

Oksana Zavalina. User Searches in IMLS DCC Collection Registry: Transaction Log Analysis. Technical Report UIUCLIS--2006/3+IMLS, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, Champaign, IL, 2006.

Introduction

Subject access to collections has been in the focus of attention of LIS field for decades. A number of catalog use studies have been conducted in attempts to better understand its role and the problems user faces while searching for the information on a particular topic, with transaction log analysis being one of the methods widely employed by these studies. However, issues of subject access in federated collections, where the “unit of analysis” is a collection rather than an item search, have not yet been investigated. This paper reports an attempt of such an analysis performed on the IMLS Digital Collection Registry transaction log dataset.

IMLS Digital Collections and Content project is being implemented at the University of Illinois at Urbana-Champaign since January 2003. Within the project framework, a registry of all National Leadership Grant collections with digital content has been created. The IMLS Collection Registry includes collection level descriptions¹ and links to homepages of over 170 digital collections, created by libraries, museums, historical societies, botanical gardens and other cultural heritage institutions with support of the National Leadership Grant administered by the Institute of Museum and Library Services since 1998. The IMLS Digital Collections Registry is indexed with the GEM (Gateway to Educational Materials) subject headings, which provide broad categories for browsing considered suitable for the educational and cultural heritage communities.

GEM project, started in September 1996, is an initiative of the National Library of Education to expand educators’ access to Internet-based lesson plans, curriculum units and other educational materials (Sutton, 1999). GEM Element Set is an extension of the Dublin Core Element Set, with eight elements added to the initial 15-element DC package. In 1997, GEM subject scheme was created as the Subject Element GEM Controlled Vocabulary to describe digital objects in Gateway for Educational Materials repository. In part due to high national and international

¹ See for example <http://imlsdcc.grainger.uiuc.edu/collections/FullDisplay.asp?cid=2404>

reputation that GEM project in general has gained since its inception, today GEM subject scheme's application goes beyond its original domain: the Gateway to Educational Materials database. GEM subject scheme is now one of the many² controlled vocabularies and subject hierarchies being used to provide subject access to online resources and digital libraries such as Everglade, Internet Scout Portal, Federal Resources for Educational Excellence (FREE), RefWorks, National Science Digital Library etc. Being a domain-specific controlled vocabulary aimed at educators, GEM subject headings are considered suitable for browsing databases in both educational and more general humanities domains.

GEM subject scheme (see Attachment 1) consists of 12 "level 1" broad subject headings: Arts, Educational Psychology, Foreign Languages, Health, Language Arts, Mathematics, Philosophy, Physical Education, Religion, Science, Social Studies, Vocational Education, each of which has from 12 to 29 narrower "level 2" headings under it. The second level subject headings for Philosophy and Religion replicate ERIC Thesaurus "Narrower Terms" for these two broad subjects. Several of the level 2 GEM subject headings – Careers, History, Informal education, Instructional issues, Process skills, and Technology – are facets applicable to each of the twelve broad subject headings.

According to Stuart Sutton (2004), the major deficiency of the digital library architecture, including GEM, is the absence of the standardization in name authority; neither name nor place subject are represented in GEM subject scheme.

Collection administrators participating in the IMLS Collection Repository project are required to provide top-level GEM subjects in collection descriptions for the registry. Use of other subject headings is not required but supported by the metadata schema. As recently collected survey and interview data show, collection administrators are not completely satisfied with GEM subject scheme use for collection level description. Most of them point at the significant drawback of GEM subject scheme – lack of breadth and depth in topic coverage, especially at the top level of the subject hierarchy.

Research Question

The major research question in this study was: How suitable is the GEM (Gateway for Educational Materials) subject scheme adopted by IMLS Collection registry for describing diverse collections in the Registry? If it does not provide appropriate subject representation,

² GEM subject scheme is one of the 129 thesauri listed by the Library of Congress list of codes for subjects <http://www.loc.gov/marc/sourcecode/subject>

would another controlled vocabulary do a better job for this particular registry?

Based on the literature for evaluating subject schemes (Cochrane 1986, Larson 1991A etc.) and my own observations I have formulated eight general criteria for measuring GEM subject scheme suitability to collection level description in IMLS registry:

1. diversity of topics covered by GEM subject headings (breadth and depth of subject coverage),
2. syndetic structure of the GEM subject scheme,
3. heading structure of GEM subject headings,
4. currency of GEM subject headings,
5. availability of links between GEM subject headings and subject terms from other controlled vocabularies
6. degree of overlap between GEM subject terms and other subject terms used in the collection level description,
7. degree of overlap between the collection level description GEM subject terms and subject headings used in item level description.
8. semantic match between the GEM subject terms and keywords used by searchers of the registry.

A preliminary analysis of a sample of 23 digital collections based on the first seven criteria demonstrated overall inability of GEM subject scheme to adequately represent breadth and depth of subjects of the diverse collections in IMLS Collection registry. For this project, I chose to focus on the last and very important criterion – the semantic match between keywords applied by users in their searches and the GEM subject headings used in collection description records. Because no research has been done yet with the focus on specifics of search types and approaches in federated collections at collection level, another area of my interest in this project is general description of the searches made by users in IMLS Digital collection registry: the weight of subject versus known-item searches, typical query profile in terms of the number of words, frequency of each query use etc. I am also interested in correlation – if any – between the type of search and the semantic match of search terms with controlled vocabulary terms.

Given all the above, the more specific research question for this project is: how similar are the IMLS Collection Registry user keywords (extracted from transaction log) and the controlled vocabulary terms from three different controlled vocabularies – GEM thesaurus and its

alternatives? Which of the three controlled vocabularies matches higher percentage of the search terms from the user queries made in the Registry?

Data Collection and Data Analysis

Based on the recent decade's research on matching user terms with controlled vocabulary terms (Collantes 1995, Dubin 1998, Greenberg 2001, Gault, Shultz & Davies 2002, Gross & Taylor 2005, Nowick & Mering 2003, Qin 2000), the following conclusions can be made regarding the typical data source, data processing and data analysis techniques applied:

Typical source of data: transaction logs, user terms submitted for mediated search

Typical data processing techniques:

- Parsing user queries into separate terms (excluding stop words) and phrases
- Extracting stems from the words in user queries

Typical analysis steps:

- Match user queries with controlled vocabulary terms: most often exact, near exact matches (with variations in spelling, endings and plurals/singulars), and synonyms (SYN), sometimes – also broader terms (BT), related terms (RT), narrower terms (NT) (the latter works only in structured thesaurus, which GEM is not).
- Run user queries (with user terms either mapped or un-mapped to particular thesaurus, including SYN, BT, NT, RT) in the same or other comparable system they originate from (e.g. OPAC, article database)

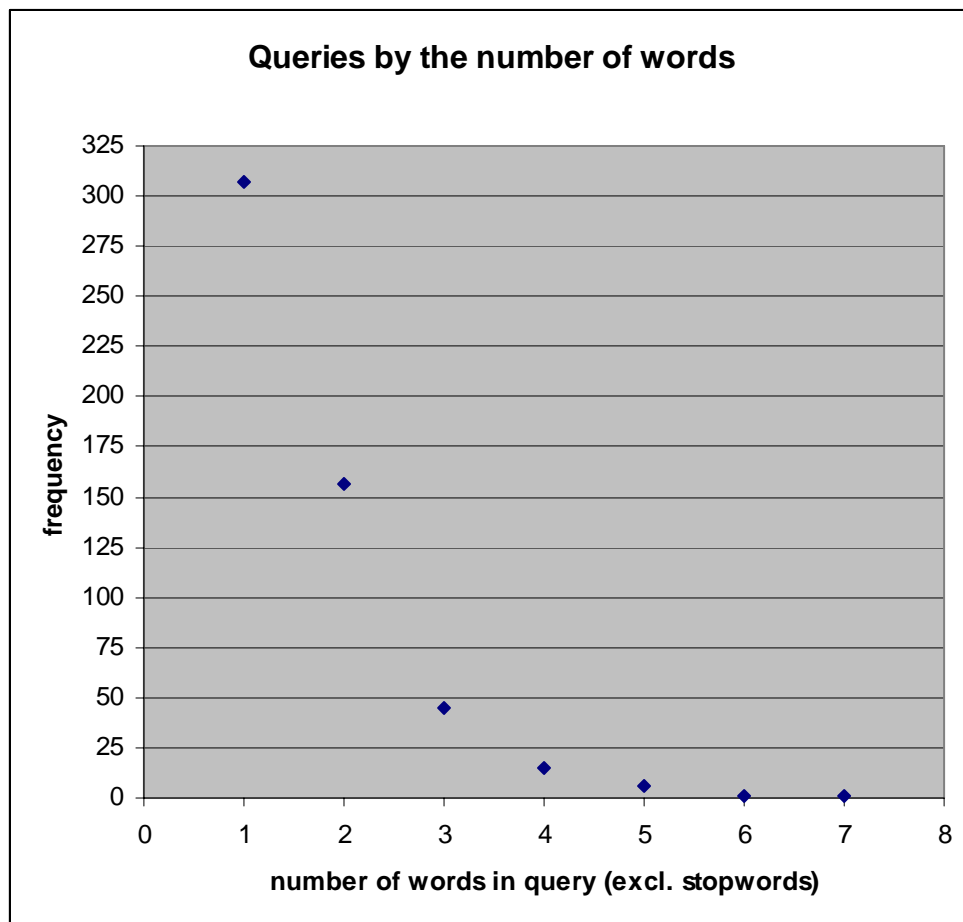
Typical data analysis techniques: Qualitative analysis and descriptive statistics.

The major dataset used in the analysis is the IMLS Collection Registry transaction log dataset – an Access file that consists of over 19,000 records and covers a period of approximately 7 months, between February 2005, when collection registry was first made publicly accessible, and September 2005. Initially transaction log file consisted of over 100,000 records, but after exclusion of the noise – searches and browsing made in the Collection Registry by web crawlers and Registry testers – the size was reduced to approximately 19,000 records. Each record/row contains information on IP address the query originated from, date and time of access, webpage visited within the Collection Registry, raw query string etc.

The transaction log was manually processed to extract all the keyword search query strings – a total of 945, which were then alphabetized (see Attachment 2). Given the time constraints of this project, a subset of 533 user queries was selected for analysis. Since the sample constitutes a large portion – over 56% – of the total dataset, it should be representative of a dataset as a whole.

Sampling procedure was conveniently applied as follows: queries that start with letters “A” to “L” were selected for analysis. The limitations of such sampling include uneven distribution of potential search terms throughout the alphabet: some letters have much more words starting with them than the others. Also, the search terms that started with numbers were not included. The rest of the dataset will be included in further analysis.

The user keyword queries vary in complexity and length. For example, the number of words in each query ranges from 1 to 7, with the vast majority consisting of one or two words, as can be seen from the chart below.



Preserving the context of a search is an important factor for analysis, especially when trying to decide on search type and finding a match with the terms in a controlled vocabulary. Therefore, the decision was made not to parse queries into separate words or even further – into stems. The minimal processing of the queries was done with noun words in queries: plurals were truncated and grouped together in the same query with the singulars of the same words (e.g. “Indians” and “Indian” became “Indian*”, “clipper ships” and “clipper ship” became “clipper ship*”). Both correct and misspelled versions of the same words were considered the instances of the same query (e.g. “Antarctica” and “antartica”, “immigration” and “imigration”).

At the **first stage of analysis**, general descriptive statistics procedures were used: search frequencies and the number of words excluding stop words in queries were calculated for each query, averaged for the whole sample and for each category separately. The stop words for these purposes included prepositions, conjunctions and articles.

The major part of the first stage of analysis was categorizing the user queries into seven broad search types or categories, derived from the Functional Requirements for Bibliographic Records (FRBR, 1998) classification of the entities in bibliographic universe. Seven out of ten FRBR entities that can be subjects of the work were used in this study's framework: *work*, *person*, *corporate body*, *concept*, *object*, *event* and *place*. The definitions of each entity and examples given by FRBR – detailed for *work*, *person*, and *corporate body*, but scarce for *object*, *concept*, *event* and *place* – were followed as guidelines for distinguishing between the categories. In essence, seven categories are characterized by FRBR as:

1. *work*: a distinct intellectual or artistic creation (FRBR, p. 16)
2. “*person*: an individual; encompasses individuals that are deceased as well as those that are living” (p. 23)
3. “*corporate body*: an organization or group of individuals and/or organizations acting as a unit; encompasses organizations and groups of individuals and/or organizations that are identified by a particular name, including occasional groups and groups that are constituted as meetings, conferences, congresses, expeditions, exhibitions, festivals, fairs, etc. The entity also encompasses organizations that act as territorial authorities, exercising or claiming to exercise government functions over a certain territory, such as a federation, a state, a region, a local municipality, etc. The entity encompasses organizations and groups that are defunct as well as those that continue to operate” (p. 24)
4. “*concept*: an abstract notion or idea; encompasses a comprehensive range of abstractions that may be the subject of a *work*: fields of knowledge, disciplines, schools of thought (philosophies, religions, political ideologies, etc.), theories, processes, techniques, practices, etc. A *concept* may be broad in nature or narrowly defined and precise” (p. 25)
5. “*object*: a material thing; encompasses a comprehensive range of material things that may be the subject of a *work*: animate and inanimate objects occurring in nature; fixed, movable, and moving objects that are the product of human creation; objects that no longer exist” (p. 26)

6. “*event*: an action or occurrence; encompasses a comprehensive range of actions and occurrences that may be the subject of a work: historical events, epochs, periods of time, etc.”(p. 27)
7. “*place*: a location; encompasses a comprehensive range of locations: terrestrial and extra-terrestrial; historical and contemporary; geographic features and geo-political jurisdictions”(p. 27).

FRBR’s *expression*, *manifestation* and *item* entities were not adopted as categories for this analysis, since it is virtually impossible to detect from transaction log what exactly the user was searching for: an abstract work, its particular expression, manifestation or item. Therefore, in my classification of Collection Registry queries, *work* is broader than FRBR’s *work* and covers any artistic creation that has a title, including the digital collections that are members of IMLS Collection Registry.

Although the FRBR *person* entity does not currently cover families, there is a provision to update FRBR model with adding *family* entity to the same group of entities that contains *person* and *corporate body*. Therefore, I tentatively expanded the *person* category in my analysis to include families (e.g. “Cushmans”), as well as ethnic groups/nationalities (e.g. “Irish Americans”) and classes of persons (e.g. “children that are abused”) that I believe belong to the same group of entities and are tightly connected with *person* entity.

The rare occasions of fictitious characters were treated on the basis of “what they would be if they really existed” (e.g., TV series’ character Alf is a creature, thus an FRBR’s *object*, as would also be a dog or a squid).

For consistency in distinguishing between types of searches in less straight-forward cases, some simple rules were developed: unspecified social and business institutions (e.g. “library”, “archive”, “can company”, “amusement park”) were classified as *concepts*, institutions for which physical structure is more important (e.g., “ballrooms”, “highways”) as *objects*, and more specifically named ones (e.g., “Icy Hot Bottle Co.”, “library+Moorhead”) as *corporate bodies*. Some queries presented a real challenge for classification: “books” was one of them, which I tentatively categorized as a *concept*, although it could as well be an *object*. As any categorization, such an approach is inevitably judgmental, which is one of the limitations of the study. Another limitation of applying FRBR framework – as probably any other – for

categorization of subject searches lies in ambiguity of actual searches, further discussed in Findings and Discussion section.

The queries that presented no clue as to what type they belong to (e.g., “aF”, “beyond”, “LU+65”) were grouped together in an eighth category – *unknown*.

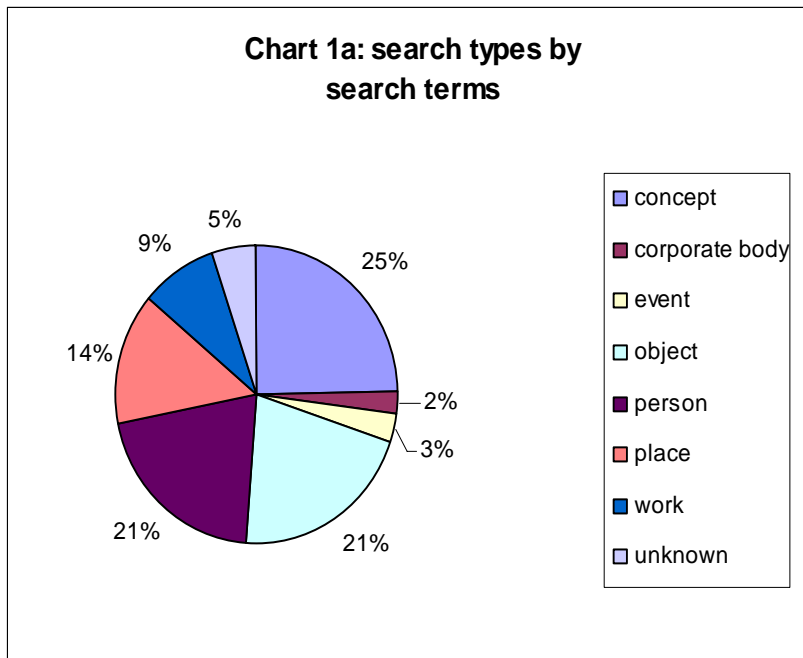
The **second stage** of analysis included searching in three controlled vocabularies – GEM, LCSH, and Art and Architecture Thesaurus – for the semantic matches of actual user queries from the IMLS Collection Registry transaction log. Library of Congress Subject Headings was selected for analysis as a controlled vocabulary that almost a half of digital collections participating in Collection Registry are using for item-level description and that is being considered by some of surveyed collections as an alternative to GEM for collection-level description. OCLC Connexion database features – LCSH authority file and Web Dewey search for editorially mapped LCSH headings – were used for matching user queries with LCSH. Art and Architecture Thesaurus (AAT) was selected as another plausible alternative for describing cultural heritage materials – and possibly collections. A number of collections participating in the registry are using AAT for their item-level description. Moreover, AAT is a controlled vocabulary of a smaller scope than LCSH, but significantly more detailed than GEM.

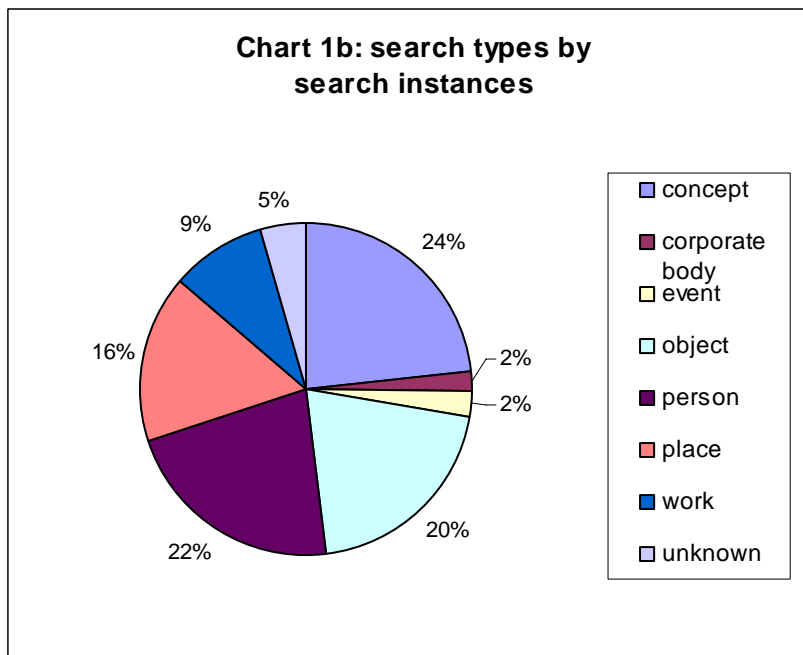
Only exact/abbreviated and synonymous matches (e.g., “inoculation” and “vaccination”) were treated as semantic matches for the purposes of this analysis. Abbreviated queries were matched with the full terms in controlled vocabularies, e.g. “ilgwu” with “International Ladies’ Garment Workers’ Union”. The order of the terms in the query, as well as presence or absence of prepositions and conjunctions was ignored for analysis. (e.g., “French art” was matched with “Art, French”; “epistemology” with “knowledge, theory of”, “children that are abused” with “abused children”). Endings of the words were also disregarded, as long as they did not affect the meaning (e.g., “automated speech recognition” was matched with “automatic speech recognition”). Both preferred terms and variant terms in controlled vocabulary were considered legitimate matches. For example, both 150 MARC field (USE) and 450 field (USE FOR) in LCSH authority records were analyzed to find a semantic match to a user query. Simple user queries were in some cases matched with compound LCSH subject headings, for instance “housing for shipyard workers” was matched with “Shipbuilding industry—Employees—Housing”.

The number of matches was totaled and averaged for the whole sample and for search types. Complete categorized listing of user query terms, along with descriptive statistics and calculations is available at <https://netfiles.uiuc.edu/zavalina/MDRTPapers/AtoLwithAAT.xls>.

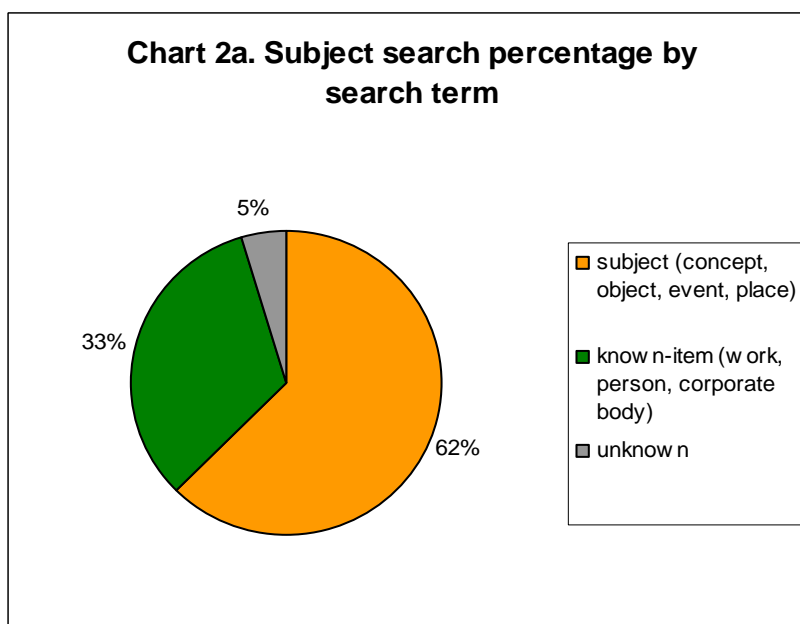
Findings and Discussions

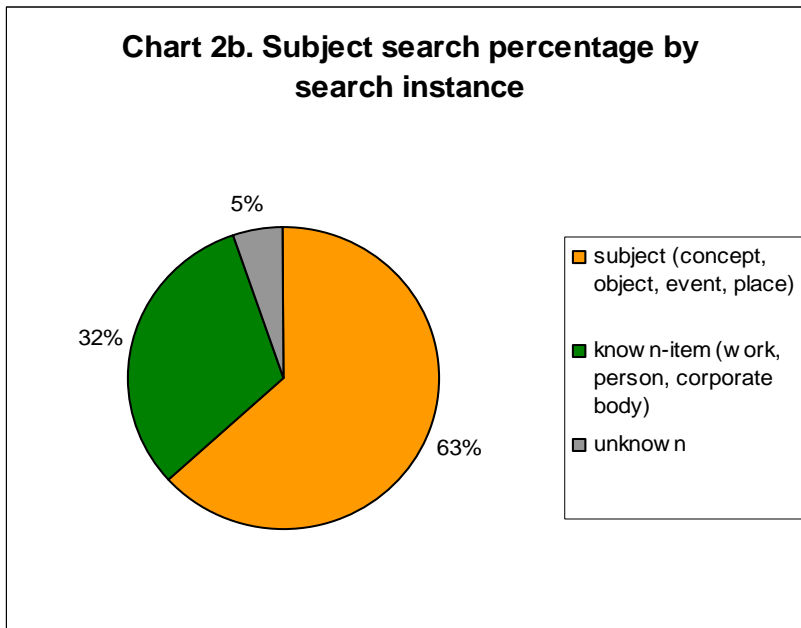
As the first stage of analysis demonstrated, two thirds of all searches made in IMLS Digital Collection Registry are spread between three broad FRBR categories: *concept*, *object*, and *person*, with *concept* search leading among both search terms and search instances. *Place* also takes significant percentage of searches, while *corporate body*, *event*, *work*, and *unknown* search types combined total below 20% of the searches. The low level of *event* searching is surprising, since most of the historical searches would be searches for *events*.



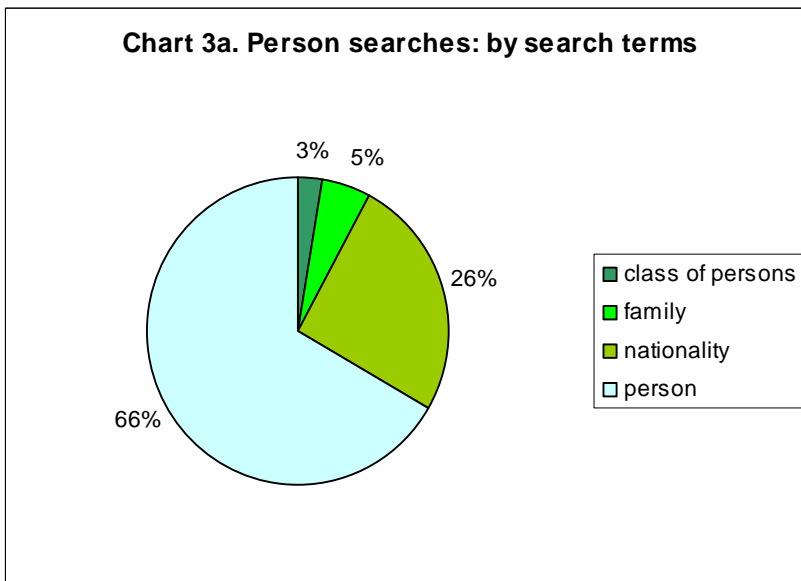


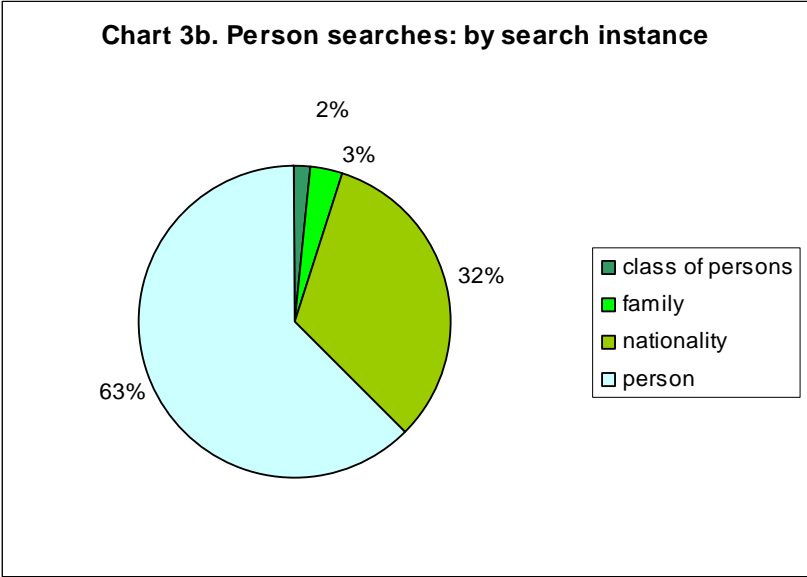
Because of the very nature of *concept*, *object*, *place*, and *event*, these cannot possibly belong to the widely-used in LIS general type of known-item searches (i.e. searches where the user knows either author or title of the work sought), and therefore search categories can be safely considered subject searches in the IMLS Collection Registry. As can be easily seen from the chart below, then subject search constitutes at least 62% of all search terms and all search instances.



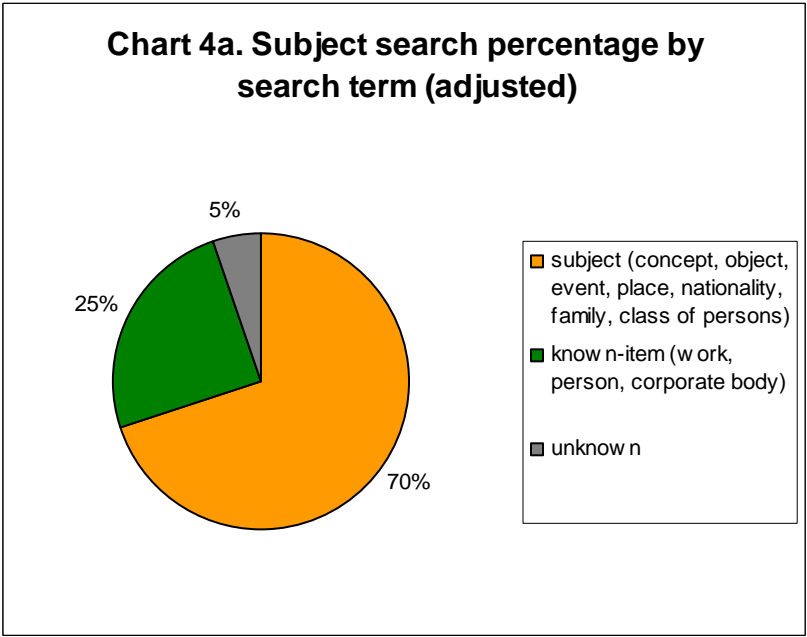


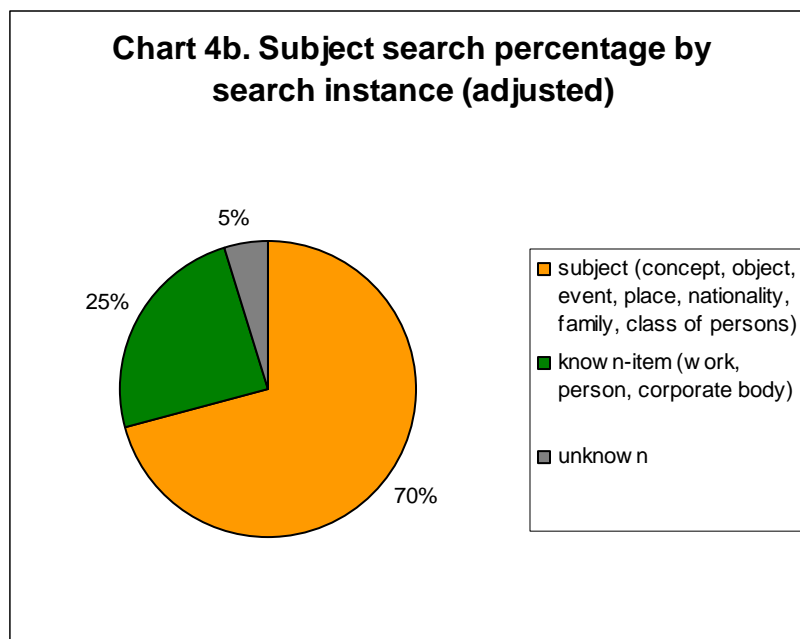
However, of course not all *person* searches will be known-item searches, since the broad *person* category includes also families, ethnic groups/nationalities and classes of persons. In the sample studied, over one third – 34% to 37% – of all searches initially assigned to the *person* search type represents these types of searches:





Thus, by adding these family, ethnic group/nationality and class of persons searches to the pool of subject searches, the percentage of subject searches made in the IMLS Collection Registry increases to 70% by both search term and search instance:





Although the number of federated digital collections has been rapidly growing recently, as did the creation and use of collection registries, no attempt has been documented in LIS literature so far to conceptualize known-item and subject searches at the collection level. In my operational definition, since in IMLS Digital Collection Registry the searches are conducted at the collection level, the known-title search in such situation will be the search, where the user knows the title of the digital collection; everything else will be a subject search, which, broadly defined, includes both controlled- and uncontrolled-vocabulary searches with an intent to find information on particular subject/topic/discipline/area.

The majority – sixty-three percent of search terms and seventy-two percent of search instances – of the *work* searches in the Registry were the searches for specific digital collection title, thus a known-item search. Although the rest of *work* searches were for the specific item-level titles, and therefore at the collection-level search can be treated as subject searches, the number of them was not significant enough to affect the distribution of two major search types – subject and known-item – as shown above.

The prevalence of subject search is obvious from the charts and remains in agreement with results of the 1982 large-scale Council for Library Research (CLR) study of online catalog use, which radically changed the conception of catalog use by finding subject search to be unexpectedly widely used by patrons – 59% of all searches. Compared to the earlier transaction log studies of online catalog use (e.g., Larson 1991B), including CLR study itself, the relative value of subject search as shown by the current study is much higher, which can be explained by at least two reasons:

- a general shift towards subject searches in a world where abundance of publication makes it less and less possible to know the title or author of the specific item
- a conceptual difference between collection-level and item-level searches, which implies a trend towards increased levels of subject search in federated collection registries.

Further research into how the search type distribution in IMLS Item-level Repository and IMLS Digital Collection Registry correlate with each other will help to answer these questions.

It should be noted here that actual searches conducted by users in the Registry rarely could be categorized “strictly” as any one of the search categories/FRBR entities, and sometimes presented a real challenge in determining which entity was the major component of a query. Below is a discussion of some of the examples found in this transaction log:

1. **“Amusement park”**. As an abstract idea of amusement parks this query might be categorized as a *concept*. On the other hand, amusement parks are physical structures created by people, which makes it an *object* in FRBR definition. There is no correct answer to this question, even asking a user what (s)he meant when making this search would not clarify the ambiguity in most cases, because a concept of amusement parks is tightly connected to the object of amusement park. If you ask a user, you might learn that the search was for a specific institution, thus a FRBR *corporate body*. Similar examples of queries from the sample studied include **“Archives”**, **“Ballrooms”**, **“Highways”**, **“detroit+historical+museums”** (the latter is also inseparable from a specific location – FRBR *place*, as is **“library Moorhead”**).
2. **“Industrial models”**. The very word “models” implies it being a *concept*, as modeling requires conceptualizing. On the other hand, industrial models are physical structures created by people to assist in specific industrial processes, therefore it can also be categorized as an *object*. **“lesson+plans”** appears to be a very similar example, only from another realm – education rather than industry.
3. **“Landscape”** is something that exists in the nature, or can be created by people, thus it seems to be a FRBR *object*. However, the possibility exists for it to be classed as a *concept* too, if a user is searching for literature on landscapes and landscaping as a discipline.
4. **“letters+from+19th+century”** is a pretty straightforward example of *object* search. However, it is qualified by a specific time period, which, in FRBR definition, is an *event*.
5. **“asian+American”** appears to be a *person* search, although often refers to a broader category nonexistent in FRBR model yet – an ethnic group. However, it is inseparable from two *places* – Asian and American continents. In my understanding, a person or ethnic group in general is in most cases defined through place. Similarly, **“children+that+are+abuse”** is also a group (or a class) of *persons* inseparable from another FRBR entity, but defined by *event* of abuse rather than by *place*.
6. **“henry+fordmuseum+and+greenfiel+village”** is a specific corporate body (the Library of Congress corporate body authority file exists for it in WorldCat). However, it is

obvious, that a *person* of Henry Ford and a *place* of Greenfield Village are integral parts of this query.

7. “*don+quijote*” is both a fictitious character created by Cervantes and a phrase widely known as a title of his book -- although in fact it is just a part of the book’s title. Categorization of this search entirely depends on the user intention, which cannot be known from the query itself. If the search was for a book, it was a *work* search, but if it was for a character it was either a *concept* (something abstract that does not exist and never physically existed), or a *person* if we follow the logic of “what it would be if it existed”.
8. “*Civil rights movement*” might be classified as an *event*, which is a tricky entity because it is, according to Functional Requirements to Subject Authority Records (Zeng and Salaba, 2005), a combination of *place* and time. But where is time and place in this query? It may equally refer to various times and place, e.g. 1950s United States, or 1960s France, or 1970s Soviet Union, or 2000s China. Does the absence of explicit or implicit qualifiers make it a *concept*? “*Census*” seems to belong to the same cluster of examples.

Studies of transaction logs typically look at the frequencies of search term use and the average number of words in the search query. For the sample of queries analyzed in this study, the average frequency of term use was rather low – 1.4. The highest search term use frequency was recorded for *place* category – 1.58 – and the lowest was recorded for *event* category – 1.08. In terms of the typical number of words in query excluding stop words, the average for the whole sample constituted 1.69 words per query. The highest average number of words per query was recorded for *corporate body* category of search – 2.78 – and the lowest was recorded for *place* – 1.35 words per query.

At the second stage of analysis, the number of matches for user search queries in three controlled vocabularies – GEM subject scheme, Library of Congress Subject Headings, and Art and Architecture Thesaurus – was compared for each search term (combination of terms in the user query), for each category of searches, and for the whole sample. A total of 10 matches – 2.6% out of 380 unique search terms – were found in GEM subject scheme. A total of 271 matches – 71.3% – were found in LCSH. Art and Architecture Thesaurus matched only 86 – 22.63% of user keywords. The only category of user searches GEM had matches to was *concept*, while LCSH had matches to all the categories, including a couple of *unknown* searches, which as the category were the worst represented in LCSH. Art and Architecture Thesaurus terms matched mostly *concepts* and *objects*, with no matches at all in *corporate body*, *place* and *work* search categories. The table below illustrates absolute and relative values of these semantic matches:

Table 1.

FRBR subject type	Unique search terms	search instances	GEM match	GEM match, %	LCSH match	LCSH match,%	AAT match	AAT match, %
concept	94	125	10	10.54	87	92.55	53	56.38
corporate body	9	10	0	0	5	55.56	0	0
event	12	13	0	0	6	50.00	2	16.66
object	79	108	0	0	51	64.56	29	36.71
person	78	117	0	0	63	80.77	1	1.28
place	55	87	0	0	51	92.73	0	0
Work	34	49	0	0	4	11.76	0	0
Unknown	19	24	0	0	4	21.05	1	5.26
TOTAL	380	533	10	2.63	271	71.32	86	22.63

The low level of matching between the user search terms and the GEM subject terms is explained by the extreme broadness of this subject scheme. There is no widely shared notion of the digital collection even among collection creators and managers (Lee 2000, Hill et al. 1999); much more confusion exists among the users of federated collection repositories. Such an ambiguity can cause sometimes unjustified preciseness and narrowness of collection-level search terms selected by Registry users, who are not making distinction between searching for items in collection and searching for collections in collection registry. Whatever is the reason, the mismatch between the GEM subject scheme and actual searches is obvious.

Surprisingly, LCSH, although matching most of the user terms, still leaves almost 30% unmatched. LCSH is the most effective in matching *places* and *concepts*, while *works* remain the least matched; only about a half of *corporate bodies* and *events* from this study's sample are covered by LCSH terms. The reason may lay in general inflexibility of LCSH – a large scheme that is extremely hard to keep up-to-date. A vivid illustration from this study is the absence of such term as “learning standards” in LCSH authority file.

However, as can be seen from the Table 2 below, compared to the other two controlled vocabularies, LCSH on its own (without overlap with AAT or GEM) covers the lion share – almost 50% of user search terms. Only 6 terms matched by AAT were not also matched in LCSH, and all the terms matched in GEM were also matched in LCSH.

Table 2.

FRBR subject type	unique search terms	matched by GEM alone	matched by GEM and LCSH	matched by GEM and AAT	matched by LCSH alone	matched by LCSH and AAT	matched by AAT alone	matched by ALL	matched by NONE
concept	94	0	3	0	32	45	1	7	6
corporate body	9	0	0	0	5	0	0	0	4
event	12	0	0	0	4	2	0	0	6
object	79	0	0	0	26	25	4	0	24
person	78	0	0	0	62	0	1	0	15
place	55	0	0	0	51	0	0	0	4
work	34	0	0	0	4	0	0	0	30
unknown	19	0	0	0	3	1	0	0	15
TOTAL	380	0	3	0	187	73	6	7	104

The most unexpected finding of the second stage of analysis was that well-developed, up-to-date, flexible and faceted Art and Architecture Thesaurus, which seems to be especially suitable for describing cultural heritage materials and possibly collections, matched such a small proportion of user search terms. The explanation can lay in the fact that AAT, just as GEM, does not include name and place authority files. However, the broader Getty Thesaurus framework, along with AAT, also includes such authority files.

Conclusions

This study results demonstrate an unusually high for catalog use / transaction log analysis studies level of subject searching made by patrons at the collection level. Further investigation is needed into the reasons of such increase in subject search proportion, including collection of data through collection registry users' interviews and observations.

Further research is also needed into which controlled vocabulary would best represent digital collections in the IMLS collection registry. Although LCSH has shown relatively good results, none of the three controlled vocabularies in this study fully represents the subjects of diverse collections in the IMLS Digital Collection Registry, or at least a user's expectations towards these subjects. For the future study, another – more flexible than LCSH – controlled vocabulary of the moderate scale, which, unlike GEM or AAT, represents a wider variety of search types – not just *concepts* and/or *objects* – should be selected for the same analysis and for comparison with GEM, LCSH, and AAT. To compensate for deficiencies of the transaction log analysis as a method that does not provide any insight into user motivations and intentions and deals only with user actions, think-aloud protocol observation of the users searching IMLS Digital Collection registry should be incorporated into further analysis.

Bibliography

- Cochrane, P. (1986). *Improving LCSH for Use in Online Catalogs*. Colorado Springs, CO: Libraries Unlimited.
- Cochrane, P. (2000). Improving LCSH for use in online catalogs revisited: what progress has been made, what issues still remain. *Cataloging and Classification Quarterly*, 29 (1/2), 73-89.
- Collantes, L. Y. (1995). Degree of Agreement in Naming Objects and Concepts for Information Retrieval. *Journal of the American Society for Information Science*, 46(2), 116-132.
<http://www3.interscience.wiley.com.proxy2.library.uiuc.edu/cgi-bin/fulltext/10050153/PDFSTART>
- Dubin, D. S. (1998) Addressing the heterogeneity of subject indexing in the ADS databases. In U. Grothkopf, H. Andernach, S. Stevens-Rayburn, and M. Gomez, editors, *Library and Information Services in Astronomy III, volume 153 of A.S.P. Conference Series*, 77-83, San Francisco. Astronomical Society of the Pacific.
<http://www.eso.org/gen-fac/libraries/lisa3/dubind.html>
- Functional Requirements for Bibliographic Records: Final report (1998). IFLA. *UBCIM Publications, New Series*, 19. <http://www.ifla.org/VII/s13/frbr/frbr.pdf>
- Gault, L. V., Shultz, M., & Davies, K. J. (2002). Variations in medical subject headings (MeSH) mapping: From the natural language of patron terms to the controlled vocabulary of mapped lists. *Journal of the Medical Library Association (JMLA)*, 90(2), 173-180.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=100762>
- Greenberg, J. (2001). Automatic Query Expansion via Lexical-Semantic Relationships. *Journal of the American Society for Information Science*, 52(5), 402-415.
<http://www3.interscience.wiley.com.proxy2.library.uiuc.edu/cgi-bin/fulltext/77002637/PDFSTART>
- Gross, T., & Taylor, R. (2005). What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results. *College and Research Libraries*, 66 (3), 212-230.
<http://vnweb.hwwilsonweb.com/hww/jumpstart.jhtml?recid=0bc05f7a67b1790e183771395b86e5aa985cedb4235ee102d23781fb323b5ec2a6cec93038a63198&fmt=P>
- Hill, L., Janée, G., Dolin, R., Frew, J., & Larsgaard, M. (1999). Collection metadata solutions for digital library applications. *Journal of the American Society for Information Science*, 50 (13), 1169-1181.
- Larson, R. R. (1991A). Between Scylla and Charybdis: Subject searching in online catalogs. *Advances in Librarianship*, 15, 175-236.
- Larson, R. R. (1991B). The decline of subject searching: long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science*, 42 (3), 197-215.
- Lee, H. (2000). What is a collection? *Journal of the American Society for Information Science*, 51(12), 1106-1113.
- Matthews, J. R., Lawrence, G. S., & Ferguson, D. K. (Eds.), (1983). *Using online catalogs: A nationwide survey: A report of a study sponsored by the Council on Library Resources*. New York, NY: Neal-Schuman.
- Nowick, E., & Mering, M. (2003). Comparisons between Internet Users' Free-Text Queries and Controlled Vocabularies: A Case Study in Water Quality. *Technical Services Quarterly*, 21(2), 15-32.
http://www.haworthpress.com.proxy2.library.uiuc.edu/store/E-Text/View_EText.asp?sid=F38AWX7KVNXE8G8K3MJ2XBKG9S4U2SX&a=3&s=J124&v=21&i=2&fn=J124v21n02%5F02
- Qin, J. (2000). Semantic Similarities between a Keyword Database and a Controlled Vocabulary Database: An Investigation in Antibiotic Resistance Literature. *Journal of American Society for Information Science*, 51(2), 166-180. <http://www3.interscience.wiley.com.proxy2.library.uiuc.edu/cgi-bin/fulltext/69501138/HTMLSTART>
- Sutton, S. A. (1999) "Conceptual design and deployment of a metadata framework for educational resources on the Internet." *Journal of the American Society for Information Science*, 50, 1182-1192.
- Sutton, S. A. (2004) Building an Education Digital Library: GEM and Early Metadata Standards Adoption. In *Metadata in Practice*. Ed. by D. I. Hillmann, E. L. Westbrooks - Chicago: American Library Association, 1-15.
- Zeng, M., & Salaba, A. (2005). Toward an International Sharing and use of Subject Authority Data. FRBR Workshop, OCLC, 2005. http://www.oclc.org/research/events/frbr-workshop/presentations/zeng/Zeng_Salaba.ppt

ATTACHMENT 1

Gateway to Educational Materials Subject Scheme

GEM Level 1	GEM Level 2
Arts	Architecture Art therapy Careers* Computers in art Dance Drama/dramatics Film History* Informal education* Instructional issues* Music Photography Popular culture* Process skills* Technology* Theater arts Visual arts
Educational technology	Audio-visual equipment Careers* Educational media History* Informal education* Instructional issues* Integrating technology into the classroom Language laboratories Multimedia education Process skills* Staff inservice Technology* Technology planning
Foreign languages	Alphabet Bilingualism Careers* Cultural awareness Grammar History* Informal education* Instructional issues* Linguistics

	<p>Listening comprehension</p> <p>Process skills*</p> <p>Reading</p> <p>Speaking</p> <p>Spelling</p> <p>Technology*</p> <p>Vocabulary</p> <p>Writing</p>
Health	<p>Aging</p> <p>Body systems and senses</p> <p>Careers*</p> <p>Chronic conditions</p> <p>Consumer health</p> <p>Death and dying</p> <p>Disease</p> <p>Environmental health</p> <p>Family life</p> <p>History*</p> <p>Human sexuality</p> <p>Informal education*</p> <p>Instructional issues*</p> <p>Mental/emotional health</p> <p>Nutrition</p> <p>Process skills*</p> <p>Safety</p> <p>Smoking</p> <p>Substance abuse prevention</p> <p>Technology*</p>
Language Arts	<p>Alphabet</p> <p>Careers*</p> <p>Debate</p> <p>Grammar</p> <p>Handwriting</p> <p>History*</p> <p>Informal education*</p> <p>Instructional issues*</p> <p>Journalism</p> <p>Listening comprehension</p> <p>Literature</p> <p>Mechanics</p> <p>Phonics</p> <p>Process skills*</p> <p>Reading</p> <p>Reading aloud</p> <p>Speech</p>

	<p>Spelling</p> <p>Story telling</p> <p>Technology*</p> <p>Vocabulary</p> <p>Whole language</p> <p>Writing (composition)</p>
<p>Mathematics</p>	<p>Algebra</p> <p>Applied mathematics</p> <p>Arithmetic</p> <p>Calculus</p> <p>Careers*</p> <p>Discrete mathematics</p> <p>Functions</p> <p>Geometry</p> <p>History*</p> <p>Informal education*</p> <p>Instructional issues*</p> <p>Measurement</p> <p>Number sense</p> <p>Number theory</p> <p>Patterns</p> <p>Probability</p> <p>Process skills*</p> <p>Statistics</p> <p>Technology*</p> <p>Trigonometry</p>
<p>Philosophy</p> <p>Note:</p> <p>2nd level = ERIC</p> <p>Thesaurus</p> <p>"Narrower Terms"</p>	<p>Aesthetics</p> <p>Careers*</p> <p>Educational Philosophy</p> <p>Epistemology</p> <p>Ethics</p> <p>Existentialism</p> <p>Hermeneutics</p> <p>History*</p> <p>Informal education*</p> <p>Instructional issues*</p> <p>Logic</p> <p>Marxism</p> <p>Phenomenology</p> <p>Platonism</p> <p>Process skills*</p> <p>Semiotics</p> <p>Technology*</p>
<p>Physical</p>	<p>Adventure and risk challenge activities</p>

<p>Education</p>	<p>Aquatics Careers* Games (educational) Gymnastics(educational) History* Individual sports Informal education* Instructional issues* Motor/movement skills Outdoor education Process skills* Rhythms and dance Skill-related fitness Team sports Technology*</p>
<p>Religion Note: 2nd level = ERIC Thesaurus "Narrower Terms"</p>	<p>Buddhism Careers* Christianity Confucianism History* Informal education* Instructional issues* Islam Judaism Process skills* Taoism Technology*</p>
<p>Science</p>	<p>Agriculture Astronomy Biological and life sciences Biology Botany Careers* Chemistry Earth science Ecology Embryology Engineering Entomology General science Geology Histology History* Informal education* Instructional issues*</p>

	<p>Metallurgy</p> <p>Meteorology</p> <p>Natural history</p> <p>Oceanography</p> <p>Paleontology</p> <p>Pharmacology</p> <p>Physical sciences</p> <p>Physics</p> <p>Process skills*</p> <p>Space sciences</p> <p>Technology*</p>
<p>Social studies</p>	<p>Anthropology</p> <p>Careers*</p> <p>Civics</p> <p>Comparative political systems</p> <p>Criminology</p> <p>Current events/issues</p> <p>Economics</p> <p>Geography</p> <p>Gerontology</p> <p>History*</p> <p>Human behavior</p> <p>Human relations</p> <p>Informal education*</p> <p>Instructional issues*</p> <p>Process skills*</p> <p>Psychology</p> <p>Social work</p> <p>Sociology</p> <p>State history</p> <p>Technology*</p> <p>Technology and civilization</p> <p>United States Constitution</p> <p>United States government</p> <p>United States history</p> <p>Urban studies</p> <p>World history</p>
<p>Vocational education</p>	<p>Agriculture</p> <p>Allied health occupations</p> <p>Business</p> <p>Careers*</p> <p>Cooperative education</p> <p>Distributive</p> <p>History*</p> <p>Informal education*</p>

	Instructional issues* Occupational home economics Process skills* School-to-work Tech prep Technical Technology* Trade and industrial
--	--

ATTACHMENT 2

User Keyword Queries in IMLS Digital Collection Registry: Alphabetic Sequence

1. 16+MM
2. 1704
3. 1704
4. 1800-1849+fashion+
5. 1800-1849+fashion+of+clothing
6. 1800-1849+fashion+of+clothing
7. 1818+
8. 1876
9. 1895
10. 1895
11. 1976
12. 19th+century+epistles
13. a%3F
14. A.J.+Small
15. a+bird+in+a+gilded+cage
16. a+streetcar+named+desire
17. aboriginal
18. accounting
19. adams
20. adult
21. aerial
22. aeruak
23. africa
24. africa
25. africa+focus
26. africa+focus
27. african
28. african
29. african+american+studies
30. agriculture
31. aircraft
32. Ajumawi
33. Ajumawi%2Fatsugewi
34. akron
35. alex+janis
36. alf
37. alferd+packer
38. alfred+packer
39. algeria
40. alternative+energy
41. american+
42. American+Centuries
43. american+history&type=text
44. american+indian
45. american+jouneys
46. american+jouneys
47. american+journeys
48. american+journies
49. american+literature
50. american+literature
51. american+natural+science
52. american+studies
53. amusement
54. amusement+park
55. amusement+parks
56. animals
57. ansil+addams
58. Antarctica
59. Antarctica
60. antarctica
61. Antarctica
62. antartica
63. antartica
64. Antifederal+Club
65. arab
66. archaeological
67. archaeology
68. archaeology
69. architecture
70. architecture
71. Architecture
72. archives
73. Arizona
74. arkansas

75. art+deco
76. artificialintelligence
77. Asia+continent
78. asian+American
79. assessing+governmental+performance%3A
+an+analytical+framework
80. astronomy&type=image
81. Atsuge
82. audio
83. autoharp
84. automated+speech+recognition%94+&type
=dataset&type=text
85. automated+speech+recognition%94+AND+
%28software+OR+system%29
86. automated+speech+recognition%94+AND+
%28software+OR+system%29&type=datase
t&type=text
87. automobile
88. automobile
89. automobile
90. Baby+Beauty+Contests+in+Pittsburgh%2C
+PA+1936-1941
91. Ballrooms
92. Bangwell+Putt
93. Bangwell+Putt
94. baseball
95. baseball
96. baseball&type=image&type=moving+image
&type=sound
97. basketball
98. baskin
99. basque
100. basque
101. battle+of+new+orleans
102. Bay+State+Belting+Co.
103. beadwork
104. beaver
105. Belle+Isle
106. berkely+university+dinosaur
107. berryman
108. beyond
109. biography+of+Enid+M.+Baa
110. birds
111. birth+announcements
112. black
113. black+studies
114. blimp
115. blue&type=sound
116. blue&type=sound
117. blue&type=sound
118. BNDIAN+SITE
119. body
120. bohemian+grove
121. bohemian+grove
122. bohemian+grove
123. bohemian+grove
124. bohemian+grove
125. bolles
126. Books
127. books
128. Boston+City+Directory+1885
129. Boston+City+Directory+1905
130. Boston+City+Directory+1935
131. bottom+trawling
132. Bozeman+area+Indians
133. BROMELIAD
134. bronx+neighborhoods
135. bronx+postcards
136. bronxart.lehman.cuny.edu%2Fpa%2Fneighb
orhood.htm
137. brooklin%2C+maine
138. brooklin+me
139. Brooklyn+Daily+Eagle+--+Dittman
140. Brooklyn+Daily+Eagle+in+1941
141. broward
142. busquets
143. california+city+directories
144. california+digital+library
145. californian+indian+art
146. cameo
147. can+company
148. Canada
149. canada
150. Canaletto

151.canallete
 152.canion
 153.cannibals
 154.cape+cod
 155.cape+may
 156.car
 157.car
 158.car
 159.car
 160.car
 161.car
 162.car
 163.car
 164.car
 165.car
 166.carbon
 167.Cars
 168.cars
 169.cars
 170.cat
 171.catherine+beecher
 172.catherine+beecher
 173.catherine+beecher
 174.census
 175.cervantes&type=text
 176.chemistry&type=unknown
 177.Cheques
 178.chess
 179.chicago
 180.chicago
 181.child+abuse
 182.child+abuse+
 183.childabuse+case+
 184.childabuse+case+in+maryland
 185.children+that+are+abuse
 186.Chile
 187.Chinese
 188.chinese
 189.Chinese
 190.chinese&type=text
 191.chinese+American
 192.chinese+language
 193.city+directories
 194.civil+rights+movement
 195.civil+war+records++illinois
 196.clark
 197.cleveland
 198.clipper+ship
 199.clipper+ship+cards
 200.clipper+ships
 201.close+quarters+in+detroit
 202.coal
 203.coal
 204.Colorad
 205.Colorado
 206.Colorado
 207.Colorado
 208.colorado
 209.Colorado
 210.Colorado+Granger
 211.Colorado+Granger
 212.columbia
 213.communist
 214.community
 215.computer
 216.comradeship
 217.concrete+music
 218.confucianism
 219.Congressional+Record
 220.Connecticut
 221.connecticut
 222.connecticut+history
 223.connecticut+history+online
 224.connecticut+teaching
 225.conservation
 226.Cook
 227.Cook+San+Francisco+Scrapbook
 228.Cook+Scrapbook
 229.cookbook
 230.cookbooks
 231.coommunity
 232.correspondence+19th+century
 233.costume
 234.county

235.Crosley
 236.cruikshank
 237.Cruikshank
 238.cruikshank
 239.cruikshank
 240.cuba+
 241.cuba+
 242.cuba+indipendence
 243.cuban+immigrants
 244.cuban+immigrants
 245.cubans
 246.cultural+competency
 247.currency
 248.cushman
 249.Cutrell
 250.Daugherty
 251.daumier
 252.Daumier
 253.deaf
 254.deaf+child
 255.Deerfield
 256.Deerfield
 257.deerfield
 258.deerfil%5Celd
 259.demography
 260.dentist
 261.dentist
 262.design
 263.Detroit
 264.Detroit
 265.Detroit
 266.Detroit+Boat+Club
 267.detroit+historical+museums
 268.detroit+river
 269.diaries+from+the+1930s+under+the+New+
 Deal+agencies
 270.digital+dress
 271.digital+dress
 272.digital+dress
 273.dinosaur
 274.dinosaurs
 275.dissertations
 276.documenting+american+south
 277.documenting+the+american+south
 278.dogs
 279.dogs
 280.dolphins
 281.dolphins%5C
 282.don+quijote
 283.don+quixote
 284.dorothea
 285.dorothea+lange
 286.dorothea+lange
 287.dorothea+lange
 288.Dorothea+Lange
 289.dorothea+lange
 290.dorothea+lange
 291.dorothea+lange&submit=Search
 292.dortha+lange
 293.dottie+long
 294.dottie+lucille+long
 295.drabik
 296.dresses+from+the+1900+to+1980
 297.durer
 298.earth+field+trip
 299.easter
 300.eastern+Europe
 301.Eastman
 302.economics
 303.economics
 304.economics
 305.edge+of+the+cedars+museum+collection
 306.education+by+design
 307.edward+curtis
 308.edward+curtis+wax+cylinder
 309.Edward+Mattis
 310.edwards
 311.Egypt
 312.eico+369
 313.Elsevier
 314.empire+state+building
 315.empire+state+building
 316.epistemology
 317.erik+satie

318.eubie+blake
 319.eubie+blake
 320.eubie+blake+scores+free
 321.eugenics
 322.exploratorium
 323.fairport+ny
 324.family
 325.family+tree
 326.farming+
 327.fashion
 328.fashion
 329.fashion
 330.fashion+for+the+1800-1849
 331.fashion+for+the+1800-1849%5C
 332.feeding
 333.feeding
 334.feeding+America
 335.feeding+america
 336.feeding+america
 337.ferrotype+Lincoln
 338.FILM
 339.find+it
 340.fire
 341.florida
 342.florida
 343.Florida&type=dataset&type=interactive+res
 ource
 344.Florida&type=dataset&type=soun
 345.florida+folklife
 346.flying+cloud
 347.folkstreams
 348.Fox+%2Cet+al+First+steps+to+accreditatio
 n+%2C+1992+gazette
 349.fragrance
 350.Frances+Lee+Pratt
 351.freemasonry
 352.french+art&type=image
 353.french+art&type=image
 354.freshwater+mussels
 355.gabriel+Moulin
 356.gambling
 357.gandhi
 358.gardener
 359.GATT
 360.gauguin
 361.GEM
 362.Genealogical
 363.genealogy
 364.genealogy
 365.genealogy&type=image&type=text
 366.george+washington
 367.gerd
 368.Giant+Squid
 369.glen+genz
 370.global+warming
 371.glopad
 372.google
 373.grainger
 374.grand+central+station
 375.graves
 376.graybar+building
 377.great+lakes
 378.Gros+Ventres
 379.Guinea
 380.Hamonic
 381.Hamonic+Fire
 382.harry+collins
 383.hartford+Connecticut
 384.haven%2C+maine
 385.haven+colony
 386.Hawaii
 387.hawaii%2C
 388.hearth
 389.heliotrope
 390.Heliotropium
 391.Heliotropium+tenellum
 392.henry+fordmuseum+and+greenfiel+village
 393.henry+fordmuseum+and+greenfiel+village
 394.Hibi
 395.higher+education
 396.Highland+Park
 397.Highways
 398.hippopotamus
 399.hippopotamus&type=image

400.hisako
 401.historic+atlas
 402.historic+atlase
 403.historic+atlases
 404.History
 405.history
 406.history+of+highways
 407.history+of+physical+education
 408.Hokusai
 409.Hollywood
 410.Hollywood
 411.holocaust
 412.holocuast
 413.homefront
 414.honore
 415.honre
 416.horse
 417.House
 418.housing
 419.Housing+for+Shipyard+Workers
 420.hungary
 421.Icy+Hot+Bottle+Co.
 422.ieee+collections
 423.ieee+publications
 424.ilgwu
 425.illinois
 426.illinois
 427.Image
 428.imigration+diaries

 429.imigration+photographs
 430.immigration+
 431.immigration+diaries
 432.impeachment
 433.indian
 434.indian
 435.indian
 436.indian
 437.Indian
 438.indian
 439.indian
 440.indian
 441.Indian+House+Door
 442.INDIAN+MOUND
 443.Indians
 444.indians
 445.Indonesian
 446.industrial+models
 447.infomine
 448.INFOMINE&type=unknown
 449.information+
 450.injection
 451.inoculation
 452.inquisition+
 453.insurance
 454.insurance
 455.insustrial+models
 456.international+pewter
 457.Internet
 458.interstate+compacts
 459.Interstate+Water+Compacts
 460.Interstates
 461.Iquique
 462.Iranian
 463.irish
 464.irish+american
 465.irish+country+people
 466.Irish+folk+tales
 467.iron
 468.iron+forge
 469.israel
 470.Israel
 471.israel
 472.italy
 473.j.+b.+priestley
 474.jabotinsky
 475.jackson
 476.jackson+davis
 477.jackson+davis
 478.jacques+louis+david
 479.jameskojack
 480.japan
 481.Japan&type=moving+image
 482.Japanese+art&type=image&type=physical+
 object

483. Japanese+art&type=moving+image%2C+physical+object

484. Japanese+art&type=moving+image&type=physical+object

485. Japanese+art&type=physical+object

486. jerusalem

487. jew

488. jewish

489. jews

490. Jews

491. john+brow

492. john+brown+invoice

493. john+cage

494. K-12

495. Kansas

496. Karachi

497. kendall+thomas

498. kendall+Thomas

499. kennywood

500. kentucky

501. keystone&type=image

502. king+county+snapshots

503. king+philip

504. king+Philip

505. King+Philipe+Augustus

506. King+Philipe+II

507. King+Phillip+II

508. klan

509. klimt

510. kmoddl

511. knowledge+wins

512. labor

513. labor

514. laboratory&type=image

515. ladies+garmet+workers+of+1900

516. lake+st+clair

517. Lakota

518. land+development

519. Landscape

520. landscape&type=image

521. landscape&type=image

522. Landscape+prints

523. learning+standards

524. lesson+plans

525. letters+from+19th+century

526. lewis

527. librarian

528. librarians

529. librarianship

530. librarianship

531. librarianship

532. libraries

533. library

534. library

535. library%2Bmoorhead

536. Lincoln

537. lincoln+blood

538. linking+florida

539. list+of+cherokee+names+

540. list+of+cherokee+registry+names+

541. list+of+cherokee+registry+names

542. liver

543. liver+disease

544. logging

545. los+angeles

546. losier

547. love+letters+

548. Lowry

549. lozier

550. LU.+65

551. lyman

552. madison+county

553. maine+memory

554. making+results-based+state+government+work

555. mambi

556. man&type=image

557. manuel+fernandez+del+casillo

558. manuel+fernandez+del+Castillo

559. maps

560. maps&type=dataset

561. maps&type=image

562. maps&type=interactive+resource

563. maps&type=moving+image

564.maps&type=physical+object
 565.maps&type=sound
 566.maps&type=text
 567.maps&type=unknown
 568.maria+thomas
 569.marianas
 570.marin
 571.marin+county
 572.mark+twain
 573.mars+hill
 574.masonic+%2Bmanuscripts
 575.massachusetts
 576.massachusetts+arms+invoice
 577.matsusaburo+Hibi
 578.mccaskey
 579.mchale
 580.meadow+brook+hall
 581.medieval+quest
 582.mesta+machine+co
 583.method+of+dating
 584.metis
 585.michael+Collins
 586.middle+east
 587.Migrant+workers
 588.milgrim
 589.milgrims
 590.mind+mode%3Bs
 591.mind+models
 592.mines
 593.mining
 594.mining
 595.mining+stocks
 596.minnesota
 597.mint
 598.MISANTHROPE
 599.mizltplec
 600.moac
 601.monsen
 602.moon&type=image
 603.moorhead
 604.motor+city
 605.motorcycle
 606.motorcycle
 607.MP3
 608.msp01047
 609.msp01047
 610.msp01047
 611.Mulholland+highway
 612.museum
 613.museum+Illinois
 614.music
 615.music+boxes%22&type=image%2C+physical+object%2C+sound%2C+text
 616.music+boxes&type=image%2C+physical+object%2C+sound%2C+text
 617.music+boxes&type=image&type=physical+object&type=sound&type=text
 618.music+therapy
 619.musique+concrete
 620.mussels
 621.Mystic
 622.naismith
 623.naismith
 624.Nakajima
 625.narraguagus+river
 626.Native+American
 627.Native+American
 628.Native+American
 629.native+american+photos
 630.native+american+settlement
 631.native+american+settlement
 632.naturalization
 633.naturalization
 634.naturalization+lesson+plans
 635.ND-10043
 636.Nevada
 637.New+Deal+agencies
 638.new+jersey
 639.new+york
 640.new+york+city+skyline
 641.new+york+picture
 642.new+york+public+library
 643.newberry
 644.newspaper

645.newspapers
 646.nietzsche
 647.noank
 648.Norman+Rockwell
 649.Norman+Rockwell
 650.Noronic
 651.north+Carolina
 652.north+Carolina
 653.north+carolina+experience
 654.north+caroline+experience
 655.Oac
 656.oakland+california
 657.octopus
 658.oklahoma
 659.olfaction
 660.olga+constantine
 661.oliver
 662.online+archive+of+California
 663.oteiza
 664.otto+perry
 665.Ottoman
 666.park+forest
 667.park+forest%2C+il
 668.pdf
 669.pee+wee
 670.Peep+into+the+Antifederal+Club
 671.pennsylvania
 672.penrose+correspondece
 673.penrose+correspondece
 674.penrose+correspondece
 675.perfum*
 676.perfum*
 677.perfum*&type=image&type=text
 678.perfume
 679.perfumer
 680.perfumes
 681.perfumes
 682.personal+correspondence
 683.philadelphia
 684.philadelphia&type=dataset
 685.philadelphia&type=interactive+resource
 686.philippines
 687.photographs
 688.Photographs
 689.photographs+of+river
 690.photographys
 691.photos+of+Matilda+Wilson+Dodge
 692.pictures+of+Enid+Baa
 693.pictures+of+Enid+M.+Baa
 694.pioneer
 695.Pit+River
 696.Pitt+River
 697.Pittsburgh
 698.Pittsburgh+And+Lake++Erie+Railroad
 699.pittsburgh+and+lake+erie+r.r.
 700.Pittsburgh+And+Lake+Erie+Railroad
 701.Pittsburgh+And+Lake+Erie+Railroad
 702.plain+Indians
 703.plains
 704.plains
 705.plant%2Blabel
 706.plant+images
 707.plate+no+28
 708.plate+no+29
 709.plate+no+29
 710.plate+no+39
 711.plate+no+39
 712.policy
 713.Polio
 714.population
 715.portfolios
 716.portraits
 717.poster
 718.Pratt%2C+Frances
 719.pre-Columbian
 720.prisoners
 re-entry%5C
 721.prisons
 722.prohibition
 723.propaganda+
 724.propaganda+techniques
 725.public+art
 726.public+art+bronx
 727.puck

728.quilt	769.scarves
729.quilts	770.Scavenger
730.R. Pullman	771.Scavengers
731.raffles	772.scent
732.raid+on+deerfield	773.scent
733.raid+on+deerfield	774.scorsasie&type=moving+image
734.raid+on+deerfield	775.sculpture
735.raid+on+Deerfield	776.Sheldon
736.raid+on+deerfield	777.Sheldon+House+Door
737.raid+on+deerfield	778.ship+images
738.raid+on+Deerfield	779.ships
739.railroad	780.Shoshone
740.real+estate+appraisal	781.singapore
741.red+sox	782.Sioux
742.re-entry	783.sioux+indian
743.registry+repair	784.sitting+bull
744.renewable+energy+	785.skyscraper
745.renewable+energy+sources	786.smell
746.rhinoceros	787.smell
747.richard+olderman	788.Smithsonian
748.Richmond++Housing+for+Shipyard+Worke rs	789.social+customs+
749.rights	790.social+security
750.riker-jaynes	791.social+work
751.riot	792.sonora
752.rivers+of+Guinea	793.south
753.rivers+of+Guinea	794.southeast+asia
754.Roads	795.soviet+union
755.rochester	796.spain
756.rubus	797.spanish+american
757.russia	798.spanish+american
758.rwanda	799.spectra&type=unknown
759.sailing+ships	800.Springer+link
760.Saint-denis	801.Springfield
761.Saint-denis+tombs	802.springfield+ymca
762.Saint-denis+tombs	803.standard+operating+procedure+for+laborato ries
763.sales+reciept	804.stanford+green+library
764.satellite	805.star+maps&type=image
765.savage+indian	806.stark+county (cd=2484)
766.savage+indian	807.stars&type=image
767.sayres	808.starvation
768.scalping	809.steel+works

810.stephen+king&type=text
 811.street
 812.streets
 813.stutler
 814.stutler+brown
 815.stutler+brown
 816.summer+drawings
 817.summer+drawings
 818.summer+landscape&type=image
 819.summer+landscape&type=image
 820.summons
 821.summons+to+comradship
 822.summons+to+comradship
 823.Sweets+Ballroom
 824.tain+bo+%22
 825.teacher+and+student+resources
 826.teaching+with+digital
 827.team+work
 828.teamwork
 829.teepee
 830.TELEVISION
 831.tepee
 832.test
 833.texas
 834.The+Great+Plains
 835.the+nazi+march+in+Skokie
 836.The+Tigers+claw
 837.the+uffizi+an+anthology
 838.the+wave
 839.Theresa+Cha
 840.thesis
 841.three+rivers
 842.tibbetts
 843.tobacco
 844.tobacco+currency
 845.tobacco+currency
 846.tom+sawyer
 847.tom+sawyer&type=moving+image
 848.tools
 849.top+religion+in+1930s
 850.Topaz
 851.topiary
 852.TRAIN&type=image
 853.training
 854.trains
 855.Transportation
 856.transportation
 857.Transportation
 858.transportation
 859.tranvias
 860.Turkish
 861.type=moving+image
 862.U.S.+History
 863.ukraine
 864.university
 865.university+collections
 866.university+of+California
 867.university+Wisconsin
 868.utah+newspaper
 869.vaccination
 870.van+gogh&type=text
 871.Van+Horn
 872.vanhorn
 873.varese
 874.victor+elford
 875.Vocational+Education
 876.voice+of+colorado
 877.volleyball
 878.von+tilzer
 879.voyager+spacecraft
 880.W.P.A.+PHOTO
 881.w.p.a.+puppets
 882.Wales
 883.walking+stick
 884.walking+stick
 885.Walter+Hawkins&type=text
 886.war
 887.war
 888.war
 889.war
 890.Washington&type=image
 891.washington+stae
 892.washington+state
 893.washington+state

894. Washington+township
895. Watkins
896. wayne+state%22
897. Welsh+language
898. western+1818
899. western+high+school
900. western+high+school
901. western+waters
902. westervelt
903. wgbh
904. whaling
905. whaling
906. white+train
907. Whitman
908. William+Letts+Oliver
909. windorpski
910. wisconsin
911. wisconsin
912. wisconsin
913. wisconsin
914. Women
915. women
916. women
917. women
918. women , 2530)
919. world+war
920. world+war+i
921. world+war+i
922. World+War+I
923. World+War+I
924. wpa+program
925. wrighting
926. wrighting
927. WW1+Posters
928. wwii
929. Wyandoch+Kansas
930. yearbook
931. ymca
932. YMCA
933. YMCA
934. YMCA
935. YMCA
936. YMCA
937. ymca
938. ymca
939. YMCA
940. yoko+ono
941. z39.50
942. zeppelin
943. Zionism
944. Zohaib+khan
945. zoo