

# Discovery Systems as an Alternative to Stand-Alone Databases. The Example of Primo at BI Norwegian Business School

*Peer Reviewed*

**Emmanuelle Dabin\*<sup>1</sup>, Michael Preminger<sup>2</sup>**

*Unit – The Norwegian Directorate for ICT and Joint Services in Higher Education and Research<sup>1</sup>, OsloMet - Oslo Metropolitan University<sup>2</sup>*

## Abstract

---

Discovery systems (DS) harvest metadata from various sources into one central index. This data can be searched through thanks to an intuitive interface, which also redirects users to full-text resources in their native databases. This paper aims at evaluating whether the DS Primo can serve as an alternative to specialized databases subscribed to by BI Norwegian Business School. Various article searches were run in Primo and four databases BI subscribes to. 1200 records were exported to EndNote. The rank order and the source of the records in Primo were kept track of. Some individual records were later checked for metadata. Most times, the record describing an article in Primo was not harvested from the article's native database. When the record source was this native database, subject field's metadata was identical. Some articles appeared twice due to metadata inconsistencies across harvested resources. Almost all records included one of the subject headings searched for. Keywords were otherwise mostly found in the records' title. Downsizing being used in various disciplines, unexpected records were retrieved. Not all databases are indexed in Primo Central Index, but its size and coverage make Primo a smart choice as a one-stop search engine, if one uses the available narrowing options. It can to a certain extent function as an alternative to specialized databases for other tasks than systematic reviews, such as exploratory searches, or to get a sense of the available content.

**Keywords:** oria, primo ex libris, discovery systems, discovery services, information retrieval, interfaces, academic libraries

**\*Contact:**

Emmanuelle Dabin

BIBSYS, Norway

e-mail: [emmanuelle.dabin@bibsys.no](mailto:emmanuelle.dabin@bibsys.no)

## Introduction

Discovery systems (DS) centralize metadata records gathered from numerous content providers, describing billions of documents in a single index. As such, they offer an answer to research libraries' increasing proportion of electronic resources, subscribed to and accessed through external databases and repositories (Breeding, 2010, p. 7). They are also a better alternative to online catalogs that were mostly a gateway to printed collections (p. 6).

Users' Google-shaped behavior and expectations is another issue academic libraries have had to face (Sadeh, 2011, p. 6). Discovery service technology addresses it as the central index can be searched seamlessly at the end-user side through an intuitive interface. The very same interface returns a single list of results ranked according to relevance, and provides further access to the full-text article when possible (Vaughan, 2011, p. 6).

The background for this paper is a bachelor thesis in Library and Information Science at the University College of Oslo and Akershus. The thesis was conducted during the first half of 2016, in cooperation with BI Norwegian Business School (BI). The main focus of the study was BI's implementation of the discovery system Primo by Ex Libris, known in Norway as "Oria". The primary aim was to understand factors that have an impact on the DS' result lists and to explore the following research question: *to what extent can Primo be used as an alternative to stand-alone databases BI-library subscribes to.*

## BI's Search Tools as a Study Object

The BI-library is a member of a consortium that gathers about 90 Norwegian academic and research libraries around the provider of library services BIBSYS (BIBSYS, n.d.-b). In 2013, BIBSYS Consortium decided to implement the discovery service Primo, provided by the "ProQuest company" Ex Libris (Ex Libris, March 19th, 2013).

BI offers a wide range of stand-alone databases to their students and academic staff on topics of relevance to the institution's areas of research such as business and management (BI Handelshøyskolen, n.d.).

Four of the databases were used in this paper. They were chosen because they primarily include scholarly articles within management, with some exceptions such as ScienceDirect (SD). Developed by the content provider Elsevier, SD covers other subject areas, for example medicine, science and technology (Elsevier, n.d.).

The databases also have search features and interfaces comparable to Primo. However, two of them enable searching with controlled vocabulary. These databases are ABI/Inform (ABI) and the EBSCOhost database Business Source Complete (BSC), respectively produced by ProQuest (ProQuest, n.d.) and EBSCO Information Services (EBSCOhost, n.d.).

Emerald Management (EM) is the last database featured in the experiment and is provided by the global publisher Emerald. With around 300 journals in its portfolio (Emerald Group Publishing, n.d.), it differs significantly in scope compared to the thousands of journals contained by ABI, BSC, and SD (ProQuest, n.d.; EBSCOhost, n.d.; Elsevier, n.d.).

## Literature Review

### Characteristics of Discovery Systems

A discovery system is composed of a central index including pre-harvested content from multiple sources. Thanks to the front-end interface, the end user can simultaneously search through the pre-harvested content (Vaughan, 2011, p. 6). This technology differs from the “federated search” that sends queries to the native databases before returning the various result lists presented as one (Narayanan & Mukundan, 2013, p. 2). The central index mainly includes content provided by publishers and aggregators, although free scholarly materials are also included (Renaville, 2016, p. 6). The central index of some DS can also contain local resources, imported from the library’s collections. Primo, on the other hand, harvests library collections into a local index. Results from both this index and Primo Central Index (PCI) are blended together in the front-end interface (Vaughan, 2011, p. 40).

The end-user interface, or discovery layer, returns and displays results according to a relevance algorithm. In order for users to access the retrieved e-books, electronic articles, and in some cases only viewing the records, the library has to subscribe to this specific content through agreements with the different content providers (Hoeppner, 2012, p. 9).

### Processes Involved in the Constitution of Primo Central Index

Prior to any search, data is harvested into the central index from multiple sources, meaning the same document can be referred to by bibliographic records from various databases. The frequency with which data is harvested varies depending on content sources (Hoeppner, 2012, p. 9). During the normalization process, metadata is mapped and converted to an XML file called “Primo Normalized XML” (PNX) . The PNX is further organized into various sections, that each play a specific role. First, what is shown to the user in the interface depends on the metadata contained in the display section. Second, the record is retrieved based on keywords listed in the search section (Ex Libris, 2016b, pp. 5, 17).

Following the normalization, PNXs are matched against each other in the *de-duplication process*. If matching records are found, a merged record is created, based on one of the records. The new record, and only it, is then indexed and loaded into the central index and displayed in the interface (Ex Libris, 2016b, p. 99).

As a consequence, fewer records are left to go through the *FRBRization process*. In the *FRBRization-process*, records that are considered similar enough by the DS but previously were too different to be merged, are grouped together and assigned a “FRBR-group ID” based on common fields. “FRBR-group” here refers to a group of records in PCI identified as representing the same article, and not to the three groups of entities conceptualized in IFLA's Functional Requirements for Bibliographic Records (OCLC, n.d.). Only one record, representing the members of the group, is displayed in the user interface. This record is either one of the members in the group or a generic record gathering information elements from all records in the group (Ex Libris, 2016b, p. 118). All members are indexed and searchable (Ex Libris Knowledge Center, n.d.-c). In Primo, the origin of the record is displayed in the interface in the “Details” section. We will refer to it as “*record source*”.

## Agreements

For a metadata record to be shown in the DS, it presupposes that an agreement was brokered between the DS vendor and content providers. The level of indexing and the target group (authenticated users or not) depends on the nature of the agreement (Hoepfner, 2012, p. 9; Vaughan, 2011, p. 9). Some providers allow Ex Libris to show their metadata to non-subscribers (Hoepfner, 2012, p. 9). The interest of entering such an agreement for providers is that they can increase access and use of their content by library users (Breeding, 2014, p. 14).

Others are reluctant to share their metadata, especially when the content is enriched with controlled vocabulary. This additional metadata enhances searching capabilities within the provider's own product but is still of great value in other search environments such as discovery services. EBSCO for example is a database provider but also a DS vendor. Their metadata is not included in their competitors' central indices (p. 14).

Users can get hits on resources in EBSCO collections as long as the same resources are available in other providers' collections and indexed in PCI (Ex Libris, 2017b, p. 1). According to Ex Libris' coverage analysis, the EBSCO database BSC is covered up to 82 % through collections such as Academic One File and General One File, freely searchable in Primo regardless of subscriptions (pp. 1, 12). Both of these databases are produced by the provider of research resources Gale, part of the company Cengage Learning (Ex Libris, April 29th, 2010). This means that even though resources from BSC are deliverable through Primo, the metadata on display will always come from a third party. This illustrates how the "record source" presented in the "Details" section is all about metadata and has little to do with content delivery.

For any metadata from the central index to be viewable and searchable in Primo, it also requires that the "resource collection" it comes from has been activated. This is usually done by each institution, in the back-office administration interface of Primo (Ex Libris Knowledge Center, n.d.-b). For consortium member libraries, some resources might be centrally activated (Ex Libris Knowledge Center, n.d.-a). This is the case with the consortium BI is a member of: almost all resources that require subscription, such as Emerald and ScienceDirect, are centrally activated by BIBSYS on behalf of all member institutions. However, only metadata records referring to documents the library has access to will be shown by default in Primo (BIBSYS, n.d.-a).

## Metadata

Metadata originates from disparate sources using different standards and practices. This makes it particularly challenging to normalize content into a "uniform database" (Calarco, Conrad, Kessler, & Vandenburg, 2014, p. 535). One of the key challenges is the way author metadata is broken down and expressed by the different providers, for which a universal format has yet to be set (p. 536). In addition to a lack of universal standards, spelling errors, wrong dates or incorrect content types may be difficult to identify among millions of records and can have an impact on content discoverability and delivery (Calarco et al., 2014, p. 535). Metadata inconsistencies can have a negative impact on user experience if it causes the DS to display duplicate results (Hanneke & O'Brien, 2016, p. 111).

The variety in "depth" of metadata may also be an issue when it comes to interpreting the search results. While some records may only include a few metadata fields, others may comprise a more thorough abstract or subject headings (Narayanan & Mukundan, 2013, p. 4), or even the full text of a document (Breeding, 2014, p. 13). Records can be retrieved even

though query terms do not match the displayed metadata. This may confuse users and decrease their confidence in the search tool (Calarco et al., 2014, pp. 538-539).

### **Relevance Ranking**

Results in DS are by default sorted by the system according to relevance. The relevance is calculated based on how well a record matches the query terms according to a specific algorithm, with the most closely matching records appearing first (Narayanan & Mukundan, 2013, p. 5).

In Primo's algorithm, some of the factors taken into account are term frequency, currency, field weighting, and peer-review status (Vaughan, 2011, p. 40). Other vendors might put emphasis on other factors – for example subject heading – as each DS provider has developed their own relevance algorithm. This implies different results from one DS to another with a similar query (Narayanan & Mukundan, 2013, p. 4).

With these algorithms being proprietary, it can be difficult for client libraries to find out how and why some materials appear on top of the result list (Kelley, 2012, p. 39).

But institutions are also given the possibility by some DS vendors to influence the algorithm, or to boost some items in particular, making them appear higher on the result list. Libraries using Primo are for example able to define the weight of some record fields (Vaughan, 2011, p. 40).

### **Discovery Services vs. Stand-alone Databases**

DS make searching in more than one place optional by harvesting content from diverse resources. However, they are not meant to replace stand-alone databases. Sadeh (2011) argues that DS-technology is somewhat rigid compared to the variety of content types and resources it has to accommodate. As such, it cannot become “the ultimate search entry point for many users” (p. 15).

As mentioned by Vaughan (2011), metadata may be obtained from a specific search environment and adapted to a particular content type or discipline such as health databases. Consequently, the additional value provided by controlled vocabularies and other search features is lost in the DS. Furthermore, DS's central indices do not necessarily comprehend the entirety of the collections owned and subscribed to by a library (p. 9).

In addition, highly educated users, such as researchers with information needs within a specific discipline, are more likely to continue using native interfaces (Breeding, 2014, p. 13). Calarco et al. came to the same conclusion: being up to date about the latest literature on their area of expertise, faculty members are aware of what and if a key article is missing from the first page of results in a DS. As a consequence, they tend to opt out the library's DS and use specific databases instead (Calarco et al., p. 539).

Rather than a replacement to stand-alone databases, DS serve other purposes. With a simplified search environment, they are particularly appealing for novice users (Breeding, 2014, p. 13). This is especially true for searchers with low information literacy, who could be tempted to find literature on unreliable web pages. They can instead benefit from an intuitive interface that returns high quality and reliable scholarly content (Vaughan, 2011, p. 8).

Hanneke and O'Brien (2016) compared three DS with the health database PubMed/MEDLINE. They were surprised to find that all three DS retrieved relevant literature that was not found otherwise with a precision search in Medline. The queries they used in the DS were simple keywords that could potentially have been used by “inexpert users”. The authors concluded that DS were particularly effective for this kind of audience and for other purposes than systematic literature reviews (p. 115).

Newcomer (2011) stated that, due to their coverage of various fields and disciplines, DS are very suitable for interdisciplinary content that may not be indexed by subject databases. They are also great as a starting point for a new search, to gain “a general sense of the information available” (p. 143). By making this information easily discoverable, DS may also contribute to making content from stand-alone databases more visible and used to a greater extent (Vaughan, 2011, p. 8).

## Research Design

A title search, a subject search, and a simple phrase search was performed in Primo. The same searches were also run in four of BI's stand-alone databases for control purposes. Documents represented in both Primo and at least one of the four databases, regardless of where the Primo record was harvested from, are referred to as “*co-references*” in this paper. All three searches were narrowed down to articles in English in order to limit the results to potentially comparable items from one search tool to another.

## Search Queries

The searches were formulated with one or several study objects in mind. The degree of overlap between Primo and other bases was one of them and motivated the use of the title field, described as a “consistent method for making comparisons among databases” by Read and Smith (2000, p. 122).

Specific terms were also selected for the title and phrase searches:

- *downsizing* is a specific term in the context of management and means reducing the number of employees in a company (BusinessDictionary, n.d.-a).
- *talent management* is the practice of attracting, developing and retaining skilled employees (BusinessDictionary, n.d.-b).

The subject search was a Boolean query using the operator OR and gathering two specific formulations of one topic: “Human Resource Management” and “Personnel Management”. Both formulations were found in respectively ABI's and BSC's thesauri and searched for with the subject field. The purpose of this search was to observe Primo's treatment of the keyword-field data harvested from stand-alone databases. As it first failed to retrieve co-references, the search was limited to articles published after 2015.

The formulation of relatively varied search queries also aimed at observing which fields were preferably searched by the DS, with or without the use of specific search filters.

## **The Data Collection Process**

The three searches were run at BI between March 9th and March 16th, 2016 (Week 10 & 11). The first 200 results in Primo (50 in each of the four databases) were retrieved for analysis and exported to the reference management software Endnote. Using Endnote made it easier to search through and organize the references. A new EndNote library was created for each search, as well as a new folder for each set of results (one per database) in each library. Co-references across the different folders and duplicate records in Primo were identified and exported to Excel for further comparison and analysis. The records came in different orders in the various search tools. This rank order was kept track of and added manually to the Excel files, as well as the record source for each item in Primo.

The record source for the top 200 highest-ranked items in Primo was not checked the first time, which is why the three searches were performed again in Primo up to a week later (Week 12). This time, the title and phrase searches were conducted off campus, which slightly biased the results as some resources require authentication to be searchable and displayable (Hoeppner, 2012, p. 9). However, these off-campus searches gave a good picture of the variety of resources represented and are therefore included in the research material. Some individual records also had to be searched again later in the process. Even though PCI is updated weekly (Vaughan, 2011, p. 40), most records could be found and examined again.

It should be mentioned that the records retrieved in Primo came exclusively from PCI and not the local index. This means that BIBSYS' customizations, such as normalization rules, do not apply. Local libraries still have an impact on results returned by Primo. A meeting was organized with librarians Kristin Askildsen and Anita Bergsvenkerud from BI, to look at BI's Primo Back Office and the parameters that have a direct influence on the result lists. This meeting took place on June 14th, 2016. Unless otherwise stated, future references to information they provided are from that meeting.

## **Results and Data Analysis**

### **Co-References: Overlap and Database Coverage**

The first subject search in Primo retrieved no less than 76 131 hits. None of the first 200 records overlapped with the records of any of the four databases. The number of hits dropped to 1520 when the search was set to exclude material published before 2015. Yet only 13 co-references, documents represented in both Primo and at least one of the databases, were found. This is less than for the two other searches, that retrieved a more similar number of hits (see table 1).

Table 1.  
Number of hits and co-references retrieved for each search in Primo

Result lists in Primo	First Boolean search	Boolean 2015-16	Phrase search	Title search
Total number of hits	76 131	1 520	1 102	2 008
Number of co-references (out of 200)	0	13	76	70

Regardless of where the metadata record in Primo originated from, the co-references were not evenly distributed between the bases. The highest overlap of articles was between Primo and SD (see table 2). Table 2 also shows that the origin of the record in Primo was not exclusively the database Primo overlapped with on a given reference. As some articles may be found in several databases, the number of times this was observed was lower than the overall number of co-references, with one exception. For the phrase search, SD was indeed found to be the record source 36 times for only 35 co-references. The “missing” co-reference was here retrieved from one of the three other databases but is necessarily indexed in SD as well. This indicates that this specific record was given different values across the databases, and ranked in each search interface, including Primo, according to various relevance algorithms. The rank order was also checked for a significant number of co-references and tends to the same conclusion: no logic relations were found as to where co-references were placed in each result list.

Table 2.  
Overlap Between Primo and Each Database \*The number of co-references for which each database was also the record source is indicated in brackets. \*\* SD was the record source for a record retrieved in Primo but not found otherwise among the first 50 results in SD.

Number of co-references (*)	Subject search 2015-16	Phrase search	Title search
ABI	0	14	7
BSC	1	22	25
Emerald	3	18 (15)	7 (7)
SD	9 (8)	35 (36**)	29 (25)

None of the co-references or other records had BSC as a record source. This includes records that linked to full-text articles in the database itself. The reason for this is that BSC is not indexed in PCI and does not appear in the list of resources to be activated in Primo Back Office.

ProQuest’s ABI, however, *is* indexed in PCI (Ex Libris, 2017a, p. 29) and in contrast to other subscription databases has to be activated by the individual institution (Risan, October 9th, 2014). This also means the resource is easy to deactivate by the library in Primo Back Office. We have reasons to believe this was the case in March, since none of the co-references had ProQuest as a record source. This was further confirmed by the verification of all 200 records per search in Primo retrieved during Week 12. ProQuest records mentioned in this study were retrieved in June.

### Record Variety in Primo

The searches conducted during Week 12 of 2016 revealed a wide range of content providers listed as the record source (see table 3), with up to 26 different sources for the title search. This number could potentially have been even higher, had the search been performed on campus. Some of these sources were primary publishers, sometimes from the same publishing group.

But for each individual search, the proportion of records from Cengage Learning was the highest, exceeding the second highest record number by far. According to an internal document provided by Askildsen and Bergsvenkerud, Cengage resources are freely searchable in PCI, but not subscribed to by BI.

Table 3.

*Number of Unique Record Sources and Proportion of Cengage Records \*Off-campus search*

<b>Week 12 Primo Out of 200 records</b>	<b>Subject Search without Date Filter</b>	<b>Subject Search with Date Filter</b>	<b>Title Search*</b>	<b>Phrase Search*</b>
<b>Number of Unique Record Sources</b>	6	8	26	21
<b>Proportion of Cengage Records</b>	82 %	51 %	38 %	35 %

The title search in particular produced a heterogeneous result list. Special attention was given to the “Details”-section of each record in order to evaluate which topics were represented. Even though *downsizing* is a specific topic within management, 65 out of 200 records referred to subject areas such as chemistry, engineering or pharmacology. These disciplines are not taught or researched at BI. These observations are not based on the original search but give a

good picture of the variety of topics in the retrieved records.

### **Search Fields**

The subject search aimed at observing Primo's treatment of metadata from stand-alone bases. A quick search through the 200 references exported to EndNote revealed that respectively 145 and 1 contained "Human Resource Management" and "Personnel Management" in their keywords, while 2 references included other keywords than those searched for. The remaining references were exported without keywords.

For records that could be retrieved again in Primo at a later time, an explanation was found in the PNX:

- A significant proportion of apparently keyword-free records was checked. Among them, 46 had Web of Science as a record source. None of the records had keywords to show in Primo, but "Human Resource Management" was consistently marked up as "subject" in the search section of the PNX, which is why they were retrieved in the first place.
- One of the two references with keywords other than those searched for was retrieved for the same reason.
- Only one record was found that did not include the search terms listed as a subject in the search section and appeared far down on the result list (item number 154).

Being inconsistent with the other results, this one exception only relatively questions the efficiency of the subject field. It might have occurred in the result list due to the presence of "Human resource management" in both the title and description fields of the search section.

The title field is given the highest value by Primo's relevance algorithm (Ex Libris, 2016a, p. 365). Askildsen and Bergsvenkerud confirmed that BI had kept Primo's default boosting levels of the various fields. The phrase search, conducted without any specific search field, confirmed that Primo strongly favors records whose titles match the search terms. A search performed in EndNote showed that the title of all references but one contained "talent management" (or "talent-management"). "Talent management" was also found in the keywords of 99 references and the abstract or "description" of 116 items. These numbers are potentially higher as we have observed that data exported to EndNote does not necessarily reflect data in the search section of the PNX.

All references were retrieved based on one or several of these three fields. As a result, and because only 200 out of 1102 items were analyzed, we could not observe records that were retrieved based on the full-text metadata only.

### **Subject Metadata**

Several times, identical keywords were found in Primo and SD, or Primo and Emerald. This happened when the record source in Primo was this specific database and was observed with co-references from all three searches.

In some cases, however, the record in Primo had extra words or codes together with the subject terms such as "G32" and "G34". These were not categorized as keywords but as classification codes in SD. Classification codes were also used in another record in ABI. Contrary to the example from SD, the additional codes were not listed in the display section of

the PNX and not viewable in the end user interface in Primo. In both cases the different types of data were marked as “subject” in the search section of the PNX. The same applies to both controlled vocabulary from ABI and ordinary keywords from SD.

However, a few attempts were made to retrieve some items in Primo by using the subject terms they were indexed with in their database of origin:

- A known-item search conducted on June 20th retrieved a record originating from Nature Publishing Group. The same search was run again and combined with one of the subject terms used in ABI for the same article, “Knowledge Management”. It retrieved first a record from Cengage with the following PNX field: “<subject> Knowledge Management – Methods </subject>”. A final search with the use of the phrase function retrieved a record from ABI, as it was the only one of the three PNX records with the exact formulation “<subject> Knowledge Management </subject>”. All three PNX records had the same FRBR-group ID, which is why the one record retrieved each time changed depending on the search query.

### Metadata Inconsistencies

The title and phrase searches presented a certain number of duplicate records within the same set of results in Primo.

Table 4.

*Number of duplicate records in the three result lists in Primo\*\*Duplicate records often overlap with co-references, \*\*2 or more occurrences of the same article*

Out of 200 items in Primo	Boolean 2015-16	Phrase search	Title search
<b>Total number of duplicate records **</b>	0	14	24
<b>Number of articles represented 3 times</b>	0	1	3

Some of them had slightly different journal titles, author names, publication year, or page numbers visible directly in Primo. More frequently, it seemed like the records were not identified as representing the same article due to small differences in the title of the article. Titles in Cengage records were often complemented with information about the nature of the article, for example “(Research article)”.

Differences in the viewable section of the PNX cannot alone explain duplicate records. Duplicate records were analyzed based on the data exported to Endnote, which on several occasions did not match the content in the display section. This illustrates the issue of inconsistent metadata in DS, not only from one record to another, but also within the same record.

Inconsistencies are the reason duplicate records are not identified as representing the

same resource. This problem is especially hard to manage, as it does not necessarily take much for records to be considered as different. This happened for example with two identical-looking records from Cengage. An explanation as to why they were not “merged” was only found after comparing their PNXs: <https://www.diffchecker.com/ge2abpqo> (link last checked February 7<sup>th</sup>, 2018).

Even though both items had Cengage as a record source, they were harvested from different Gale databases, most probably Academic One File (“ofa”) and Health Reference Center Academic (“hrca”) (EBSCO Support, n.d.). Apart from distinct information and metadata as to where the records came from, the only difference between the two PNX files seemed to lie in the additional data section. This section “contains data elements that are required for a number of functions in Primo that cannot be extracted from other sections of the PNX” (Ex Libris, 2016b, p. 32). One of the title subfields included the additional “Research article” (<https://www.diffchecker.com/ge2abpqo>, line 93-94). Both records were assigned different FRBR-group ID (line 67-68), supposedly as a consequence.

The inconsistency in the way author names are expressed is also a problem. Contrary to other databases, records retrieved in Emerald and Emerald-originating records in Primo used the non-inverted form. An author search using the phrase function retrieved fewer hits than otherwise, leaving out occurrences of the other forms of writing. Refraining from using the phrase function may help recall records with both forms of writing but presents another problem: searching for the author “Doug Williamson” retrieved a record whose authors were Doug [Surname] and [First name] Williamson.

The same attempt using BSC’s metadata failed for reasons previously explained. The one reference with “Personnel Management” as a keyword had MEDLINE as a record source. This subject term was not even used in BSC, which the record in Primo happened to be linked to.

## Discussion

### To Be or Not to Be in PCI

The results presented above highlighted the variety of resources contained in PCI but also the variety of agreements making – or not – these resources discoverable. Cengage, which BI does not subscribe to, was found to be the record source for numerous records in Primo. Whether or not Cengage records were chosen over other records due to the higher quality of their metadata is subject to discussion. However, the frequency with which they occurred gives an indication of the total amount of Cengage metadata contained in PCI.

Their amount also illustrates the importance of agreements between Ex Libris and third-party content providers. In some cases, where metadata describing subscribed and deliverable content is not available, providers such as Cengage are important for the discoverability of a library’s resources.

EBSCO on the other hand, is one of the providers whose metadata is not available in Primo. The use of BSC’s “Personnel management” showed that there is little purpose in using the subject terms of non-indexed databases, unless the terms are used elsewhere. Having a general idea of the resource coverage of the DS is useful for librarians using Primo. The coverage should especially be kept in mind if a specific database’s metadata is desirable and

expected in a blended result list.

Not all metadata from databases who have entered agreements with Ex Libris are indexed in PCI. Several times, we came across examples where the representing record would change depending on the search query. However, this is only valid for “FRBRized” records, where all members of a FRBR-group are indexed in PCI and retrievable in Primo (Ex Libris Knowledge Center, n.d.-c). For records going through the de-duplication process where only the merged record is indexed (Ex Libris, 2016b, p. 99), the keywords in the left-out records cannot be used to recall the represented article.

Yet, the number of indexed resources and their variety is considerable, as suggested by the number of providers listed as record sources for the three searches and by the low degree of overlap between Primo and the four databases. This makes Primo particularly suited when trying to get an overview of which information is published or for exploratory searches.

### **Challenges**

The size and diversity of resources in PCI also raises some challenges. Any item may appear in Primo regardless of its actual relevance to the average user at a given institution, as long as it fulfills these two conditions: (1) being indexed in PCI and (2) occurring in a free repository or a database the library subscribes to. Therefore, the title-word downsizing retrieved items within disciplines that were not taught or researched at BI. BI’s core areas were otherwise reflected in each result list. As pointed out by Bergsvenkerud on June 14th, 2016, this emphasis is due to the coverage of the databases BI subscribes to, not to any customization by the library. This, and the fact that downsizing is a term used in multiple disciplines, explains the presence of “off-topic” articles in the result list in Primo.

Another challenge brought about by the variety of resources indexed in PCI relates to the quality and homogeneity of metadata. As demonstrated by the use across the control databases of different author-name forms, resources in PCI relate to various metadata standards. In some cases, Primo is able to identify records as representing the same article, as shown among others with the occurrence of records with the same FRBR-group ID. However, the presence of duplicates in each result list in Primo illustrates the difficulty in overcoming inconsistent or insufficient metadata.

We also found examples where metadata was inconsistent within the same PNX, which in some situations may be due to the agreement between Ex Libris and a particular provider. None of the examined items from Web of Science showed keywords in the display section - and the interface -, when retrieved based on the keywords listed in the PNX’s search section. This illustrates the structure and the role of each section in a PNX, but also the challenges met by users when faced with items retrieved for no apparent reason.

In addition, it is likely that the diversity of resources harvested into PCI makes it difficult for Ex Libris to accommodate all their specificities when converting them to PNX. The treatment of classification codes in some of the records we encountered, suggests a certain simplification in the semantic markup of the metadata. Primo did not support searching with controlled vocabulary when the data was collected. Examples involving metadata from ABI suggest that subject terms are treated regardless of their nature during the normalization process and assigned the same field type, here the subject field. Consequently, controlled subject

headings lose their value when converted and loaded into the DS, especially when blended with metadata from other resources. Comprehensive searches such as systematic reviews are a difficult task in a DS. Instead of Primo, specific databases should therefore be used for searches requiring a high level of precision.

### **On the Use of Search Filters and Future Research**

Some of the co-references contained the exact same subject metadata in both search tools. Some articles may therefore be found in Primo using the same keywords (including controlled vocabulary) as in their database of origin. This also suggests that the content and structure of the keyword field normally is kept intact when converted and exported into the PNX.

In addition, searching with Primo's subject field proved to be quite efficient as the vast majority of the items retrieved with the subject search were found based on their keyword metadata. Primo's relevance algorithm favors the title over the subject field, as confirmed by the outcome of the phrase search. Making use of this specific filter is therefore a good strategy, especially as the topic of an article is not always mentioned in the title. End users cannot be expected to have a deep understanding of how DS work but should be encouraged by librarians to adopt effective search behaviors. Any new knowledge about DS-mechanisms is therefore useful for librarians, who in turn can help students and staff optimize their use of the search system.

Whether other filters are as effective could be worth testing as well. The present experiment focused on records that were assigned the content type "article" by the DS. When checking individual items, we sometimes ran into other records representing the same article but defined as "text resource". A few quick searches were run with the author field. Depending on the writing, a search with the same author returned different sets of results. Both examples show that the way a document is described or a metadata element expressed can greatly impact recall in the DS. As a result, even though search fields and other filters are an effective way of narrowing a search, a DS user potentially misses out on a lot of relevant resources when applying one.

The way Primo handles known-item searching in general is also an important question. Breeding (as cited in Namei & Young, 2015) declared that this type of search was problematic when using common words, especially for one-word titles (p. 523). A new experiment designed to test known-item retrieval could help evaluating this particular aspect and adapting search strategies accordingly.

### **Conclusion**

The experiment described in this paper showed that Primo may be less adapted than stand-alone databases for high-precision searches where special search features or a discipline-specific coverage are an advantage. Primo is nonetheless a smart choice as a one-stop search engine, as it potentially gathers a great number of relevant documents on various topics in one place. Indeed, the size and coverage of PCI makes it a tool with great potential, provided one uses the search and narrowing features in the interface. Those examined for the purpose of the experiment showed the semantic structure of original records was respected quite well during the normalization process. Even though controlled subject terms from native databases are treated as ordinary keywords in the DS, Primo can (to a certain extent) retrieve results

equivalent to these databases' when using similar subject terms. As a consequence, Primo can be used as an alternative to stand-alone databases for tasks other than comprehensive literature searches, for example exploratory searches, or in order to get a sense of the available content.

### Disclaimer:

The background for this paper is a bachelor thesis in Library and Information Science at the Oslo and Akershus University College of Applied Sciences. At the time it was written, the first author did not work for BIBSYS, which delivers Primo to Norwegian research.

N.B.: From January 1<sup>st</sup>, 2018, BIBSYS, CERES and parts of UNINETT are gathered in a new administrative agency, ICT Services for Education and Research:

<http://www.bibsys.no/kunnskapsdepartementets-tjenesteorgan/>

Oslo and Akershus University College of Applied Sciences is now known as OsloMet - Oslo Metropolitan University: <https://www.hioa.no/eng/HiOA-is-now-known-as-OsloMet-Oslo-Metropolitan-University>

## References

- BI Handelshøyskolen. (n.d.). Find databases. Retrieved from <https://www.bi.no/sok/#%7B%22ActiveFacets%22%3A%5B%5D%2C%22SearchType%22%3A%22LibraryDatabase%22%7D>
- BIBSYS. (n.d.-a). Aktivering av søkekilder. Retrieved from <http://www.bibsys.no/oria-informasjon/aktivering-av-sokekilder/>
- BIBSYS. (n.d.-b). Biblioteksystemkonsortiet. Retrieved from <http://www.bibsys.no/produkter-tjenester/tjenester/bibsys-biblioteksystem/biblioteksystemkonsortiet/>
- Breeding, M. (2010). *Next-gen library catalogs* (Vol. 1). London: Facet.
- Breeding, M. (2014). Library resource discovery products: context, library perspectives and vendor positions. *Library Technology Reports*, 50(1), 1-32.
- BusinessDictionary. (n.d.-a). downsizing. Retrieved from <http://www.businessdictionary.com/definition/downsizing.html>
- BusinessDictionary. (n.d.-b). talent management. Retrieved from <http://www.businessdictionary.com/definition/talent-management.html>
- Calarco, P., Conrad, L., Kessler, R., & Vandenburg, M. (2014). *Metadata Challenges in Library Discovery Systems*. Paper presented at the Proceedings of the Charleston Library Conference.
- EBSCO Support. (n.d.). What is Gale Cengage Learning? Retrieved from [http://support.epnet.com/knowledge\\_base/detail.php?id=6930](http://support.epnet.com/knowledge_base/detail.php?id=6930)
- EBSCOhost. (n.d.). Business Source Complete. Retrieved from <https://www.ebscohost.com/academic/business-source-complete>
- Elsevier. (n.d.). ScienceDirect®. Retrieved from <https://www.elsevier.com/solutions/sciencedirect>
- Emerald Group Publishing. (n.d.). Emerald Journals. Retrieved from <http://www.emeraldgrouppublishing.com/products/journals/index.htm>
- Ex Libris. (2016a). Primo - Back Office Guide. Retrieved from [https://knowledge.exlibrisgroup.com/@api/deki/files/42350/Primo\\_Back\\_Office\\_Guide.pdf?revision=1](https://knowledge.exlibrisgroup.com/@api/deki/files/42350/Primo_Back_Office_Guide.pdf?revision=1)

- Ex Libris. (2016b). Primo - Technical Guide. Retrieved from [https://knowledge.exlibrisgroup.com/@api/deki/files/42351/Primo\\_Technical\\_Guide.pdf](https://knowledge.exlibrisgroup.com/@api/deki/files/42351/Primo_Technical_Guide.pdf)
- Ex Libris. (2017a). Primo Central Index Collection List. Retrieved from [https://knowledge.exlibrisgroup.com/Primo\\_Central/Product\\_Documentation/Primo\\_Central\\_Indexing](https://knowledge.exlibrisgroup.com/Primo_Central/Product_Documentation/Primo_Central_Indexing)
- Ex Libris. (2017b). Primo Central Index Collection List - Alternative Coverage Analysis. Retrieved from [https://knowledge.exlibrisgroup.com/@api/deki/files/53346/Primo\\_Central\\_Index\\_Collection\\_List\\_-\\_Alternative\\_Coverage\\_%5BApr\\_2017%5D.pdf](https://knowledge.exlibrisgroup.com/@api/deki/files/53346/Primo_Central_Index_Collection_List_-_Alternative_Coverage_%5BApr_2017%5D.pdf)
- Ex Libris. (April 29th, 2010). Gale and Ex Libris Partner to Offer Online Access to Authoritative Electronic Books and Journals via Primo Central. Retrieved from [http://www.exlibrisgroup.com/default.asp?catid=%7B45B7B8F9-C9B2-4C1F-AFCE-CCD62D4950D7%7D&details\\_type=1&itemid=%7B47F42449-10DA-444B-8B42-6E2C3D79F3F0%7D](http://www.exlibrisgroup.com/default.asp?catid=%7B45B7B8F9-C9B2-4C1F-AFCE-CCD62D4950D7%7D&details_type=1&itemid=%7B47F42449-10DA-444B-8B42-6E2C3D79F3F0%7D)
- Ex Libris. (March 19th, 2013). BIBSYS Consortium Selects Ex Libris Primo and SFX. Retrieved from [http://www.exlibrisgroup.com/default.asp?catid={920AEE41-B7C6-496D-A64F-A020BECC60FB}&details\\_type=1&itemid={13A5C6AC-010E-409B-A67C-4C3E50848C94}](http://www.exlibrisgroup.com/default.asp?catid={920AEE41-B7C6-496D-A64F-A020BECC60FB}&details_type=1&itemid={13A5C6AC-010E-409B-A67C-4C3E50848C94})
- Ex Libris Knowledge Center. (n.d.-a). Consortium Activation in Primo Central. Retrieved from [https://knowledge.exlibrisgroup.com/Primo\\_Central/Product\\_Documentation/PC\\_Index\\_Configuration\\_Guide/Primo\\_Central\\_Index\\_Registration/Consortium\\_Activation\\_in\\_Primo\\_Central](https://knowledge.exlibrisgroup.com/Primo_Central/Product_Documentation/PC_Index_Configuration_Guide/Primo_Central_Index_Registration/Consortium_Activation_in_Primo_Central)
- Ex Libris Knowledge Center. (n.d.-b). Discovery and Delivery of Collections in Primo. Retrieved from [https://knowledge.exlibrisgroup.com/Primo/Product\\_Documentation/How\\_To\\_Documents\\_and\\_Presentations/Discovery\\_and\\_Delivery\\_of\\_Collections\\_in\\_Primo](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/How_To_Documents_and_Presentations/Discovery_and_Delivery_of_Collections_in_Primo)
- Ex Libris Knowledge Center. (n.d.-c). Overview of the Publishing Process. Retrieved from [https://knowledge.exlibrisgroup.com/Primo/Product\\_Documentation/System\\_Administration\\_Guide/System\\_Architecture/Overview\\_of\\_the\\_Publishing\\_Process](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/System_Administration_Guide/System_Architecture/Overview_of_the_Publishing_Process)
- Hanneke, R., & O'Brien, K. K. (2016). Comparison of three web-scale discovery services for health sciences research. *Journal of the Medical Library Association : JMLA*, 104(2), 109. <https://org.doi/10.3163/1536-5050.104.2.004>
- Hoepfner, A. (2012). The ins and outs of evaluating web-scale discovery services: librarians around the world are trying to learn what WSD services are and how they work. *Computers in Libraries*, 32(3), 6.
- Kelley, M. (2012). Coming into focus: Web-scale discovery services face growing need for best practices. *Library Journal*, 137(17), 34-40.
- Namei, E., & Young, C. A. (2015). Measuring Our Relevancy: Comparing Results in a Web-Scale Discovery Tool, Google & Google Scholar *ACRL 2015 Proceedings – March 25–28, 2015, Portland, Oregon* (pp. 522-535).
- Narayanan, N., & Mukundan, R. (2013). *Cloud Web Scale Discovery services Landscape: An overview*. Paper presented at the International conference on Academic Libraries, New Delhi. [https://www.researchgate.net/profile/Nikesh\\_Narayanan/publication/281811595\\_Cloud\\_Web\\_scale\\_discovery\\_landscape\\_an\\_overview/links/55f92b3208aeba1d9f181615.pdf](https://www.researchgate.net/profile/Nikesh_Narayanan/publication/281811595_Cloud_Web_scale_discovery_landscape_an_overview/links/55f92b3208aeba1d9f181615.pdf)

- Newcomer, N. L. (2011). The Detail Behind Web-Scale: Selecting and Configuring Web-Scale Discovery Tools to Meet Music Information Retrieval Needs. *Music Reference Services Quarterly*, 14(3), 131-145. <https://org.doi/10.1080/10588167.2011.596098>
- OCLC. (n.d.). OCLC Research Activities and IFLA's Functional Requirements for Bibliographic Records. Retrieved from <http://www.oclc.org/research/activities/frbr.html>
- ProQuest. (n.d.). ABI/INFORM Collection. Retrieved from <http://www.proquest.com/documents/ABIINFORM-Collection-Brochure.html>
- Read, E. J., & Smith, R. C. (2000). Searching for library and information science literature: A comparison of coverage in three databases. *Library computing*, 19(1-2), 118-126.
- Renaville, F. (2016). Open Access and Discovery Tools: How do Primo Libraries Manage Green Open Access Collections? In K. Varnum (Ed.), *Exploring Discovery The Front Door to Your Library's Licensed and Digitized Content* (pp. 233-256). Chicago: ALA Editions. Advance online publication. Retrieved from <http://arxiv.org/abs/1509.04524>
- Risan, A. (October 9th, 2014). Aktivering av Proquest databaser for søk i Oria. Retrieved from <http://discovery.bibsys.no/2014/10/09/aktivering-av-proquest-databaser-for-sok-i-oria/>
- Sadeh, T. (2011). Discovery and management of scholarly materials: new-generation library systems. *ProInflow*(2), 4-22.
- Vaughan, J. (2011). *Web Scale Discovery Services : A Library Technology Report*. Chicago: ALA Editions.