tr>
<tr>
<td>5 47th MA Infantry</td>
<td>1,141</td>
<td>1,136</td>
<td>99.39</td>
<td>0.00</td>
<td>0.17</td>
<td>0.09</td>
<td>0.00</td>
<td>3.07</td>
<td>4.82</td>
<td>19.62</td>
<td>3.58</td>
<td>67.65</td>
<td>0.60</td>
<td>0.09</td>
</tr>
<tr>
<td>6 88th MA Infantry</td>
<td>2,083</td>
<td>2,082</td>
<td>91.07</td>
<td>8.50</td>
<td>0.37</td>
<td>11.38</td>
<td>2.16</td>
<td>2.39</td>
<td>20.30</td>
<td>18.15</td>
<td>7.20</td>
<td>16.89</td>
<td>8.83</td>
<td>10.28</td>
</tr>
<tr>
<td>7 11th MA Infantry</td>
<td>2,068</td>
<td>2,055</td>
<td>80.71</td>
<td>8.12</td>
<td>10.54</td>
<td>7.20</td>
<td>1.40</td>
<td>2.43</td>
<td>18.12</td>
<td>17.83</td>
<td>10.10</td>
<td>31.96</td>
<td>5.70</td>
<td>2.37</td>
</tr>
</tbody>
</table>
User wants to cross search and cross link

Illustrated Civil War Newspapers

Charles W. J.—is a member of Company K, One Hundred and Twenty-First New York Regiment, and he has been transferred to this company and regiment from Company F of the Sixteenth New York. He has been thus transferred for the reason that the Sixteenth New York is a two years' regiment, whose time has expired, while he is a three years' recruit, who has a year or two more to serve. Now he claims that pay is due him from November 1, 1862, to August 1, 1864, and

American Civil War Research Database

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Rank</th>
<th>Enlist.</th>
<th>Discharge</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>James B Jencks</td>
<td>K</td>
<td>08/23/62</td>
<td>11/10/64</td>
<td>died wounds</td>
</tr>
<tr>
<td>77</td>
<td>Charles W Johnson</td>
<td>K</td>
<td>05/11/63</td>
<td>08/24/65</td>
<td>Mustered Out Hospital, Washington, DC</td>
</tr>
<tr>
<td>78</td>
<td>Michael Kellar</td>
<td>K</td>
<td>03/28/65</td>
<td>06/24/65</td>
<td>Transferred</td>
</tr>
</tbody>
</table>
Resource Description Framework (RDF)

Make the system understand *how* things are connected
Reconcile the semantic entities

Civil War Research Database
5,441 ‘battles’
Includes skirmishes, naval encounters
Individual details
July 21st - Bull Run

US National Park Service Database
382 ‘battles’
Land based only
Major developments
No individual details
July 21st – Manassas

What is a battle?
When does a battle begin?
Where is a battle actually fought?
Inference issues: if Johnson was in the 124th regiment and the 124th fought at Bull Run can we assume he fought at Bull Run?
Etc….
Entity matching isn’t easy…

Place (geocodes) and Date are
- Universal
- Authoritative
- Comprehensive
- Standardizable

Other vocabularies not so straightforward
- UMLS unifies over 150 medical subjects such as ICD-10, MeSH, SNOMED CT, DSM-IV and more
- Requires Continuous error correction, updating and improvements.
Add lots more dbs and content

How to build links between lots of relevant but different files?
Automated entity/property extraction?

- Machine aided indexing tags subjects, names, places
- Struggles with relationships
- Work has been going on since the 1990’s
- Some success, but most had to be done manually
November 6, Tuesday, 1860

Abraham Lincoln was elected sixteenth President of the United States, with Hannibal Hamlin of Maine as his Vice-President. The Republican ticket of Lincoln and Hamlin received 1,866,452 votes and 100 electoral votes in 7 of the 31 states. The Southern Democratic ticket of Stephen A. Douglas of Illinois and Hannibal Hamlin of Maine received 1,760,971 votes, but only 31 electoral votes. The Free-Soil ticket of Horace Greeley of New York and Gideon Welles of Connecticut received 490,364 votes from 5 of the 14 slave states. The Democratic ticket received 844,949 votes from 5 of the 14 slave states, for 73 electoral votes. Constitutional Unionists John Bell of Tennessee and Edward Everett of Massachusetts received 696,057 votes, for 59 electoral votes in Kentucky, Tennessee, and Virginia. Because of human cruelty and political differences, none of the candidates won an absolute majority. The death of Lincoln was a minority President in the popular vote with just over a third of the ballots, but...
Goal: Inbound and outbound links to all

Newspapers

Websites

Music

Primary Works

Monographs

Journals

90 databases
25,000 video titles
>1m tracks of music
>5m pages of text
1.2m digital objects
Steps to integration

All linkable content

All Alexander Street Content

The American Civil War Online

The American Civil War Letters and Diaries

Not to scale ;)
• Unified, multi-database, multi-discipline cross-search
• Preserving and opening controlled vocabularies
• Single Open API for third-party access
• Launched in July 2013
Inbound linking
Challenges to building semantic links

Philosophical issues
- Many don’t want to link to ‘for fee’ objects
- Charge for traffic generated or to generate traffic?
- Building links is expensive. Who pays?
- When things change who’s responsible?
- Vested interests

Technical
- Authentication
- Permissions
- Automated tools lacking
If the data isn’t there
Not that useful
Unhappy Students…
If the data is there…

3 search examples keyed into one box…

‘Churchill’

- **Narrow your search**
  - Content Type
    - Documentary (255)
    - Newsreel (178)
    - Interview (28)
  - More…

  - Topic / Theme
    - Depression & World War II (1929–1945) (138)
    - Post-war Era (1945–1960) (134)
    - 20th Century in World History (1914-2000) (82)
  - More…

  - Speaker / Narrator
    - Robert Powell (16)
    - Winston Churchill (8)
    - Stan Watt (7)
  - More…

- **Date Written / Recorded**
  - 150
  - more…

‘Mozart’

- **Narrow your search**
  - Time Period
    - Classical (623)
    - Romantic (26)
    - 20th century (23)
  - More…

  - Composer
    - Mozart, Wolfgang Amadeus (826)
    - Beethoven, Ludwig van (30)
    - Bach, Johann Sebastian (15)
  - More…

  - Composition
    - Don Giovanni, K. 527 (74)
    - Le Nozze di Figaro, K. 492 (68)
    - Die Zauberflöte, K. 620 (64)
  - More…

  - Soloist / Performer
    - Various Artists (81)

‘Child Therapy’

- **Narrow your search**
  - Therapist / Psychologist
    - Christine A. Nelson (11)
    - Allen E. Ivey (10)
    - Derald Wing Sue (8)
  - More…

  - Presenting Condition
    - Anxiety (21)
    - Anger (17)
    - Depression (emotion) (14)
  - More…

  - Therapeutic Approach
    - Role play (29)
    - Multicultural counseling (22)
    - Crisis intervention (16)
  - More…

  - Health Subject
    - Populations (133)
    - Diagnosis (119)
If the data’s there and the usage is there...

### Thing > Intangible > JobPosting

A listing that describes a job opening in a certain organization.

<table>
<thead>
<tr>
<th>Property</th>
<th>Expected Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Properties from Thing</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>additionalType</td>
<td>URL</td>
<td>An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax - the <code>typeOf</code> attribute - for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.</td>
</tr>
<tr>
<td>description</td>
<td>Text</td>
<td>A short description of the item.</td>
</tr>
<tr>
<td>image</td>
<td>URL</td>
<td>URL of an image of the item.</td>
</tr>
<tr>
<td>name</td>
<td>Text</td>
<td>The name of the item.</td>
</tr>
<tr>
<td>sameAs</td>
<td>URL</td>
<td>URL of a reference Web page that unambiguously indicates the item’s identity. E.g. the URL of the item’s Wikipedia page, Freebase page, or official website.</td>
</tr>
<tr>
<td>url</td>
<td>URL</td>
<td>URL of the item.</td>
</tr>
<tr>
<td><strong>Properties from JobPosting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseSalary</td>
<td>Number</td>
<td>The base salary of the job.</td>
</tr>
<tr>
<td>benefits</td>
<td>Text</td>
<td>Description of benefits associated with the job.</td>
</tr>
<tr>
<td>datePosted</td>
<td>Date</td>
<td>Publication date for the job posting.</td>
</tr>
<tr>
<td>educationRequirements</td>
<td>Text</td>
<td>Educational background needed for the position.</td>
</tr>
<tr>
<td>employmentType</td>
<td>Text</td>
<td>Type of employment (e.g. full-time, part-time, contract, temporary, seasonal, internship).</td>
</tr>
</tbody>
</table>
Outbound links
Experience so far
Oral History Online

• Higher value links
• Semantic indexing and keyword searching of more than 3,000 oral history collections
• Represents the personal histories of some 300,000 people
• Value:
  – Context
  – Selection
  – Search power
  – Licensed material
  – Integration
### Table of Contents: Geographical Terms (Interviews)

<table>
<thead>
<tr>
<th>Term</th>
<th>Place Discussed</th>
<th>Place of Birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>65</td>
<td>19</td>
</tr>
<tr>
<td>Asia</td>
<td>1448</td>
<td>206</td>
</tr>
<tr>
<td>Atlantic Ocean</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>1462</td>
<td>2020</td>
</tr>
<tr>
<td>North America</td>
<td>7149</td>
<td>2015</td>
</tr>
<tr>
<td>Bahamas</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Barbados</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Canada</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Cayman Islands</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table of Contents: Historical Events (Interviews)

<table>
<thead>
<tr>
<th>Term</th>
<th>All</th>
<th>Text</th>
<th>Audio</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestant Reformation, 1400-1660</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Sand Creek Massacre, CO, 1864</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armenian Massacre, 1915-1920</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russo-Japanese War, 1904-1905</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mexican Revolution, 1910-1920</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Titanic sinks, April 1912</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balkan Wars, 1912-1913</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>World War I, 1914-1918</td>
<td>277</td>
<td>230</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Armenian Massacre, 1915-1916</td>
<td>11</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easter Uprising, April 14, 1916</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian Revolution, 1917-1921</td>
<td>35</td>
<td>35</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Prohibition, 1919-1933</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harlem Renaissance, 1920-1939</td>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Great Depression, 1939-1941</td>
<td>505</td>
<td>445</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Spanish Civil War, 1936-1939</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assassination of John F. Kennedy, November 22, 1963</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sino-Japanese War, 1937-1945</td>
<td>93</td>
<td>85</td>
<td>70</td>
<td>45</td>
</tr>
<tr>
<td>Holocaust, 1939-1945</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kristallnacht, November 9-10, 1938</td>
<td>1032</td>
<td>901</td>
<td>157</td>
<td>52</td>
</tr>
<tr>
<td>World War II, 1939-1945</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Winter War, 1939-1940</td>
<td>1500</td>
<td>837</td>
<td>516</td>
<td>61</td>
</tr>
</tbody>
</table>
Search Interviews

This screen searches all interviews in the database by specific criteria.

<table>
<thead>
<tr>
<th>Keyword(s) in Interview:</th>
<th>(e.g., Brighton*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-Text</td>
<td>Audio</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Narrator:</th>
<th>(e.g., Thomas)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Age at Interview:</th>
<th>(e.g., 42-65)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Year of Birth:</th>
<th>(e.g., 1950-1960)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Place of Birth:</th>
<th>(e.g., Vietnam)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Gender:</th>
<th>ALL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Occupation:</th>
<th>(e.g., Teacher)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Race:</th>
<th>(e.g., Black)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>All Subjects:</th>
<th>(e.g., Labor)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Years Discussed:</th>
<th>(e.g., 1920)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Places Discussed:</th>
<th>(e.g., Kiev)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>People Discussed:</th>
<th>(e.g., Clinton, Bill)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Historical Events Discussed:</th>
<th>(e.g., World War II)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Organizations Discussed:</th>
<th>(e.g., IBM Corporation)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Interviewer:</th>
<th>(e.g., Rauff, Marc)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Date of Interview:</th>
<th>From Year</th>
<th>Month</th>
<th>ALL</th>
<th>To Year</th>
<th>Month</th>
<th>ALL</th>
<th>Include Undated</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Place of Interview:</th>
<th>(e.g., Bethesda, MD)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Previously Unpublished:</th>
<th>ALL</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Collection Name:</th>
<th>(e.g., Cold War)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Repository Name:</th>
<th>(e.g., Nunavut Arctic College)</th>
</tr>
</thead>
</table>
Context for results

Results for Text Search

Search results: Found 298 paragraphs.
Search string: word=castro

Sort By: Context | Line By Line | Narrator | Interview Date | Summary by Collection

298 results | page 1 | 2 | 3 | jump to page

1. Interview of Miller, John G. 26-Jun-2004 in Adams Center (VMI) Cold War History Project [Interview Details][Collection Details][Link to Text]

These are nations that have no desire at all to hook into a global network because if they did, that they would lose control of their people; these are controlled by dictators and despots who need absolute control of their people, like Castro, Saddam Hussein, Kim Jong Il and Korea. And they have to keep them backward and in the dark to maintain control. A good part of this non-Core nation is the radical Islamic movement, the Wahabis, in Saudi Arabia is a great example, who keep screaming, "Death to infidels" and keep trying to carry it out. Professor Barnett argues that the only way to bring the non-Core nations into the Core and into fruitful exchange with the rest of the world is to depose that leadership and look for some alternative that can start to hook them into the world.

2. Interview of Arceneaux, Grace Palacio, 1920- 01-Jun-1977 in Agricultural History Santa Cruz County [Interview Details][Collection Details][Link to Text]

After that he settled with Julia Castro, who was just the salt of the earth. She raised him for three years. She was very good to him. Leo... there's nothing to say bad about him. When he was in town he visited me often. He was there every Sunday, every Wednesday, some evenings. Knaster: Did he help support the baby? Arceneaux: Oh yes, he supported him. He paid people to take care of him. Knaster: When you got out of the hospital where did you go? Arceneaux: Back... Knaster: You came back to town and lived here with the baby? Arceneaux: Well no. They wanted me to have a home that had two bedrooms.

3. Interview of Arceneaux, Grace Palacio, 1920- 01-Jun-1977 in Agricultural History Santa Cruz County [Interview Details][Collection Details][Link to Text]

"You don't write, "White kills wife," or Englishman stabs..." things like that. It was an effort to stand up for themselves. I joined that and I felt pretty good about it. I didn't have anything to do with the Mexicans. I didn't know where to begin, because still I had the child that was half-Filipino. I didn't know where to go so I stayed and it was nice, very, very nice. They were gentle with me and very kind, and it was easy to raise Lenny with people like that that had watched him being raised in the community in many foster homes and then they knew that he ended up at Julio Castro's house, the lady that raised him the last three years.

4. Interview of Soto, Joe, 1937- 25-Jan-1994 in Barrios Oral History Project [Interview Details][Collection Details][Link to Text]
INTERVIEW WITH WALTER CRONKITE

(Preliminary talk not transcribed)

INTERVIEWER: Mr Cronkite, can I ask you first off: at the beginning of the Sixties, how do you think America perceived Russia, the Soviet Union?

WALTER CRONKITE: Well, at the beginning of the Sixties we were long past any thought of the Russians as being allies. The Soviet Union, as one of the big powers, it was a bad actor. It was a ally of China, it was a threat to everybody. It was seen as an aggressor, a threat to the free world. It was not seen as a potential ally.
Challenge - Linking behaviors

Rather than…

• Change links constantly
• Forbid others to link to you
• Refuse to license for crawlers
• Insist on one interface

Do the following…

• Have permanent links
• Invite others to link to you
• Exchange links in batch through protocols like OpenURIs/URLs
• Make citations visible to the outside world
• If necessary charge for links/crawlers/APIs
• Engineer for multiple interfaces
• Ask others (e.g. Google) to do the same!
Challenge - Library Perceptions

• ‘We want full-text content, not indexes’
• ‘Why should we pay for content that’s free?’
• ‘It’s not our job to index content we don’t own’
• ‘Should it be in the OPAC?’
• Not sure what to do about MARC records…

• Metrics and development funds prioritized towards
  1. Content (look what we own)
  2. Usage (look how much it’s being used)
  3. Utility (look how much is learned/discovered)
Business Issues

• Who pays for the links?
  • Connections don’t lend themselves to advertising
  • Utility is largely in the background
  • OCLC aside, discovery services don’t pay for metadata
• Permissions control is complex
  • Authentication
  • Determining ownership before providing access
• Competitive forces
  • Major players (e.g. Google) won’t allow data mining
  • Discovery services
<table>
<thead>
<tr>
<th>Feature</th>
<th>ASP</th>
<th>Discovery Service</th>
<th>Google</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality content</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Discipline relevance: comprehensive, covers rare, small sites</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Stability: automated re-directs and indicators of broken links</td>
<td>Y</td>
<td>Some</td>
<td>Some</td>
</tr>
<tr>
<td>Integration – playlist integration, easy citation, annotation.</td>
<td>Y</td>
<td>Some</td>
<td>Some</td>
</tr>
<tr>
<td>Usage statistics</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Broken link user alerting</td>
<td>Y</td>
<td>?</td>
<td>N</td>
</tr>
<tr>
<td>Deep integration of selected items – semantic indexing</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Adding discovery to Academic Video Online

Target Sites for Connectors

<table>
<thead>
<tr>
<th>Source</th>
<th># of Titles</th>
<th>What is it?</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-Span</td>
<td>400,000</td>
<td>News broadcasts since 1983.</td>
<td>Cross-searched</td>
</tr>
<tr>
<td>TED Talks</td>
<td>700</td>
<td>Unique lectures</td>
<td>Cross-searched</td>
</tr>
<tr>
<td>WGBH Open Vault</td>
<td>1,025</td>
<td>Interviews with famous people</td>
<td>Title by title</td>
</tr>
<tr>
<td>Annenberg Media</td>
<td>300</td>
<td>Scientific Videos</td>
<td>Title by title</td>
</tr>
<tr>
<td>Science Cinema</td>
<td>250</td>
<td>Experimental Videos</td>
<td>Title by title</td>
</tr>
<tr>
<td>Civil Rights Movement</td>
<td>125</td>
<td>Oral Histories</td>
<td>Title by title</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50 Sites, over 500k titles</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are publishers and librarians making progress quickly enough?
Some great sites...
Too many Mash-ups to mention…
Growing number of targets…

**Distribution of triples by domain**

**Distribution of links by domain**

http://lod-cloud.net/state
Freebase – entity definitions

A community-curated database of well-known people, places, and things

Explore Freebase Data

<table>
<thead>
<tr>
<th>Domain</th>
<th>ID</th>
<th>Topics</th>
<th>Facts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Music</td>
<td>/music</td>
<td>26M</td>
<td>173M</td>
</tr>
<tr>
<td>Media</td>
<td>/media_common</td>
<td>7M</td>
<td>21M</td>
</tr>
<tr>
<td>Books</td>
<td>/book</td>
<td>6M</td>
<td>32M</td>
</tr>
<tr>
<td>People</td>
<td>/people</td>
<td>3M</td>
<td>15M</td>
</tr>
<tr>
<td>TV</td>
<td>/tv</td>
<td>2M</td>
<td>15M</td>
</tr>
<tr>
<td>Film</td>
<td>/film</td>
<td>1M</td>
<td>15M</td>
</tr>
<tr>
<td>Location</td>
<td>/location</td>
<td>1M</td>
<td>17M</td>
</tr>
<tr>
<td>Organization</td>
<td>/organization</td>
<td>789K</td>
<td>4M</td>
</tr>
<tr>
<td>Fictional Universes</td>
<td>/fictional_universe</td>
<td>743K</td>
<td>1M</td>
</tr>
<tr>
<td>Business</td>
<td>/business</td>
<td>733K</td>
<td>3M</td>
</tr>
<tr>
<td>Sports</td>
<td>/sports</td>
<td>371K</td>
<td>3M</td>
</tr>
</tbody>
</table>

How can you get started?

Learn how it works
Discover what kind of information Freebase contains, how it’s organized, and how Freebase allows you to uniquely identify identities anywhere on the web.

Keep reading.

Use Freebase data
Freebase data is free to use under an open license. You can:
- Query Freebase using our Search, Topic, or MQL APIs
- Download our weekly data dumps

Join the Community
- Follow Freebase on G+
Knowledge Graph Search for Singapore…

Google: – 715 million hits

About 715,000,000 results (0.23 seconds)

Singapore - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Singapore
Singapore, officially the Republic of Singapore, is a Southeast Asian island city-state off the southern tip of the Malay Peninsula, 137...
Four Asian Tigers - World’s busiest port - City-state - Financial centre

Official Singapore Tourism Website - YourSingapore.com
www.yoursingapore.com/
Find tourist information and create your travel guide for events and attractions in Singapore. The official website by the Singapore Tourism Board.

Welcome To Singapore Airlines
www.singaporeair.com/
National flight carrier which includes online booking, sites and offices, schedules and route maps

gov.sg Singapore Government Information
www.gov.sg/
The official website of the Government of Singapore.

News for singapore
Singapore Cuts Export Forecast as China Slowdown Damps Recovery
Businessweek - 36 minutes ago
Singapore lowered its forecast for exports this year as a slowing expansion in China crimps demand for the nation’s goods, even as services...

Singapore’s medical risk takers
BBC News - 2 days ago
Singapore Stocks - Factors to watch on August 12
Reuters UK - 2 hours ago
Progress could be faster…

Semantic Web

Inference ‘knowledge’

Meaning based search

Automated Data interactions/Mash-ups

Discovery
• Ted Nelson outlined Hypertext in 1963.


• In 1999 Ted Nelson opined

“HTML is precisely what we were trying to PREVENT— ever-breaking links, links going outward only, quotes you can't follow to their origins, no version management, no rights management"
Where we’re headed

New ideas...

Nano-publication...

Assertion: NG_000007.3:g.70628G>A
Condition: Sardinian
Provenance: Giardine et al.

Frequency: 0.25%
Nano-publication...

1. Publishing
- Nano-Publication (doi: 10.1234/NP5678)
  - isoproterenol binds Alpha-2 adrenergic receptor
  - Provenance information

2. Knowledge mash-up
- External resources in RDF
  - Chebi, Uniprot, annotated literature, other Nano-publication, etc...

3. Query result or feed
- Hypothesis: Catecholamine binds Alpha-2 adrenergic receptor?
- Adrenaline binds Alpha-2 adrenergic receptor!
- Machine
  - Which chemicals bind Alpha-2 adrenergic receptor?
  - Is there any common chemical family binding this receptor?
Elsevier’s Executable Article Challenge

- Equations, tables and graphs interactive so reviewers, readers and machines can check, manipulate and explore the result space
- Long term technical compatibility
- Data and code validated easily and automatically
- Balances author rights, institutional rights, and open access
- Registration and tracking of actions for provenance
- Viruses and code contamination, plagiarism proof
- Open API to enable automated queries by outsiders
Utopia Suite – selected tools

Search Tool

3D Structure Tool

PRECIS results with citations

Sequences alignment editor