FAST

Faceted Application of Subject Terminology

Jeff Mixter
Research Support Specialist, OCLC Research

11-21-14
NOTSL Fall 2014 Meeting, co-sponsored by NORASIST
#JeffMixter
Background

- Started in 1999-2000 as an OCLC Research Project
- Edward T. O’Neill and Lois Mai Chan
- Stemmed from the CORC (Cooperative Online Resource Catalog) Project – simple way to describe things on the Web using Dublin Core
  - Problem: there was no easy to use/learn subject vocabulary
Today

• Still being worked on…
  – Ed recently retired from OCLC Research and the project is now headed by Diane Vizine-Goetz

• Most recent development is applying FAST headings to WorldCat
Mechanics

- FAST uses LC vocabularies as the backbone for generating faceted subject terms.
- Since FAST is derived from LC vocabularies, it is very complimentary to but also limited by the scope/coverage of the LC vocabularies.
- The headings are derived in a few different ways:
  - The MARC Authority records
  - Rules for creating complex LC headings
    - *Cleveland (Ohio)--History--Civil War, 1861-1865*
Mechanics

- FAST team enhances FAST authority records
  - Add human readable geographic coordinates
  - Usage statistics
  - RDA data – birthdate and deathdate
Creating FAST terms

• Derived from LC vocabularies
• OCLC gets copies of LC Authority files (MARC) which are then converted into FAST Authority files
• Conversion is based on a variety of parameters
  – MARC field tags
  – Subfield codes
  – Set rules where opinions differ
    • History (Form vs. Topic)
• Faceting has benefits and drawbacks
  – Ease of use
  – Loss of semantics
Creating FAST terms

MARC with LC Headings:

600 $aWashington, George, $d 1732-1799
650 $aTrenton, Battle of, Trenton, N.J., 1776 $v Juvenile literature.
651 $aDelaware River (N.Y.-Del. and N.J.).

FAST Converted Headings:

600 17 $a Washington, George, $d 1732-1799 $2 fast $0 (OCoLC)fst00178100
611 07 $a Trenton, Battle of (New Jersey : 1776) $2 fast $0 (OCoLC)fst01404297
648 7 $a 1776 $2 fast
651 7 $a New Jersey $z Trenton $2 fast $0 (OCoLC)fst01207908
651 7 $a United States $z Delaware River $2 fast $0 (OCoLC)fst01310316
655 7 $a Juvenile works $2 fast $0 (OCoLC)fst01411637
Facets

• FAST is divided into 8 facets:
  – Topical
  – Personal Names
  – Corporate Names
  – Event
  – Geographic
  – Uniform Title
  – Form
  – Period
Topical

• Derived from the 150 field in MARC
• Topics are separated from non-topical subfield headings
  – Geographic subfield heading is split into a separate Geographic heading
• Most popular facet by organizations that have adopted FAST
• Sailboats--Design and construction
Personal Names

• Derived from the 100 field in MARC
• Includes the largest number of terms
  – Must be used as a subject before it is created for use in FAST
• Enhanced with links to VIAF and Wikipedia.org
• Follow same general pattern as LCNAF names
• Washington, George, 1732-1799
Corporate Names

- Derived from 110 field in MARC
- Like Person headings, these are linked to VIAF and Wikipedia.org when possible
- *General Motors Corporation*
Event

- An event is any activity that occurs during a limited period of time
- LC focused on just meetings whereas FAST includes meetings and events
- 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)
Geographic

- 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

- Geographic headings are enhanced with geographic coordinates
  - Based on GeoNames
  - This data drives the mapFAST service
Uniform Title

- 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)
Form

- Derived from the 155 field in MARC and also 150 field when appropriate
- This facet is currently undergoing some refinement
Period (Chronological)

• Periods are unique in that they are not related to any LCSH Authority records
  – They are mined out of bibliographic records
• As a result the Period headings are not comprehensive in coverage.
Result of Faceting

26.5 million unique LCSH headings
→ 1.7 million FAST headings
Recent developments

• FAST is now being added to WorldCat records
  – FAST has been added to ~80 Million records
• Perspective:
  – ~300 Million records in WorldCat
  – ~100 Million have LCSH headings
    • Not all are authority headings
    • Not all are valid
• Evaluation of how Form headings are created
  – Transition from subfield ‘v’ to use of LC Genre/Form Terms vocabulary
FAST Adoption

• Survey conducted 2013 to see who was using FAST
  – 10 organizations were using it
  – This did not include Cornell!
• Organizations using FAST:
  – From six countries
  – Variety of types – publishers, national libraries, special libraries and other non-library organizations
FAST Usage

• Facets that are most frequently used:
  – Topical
  – Geographic

• Reasons for adopting FAST:
  – Ease of use
  – Simple syntax
  – Non-cataloger use
  – All headings are linked
  – Rich vocabulary

  Work-flow efficiency

  Linked Data

  Based on LCSH
Questions?
FAST Features and Services
Services

- [http://fast.oclc.org/](http://fast.oclc.org/)

**FAST Projects**

- **searchFAST**
  A full feature search interface to the FAST database.

- **original FAST search**
  The original full feature search interface to the FAST database.

- **mapFAST**
  mapFAST is a Google Maps mashup prototype designed to provide map based access to bibliographic records using FAST geographic and event authorities. Also available in Mobile versions.

- **FAST Converter**
  The FAST Converter is a Web interface for the conversion of LCSH headings to FAST headings.

- **FAST Linked Data**
  FAST as a Linked Data service to interact with the Semantic Web.

- **assignFAST**
  A Web service that automates the manual selection of FAST Subjects based on autosuggest technology.

**FAST is a project of OCLC Research.**

Learn more about the FAST project.
Project background, history, and bibliography.

The FAST Dataset is available for download.

The most recent update for FAST was 10/14/2014. The FAST database will be updated periodically. Our target is to update FAST at least twice each year.

For comments on FAST please contact the FAST Team.
SearchFAST
• Lookup service for FAST headings

Search Keyword, headings, LCCN, etc.

Limit by facet

Type in search
mapFAST

• Browse headings based on geographic location
FAST Converter

- Converts MARC encoded LC headings into FAST headings

**FAST Converter**

**Enter Library of Congress Subjects:**

- 600 $a Washington, George, $d 1732-1799
- 650 $a Trenton, Battle of, Trenton, N.J., 1776 $v Juvenile literature.
- 651 $a Delaware River (N.Y.-Del. and N.J.).

**FAST Result:**

- 600 17 $a Washington, George, $d 1732-1799 $2 fast
- 611 7 $a Trenton, Battle of (New Jersey : 1776) $2 fast
- 640 7 $a 1776 $2 fast
- 651 7 $a New Jersey $z Trenton $2 fast
- 651 7 $a United States $z Delaware River $2 fast
- 655 7 $a Juvenile works $2 fast

**File Upload:**

- Choose File: No file chosen

Browse for a Marc file with LCSH headings. By clicking Convert, a resulting file with FAST headings will be available for download to the right.

**Convert**

Delimeter selection: ₨ Dollar sign ⚫ Double Dagger ⚫ Alveolar Click
assignFAST

- Create FAST headings that can be copied and pasted into MARC
Questions?
FAST Linked Data
Introduction

• **FAST Linked Data** was first published **December of 2011**
  • Derived from MARC
• It was developed using **SKOS (Simple Knowledge Organization Schema)***
  • Similar to **Library of Congresses Linked Data project**
• 2014 FAST Linked Data was remodeled using **Schema.org**
• FAST headings link to their respective Library of Congress heading(s)
• FAST Geographic headings are linked to **GeoNames**
Old Environment

Z693.5  Chan, Lois Mai.
U6C48    Cataloging and classification: an
          introduction / Lois Mai Chan. -- 2nd ed.
          xxii, 519 p. : ill. ; 24 cm.

          Includes bibliographical references (p.
          493-502) and index.

Links: to & from

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

Embedding

Subject headings

Linking

sh 85129426
Linking:
Is Full Enumeration Required?

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

**Enumerative:** All subject headings are established and included in the authority file. (FAST)
Linking as MARC Fields

LCSH:

650 0 $aSubject headings $0(DLC)sh 85129426

↑  
Source

650 7 $aSubject headings $2fast $0(OCCLC) fst01136458

↑  
ID

FAST:
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 $a Subject headings $b (DLC) $d sh 85129426
700 1 O'Neil, Edward T.
LCSH Linking; 3 cases

**Simple Link 5.8%**

- Advertising—Automobiles
- Advertising—Automobiles

**No Link**

- Love—Religious aspects—Sikhism
- Love—Religious aspects—Buddhism, [Christianity, etc.]

**Multiple Links**

- Burns and scalds—Patients
- Burns and scalds
- Patients

*All Statistics as of 1/1/2013*
FAST Linking; only simple links

Advertising—Automobiles

Bibliographic

Love—Religious aspects—Bahai Faith

fst00797529

fst01002784

fst00841932

Advisers

Authorities

Burns and scalds—Patients
Options for Simple Links

- Faceting.
- Validation records (Create authority records for all valid headings)
- Hybrid (Link when possible, embedded otherwise)
FAST URIs in MARC Bibliographic Records

• MARC is currently the data standard
  • Should not prevent libraries from accommodating Linked Data URIs
• There is no way to actually imbed the FAST URIs into MARC
  • It is possible to add all of the needed information to generate a URI
  • Use of canonical identifiers
• The MARC $0
  • Works well with FAST but is sometimes problematic for LCHS
canonical ID in the $0 ≠ http://id.worldcat.org/fast/1204623$
5.8% of LCSH Headings are Established
24.8M are Unestablished

- Persons (600)
- Corporates (610)
- Conf. & Meetings (611)
- Titles (630)
- Topicals (650)
- Geographics (651)
LCSH Headings are Growing Rapidly

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

FAST vs. LCSH

No. of Established Headings

- Persons (600)
- Corporates (610)
- Conf. & Meetings (611)
- Titles (630)
- Topicals (650)
- Geographics (651)

FAST

LCSH
Canonical URIs

• 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
• Of the 30 works that had Stein, Jock  (n 78081636) as either a 100 or 700 entity, only 3 were changed to Stein, Jock (Cleric)
• LC practice prevents a 400 field from being added to n78081636 because it is now the valid 100 field heading in no2012157653
• It is now impossible to differentiate the two names
• This change pattern occurred 840 times in 2013 alone
Questions?
Thank You!

Jeff Mixter
mixterj@oclc.org
@JeffMixter