FAST: Subject Headings for the Semantic Web

Ed O’Neill

January 27, 2014
Midwinter SAC Program
ONeill@OCLC.org
Changing Environment

19th Century 21st Century

Card Based
Small Collections
Local Catalogs
Intermediaries

Computerized
Global Collections
Automated
Semantic Web
End User Searches
FAST’s Origins

The SAC Subcommittee on Metadata and Subject Analysis (1997-2001) recognized that a new schema is required for Internet resources and other non-traditional materials. The options identified:

- Develop a new schema,
- Use an existing schema(s),
- Adapt or modify an existing schema.
Requirements

- Compatible with LCSH,
- Simple in structure and syntax,
- Usable by non-catalogers and in non-library environments,
- Semantic interoperability,
- Compatible with MARC, Dublin Core, and other popular metadata schemas,
- Amenable to authority control,
- Linkable.
Joint Project

- FAST was developed jointly by the Library of Congress and OCLC. The development team included:
  - OCLC staff,
  - Lois Mai Chan (Univ. of Kentucky, Consultant)
  - Lynn El-Hoshy (Library of Congress)
- SAC established the FAST Subcommittee to provide assistance and guidance.
- OCLC and LC have unrestricted non-exclusive rights to FAST except neither can sell FAST.
The FAST Approach

- Retain LCSH’s rich vocabulary
- Eight facets
- Established all headings
Precoordination vs. Faceted

LCSH is precoordinated:
Cataloging—United States

FAST is faceted (postcoordinated):
Cataloging
United States
Enumerative vs. Synthetic Languages

**LCSH is Synthetic:** Only a set of core terms is established but those terms can be combined or extended following the synthetic rules.

**FAST is Enumerative:** All terms are established and included in the authority file.
Eight Facets

Topical
  Subject headings
    —Evaluation

Geographic
  Ohio—Dublin

Events
  Olympic Games

Chronological
  1939 - 1945

Personal Names
  Kilgour, Frederick G.

Corporate Names
  OCLC

Uniform Titles
  Dead Sea Scrolls

Form/Genre
  Guidebooks
Geographic Facet

- Geographic names will be established and applied in indirect order and abbreviations are not used:
  Pennsylvania—Philadelphia

- First level geographic names are limited to continents, countries, and other high level names (e.g., Ohio, Victoria, China, Europe, etc.). Other names are entered as subdivisions under the smallest first-level name in which it is fully contained:
  Europe—Curzon Line

- Parts of large bodies of water (bays, gulfs, etc.) are established under the larger body of water:
  Atlantic Ocean—Chesapeake Bay
Event Headings

- An event is any activity that occurs during a limited period on time,
- Both single and reoccurring events are included,
- Non-reoccurring events are generally qualified by place and date.

Great War (Uruguay: 1843-1852)
International Youth Year (1985)
Rose Bowl (Football game)
Chronological (Period) Facet

Headings consist of either a single date or a starting and ending date but are formatted for display:

1945
1942 – 1945
Since 1987
221 B.C. - 220 A.D.

Chronological headings are an exception; they are only established when needed for cross-references.
Uniform Title Facet

Uniform titles are entered directly and qualified by author:

Anonymous
  Beowulf
  Dead Sea scrolls

Personal Authors
  Paradise regained (Milton, John)
  Killers (Hemingway, Ernest)

Corporate Authors
  Job Training Partnership Act (United States)
  Basic law (Hong Kong)
Most LCSH Headings are **Synthetic**

5.8% of LCSH Headings in WorldCat are established. 24.8M are not established but synthetically built.

All statistics are from WorldCat as of 1/1/13.
LCSH Headings Stats

- 26,423,651 unique LCSH headings in WorldCat,
- 1,490,619 new LSCH headings were added to WorldCat in 2012,
- 1,586,961 of the unique LCSH headings are established,
- 59,895 of the LCSH headings were established in 2012.
Result of Faceting

26.5 million LCSH headings
\[\rightarrow\] 1.7 million FAST headings
Links: to & from

Bibliographic Record → Authority Record → Other Authorities / Sources
Embedding vs. Linking
Linking as MARC Fields

LCSH:

650 0 $aWorld Wide Web $x Subject access$0(DLC)sh 0003176

Source

Normalized Text as an Identifier
→ world wide web$subject access

FAST:

650 7 $aWorld Wide Web $x Subject access

$2fast $0(OCoLC) fst01181332

Source

ID
Linking; A Simplified Example

010 2010015675
050 00 Z695.Z8 $b F373 2010
100 1 Chan, Lois Mai.
245 10 FAST : $b Faceted Application of Subject Terminology : principles and applications / $c Lois Mai Chan and Edward T. O'Neill.
300 xvii, 354 p. : $b ill. ; $c 26 cm.
650 0 $0(DLC)sh 0003176 World Wide Web $x Subject access
700 1 O'Neill, Edward T.
LCSH Linking; 3 cases

Simple Link
5.8%*

- Advertising—Automobiles
  - sh 85001092
  - Advertising—Automobiles

No Link

- Love—Religious aspects—Bahai Faith
  - sh 85078522
  - Love—Religious aspects—Buddhism, [Christianity, etc.]

Multiple Links

- Burns and scalds—Patients
  - sh 85018164
  - Burns and scalds
  - Patients
  - sh 00006930
  - Patients
FAST Linking; only simple links

Bibliographic → Authorities

Advertising—Automobiles
fst00797529
Advertising—Automobiles

Love—Religious aspects—Bahai Faith
fst01002784
Love—Religious aspects—Bahai Faith

Burns and scalds—Patients
fst00841932
Burns and scalds—Patients
Identifiers: Strings vs. Numeric

Traditionally normalized versions of the headings have been used as the link between headings in bibliographic records and the authority file:

Academic libraries $x$Circulation analysis
→ academic libraries$circulation analysis
→ sh 85076590

O’Neill, Edward T.
→ oneill, edward t
→ n 87146135
Indentifying Jock Stein

On 2013-01-16 LC made the following changes to two name authority records:

n 78081636
   Stein, Jock → Stein, Jock (Cleric)
no2012157653
   Stein, Jock, Pulp fiction writer → Stein, Jock

stein, jock = n 78081636   no2012157653
Options for Simple Links

- Faceting
- Validation records (Create authority records for all valid headings)
- Hybrid (Link when possible, embedded otherwise)
Problem with Strings as Identifiers

<table>
<thead>
<tr>
<th>Name</th>
<th>Bibs on 1/1/2012</th>
<th>Bibs on 1/1/2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stein, Jock</td>
<td>30</td>
<td>28 (1 correct, 27 erroneous)</td>
</tr>
<tr>
<td>Stein, Jock, Pulp fiction writer</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>Stein, Jock (Cleric)</td>
<td>—</td>
<td>3</td>
</tr>
</tbody>
</table>

In 2012, 840 name authority records were changed such that the established form of a personal name was changed and the previous form reused for a different person.
FAST Identifiers

- FAST initially used normalized strings as links to the authority file.
- FAST recently has switched to using identifiers.
- The recommended form of a FAST heading in a MARC bibliographic record is to include the authority record number.

650 7 $aWorld Wide Web $x Subject access $2fast $0(OCoLC) fst01181332

Obsolete headings will be retained in the FAST authority file and FAST records numbers will not be reused for different entities.
Adding FAST to WorldCat

- Beginning in September 2013, OCLC is systematically adding FAST headings to WorldCat bibliographic records that already have LCSH assigned.
- Initially, records will be excluded from FAST enrichment until at least six months from date of entry have elapsed.
- The processes performing this FAST enrichment work in WorldCat will run as background processes and should not adversely impact system performance.
- It’s anticipated that the FAST enrichment of record batch process will take at least one year, and that FAST headings will be added to approximately 100 million records.
- WorldCat records containing member-assigned FAST headings will not be modified by the FAST enrichment process.
LCSH to FAST Conversion

LCSH

600 Lincoln, Abraham, $d 1809-1865
650 Political leadership $z United States $v Case studies
650 Genius $v Case studies
600 Lincoln, Abraham, $d 1809-1865 $x Friends and associates
650 Presidents $z United States $v Biography
651 United States $x Politics and government $y 1861-1865

FAST

600 Lincoln, Abraham, $d 1809-1865
648 1861 - 1865
650 Political leadership
650 Genius
650 Friendship
650 Presidents
650 Political science
651 United States
655 Case studies
655 Biography
Searching FAST in Connexion

ho:fast and ho:academic libraries

The relevant index labels:
**ho:** for the keyword index,
**ho=** for the phrase index,
**how=** for the whole phrase index.
The authority file is first searched to find the appropriate subject heading(s).

The bibliographic file is then searched to identify the resources with the subject heading(s).
mapFAST

- Is a prototype that uses an authority file as the source of subject headings rather than the headings from bibliographic records.
- Uses the FAST subject headings.
- Uses the WorldCat.org bibliographic records.
- Complements other access methods.
- Is freely available at fast.oclc.org/mapFAST.
mapFAST

- When looking for information about a particular place, it is often useful to check surrounding locations as well.
- FAST geographic subjects provide access to information with a geographic aspect based on their latitude and longitude.
- A mashup with Google Maps allows users to see surrounding locations that have been FAST subjects.
- The map interface allows for simple selection of a location, with links to enter it directly as a search into either WorldCat.org or Google Books.
Find FAST Geographic Subject Headings Near:

- Las Vegas
- Nevada - Las Vegas Valley
- Nevada - Las Vegas
- Nevada - Las Vegas Metropolitan Area
- Nevada - Las Vegas Region
- Nevada - Las Vegas Suburban Area
- Nevada - Las Vegas Ranch
- Nevada - Henderson
- Nevada - Blue Diamond
- Nevada - North Las Vegas
- Nevada - Spring Mountain Ranch State Park
- Nevada - Red Rock Canyon National Conservation Area
- Nevada - Clark County
- Nevada - Mountain Springs

Find FAST Geographic Subject Headings Near:

Radius: 100 km
Display: Land Feature

- Nevada -- Las Vegas Valley
- Nevada -- La Madre Mountain
- California -- Mesquite Lake
- United States -- Pahrump Valley
- Arizona -- Detrital Valley
- California -- Kingston Range
- Nevada -- Arrow Canyon Range
- Nevada -- Valley of Fire
- United States -- Mojave Desert
- Nevada -- Spring Mountains
- Nevada -- Coyote Spring Valley
- United States -- Castle Mountains
- Arizona -- White Hills (Mohave County : Mountains)
### Nevada--Valley of Fire

<table>
<thead>
<tr>
<th>INFORMATION ABOUT THE AUTHORITY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leader</strong></td>
</tr>
<tr>
<td><strong>001</strong></td>
</tr>
<tr>
<td><strong>003</strong></td>
</tr>
<tr>
<td><strong>005</strong></td>
</tr>
<tr>
<td><strong>008</strong></td>
</tr>
<tr>
<td><strong>016</strong></td>
</tr>
<tr>
<td><strong>034</strong></td>
</tr>
<tr>
<td><strong>040</strong></td>
</tr>
<tr>
<td><strong>043</strong></td>
</tr>
<tr>
<td><strong>151</strong></td>
</tr>
<tr>
<td><strong>451</strong></td>
</tr>
<tr>
<td><strong>550</strong></td>
</tr>
<tr>
<td><strong>551</strong></td>
</tr>
<tr>
<td><strong>670</strong></td>
</tr>
<tr>
<td><strong>688</strong></td>
</tr>
<tr>
<td><strong>688</strong></td>
</tr>
<tr>
<td><strong>751</strong></td>
</tr>
<tr>
<td><strong>751</strong></td>
</tr>
</tbody>
</table>

### DIFFERENT FORMATS AVAILABLE

**RECORD VIEWS**

- Generic Document
- MARC-21 (MARC-21 raw data)
- RDF/XML (RDF/XML raw data)
Advantages of Authorities

- Increases the number of access points;
  - Latitude and longitude,
  - Feature type,
  - Cross references,
  - Subject classification,
  - Links to additional information,
  - etc.

- Easier to maintain the authorities than to embed the information directly in bibliographic records.
FAST applications


The study examined:

- How FAST is being utilized,
- Why those using FAST selected it,
- What benefits FAST provides,
- What can be done to enhance the value of FAST.
Thank You!

Ed O’Neill
oneill@oclc.org