



World Library and Information Congress: 69th IFLA General Conference and Council

1-9 August 2003, Berlin

Code Number: 010-E
Meeting: 126. Classification and Indexing
Simultaneous Interpretation: -

FAST (FACETED APPLICATION OF SUBJECT TERMINOLOGY): A SIMPLIFIED LCSH-BASED VOCABULARY

Edward T. O'Neill

OCLC, Dublin
USA

and

Lois Mai Chan

University of Kentucky, Lexington
USA

ABSTRACT:

The Library of Congress Subject Headings (LCSH) schema is by far the most commonly used and widely accepted subject vocabulary for general application. It is the de facto universal controlled vocabulary and has been a model for developing new subject heading systems around the world. However, LCSH's complex syntax and rules for constructing headings restrict its application by requiring highly skilled personnel and limit the effectiveness of automated authority control.

Recent trends, driven to a large extent by the rapid growth of the Web, are forcing changes in bibliographic control systems to make them easier to use, understand, and apply, and subject headings are no exception. The purpose of adapting the LCSH in a faceted schema with a simplified syntax is to retain the very rich vocabulary of LCSH while making it easier to understand, control, apply, and use. The FAST schema

maintains upward compatibility with LCSH, and any valid set of LC subject headings can be converted to FAST headings.

FAST consists of eight distinct facets. Authority records have been created for all established headings except for the chronological facet. The initial version of the FAST authority file will contain approximately two million authority records.

INTRODUCTION

With the phenomenal growth of electronic resources and the emergence of numerous metadata schemes for their description, there is a need, particularly, for subject access methods that can handle a large volume of materials without incurring the same amount of effort and cost as in the treatment of traditional library materials. In 1998, OCLC, in search for a subject access system that optimizes the use of technology for Dublin Core metadata records, began exploring a new approach to subject vocabulary¹. In keeping with the premises of the Dublin Core, it was determined that a subject vocabulary suitable for the web environment should meet the following requirements:²

- It should be simple in structure (i.e., easy to assign and use) and easy to maintain;
- It should provide optimal access points;
- It should be flexible and interoperable across disciplines and in various knowledge discovery and access environments including the online public access catalog (OPAC).

In developing a subject schema to meet these requirements, two key decisions are required: (1) defining the semantics (the choice of vocabulary); and, (2) formulating the syntax (pre-coordination vs. post-coordination). Regarding the semantics, OCLC decided to retain the LCSH vocabulary. By adapting the LCSH vocabulary, the compatibility with LCSH is retained. As a subject vocabulary, LCSH offers several advantages:³

¹ FAST is being developed by a team based in the OCLC Office of Research with support from the Library of Congress. The members of the FAST development team are: Eric Childress, Rebecca Dean, Anya Dyer, Kerre Kammerer, Edward T. O'Neill, Diane Vizine-Goetz (OCLC, Dublin, OH, USA); Lois Mai Chan (University of Kentucky, Lexington, Kentucky, USA); Lynn El-Hoshy (Library of Congress, Washington D.C., USA).

² Chan, Lois Mai, Eric Childress, Rebecca Dean, Edward T. O'Neill, and Diane Vizine-Goetz. 2001. A Faceted Approach To Subject Data In The Dublin Core Metadata Record. *Journal of Internet Cataloging* 4(1/2): 35-47.

³ *Subject Data in the Metadata Record Recommendations and Rationale: A Report from the ALCTS/SAC/Subcommittee on Metadata and Subject Analysis.* 1999. <http://www.govst.edu/users/gddcasey/sac/MetadataReport.html>

- It is a rich vocabulary covering all subject areas;
- It is the largest general indexing vocabulary in the English language;
- There is synonym and homograph control;
- It contains rich links (cross references) among terms;
- It is a de facto universal controlled vocabulary and has been translated or adapted as a model for developing subject heading systems by many countries around the world;
- It is compatible with subject data in MARC records;
- With a common vocabulary, automated conversion of LCSH to the new schema is possible;
- The cost of maintaining the new schema is minimized since most of the changes to LCSH can be automatically incorporated into the new schema.

While the rich vocabulary and semantic relationships in LCSH provide subject access beyond the capabilities of keywords, its complex syntax often proves to be a stumbling block and runs counter to the basic premises of simplicity and semantic operability of the Dublin Core. OCLC decided to devise a simplified syntax to be used with the LCSH vocabulary. The resulting schema is a controlled vocabulary built on the terminology and relationships already established in LCSH but structured with a different syntax and applied with different policies and procedures that are more inclined towards post-coordination. By separating syntax from semantics, the application process can be simplified while retaining the richness of vocabulary in LCSH thus making the schema easier to use and maintain. Furthermore, with the simplified syntax and application rules, computer technology can be used to greater advantage in both the assignment and the maintenance of subject data as well as in subject authority control.

While LCSH has served libraries and their patrons well for over a century, its complexity greatly restricts its use beyond the traditional cataloging environment. It was designed for card catalogs and excelled in that environment. However, because space on a 3x5 card was limited and each printed subject heading required a new card, the number of headings per item that can be assigned was severely restricted. Since the card catalog was incompatible with post-coordination, the pre-coordinated headings were the only option available in that environment.

LCSH is not a true thesaurus in the sense that it is not a comprehensive list of all valid subject headings. Rather LCSH combines authorities, now five volumes in their printed form, with a four-volume manual of rules detailing the requirements for creating headings that are not established in the authority file and for the further subdivision of many of the established headings.

The rules for using free-floating subdivisions controlled by pattern headings illustrate some of these complexities. Under specified conditions, these free-floating subdivisions can be added to established headings. The scope of application is often limited to particular types (patterns) of headings. For example, **Burns and scalds—Patients—Family relationships** is a valid heading formed by adding two free-floating

subdivisions to the established main heading **Burns and scalds**. The subdivision **Patients** is one of several hundred subdivisions that can be used with headings for diseases and other medical conditions. Therefore it can be used as a subdivision under the main heading **Burns and scalds**. However, the addition of **Patients** changes the meaning of the heading from a medical condition to a class of persons. Now, since **Family relationships** is authorized under the pattern for classes of persons, it can also be added to complete the heading.

Other examples of some of the complexities are illustrated by a type of authority records known as ‘multiples.’ In LCSH a group of similar headings are sometimes established as a “multiple subdivision,” i.e., “a subdivision in the subject authority file that incorporates bracketed terms, generally followed by the word **etc.** used to suggest the creation of similar subdivisions under the heading in question.”⁴ In other words, multiples are headings that establish a pattern of use. For example, the multiple **Love—Religious aspects—Buddhism, [Christianity, etc.]**, indicates that the names of other religions can replace ‘Buddhism’ to form new headings. The ‘multiple’ heading that actually appears in the 1xx field of an authority record should never be used in its multiple form in a bibliographic record. All the possible headings that can be created using ‘multiples’ are not included in LCSH and there is no comprehensive list of religions.

A third area that illustrates the complexities is music. Some of the complexities include: determining the group for each solo instrument (e.i., wind instruments), the ordering of instruments within the individual group, and when a heading should or should not be qualified (e.i., Concertos). Overall, music headings account for the largest number of valid headings that were not established or listed in LCSH.

While the rich vocabulary and semantic relationships in LCSH provide subject access far beyond the capabilities of keywords, its complex syntax has inhibited its application beyond the traditional cataloging environment. Not only are the rules for pattern headings complex, their application requires extensive domain knowledge since there is no explicit coding that identifies which free-floating subdivisions are appropriate for particular headings.

Although FAST will retain headings authorized under these rules, they will be established in an authority file, effectively hiding the complexity of rules under which they were created. Thus, FAST is based on the existing vocabulary in *Library of Congress Subject Headings* (LCSH), but applied with a simpler syntax than that currently used by libraries according to Library of Congress application policies.

⁴ Library of Congress, Cataloging Policy and Support Office, *Subject Cataloging Manual: Subject Headings*, 5th ed., 2000 cumulation (Washington, D.C.: Library of Congress, 2000), H1090, p. 1.

SYNTAX

For the sake of simplicity and semantic interoperability, the post-coordinate approach was chosen for FAST, because it is more in line with the basic premises and characteristics of the automated retrieval systems. It is also in keeping with the primary intent of the Dublin Core to make it "usable by non-catalogers as well as by those with experience with formal resource description models."

With these considerations in mind, the FAST schema is:

- A controlled vocabulary with all headings established in the authority file, with the exception of headings containing numeric values only;
- Based on the LCSH vocabulary;
- Designed for an online environment;
- A post-coordinated faceted vocabulary;
- Usable by people with minimal training and experience,
- Compatible with automated authority control.

FACETS

FAST consists of eight distinct facets: Topical, Geographic (Place), Personal Name, Corporate Name, Form (Type, Genre), Chronological (Time, Period), Title, and Meeting Name. The personal and corporate name facets are limited to their use as subjects and do not include name-title entries. The title and meeting facets are not included in the initial phase of the development. It is recognized these facets appear as subject access points and are necessary, and they will be included in the next phase of the FAST development.

Literary warrant is the basis for determining which headings will be established. In theory, there is an infinite number of valid LCSH headings that can be created. For example, there were 175 distinct musical instruments identified. Therefore, by taking all the different combinations of three instruments, in theory almost a million unique headings for trios could be formed. For nonets, over a billion billion (10^{18}) combinations of nine instruments are possible. Obviously, most of combinations will never be needed—many would not even be logical. Therefore, it is neither necessary nor feasible to create headings for every possible combination. **The establishment of a particular FAST heading is determined by its usage in OCLC's WorldCat, which also includes all of the headings assigned by the Library of Congress. Headings that have never been assigned in WorldCat will not be established in FAST even though they may be valid.**

FAST continues the use of subdivisions and retains the hierarchical structure of LSCH. Its major difference from LCSH is that, in a particular FAST headings, subdivisions must belong to the same facet as the main heading. Topical headings can be subdivided by other topicals, geographic headings by other geographics, etc. That is, a particular main heading may not be subdivided by subdivisions from a different facet.

Topical

FAST topical headings consist of topical main headings with appropriate topical subdivisions, including those topical subdivisions found under name and geographic headings in LCSH. The FAST topical headings look very similar to the established form of LCSH topical headings, for example,

Education
Natural gas pipelines—Economic aspects
Photoconductivity—Measurement
Travel—Safety measures
Urbanization
Hospitals—Staff—Labor unions—Organizing

A difference between LCSH and FAST practice is that all free-floating topical subdivisions will be part of the established form of the headings and all multiple subdivisions will be expanded. Elements in the same facet are pre-combined to the extent allowed by Library of Congress application policies. However, only those that have actually been used will be established. For example, headings based on the following heading with multiple subdivisions are established in the Subject Authority File. For example:

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

For FAST, multiples are not used. Each combination of **Love—Religious aspects** and a religion that has been used in WorldCat will be individually established such as:

Love—Religious aspects—Buddhism
Love—Religious aspects—Christianity
Love—Religious aspects—Islam
Love—Religious aspects—Hinduism
etc.

However, headings will not be established for every known religion—only those combinations that have actually been used.

Geographics

Geographic names are established and used in indirect order. For example, **Germany—Berlin** is the form used rather than the direct order form, **Berlin (Germany)**. In LCSH, place names used as main headings are entered in direct order, but when they are used as subdivisions, they appear in indirect order. First level geographic names in FAST are far more limited than in LCSH and are restricted to names from the

Geographic Area Codes table.⁵ Linking the first level entries with the Geographic Area Codes also provides additional specificity and a hierarchical structure to the headings. In this way, the Geographic Area Codes can be used to limit a search.

Some geographic names appear significantly different in their direct and indirect forms. In LCSH, North Carolina as a first level entry or as a subdivision, is spelled out but, as a qualifier, it is abbreviated as N.C. To ensure all occurrences of a name is found, users frequently must search for both forms of the name. A comprehensive search for Washington, North Carolina, for example, requires searching for both **Washington (N.C.)** and **North Carolina—Washington**. In a simple search using only the city name, Washington, North Carolina will be lost in among the far more numerous material on Washington, D.C. On the other hand, in FAST, precise searches are simpler since only one form of the geographic name is required to retrieve all material.

Second level names are entered as subdivisions under the name of the smallest first level geographic area in which it is fully contained. This is done to reduce the number of first level names, to ensure that the first levels are generally recognizable, and to place the second level into a broader context. For example, the Curzon line (the proposed line of demarcation between Poland and Soviet Russia during the Russo-Polish War of 1919-20) would be established in FAST as **Europe—Curzon Line**. In LCSH, on the other hand, since the Curzon line crosses national boundaries, it is established without qualification simply as **Curzon Line**.

Geographic names are generally limited to two levels except for place names within cities. For example, the Dravlje section of Ljubljana is established as **Slovenia—Ljubljana—Dravlje**. Qualifiers are used both to specify the type of geographic name (County, Lake, Kingdom, Princely State, etc.) and, when the name is not unique, to identify the particular place. For the United States, County names are the most common means for identifying a particular place name when the name is not unique within the state. For example, there are two Beaver Islands in Michigan; the larger and better-known island is in Lake Michigan, but another Beaver Island also exists in Lake Superior. To uniquely specify the island in Lake Michigan, it would be qualified by the county as **Michigan—Beaver Island (Charlevoix County)**. When different types of geographic entities have the same name, the name is qualified to reflect the type of entity. For example, Otsego Lake is both a town and a lake in Michigan. To distinguish between them, the names are established as **Michigan—Otsego Lake** and **Michigan—Otsego Lake (Lake)**. Following the practice of the LCSH, populated places are the default and need not be qualified.

Some examples of FAST geographic headings and their corresponding Geographic Area Codes are:

⁵ Library of Congress, Network Development and MARC Standards Office. MARC Code List for Geographic Areas. Web Version. Last updated on March 11, 2003. <http://www.loc.gov/marc/geoareas/gacshome.html>

Great Lakes [nl]
Germany [e-gx]
Mars [zma]
Maryland—Worcester County [n-us-md]
Slovenia—Maribor [e-xv]
France—Loir River Valley [e-fr]
England—Chilton (Oxfordshire) [e-uk-en]
India—Limdi (Princely State) [a-ii]
Califorina—San Francisco—Chinatown [n-us-ca]

Personal and Corporate Names

Personal names, including family names, and corporate names are both derived from the NACO Name Authority File. The requirements for either a personal or a corporate name to be included in FAST is (1) the name is used as a subject in at least one WorldCat record and (2) the name is established in the Name Authority File and is valid for subject use. For personal names, in addition to the name, the numeration, titles, dates, and the fuller form of the name are also used to establish the name. For a corporate body, name, the name and all subordinate units are used to establish the FAST corporate name heading. These corporate names include those of jurisdictions as well as corporate bodies. Name-title entities such as **Smollett, Tobias George, 1721-1771. Expedition of Humphry Clinker** are excluded from both the personal and corporate names facets. These name-title headings will be included in the second phase of the FAST development.

Some examples of FAST personal name headings are:

Woodward, Bob
Dewey, Melvil, 1851-1931
Kennedy family
Edward II, King of England, 1284-1327
Bush, George W. (George Walker), 1946-

and FAST corporate name headings include:

OCLC
Bayerische Motoren Werke
United States. Coast Guard
Bodleian Library

Form

Form or genre data are treated as a distinct facet. Form headings for FAST were identified from a variety of sources. Many of the forms were identified by extracting form subdivisions from LCSH authority records and assigned headings from cataloging

records. The details of that effort was described in detail by O'Neill et. al⁶. Some examples of form headings include:

Case studies
Abstracts
Census
Rules
Dictionaries—Portuguese
Folklore
Bibliography—Union lists
Periodicals

Chronological

FAST chronological headings follows the practice recommended by the SAC/ALCTS Subcommittee,⁷ and the recommendations discussed at the Airlie Conference.⁸ FAST chronological headings reflect the actual time periods of coverage for the resources and are not limited to specific periods associated with particular events. Chronological headings will be expressed as either a single numeric date or as a date range. In cases where the date is expressed in LCSH as a century, such as **20th century**, the heading is converted to the date range: 1900-1999. Similarly, periods related to geological eras would be expressed as dates in addition to the name of the period. For example, the Jurassic period would be expressed both as **Jurassic** and **From 140 to 190 million years ago**.

The only general restriction on FAST chronological headings is that when a date range is used, the second date must be greater than the first. Therefore, there is no need to routinely create authority records for chronological headings. For example, no authority record is need for chronological period **1900-1999** corresponding to the 20th century. Authority records for chronological headings will only be created when necessary for cross references.

⁶ O'Neill, Edward T., Lois Mai Chan, Eric Childress, Rebecca Dean, Lynn El-Hoshy, Kerre Kammerer, and Diane Vizine-Goetz. "Form Subdivisions: Their Identification and Use in LCSH". 2001. *Library Resources & Technical Services* 45, No. 4: 187-197.

⁷ *Subject Data in the Metadata Record Recommendations and Rationale: A Report from the ALCTS/SAC/Subcommittee on Metadata and Subject Analysis*. 1999. <http://www.govst.edu/users/gddcasey/sac/MetadataReport.html>

⁸ *The Future of Subdivisions in the Library of Congress Subject Headings System: Report from the Subject Subdivisions Conference Sponsored by the Library of Congress, May 9-12, 1991*, edited by Martha O'Hara Conway (Washington, DC: Cataloging Distribution Service, Library of Congress, 1992).

VALIDATING FAST HEADINGS

FAST headings are established by faceting established LCSH headings and/or headings extracted from MARC records in WorldCat. For example, faceting the following LCSH heading,

Architecture, Modern \$y 20th century \$z United States \$v Bibliography⁹

results in the following FAST headings:

Topical: **Architecture, Modern**
Geographic: **United States**
Chronological: **1900-1999**
Form: **Bibliography**

Where a heading such as **Heart \$x Diseases \$x Diet therapy \$v Recipes** has multiple subdivisions within the same facet, the hierarchy is retained so that the resulting FAST headings are:

Topical: **Heart \$x Diseases \$x Diet therapy**
Form: **Recipes**

In the validation process, a file containing all unique LCSH topical and geographic subject headings extracted from OCLC's WorldCat was first created. This file contained 6,912,980 unique topical and 1,471,023 geographic headings, representing over 50 million individual subject heading assignments in MARC records. These headings were then faceted to create the initial versions of the FAST topical, geographic, chronological, and form facets. Additional form headings were identified in LCSH authority records. A variety of algorithms for validating the headings automatically have been developed. This initial set of headings underwent extensive validation to minimize the number of erroneous entries. The entries remaining after this validation step were then established as FAST headings.

AUTHORITIES

The final step in developing FAST was creating an authority record for each established heading. Because of its wide acceptance, the MARC 21 format for authority data¹⁰ was selected. That format is very comprehensive and meets most, if not all of the FAST requirements. However, neither the authorities nor the bibliographic formats provided

⁹ In MARC21 formats, \$v=form subdivision, \$x=topical subdivisions, \$y=chronological subdivision, and \$z=geographic subdivision.

¹⁰ *MARC 21 Format for Authority Data: Including Guidelines for Content Designation*, prepared by Network Development and MARC Standards Office (Washington: Cataloging Distribution Service, Library of Congress, 1999).

for chronological headings—only for chronological subdivisions. A proposal¹¹ was submitted to MARBI (Machine-Readable Bibliographic Information), the interdivisional committee of the American Library Association charged with maintaining standards for the representation in machine-readable form of bibliographic information. That proposal was accepted with some minor enhancements in June of 2002. As a result, additional fields have been added to the MARC21 formats to accommodate unique requirements of FAST. Examples of FAST authority records are shown in figures 1 and 2.

CONCLUSIONS

In the networked, electronic environment, bibliographic and subject access tools must meet the following criteria:

- Efficiency and capacity for handling large quantities of resources
- Scalability and extensibility
- Interoperability

The advantages of FAST, based on faceted LC subject headings can be summarized in the following terms:

- It will facilitate computer-assisted authority control
- It will be easier and more economical to maintain than a highly enumerated vocabulary
- It will be more amenable to computer-assisted indexing
- It will enable a tiered approach to allow different levels of subject representation
- It will be able to accommodate both precoordinate and postcoordinate indexing and retrieval
- It will be able to accommodate different retrieval models
- It will facilitate mapping of subject data and cross-domain searching

¹¹ Changes for Faceted Application of Subject Terminology (FAST) Subject Headings, Proposal 202-13, May 8, 2002. <http://lcweb.loc.gov/marc/marbi/2002/2002-13.html>. (accessed March 10, 2003).

```

LDR      nz n
001      fast 611370
003      OCoLC
005      20021209141434.0
008      021209nneanz||babn n ana d
040      OCoLC $b eng $c OCoLC $f fast
050      RC684.D5
150      Heart $x Diseases $x Diet therapy
550      Heart $x Diseases $x Nutritional aspects
550      Heart $x Diseases $x Treatment
688      LC usage 64 (1999)
688      OCLC usage 394 (1999)
750      0 Heart $x Diseases $x Diet therapy $0 (DLC)sh
85059656

```

Figure 1.

MARC 21 FAST Topical Authority Record for **Heart—Diseases—Diet therapy**

```

LDR      00661nz      2200181n      4500
001      fast 522597
003      OCoLC
005      20030321133146.0
008      030321nneanz||babn          n ana      d
040      $a OCoLC $b eng $c OCoLC $f fast
043      $a n-us-ga
151      $a Georgia $z Saint Simons Island
451      $a Georgia $z Saint Simons
451      $a Georgia $z St. Simons Island
451      $a Georgia $z Saint Simon Island
451      $a Georgia $z Saint Simons Village
670      $a GNIS, Feb. 12, 2002 $b (Saint Simons Island,
PPL, 31° 09' 01" N, 81° 22'11" W, Glynn County,
variants: Saint Simon Island, Saint Simons, Saint
Simons Village)
751      0 $a Saint Simons Island (Ga.) $0 (DLC)n      82023244

```

Figure 2.

MARC 21 FAST Geographic Authority Record for **Georgia—Saint Simon Island**