Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

297

Subject Access in Swedish Discovery Services†

Koraljka Golub

Linnaeus University, School of Cultural Sciences, Faculty of Arts and Humanities,
Department of Library and Information Science, 351 95 Växjö, Sweden, <koraljka.golub@lnu.se>

Koraljka Golub is an associate professor in library and information science at Linnaeus University, Sweden. Her research interests focus on knowledge organization, primarily in the context of information retrieval. Research projects she has worked on have explored the potential of social tagging when enhanced by suggestions from controlled vocabularies, automatic subject indexing and evaluation of subject indexing in the context of retrieval. She would like to examine to what degree automatic full-text indexing, end-user tagging, author tagging, professional subject indexing, and automatic assigned indexing, or any combination thereof, contribute to successful retrieval.

Abstract: While support for subject searching has been traditionally advocated for in library catalogs, often in the form of a catalog objective to find everything that a library has on a certain topic, research has shown that subject access has not been satisfactory. Many existing online catalogs and discovery services do not seem to make good use of the intellectual effort invested into assigning controlled subject index terms and classes. For example, few support hierarchical browsing of classification schemes and other controlled vocabularies with hierarchical structures, few provide end-user-friendly options to choose a more specific concept to increase precision, a broader concept or related concepts to increase recall, to disambiguate homonyms, or to find which term is best used to name a concept. Optimum subject access in library catalogs and discovery services is analyzed from the perspective of earlier research as well as contemporary conceptual models and cataloguing codes. Eighteen proposed features of what this should entail in practice are drawn. In an exploratory qualitative study, the three most common discovery services used in Swedish academic libraries are analyzed against these features. In line with previous research, subject access in contemporary interfaces is demonstrated to less than optimal. This is in spite of the fact that individual collections have been indexed with controlled vocabularies and a significant number of controlled vocabularies have been mapped to each other and are available in interoperable standards. Strategic action is proposed to build research-informed (inter)national standards and guidelines.

1.0 Introduction

While support for subject searching has been traditionally advocated for in library catalogs, notably since Cutter's objectives for library catalogs (1876), research shows that subject access in online library catalogs has not been satisfactory. Developments and adoption of web-based discovery services (henceforth: discovery services) which serve as one-stop-for-all resources to which a library has access, try to match users' expectations by implementing Google-like single search box interfaces. However, it seems that efficient mechanisms such as ranking algorithms used by Google, or, exploitation of intellectual effort that has been invested into subject indexing and classification, are still missing from these services, leading to retrieval failures.

Based on an exploratory study, I aim to establish the picture of the current state of affairs related to subject access in Swedish discovery services (online library catalogs are not the specific focus here), in order to inform future developments. Based on previous research, a list of desirable features for subject access is drawn. The three most common discovery services used in twenty academic libraries of Sweden's largest universities are analyzed against these features.

The article is structured as follows. In the background section, a stage is set to provide context of what objectives regarding subject access contemporary catalogs and discovery services should meet; this includes an overview of related research. The next section provides a desirable list of functionalities for subject access. Section 4 describes the methodology and results of the exploratory study. In the conclusion, a summary of the results is given with implications for future research and development.

298

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

2.0 Background

2.1 Subject searching

Subject searching is a common type of searching in library catalogs (Hunter 1991; Villén-Rueda and De Moya-Anegón 2007) and discovery services (Meadow and Meadow 2012). However, in comparison to known-item searching (finding an information object whose title, author, etc., is known beforehand), searching by subject is much more challenging. This is due to difficulties in query formulation including lack of knowledge of the subject matter at hand and of information searching, ambiguities of the natural language and related problems. In order to alleviate these problems, library catalogs and related information retrieval systems (could) employ:

1) Interactive online help and instruction on information searching, in order to teach users about search strategies, search techniques and query formulation;
2) Hierarchical browsing of classification schemes and other controlled vocabularies with hierarchical structures, which help the user further her understanding of the information need and provide support to formulate the query more accurately; and,
3) Controlled subject terms from vocabularies such as subject headings systems, thesauri and classification systems, to help the user to choose, for example, a more specific concept to increase precision, a broader concept or related concepts to increase recall, to disambiguate homonyms, or to find which term is best used to name a concept.

2.2 Cataloging for subject access

Objectives of library catalogs in relation to subject access have been traditionally anchored in Cutter's "objects," as he called them, which are to: 1) enable finding an item of which the subject is known; 2) show what the library has on a given subject; and, 3) assist in the choice of a book as to its topical character (Cutter 1876, 5). These objects have been an integral part of cataloguing codes ever since and continue to be so in the contemporary *FRBR* (*Functional Requirements for Bibliographic Records*) family of conceptual models for catalog functionality. The *FRBR* family includes:

– *Functional Requirements for Bibliographic Records* (*FRBR*);
– *Functional Requirements for Authority Data* (*FRAD*); and,
– *Functional Requirements for Subject Authority Data* (*FRSAD*).

In 2017, these three models were consolidated into the IFLA *Library Reference Model* (IFLA 2017). The consolidated model prescribes five user tasks, which then need to be translated into cataloging rules to account for relationships between works, expressions, manifestations and items, as well as for relationships between topics and these corresponding works, expressions, manifestations and items. In the context of subject access, IFLA *LRM* and *FRSAD* (IFLA *FRSAR* 2011) tasks of finding, identifying, selecting, obtaining and exploring, could be applied as:

– Find: to find resources embodying works that are described by a given subject label, for example, search using a *nomen* that is used in a subject headings system or a classification scheme;
– Identify: to clearly understand the nature of the resources found and to distinguish between similar resources, e.g., those that are indexed by homonyms, or those with the same topic but from a different perspective (e.g., different branches of a classification system like virus from a zoological perspective versus virus from a medical perspective);
– Select: to determine the suitability of the resources found and to choose (by accepting or by rejecting) specific resources that seem the most relevant, e.g., due to certain aspects, facets or approach to the subject described;
– Obtain: to access the content of the resource;
– Explore: to use the subject relationships between one resource and another to place them in a context, e.g., to browse around related topics such as through using related terms in a thesaurus, or to see narrower and broader terms or classes, in order to understand the relationships between various *nomens* for an entity such as: examine the variant names for a subject within a controlled vocabulary, survey the variant terms used in different contexts of use, which may include different languages; explore correlations between *nomens* for the same entity in different controlled vocabularies, e.g., finding a thesaurus descriptor which corresponds to a classification number.

While previous cataloging codes, such as *AACR2* (*Anglo-American Cataloging Rules, 2nd ed.*) did not mention subject cataloging, the most recent cataloguing principles, Resource Description and Access (RDA) makes an effort to point out that subject representation or relationship to the subject of a work is needed: "The RDA element for the subject relationship generally reflects the relationship associated with the entity work as defined in FRSAD" (Kuhagen 2015, 3). Section 7 covers the relationships that are used to find works on a particular subject and Chapter 23 is given the title "General Guidelines on Recording Relationships Between Works and Subjects."

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

299

In spite of over 140 years passing since Cutter's objects were published, it has been said that the catalog has never lived up to his original ideal (see, e.g., Salaba and Zhang 2007). Furthermore, Cutter's objectives were not founded on an empirical ground of user search behavior (Borgman 1996). Today, although both the *FRBR* family of standards and *RDA* have put more emphasis on the end user, these aspects still remain insufficiently studied (Cossham 2013).

2.3 Subject access in online public access catalogs

In addition, many researchers have addressed problematic (subject) access to information in online catalogs, pointing to continuing challenges for end users (e.g., Casson, Fabbrizzi and Slavic 2011). An overview through a discussion of three generations of online library catalogs (a framework set by Hildreth 1984), is given by Barton and Mak (2012). Key points are briefly presented here. First generation online public library catalogs (OPACs) were developed with a focus on efficiency resulting from automation, rather than having service to end users in mind. Their functionalities were restricted to exact matching of known-item searches by author, title, or control number; effectively, this was a card catalog in the online form. Second-generation online catalogs supported post-coordinate subject searching using Boolean operators, which, while an improvement in terms of functionalities, proved counterintuitive and hard to use. Third-generation catalogs were developed as experimental systems, Okapi and Cheshire, and research concluded that the functionalities should include, among others, post-Boolean probabilistic searching, automatic spelling correction, term weighting, relevance feedback, output ranking and support for finding strategies. Markey (2007) provides ten reasons why these solutions were not applied to online library catalogs, among them: the failure of library system vendors to monitor shifts in information-retrieval technology and respond accordingly with system improvements; the failure of the research community to arrive at a consensus about the most pressing needs for online catalog system improvement; and, decreasing funding and at the same time the high cost of integrated library systems.

As a result, by the time the World Wide Web became prevalent, OPACs were still second-generation catalogs, and the demand to implement functionalities of global search engines such as Google and other commercial services like Amazon, was increasing. These included single search box, attractive web design, relevance ranking of results, recommendations, and access to a wide range of resources. However, Markey (2007) argued that the new direction of developments towards simplification would not attract users back to the online catalog. In integrated library catalogs, each search would result in "millions of hits with no guarantee that the top-ranked ones will address your desired topic in depth or at your level of understanding." Instead, she called for a redesign of an online library catalog that embraces:

1) post-Boolean probabilistic searching on full text;
2) subject cataloging, to help end users define the query, but also improve ranking algorithms by assigning high weights to subject headings, class numbers, as well as back-of-the-book indexes and entries from tables of contents; and,
3) "qualification cataloging," as she calls it, i.e., adding metadata like genre, purpose, reviews, academic level etc., which would allow end users to customize retrieval according to their level of understanding; such metadata could be in part contributed by end users through web 2.0 functionalities.

2.4 Web-based discovery services

To clarify terminology related to discussions so far, what Hildreth (1984) called a third-generation catalog is also known as the next generation catalog. In addition, because such a catalog may also include resources from outside the library like e-books, journal articles from commercial databases and pre-prints, it has been referred to as an integrated catalog or a web-scale discovery service. Discovery services, discovery layers, discovery interfaces and discovery tools are also common terms. In this article, terms third generation catalog, next generation catalog, integrated catalog and discovery service are used depending on the context of the author or topic discussed.

Discovery services today predominantly operate on one integrated index of metadata from all resources involved. A single index provides faster retrieval compared to distributed searching which compiles information from different databases on the fly (Barton and Mak 2012). In order for this one central index to operate well, contributing metadata elements and values need to be interoperable. While metadata are standardized for many uses today, when brought together, they have to be mapped to all other metadata standards used in the integrated index. Furthermore, values such as author names, place names and topics need to be mapped too. Lastly, metadata policies at different institutions need to be harmonized too; for example, large research libraries may have subject indexing policies aimed at a greater level of specificity and exhaustivity than do some more general collections for the use of the general public; the same holds for the choice of metadata elements—different collections may use a different subset of elements from the same metadata standard, or they may implement them with a certain level of difference.

300

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

Harmonizing this mix of metadata elements, their values and indexing policies across collections of resources would ensure that discovery services could fulfill established objectives of a library catalog, ensuring control over search (see above). Ellero (2013), in her analysis of forty-five studies of discovery services, concludes that they are "only as effective as the quality and completeness of the metadata they ingest, process, and index." Indeed, the most common issues regarding subject searching are those of inconsistent and incomplete metadata and blending of controlled vocabularies, free keywords and full-text automatic indexing (Dempsey 2012; Fagan 2011). Majors (2012) conducted a task-based usability test of five next-generation catalog interfaces and discovery tools, with undergraduates across all academic disciplines. Major findings related to subject access show the need to provide context of what has been searched and what is not included. Lee and Chung (2016) studied search effectiveness of discovery services, comparing web-scale discovery services against four individual databases in the fields of education and library and information science by EBSCO. Based on a small sample of queries and evaluators, it was concluded that the discovery service was less effective than individual databases.

Tarulli (2016) addresses problems of integrating metadata from sources beyond library catalogues and issues which arise from reliance on vendors. A key point emphasized is the need for transparency on how integrated indexes function, in particular when it comes to ranking and facet creation. Yang and Hoffman (2011) who surveyed academic libraries from 260 colleges and universities, showed that the circulation statistics was not part of the algorithm. If success of Google is attributed to ranking based on popularity, it is important for libraries to mimic good ranking, too, and not just the simple-search-box interface. Faceted navigation has become a standard feature in discovery tools and subjects seem often to be seen as one of the facets (Chickering and Yang 2014); however, studies point to confusion arising among end users and their lack of understanding of how facets work and the type of terms included in them (Emmanuel 2011; Osborne and Cox 2015).

Prerequisites for harmonization exist to a certain level: many cross-walks of metadata elements as well as controlled vocabularies are already available. Furthermore, a significant number of metadata standards and controlled vocabularies with their mappings have made it into linked data and the semantic web; see, for example, Library of Congress Linked Data Service, or FAST (Faceted Application of Subject Terminology) which links real-world entities to DBpedia, VIAF (Virtual International Authority File) and GeoNames.

Therefore, a question arises as to whether libraries place requirements on vendors of discovery services in order to preserve established objectives of library catalogs. When selecting a discovery system, Olson (2010) found that libraries often do not approach the decision-making process based on well laid-out arguments for needed features. Instead, reasons for a decision include saving money, facilitating a departmental reorganization, or improving the public perception of the library by implementing something new. One move towards standardization in order to bridge issues preventing unified search is the NISO Open Discovery Initiative (ODI) (National Information Standards Organization 2018; Walker 2015). ODI creates a technical recommendation and model for data exchange, which serves as a way for libraries as content providers to work with discovery service vendors. Apart from simplifying the data exchange, it ensures that the vendors follow fair and unbiased indexing and linking practices.

3.0 Desirable functionalities for subject access in discovery services

Based on research related to the first three generations of online library catalogs, an analysis of desired features with focus on subject access was conducted and discussed by Golub (2003) who provided a compiled list of features as a result of her study of WebPACs at the time. A number of these are also discussed in related research presented earlier as well as a number of others (see, e.g., Balíková 2011; Landry et al. 2011). Now aligned with user tasks related to subject access from the *FRBR* family of standards, and updated with findings on discovery services, the following is the proposed combined list of desirable functionalities of library catalogs and discovery services in relation to subject access:

1) Browsing by subject access points: subjects from controlled vocabularies, like subject headings, captions from classifications systems, free keywords.
2) Searching by subject access points from controlled vocabularies, including by individual words.
3) Browsing by facets, aspects and individual concepts from controlled vocabularies, such as individual terms from subject headings, as well as captions and notations representing individual concepts from synthesized classmarks (e.g, in Universal Decimal Classification).
4) Searching by any combination of individual concepts and facets (as above).
5) Searching by major and minor themes represented by controlled vocabularies, if supported by the indexing policy.

Knowl. Org. 45(2018)No.4

Golub, Koraljka. Subject Access in Swedish Discovery Services

301

6) Presenting and browsing excerpts of concept hierarchies (e.g., a classification scheme, a thesaurus), matching words and phrases from search terms, including for disambiguation, narrow, broader and related searching.
7) Auto-completing search terms once the user begins typing.
8) Auto-suggesting of authorized controlled versions of entered search terms, presenting all the relationships and allowing further choice on browsing or searching the controlled vocabularies.
9) Suggesting corrected versions of mistypes.
10) Searching by words from various metadata elements and full-text.
11) Combining controlled subject searching with searching by other bibliographic fields.
12) Highlighting search terms in retrieved metadata and resources.
13) Advanced searching by Boolean and proximity operators, truncation, wildcard.
14) Linking each subject access point to its resources.
15) Linking subject access points from one controlled vocabulary to corresponding concepts in others.
16) Adding, browsing and searching end user tags.
17) Combining previous search formulations.
18) Help on searching.

4.0 Subject access in Swedish discovery services

4.1 Methodology

An exploratory study of Swedish discovery services was conducted to determine the level to which they provide quality subject access. Since no detailed studies on the topic had been published earlier, this approach was chosen in order to identify major issues, which could then serve as a basis on which to provide research foci and inform the design of future in-depth studies. The analysis was conducted by accessing the discovery services and examining possible searching and browsing options and comparing them against the list of eighteen functionalities outlined above.

As seen from Table 1 below, in total, twenty of the biggest Swedish university libraries (counted by the number of full-time students at undergraduate and graduate levels) were examined as to which discovery service they use. Primo by ExLibris is used by ten libraries, EDS (EBSCO Discovery Services) is used by seven libraries, and Summon by ProQuest is used by three libraries.

Next, the library of the largest university using each of the three discovery services was compared against the list of eighteen functionalities, by running different queries and noting which characteristics are present, and to what degree. One complex, ambiguous topic was chosen as the main search query term, "Macedonia," because it can refer to: 1) the Republic of Macedonia, the country of the south-central Balkans; 2) FYROM (Former Yugoslav Republic of Macedonia), referring to the same Republic of Macedonia but under a different name due to its contested nature; 3) the region of Macedonia, today covering the Republic of Macedonia as well as parts of Greece and Bulgaria; and, 4) the ancient kingdom in the northeastern corner of the Greek peninsula. Provisions to disambiguate the term can easily be made by controlled vocabularies and help the searcher to define her query. Determining to what degree this well-recognized role for controlled vocabularies is used in today's most modern discovery services would help illuminate any challenges involved. As an exploratory study, the methodology is limited to the one search query. Further, the assessment is descriptive only and does not apply any other measures such as precision and recall. In all of them, guest access interface in English was chosen. The study was conducted in the period between 25 November and 10 December 2017.

	Primo by Ex Libris	EDS (EBSCO Discovery Services)	Summon by Proquest
1	Gothenburg University	Stockholm University	Uppsala University
2	Umeå University	Lund University	Chalmers University of Technology
3	KTH Royal Institute of Technology	Linköping University	Dalarna University
4	Örebro University	Malmö University	
5	Jönköping University	Luleå University of Technology	
6	Linnaeus University	Karlstad University	
7	Mälardalen University	University of Gävle	
8	Mid Sweden University		
9	University of Borås		
10	Södertörn University		
Total			

Table 1. An overview of discovery services used in twenty Swedish university libraries.

302

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

4.2 Results

4.2.1 Primo by ExLibris

The library of the largest university in the sample using Primo is the Gothenburg University Library (http://www.ub.gu.se). The home page offers a "superSearch" tab with an instruction that it searches "articles, e-books, and more." There is no further help stating which fields will be searched or similar. When using this Google-like simple-search box, by entering a simple search word, in this case "Macedonia," many results are retrieved—98,009 resources. Of facets offered to narrow down the result set, none of them are related to subject. Advanced search offers search by "subject," which retrieves 4,866 results for the same query. There does not seem to be any help file or instruction to clarify what this field search entails, which controlled vocabularies are used, whether they are mapped, and how to search on them for best results.

However, on the top of the interface with results, both of simple and advanced search, there is a "browse" option. This offers a search box into which a string must be entered before any browsing is offered. Once a query is entered, an alphabetical listing of subjects matching the query is given. Some seem to have the form of pre-coordinate subject headings, but no information is given in this regard. Clicking on a subject "mac" results in a list of two metadata records, each listing "mac" as one of its "subjects." When clicking on "mac" as "subjects" in any of these two metadata records, 85,128 results are retrieved. This demonstrates how these links are misleading. Also, it remains unclear how "subjects" in metadata records, in the "browse" option, and the "search" option relate to one another. Close to where the "browse" option is found, there is also a "tag" option, although this seems to be just recently implemented or hardly used, as it had, in total, nine instances of tags.

In all, this discovery service has implemented ten of the eighteen recommended features, albeit with restrictions and lack of clarity about what they entail:

1) Browsing by subject access points from controlled vocabularies, although it is not clear which ones, and how widely applied they are across all the resources; it is only alphabetical, not hierarchical.
2) Searching by subject access points from controlled vocabularies, although it is not clear which ones, and again how widely they are applied.
7) Auto-completing search terms once the user begins typing.
9) Suggesting corrected versions of mistypes.
10) Searching by words from various metadata elements and full-text.
12) Highlighting search terms in retrieved metadata and resources.
13) Advanced searching by Boolean and proximity operators, while it is not certain whether truncation and wildcard searching is supported as there is no help file at all.
14) Linking each subject access point to its resources, both via alphabetical browsing and from individual metadata records, although they lead to vastly different results.
16) Adding, browsing and searching end user tags (though less than a dozen tags in total).
17) Combining previous search formulations.

4.2.2 EBSCO Discovery Services (EDS)

The library of the largest university in the sample using EDS is Stockholm University Library (http://su.se/english/library/). The home page offers a tab to search for journal articles in the EDS discovery service. An image of a question mark leads to the help file describing the differences between the two tabs, with no other instruction on how to perform searches.

Using the simple search of the EDS tab on the "home" page, the query "Macedonia," retrieves 69,165 resources. The resulting interface has a search box with the original query, now showing that the search was conducted on Macedonia as a "keyword;" also "title" and "author" are possible to select. Of options to further clarify the meaning of the query, facet "subject" is provided. One can select a term from this facet as a search term by checking the box next to it.

As seen from the example in Figure 1, top retrieved facets still contain a very large number of items, and do not make it possible to specify further topical granularity within each of the subjects. Clicking on "show more" results in the total of fifty subjects, which can be ordered alphabetically or by the number of items. The top one by number of hits is also "Macedonia," this time with a smaller number of 2,265, and at the bottom is "political science" with 119 hits. These differences and the origin of the subjects are not explained in help or anywhere else.

When choosing advanced interface, "keyword" as a search field is no longer an option, but "subject" is. The difference is not explained anywhere. When entering the same search term there, 19,610 results are retrieved. In the "subject" facet, top facets are different than in the previous (Figure 2). Again, reasons for these differences are not clarified.

Once a chosen metadata record is opened, values of the element "subjects" are clickable, and lead to other records with the same subject. Looking at the top results, one example of "subjects" includes "history / Europe / general."
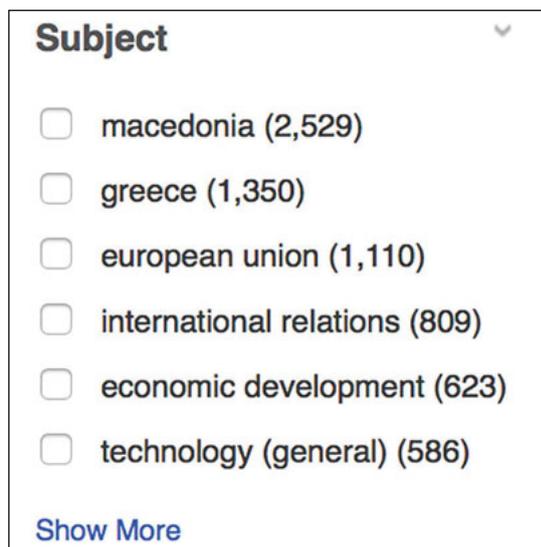
Knowl. Org. 45(2018)No.4

Golub, Koraljka. Subject Access in Swedish Discovery Services

303

Figure 1. Facet "subject" after searching "Macedonia" in the "keyword" field.

Figure 2. Facet "subject" after searching "Macedonia" in the "subject terms" field.

Figure 3. An extract from metadata record (above), transformed in the result set (below).

When clicking on it, other records with the same subject are retrieved. Automatically the search box contains a field name followed by the subject: "ZK 'history / Europe / general.'" The help file contains information on field codes, where it is stated that they are database specific. No list of codes and their usage is given for the interface of the discovery service.

There do not seem to be mappings between controlled vocabularies used. Some metadata records have "subjects" and "categories," without the difference explained anywhere, which are merged into "subjects" in the listing of results; for an example, see Figure 3.

This discovery service has implemented nine out of the eighteen recommended features:

2) Searching by subject access points from controlled vocabularies, including by individual words, although it is not clear which ones, and how universally or systematically applied they are across the resources.

7) Auto-completing search terms once the user begins typing.

9) Suggesting corrected versions of mistypes.

10) Searching by words from various metadata elements and full-text.

12) Highlighting search terms in retrieved metadata and resources.

13) Advanced searching by Boolean and proximity operators and truncation.

14) Linking each subject access point to its resources, although indirectly by having to run search on them, or by opening a metadata record and clicking on the subject there. However, they are not mapped across.

17) Combining previous search formulations.

18) Help on searching.

304

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

4.2.3 Summon by ProQuest

Uppsala University (http://ub.uu.se) is the largest in the sample with a library using Summon discovery service. The initial interface offers simple search box, with a default being search on "all," and options to delimit by title and by author. Directly there is a link to "advanced Search" and to "help," the latter being a brief sheet on the basics of searching.

Using default values, a search on "Macedonia" retrieves 85,474 results. An option to add results beyond the library collection results in the total of 442,408 items. Of facets most related to topical searching, there are two: "discipline," which offers five instances, ordered by the descending number of items per each; and "subject terms," ordered in the same way (Figure 4). Choosing one discipline in the former will reduce the number of results in the latter, probably restricting "subject terms" to categories found in the selected discipline.

Clicking on "more" in the "discipline" facet leads to an alphabetical listing of all disciplines, fifty-nine in our search on "Macedonia," many of which contain over 1,000 items. Clicking on "more" in the "subject terms" facet leads to an alphabetical listing of subject terms, 102 in our search on "Macedonia," many of which contain over 1,000 items. They also included genre, such as "article," "ebrar," "ebsco ebook academic collection," "electronic books" and "electronic books—local" (sic). The last three "subject" terms seem to be duplicates, pointing to the fact that no mappings have been conducted in the background.

When clicking on a result, the metadata record contains "subjects" although with no instruction anywhere on their origin or how to use them in searching. Clicking on a value found in "subjects" would result in other resources which have some post-coordinate combination of its words. For example, searching for an e-book with subject "Women—Macedonia" results in an automatic query that reads: "Subject terms: 'Women'AND subject terms: 'Macedonia.'" This retrieves thirty-one resources; opening one journal article shows that it has the subject "Women" while the word "Macedonia" does not exist in the metadata but does in the full-text of the article. So, "subject terms" seems to include automatic full-text indexing.

"Advanced search" lists the following subject related fields: "subject terms," "dewey," "call number" (in addition to "title," "abstract," "full text"). "Help" neither explains these further nor provides information on relationships between them or on existence of mappings between, e.g., "subject terms" and "dewey." Searching on "dewey" using class number for Macedonia, "(DEWEY:(949.76))," retrieves thirty-three results. This shows that mappings do not exist as there are more than thirty-three resources on Macedonia in the discovery service, as seen from previous queries. Searching on captions or "relative index" terms is not supported, as queries "(DEWEY:(Macedonia))" in English or "(DEWEY:(Makedonien))" in Swedish result in zero hits. After zero hits, an instruction is given to try resources outside the library by checking the box, but also zero results are retrieved as a result.

This discovery service has implemented seven of the eighteen recommended features:

2) Searching by subject access points from controlled vocabularies, including by individual words, although it is not clear which ones and how systematically they are applied across included resources.

7) Auto-completing search terms once the user begins typing.

DISCIPLINE	SUBJECT TERMS
history & archaeology (15,982)	history (8,569)
political science (10,922)	political science (8,004)
economics (9,710)	electronic books (7,144)
engineering (7,792)	article (6,438)
medicine (7,300)	europe (5,931)
More...	More...

Figure 4. Facets related to topics, resulting from a search on "Macedonia."

Knowl. Org. 45(2018)No.4

305

Golub, Koraljka. Subject Access in Swedish Discovery Services

10) Searching by words from various metadata elements and full-text.
12) Highlighting search terms in retrieved metadata and resources.
13) Advanced searching by Boolean and proximity operators and truncation.
14) Linking each subject access point to its resources, although indirectly by having to run search on them, or by opening a metadata record and clicking on the subject there, when they are being automatically post-coordinated into individual words from the clicked on phrase. Furthermore, they are not mapped across.
18) Help on searching.

4.2.4 Summary

The results imply that quality-controlled subject access in examined discovery services seems severely hindered. This is in spite of the fact that huge resources have been allocated to adding index terms from subject indexing systems to library catalog records. Little of this is adding value to existing interfaces. While imitating Google's black box approach, the task to retrieve relevant resources to a search query is addressed without making use of the existing index terms, relationships and structures of applied subject indexing languages.

As seen from Table 2 below, summarizing the features across the three systems, of the guidelines from the literature, only a small portion has been implemented. Those largely lacking are:

1) Browsing by subject access points from controlled vocabularies. For example, instead of generating facets randomly (at least seemingly so), they could be taken out from existing controlled vocabularies; or, even better, merged for the purposes of the discovery service at hand (such as UMLS, Unified Medical Language System). Also, entire hierarchical browsing structures could be made available, like those based on classification systems (see, e.g., Swedish union catalog LIBRIS, http://libris.kb.se/subjecttree.jsp).
2) Searching by subject access points from controlled vocabularies, including by individual words, whereby the user needs to know that controlled vocabularies or "subject" field values are applied to all the resources being searched, and consistently so, at the same level of specificity and exhaustivity.
3) Browsing by facets, aspects and individual concepts from controlled vocabularies, such as individual terms from subject headings, as well as captions and notations representing individual concepts from synthesized classmarks (e.g., in Universal Decimal Classification), again whereby the user needs to know that con-

Criteria	Primo (Gothenburg)	EDS (Stockholm)	Summon (Uppsala)	*Total*
1	√			*1*
2	√	√	√	*3*
3				*0*
4				*0*
5				*0*
6				*0*
7	√	√	√	*3*
8				*0*
9	√	√		*2*
10	√	√	√	*3*
11				*0*
12	√	√	√	*3*
13	√	√	√	*3*
14	√	√	√	*3*
15				*0*
16	√			*1*
17	√	√		*2*
18		√	√	*2*
Total	*10*	*9*	*7*	*26*

Table 2. An overview of 18 features in 3 discovery systems

306

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

trolled vocabularies are applied to all the resources in the discovery service.

4) Searching by any combination of individual concepts and facets (as above).

5) Searching by major and minor themes represented by controlled vocabularies, if supported by the indexing policy.

6) Presenting and browsing excerpts of concept hierarchies (e.g., a classification scheme, a thesaurus), matching words and phrases from search terms, including for disambiguation ("did you mean …"), and presenting narrower, broader and related concepts ("see also," but based on vocabulary control).

7) Auto-suggesting of authorized controlled versions of entered search terms, presenting all the relationships and allowing further choice on browsing or searching the controlled vocabularies.

8) Combining controlled subject searching with searching by other bibliographic fields, whereby the prerequisite is also that controlled vocabularies are applied to all the resources being searched on, and consistently so, at the same level of specificity and exhaustivity.

9) Linking each controlled subject access point to its resources.

10) Linking subject access points from one controlled vocabulary to corresponding concepts in others.

11) Adding, browsing and searching end user tags.

Terms like "subject," "keyword," and "category" are used, but it is not stated anywhere what kind of controlled vocabulary it is, if any, or what the differences are between them. The end user is not informed about the lack of mappings. This prevents truly integrated cross-searching in that resources on a certain subject from one controlled vocabulary that have been indexed using terms from another controlled vocabulary will not be retrieved in a query in which the searcher only uses terms from the first.

Furthermore, there is an obvious loss of the specificity and granularity that controlled vocabulary traditionally used by libraries has provided, for example, in subject headings. Unlike when we search on "Macedonia" in *Library of Congress Subject Headings* (*LCSH*) (Figure 5), no obvious disambiguation is immediately provided; neither are specific approaches or subtopics given, in contrast with the examples from Figure 5: "20th century," "biography," "administrative and political divisions," "maps."

5.0 Conclusion

This exploratory study confirms findings of related research, in which discovery services are criticized for the lack of transparency concerning the processes behind the

scenes, lack of mappings between metadata elements and values thereof and overwhelming numbers of results. The fact that results of test searches appear to be complex and confusing is in part due to merging of a number of resource collections, each using different indexing systems. This implies that providing widened search in loosely-controlled discovery services as opposed to traditional OPACs or individual databases of journal articles is not necessarily an advantage.

In terms of *LRM* and *FRSAD*, the potential of controlled vocabularies has not been utilized to address the following user tasks:

1) To find, as different resources are indexed using different controlled vocabularies, and also most probably following different indexing policies as they come from different collections of resources;

2) To identify, as homonyms are not disambiguated, different perspectives are not disambiguated, at least not systematically by taking advantage of controlled vocabularies;

3) To select, as aspects, facets or approach to the subject are not accounted for;

4) To explore, as it is not possible to, e.g., browse around related topics such as through using related terms in a thesaurus, or see narrower and broader terms or classes, in order to understand the relationships between various nomens for an entity; and, as it is not possible to explore correlations between nomens for the same entity in different controlled vocabularies, e.g., finding a thesaurus descriptor which corresponds to a classification number.

The topic of support for subject searching in contemporary discovery systems is particularly timely. Problems that have since long been addressed in the design of controlled vocabularies are not applied to the design of user interfaces in the examined discovery systems. As such, this work also provides guidelines for the design of relevant discovery systems which should make use of the intellectual effort and resources invested into creating controlled subject index terms and indexing languages.

The exploratory nature of the study warrants the need for an extensive study of real end-user behavior in order to find answers to the following research questions: 1) for which real end-user tasks discovery services do (not) work and why; 2) which (semi-)automated query reformulation mechanisms work best and why; and, 3) which elements of metadata records, or combination thereof, contribute to successful retrieval and which ones to failures; to name a few.

Since many collections have invested a lot of resources to assign index terms from subject headings and thesauri

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

307

Figure 5. LCSH example of disambiguation of the word "Macedonia," and levels of granularity. This example lists only the top alphabetically ordered subject headings.

and classes from classification schemes, and since mappings exist between many controlled vocabularies, the question arises why they are not utilized in discovery services. In addition to the US NISO Open Discovery Initiative, international cross-sector initiatives which would secure a sufficiently significant impact on the design of discovery services world-wide are warranted. ISO—International Standards Organization and IFLA Section on Subject Analysis and Access seem well placed to create guidelines in collaboration with a community of discovery services vendors. Also, national strategies for subject access are most probably a must in order to ensure implementation and sustainability of these efforts.

In addition, options which might also help alleviate issues of subject access, include social tagging and automated subject indexing. Further research is needed to determine the level to which it is possible to apply automated subject indexing in the library contexts, as well as to determine the value of those automatically assigned index terms, in combination and comparison with end-user assigned index terms as well as catalogers' assigned index terms in the process of information retrieval by end users. All these and the recommended functionalities for subject access, need to be studied in the context of actual end-user search behavior when it comes to their interaction with discovery services.

References

Balíková, Marie. 2011. "Focusing on User Needs: New Ways of Subject Access in Czechia," In *Subject Access: Preparing for the Future*, ed. Patrice Landry, Leda Bultrini, Edward T. O'Neill, and Sandra K. Roe. IFLA Series on Bibliographic Control 42. Berlin: De Gruyter Saur, 7-24.

308

Knowl. Org. 45(2018)No.4
Golub, Koraljka. Subject Access in Swedish Discovery Services

Barton, Joshua and Lucas Mak. 2012. "Old Hopes, New Possibilities: Next-Generation Catalogues and the Centralization of Access." *Library Trends* 61: 83-106.

Borgman, Christine L. 1996. "Why Are Online Catalogs Still Hard to Use?" *Journal of the American Society for Information Science* 47: 493-503.

Casson, Emanuela, Andrea Fabbrizzi, and Aida Slavic. 2011. "Subject Search in Italian OPACs: An Opportunity in Waiting?" In *Subject Access: Preparing for the Future*, ed. Patrice Landry, Leda Bultrini, Edward T. O'Neill, and Sandra K. Roe. Berlin: De Gruyter Saur, 37-50.

Chickering, F. William and Sharon Q. Yang. 2014. "Evaluation and Comparison of Discovery Tools: An Update." *Information Technology and Libraries* 33, no. 2: 5-30.

Cossham, Amanda F. 2013. "Bibliographic Records in an Online Environment." *Information Research* 18, no. 3: paper C42. http://www.informationr.net/ir/18-3/colis/paperC42.html

Cutter, Charles A. 1876. *Rules for a Printed Dictionary Catalogue*. Washington: Government Printing Office.

Dempsey, Lorcan. 2012. "Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention." *EDUCAUSE Review*. https://er.educause.edu/articles/2012/12/thirteen-ways-of-looking-at-libraries-discovery-and-the-catalog-scale-workflow-attention

Ellero, Nadine P. 2013. "Integration or Disintegration: Where Is Discovery Headed?" *Journal of Library Metadata* 13: 311-29.

Emanuel, Jennifer. 2011. "Usability of the VuFind Next-generation Online Catalog." *Information Technology and Libraries* 30, no. 1: 44-52. https://ejournals.bc.edu/ojs/index.php/ital/article/view/3044/2666

Fagan, Jody Condit. 2011. "Discovery Tools and Information Literacy." *Journal of Web Librarianship* 5: 171-8.

Golub, Koraljka. 2003. *Predmetno pretraživanje u knjižničnim katalozima s web-sučeljem*. Master's thesis, University of Zagreb. http://koraljka.info/publ/Magisterij-hrv.pdf

Hildreth, Charles. 1984. "Pursuing the Ideal: Generations of Online Catalogs." In *Online Catalogs, Online Reference, Converging Trends: Proceedings of a Library and Information Technology Association Preconference Institute, June 23-24, 1983, Los Angeles,* ed. Brian Aveney and Brett Butler. Library and Information Technology Series 2. Chicago: American Library Association, 31-56.

Hunter, Rhonda N. 1991. "Successes and Failures of Patrons Searching the Online Catalog at a Large Academic Library: A Transaction Log Analysis." *RQ* 30: 395-402.

IFLA Working Group on the Functional Requirements for Subject Authority Records (FRSAR). 2011. *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model*, ed. Marcia Lei Zeng, Maja Žumer, and Athena Salaba. IFLA Series on Bibliographic Control 43. Munich: De Gruyter Saur. http://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf. Errata for section 5.4.2, October 2011, http://www.ifla.org/files/assets/cataloguing/frsad/FRSADerrata2011.pdf

Kuhagen, Judith A. 2015. Subject Relationship Element in RDA Chapter 23. http://www.rda-jsc.org/archivedsite/docs/6JSC-ALA-31-rev-Sec-final.pdf

Landry, Patrice, Leda Bultrini, Edward T. O'Neill, and Sandra K. Roe, eds. 2011. *Subject Access: Preparing for the Future*. Berlin: De Gruyter Saur.

Lee, Boram and Eunkyung Chung. 2016. "An Analysis of Web-scale Discovery Services from the Perspective of User's Relevance Judgment." *Journal of Academic Librarianship* 42:529-34.

Majors, Rice. 2012. "Comparative User Experiences of Next-Generation Catalogue Interfaces." *Library Trends* 61, no. 1: 186-207.

National Information Standards Organization. 2018. "ODI: Open Discovery Initiative." http://www.niso.org/standards-committees/odi

Markey, Karen. 2007. "The Online Library Catalogue: Paradise Lost and Paradise Regained?" *D-Lib Magazine* 13, no. 1/2. http://www.dlib.org/dlib/january07/markey/01markey.html

Meadow, Kelly and James Meadow. 2012. "Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis." *College & Undergraduate Libraries* 19: 163-75.

Olson, Nasrine. 2010. *Taken for Granted: The Construction of Order in the Process of Library Management System Decision Making*. Skrifter från Valfrid 45. Borås: Valfrid.

Osborne, Hollie M. and Andrew Cox. 2015. "An Investigation into the Perceptions of Academic Librarians and Students towards Next-generation OPACs and Their Features." *Program: Electronic Library and Information Systems* 49: 23-45.

Riva, Pat, Patrick LeBoeuf, and Maja Žumer. 2017. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf

Salaba, Athena and Yin Zhang. 2007. "Functional Requirements for Bibliographic Records: From a Conceptual Model to Application and System Development." *Bulletin of the American Society for Information Science and Technology* 33, no. 6: 17–23.

Tarulli, Laurel. 2016. "Managing Outsourced Metadata in Discovery Systems." In *Managing Metadata in Web-scale Discovery Systems,* ed. Lousie F. Spiteri. London: Facet. 137-64.

Knowl. Org. 45(2018)No.4

Golub, Koraljka. Subject Access in Swedish Discovery Services

309

Walker, Jenny. 2015. "The NISO Open Discovery Initiative: Promoting Transparency in Discovery." *Insights* 28, no. 1: 85–90.

Villén-Rueda, Luis, Jose A. Senso, and Félix De Moya-Anegón. 2007. "The Use of OPAC in a Large Academic Library: A Transactional Log Analysis Study of Subject Searching." *Journal of Academic Librarianship* 33: 327-37.

Yang, Sharon Q. and Melissa A. Hofmann. 2011. "Next Generation or Current Generation? A Study of the OPACs of 260 Academic Libraries in the USA and Canada." *Library Hi Tech* 29: 266-300.