

TDG 5: What Can the Library Community Offer in Support of Semantic Interoperability?

Facilitator: Mary Charles Lasater, Authorities Coordinator, Vanderbilt University

Recorder: Lynn El-Hoshy, Library of Congress

Members: Christine Borgman, Lois Mai Chan, Jane Greenberg, John Kunze, Thomas Mann, Joan Mitchell, William Moen, Arlene Taylor, Diane Vizine-Goetz, Bella Hass Weinberg

The Assignment

Interoperability is a key issue in the identification and retrieval of networked resources, as several of the Conference papers and the National Academy of Sciences report *LC 21* point out. The Topical Discussion Group's (TDG) assignment was to recommend approaches that the library community from its unique perspective could offer towards advancing semantic agreement among multiple domain-specific metadata schemes. Such approaches might exploit library tools, resources, and practices. They might also iterate the use of mechanisms to further knowledge representation or to extend entity relationship modeling.

Report to the Conference

Mary Charles Lasater explained that the TDG had decided to concentrate on recommendations that would advance semantic *cohesion*, rather than agreement. The TDG identified the following *assumptions* as the framework for its recommendations: Information systems must interoperate in an environment of distributed collections and access; and Library-based information systems must achieve mutual semantic interoperability with other information systems. The TDG identified two *goals* for the library community as it strives to advance semantic cohesion:

Interoperability between library systems

Interoperability between libraries and other communities as well as the Open Internet

The TDG identified five salient features of the *environment* for bibliographic control of resources:

Multiple metadata schemes

Multiple controlled vocabularies

Multiple domains (formats, for example archives, art works, etc.)

Multiple subjects (i.e. subject areas)

Multiple languages

Recommendations

- 5.1. Encourage coordination of metadata schemes, e.g. by establishing or monitoring *registries* of metadata schemes and by promoting the *consistent labeling of fields* (for example, if “creator”

and “author” are the same, encourage the use of the same term).

- 5.2. Study impacts and options to develop best practices and guidelines for machine indexing to improve cross-system searching, including guidelines for *definitions of words, character sets, diacritics, punctuation, etc.*
- 5.3. Encourage use of controlled access points in resource description. Specific steps include:
 - 5.3a. Encourage use of established schemes at the general level but recognize that more specific schemes may be necessary for detailed indexing.
 - 5.3b. Make *Library of Congress Subject Headings* (LCSH) more viable in the Internet environment while retaining present OPAC functionality; in particular, develop a simplified user-friendly view or overlay that is easy to learn, use and apply.
 - 5.3c. Make Library of Congress Classification and LCSH freely available on the Internet.
 - 5.3d. Create a CIP program for research quality Websites, producing records in which catalogers apply name authority control and subject analysis.
 - 5.3e. Encourage the linking and international sharing of authority records, including names, subjects, etc. Work with the vendor community to extend NACO and optimize vendor information in authority records, as well as to expedite authority record creation.
- 5.4. Encourage the library community to identify and participate in key projects to study general knowledge organization tools for cross-mapping and browsing.

Post Conference Comments from Participants

Suggested additions (from Bella Hass Weinberg) to Recommendation 5.3 above:

- 5.3f. The source vocabulary or classification and its version should be identified in the subject fields of metadata.
- 5.3g. Encourage the development of standards (such as XML and RDF) that code semantic relationships and facilitate linking from metadata to the source vocabulary.
- 5.3h. In recognition of the fact that automatic term switching often yields suboptimal results, design interfaces to convey vocabulary mapping to users, and allow them to select the terms of interest.

From Tom Mann: “A bit of ‘legislative history’ for ... [the] recommendations may be worthwhile. One proposal before the Group was ‘that future development of LCSH be brought more in line with principles of modern indexing languages and closer adherence to national and international standards for thesaurus construction.’ This proposal was rejected by the Group, largely because other

thesauri are not dependent on precoordination to the great extent that LCSH is, nor do others need to serve as indexes to classification schemes. That is why the Group recommended ‘retaining present OPAC functionality’ of LCSH. [I distributed] a background paper ... [see below] before the meeting to the members of the Group ..., specifying just what that OPAC functionality entails:

that LCSH meanings and extensive formal links to the Library of Congress Classification scheme (LCC) are dependent on the word order of precoordinated phrases;

that a vast network of existing cross-references in LCSH is heavily dependent on precoordination;

that browse displays of precoordinate phrases in the LCSH thesaurus are needed for recognition of further links not defined by cross-references; and

that browse displays of precoordinated subdivisions in the OPAC are needed for *recognition* of search options within a topic that could never be specified in advance through postcoordinate Boolean combinations.

“Examples of all of these points are included in the ... [background paper], which in turn is a digest of my longer paper, ‘Is Precoordination Unnecessary in LCSH?...',’ available on the Conference’s Web site, <<http://lcweb.loc.gov/catdir/bibcontrol>>.

“The Group did recommend that LCSH be made ‘more viable in the Internet environment’--but the specific wording of the recommendation is important. The Group did *not* recommend that Internet viability be increased through the expedient of lessening or diminishing precoordination--and it had the option to suggest this, but did not. On the contrary, a quite different expedient *was* recommended, that of developing ‘a simplified user-friendly view or overlay that is easy to learn, use and apply’ (Recommendation 5.3b). In other words, a simplified *front-end system* or interface *leading to* LCSH is called for, rather than fundamental changes *in LCSH itself* (i.e., particularly modifications or ‘deconstructions’ of its precoordinate nature).”

Background Paper from Mann

LCSH Meanings and Links to LCC Dependent on Word Order

Indian women is not the same as Indian AND Women

Indian women–Mexico is linked to F1219.3.W6

Indian Women–North America is linked to E98.W8

Indian Women–South America is linked to F2230.1.W6

Jewish women (linked to HQ1172) is not the same as Jewish AND Women

Women alcoholics (linked to HV5137) is not the same as Women AND Alcoholics

Women clergy (linked to BV676) is not the same as Women AND Clergy

Motion pictures for women (linked to PN1995.9.W6) is not the same as Motion pictures AND Women

Photography of women (linked to TR681.W6) is not the same as Photography AND Women

Sexual ethics for women is not the same as Sexual ethics AND Women

Social work with women is not the same as Social Work AND Women

Violence in women is not the same as Violence AND Women

Women, Black, in art is not the same as Women AND Black AND Art

Women in advertising is not the same as Women AND Advertising

Women in art (linked to N7629-N7639) is not the same as Women AND Art

Women in communication (linked to P96.W6) is not the same as either **Women–Communication** or Women AND Communication

Women in development (linked to HQ1240) is not the same as Women AND Development

Women in the Bible (linked to BS57.5) is not the same as Women AND Bible

Women in Church work is linked to BV4415

Church work with women is linked to BV 4445

Church work with women–Catholic Church is linked to BX2347.W6

Cross-references Dependent on Precoordination

The precoordinated phrase **Women-Psychology** (which is explicitly tied to HQ1206-HQ1216 in LCC) is linked by cross-references to many other precoordinated terms:

- RT Women–Mental health
- NT Achievement motivation in women
 - Animus (Psychology)
 - Anxiety in women
 - Assertiveness in women
 - Body image in women
 - Cooperativeness in women
 - Helplessness (Psychology) in women
 - Leadership in women
 - Self-esteem in women
 - Self-perception in women

This entire network of relationships would be lost if users could search only Women AND Psychology. Researchers could find only isolated information, not a web of knowledge relationships.

The precoordinated phrase **Afro-Americans–Education** (which is explicitly tied to LC2701-LC2853 in LCC) is linked by cross-references to numerous other precoordinated terms:

- BT Education–United States
- RT School integration–United States
- NT Afro-American students
 - Afro-American women–Education
 - Afro-Americans–Professional education
 - Afro-Americans–Scholarships, fellowships, etc.
 - Afro-Americans–Vocational education
 - English language–Study and teaching–Afro-American students
 - Segregation in education–United States
 - Segregation in higher education–United States

Many of these links would simply vanish without precoordination; they would not be retrieved by combining Afro-Americans AND Education, and those that are retrieved would be effectively buried. Readers would not be able to simply recognize the more specific research options that use different terminology.

Browse Displays in the LCSH Thesaurus Needed for Recognition of Links Not Defined by Cross-references

Whole columns of headings related to **Monasteries**—which will lead researchers in many directions—are not linked to each other by cross-references; but they are linked nonetheless by alphabetical proximity in the *browse displays* of the LCSH thesaurus. A very brief listing of only some of these contiguous related headings includes the following:

Monasteries

(linked to BX2460-BX2749 Catholic Church and NA4850 Architecture)

Monasteries, Armenian

Monasteries, Buddhist

Monasteries, Hindu

(linked to BL1243.72-BL1243.78)

Monasteries, Jaina

(linked to BL1378)

Monasteries, Syrian Orthodox

Monasteries and state

Monasteries in art

Monastery gardens

Monastic and religious life

(linked to BX2435)

BT Spiritual life—Christianity

RT Vows

SA *subdivision* Spiritual life *under names of individual religious orders*

NT Celibacy—Christianity

Eremitic life

Evangelical counsels

Retreats for members of religious orders

Spiritual direction

Superiors, religious

—History—Early Church, ca. 30-600

(linked to BX2465)

Monastic and religious life (Buddhism)

Monastic and religious life (Hinduism)

(linked to BL12266.85)

Monastic and religious life (Zen Buddhism)

Monastic and religious life in art

Monastic and religious life in literature

Monastic and religious life of women

(linked to BX4210-BX4216)

—Psychology

(linked to BV4205)

Monastic guest houses
USE Monasteries–Guest accommodations
Monastic libraries
(linked to Z675.M7)
Monastic profession
USE Profession (in religious orders, congregations, etc.)
Monasticism and religious orders
(BX385 Greek church)
(BX580-BX583 Russian church)
(BX2410-BX4560 Catholic church)

All of these displayed relationships and linkages—and scores more not listed here—would be lost without both precoordination and a browse display *of* the precoordinated terms in the LCSH thesaurus.

Browse Displays in the OPAC Needed for Recognition of Search Options Not Perceptible in LCSH Thesaurus

While browse displays of precoordinated strings in the OPAC itself make cross-references harder to see, they also include displays of free-floating subdivisions that cannot be seen in the LCSH thesaurus:

Yugoslavia–Antiquities
Yugoslavia–Boundaries [free-floater]
Yugoslavia–Civilization [free-floater]
Yugoslavia–Description and travel
Yugoslavia–Economic conditions
Yugoslavia–Ethnic relations [free-floater]
Yugoslavia–Foreign relations
Yugoslavia–Intellectual life
Yugoslavia–Politics and government
Yugoslavia–Rural conditions [free floater]
Yugoslavia–Social life and customs [free floater]

Such displays enable researchers to recognize whole arrays of options within their topics that they could never otherwise specify in advance. All of these vanish from sight in the postcoordinate combination Yugoslavia AND History, and yet all could be of interest to an historian of Yugoslavia.

The advantages of OPACs over card catalogs lie *not only* in keyword and Boolean search capabilities, but in creating browse displays of subdivisions that we could never before see in convenient arrays. With cards, we couldn't get good *overviews* of the *extent* of subdivisions.

Keyword and Boolean search capabilities create just as many problems as they solve (cf. p. 1); the computerized browse displays of subject strings, however, solve many of just these problems. This fact seems generally to be overlooked in most literature on OPACs.

Abandon Precoordination and Browse Displays?

- It is not logical to abandon precoordination when the *very meaning* of so many LCSH terms is dependent on the word-order of their phrasing, in ways that cannot be recaptured by either postcoordinate Boolean combinations or by word-proximity searches that drop out relational prepositions as stopwords.
- It is not logical to abandon precoordination when to do so would uproot tens of thousands of LCSH strings from a *vast web of specific linkages to LCC*—i.e., changes in the word order of the subject strings also changes the classification areas to which they point.
- It is not logical to abandon precoordination when browse displays of subject-string phrases *enable researchers simply to recognize* whole ranges of options that they could never specify in advance through postcoordinate combinations (e.g., **Yugoslavia–Antiquities** rather than just **Yugoslavia AND History**; **Afro-American whalers** rather than just **Afro-Americans AND History**; **Greek language–Onomatopoeic words** rather than just **Animal sounds**). The larger the file, the more researchers are dependent on *recognition* of options that they cannot articulate beforehand.
- It is not logical to abandon precoordination when the existence of the *vast cross-reference structure* between and among headings is so heavily dependent on the retention of ordered strings (e.g, **Women-Psychology NT Leadership in women**; **Afro-Americans–Education NT Segregation in higher education–United States**).
- It is not logical to abandon precoordination and browse displays when the *relationships of alphabetically-adjacent headings* within the thesaurus would be entirely lost without them (e.g., **Monasteries** is linked to scores of precoordinated neighbor headings such as **Monasteries and state** and **Monastic and religious life of women** simply by their displayed contiguity rather than by any formal cross-references).
- It is not logical to abandon precoordination when LCSH, unlike any other thesaurus, must simultaneously cover *all subject areas*—not just one, as other thesauri do—and show *relationships among them* that readers could not specify in advance.

A Possible Way to Integrate Web Site References into OPACs

Women–Services for

Women–Services for–Bolivia–Directories

Women–Services for–Caribbean area–Case studies

Women–Services for–Ethiopia–Congresses

Women–Services for–Germany–History

Women–Services for–Michigan–Evaluation

Women–Services for–New Zealand–Bibliography

Women–Services for–North Carolina–Finance

Women–Services for–Study and teaching–United States

Women–Services for–Study and teaching–United States–Web sites (.edu)

Women–Services for–United States–Directories

Women–Services for–United States–Web sites (.com)

Women–Services for–United States–Web sites (.edu)

Women–Services for–United States–Web sites (.edu)–Data archives

[This “Data archives” subdivision may not be appropriate for this particular subject; It is offered here just as a pattern example.]

Women–Services for–United States–Web sites (.edu)–Discussion lists

Women–Services for–United States–Web sites (.edu)–Portals

[“–Portals” is used here; “–Site directories” might be an alternative, in which case a cross-reference is needed: Site directories USE Portals]

Women–Services for–United States–Web sites (.gov)

Women–Services for–United States–Web sites (.org)

Women–Services for–Wisconsin–Periodicals

Women–Services for–Zambia–Directories

12/29/00