

## Subject Cataloging by Norwegian Cataloging Agencies

Marit Kristine Ådland, Michael Preminger, David Massey, Karoline Kornkveen Hoff & Anne-Stine Ruud Husevåg

To cite this article: Marit Kristine Ådland, Michael Preminger, David Massey, Karoline Kornkveen Hoff & Anne-Stine Ruud Husevåg (29 Dec 2023): Subject Cataloging by Norwegian Cataloging Agencies, *Cataloging & Classification Quarterly*, DOI: [10.1080/01639374.2023.2294279](https://doi.org/10.1080/01639374.2023.2294279)

To link to this article: <https://doi.org/10.1080/01639374.2023.2294279>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 29 Dec 2023.



Submit your article to this journal [↗](#)



Article views: 30




View related articles [↗](#)



View Crossmark data [↗](#)

## Subject Cataloging by Norwegian Cataloging Agencies

Marit Kristine Ådland<sup>a</sup>, Michael Preminger<sup>a</sup> , David Massey<sup>a</sup>, Karoline Kornkveen Hoff<sup>b</sup> and Anne-Stine Ruud Husevåg<sup>a</sup>

<sup>a</sup>Oslo Metropolitan University, Oslo, Norway; <sup>b</sup>University of Oslo, Oslo, Norway

### ABSTRACT

This article reviews the practices of subject assignment by the two main Norwegian cataloging agencies serving the public library domain, *Biblioteksentralen* and *Bokbasen*, analyzing 47,235 records representing media cataloged by both agencies, published between 2012 and 2019. In addition to descriptive statistics representing these practices, we apply the Panofsky/Shatford model, previously used in the analysis of artworks and images, to distinguish aspects of these practices associated with levels of meaning. We find that Biblioteksentralen tends to use more abstract terms in their descriptions, while Bokbasen tends to use more general terms.

### ARTICLE HISTORY

Received August 2023  
Revised December 2023  
Accepted December 2023

### KEYWORDS

Subject cataloging; catalog indexing; the Panofsky/Shatford model; public libraries; school libraries; OPACs; books

## 1. Introduction

Currently (2023), Norwegian public libraries obtain bibliographic records predominantly from two sources: Bokbasen<sup>1</sup> (Den norske bokdatabasen) and Biblioteksentralen<sup>2</sup> (Bibbi-data). As a part of a research project,<sup>3</sup> we carried out a partial comparison of these two agencies, and during the study, we observed a difference in the way the agencies assigned subject terms to the records they prepare and distribute. The purpose of the current paper is to analyze the assignment of subject terms by the same agencies, and their respective vocabularies, as manifest in the bibliographic records.

The research questions are:

- How do the indexing practices and the underlying vocabularies of the agencies differ across domains and time?
- How do the subject terms align with the Panofsky/Shatford categories?

To answer the first question we used a quantitative method entailing calculating relative frequencies of subject terms in subdivisions of

**CONTACT** Michael Preminger  [michaelp@oslomet.no](mailto:michaelp@oslomet.no)  Oslo Metropolitan University, Oslo, Norway.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

record-pairs. To answer the second question we carried out a qualitative study using the Panofsky/Shatford categories.

To enable this analysis we have downloaded bibliographic records created by the agencies over an eight-year period (2012–2019). We compared the subject terms assigned to parallel publications, that is, publications that have been cataloged by both agencies, and identified by common ISBNs.

## 2. Theory and related work

### 2.1. Subject indexing

Subject indexing is the practice of describing literature with subject terms taken from controlled vocabularies.<sup>4</sup> Such vocabularies can have different forms: alphabetic-subject languages and classification languages.<sup>5</sup> In this paper, we study two alphabetic-subject languages: one thesaurus and one subject authority list where terms are combined according to a set of syntax rules.

Controlled vocabularies aid users in performing subject searches. They are often employed in situations where high recall is paramount.<sup>6</sup> Vocabularies that have been studied include the Library of Congress Subject Headings (LCSH),<sup>7</sup> the Australian Education Index (AEI),<sup>8</sup> and Medical Subject Headings (MeSH).<sup>9</sup> The automatic assignment of subject terms has also been the focus of research, most notably MeSH-terms.<sup>10</sup> A controlled subject vocabulary includes terms from three sources. Firstly, from the vocabulary in the literature, it is intended to describe. Secondly, from terms that real users (and librarians) use for searching. Finally, from terms that have a structural function, for example, to group a set of more specific terms. In the literature, these three sources of terms are referred to as *literary warrant*,<sup>11</sup> *use warrant*, and *structural warrant*,<sup>12</sup> respectively.

Our two vocabularies share similar literary and use warrants. But because their structures differ—one is a thesaurus and the other one a synthetic language—their structural warrant differs. In the subject vocabulary of Biblioteksentralen, compound subjects are precoordinated. The pre-coordinated subject headings are created according to Hjortsæter<sup>13</sup> and share similarities with the Sears List of Subject Headings, a controlled vocabulary with subject headings for small and medium sized libraries mainly in the USA.<sup>14</sup> Bokbasen assigns post-coordinated terms from their thesaurus when indexing documents. They also supplement the thesaurus with educational terms from the Udir<sup>15</sup> dictionary.<sup>16</sup>

When it comes to how terms are formulated, both vocabularies follow the same rules given in Hjortsæter.<sup>17</sup> Most subject terms are nouns or noun phrases. The terms should describe the subject of the document as a whole, neither broader nor narrower.

## 2.2. Categorizing subject terms

In this study, we will use the Panofsky/Shatford model to categorize subject terms. The model has been used for categorizing subject indexing of many visual collections.<sup>18</sup>

Panofsky identified three levels of meaning in Renaissance art: the pre-iconographical description, the iconographical analysis, and the iconological interpretation.<sup>19</sup> Panofsky's model, as interpreted by Markey,<sup>20</sup> Shatford,<sup>21</sup> and others, has been influential in the development of systems for subject access to images.<sup>22</sup> Shatford<sup>23</sup> extended and revised Panofsky's model. She categorized the subjects of pictures as *Generic of*, *Specific of*, and *Abstract*. Shatford also added four facets: *who*, *what*, *where*, and *when*. These correspond to Ranganathan's fundamental categories Personality, Matter, Energy, Time, and Space, although Shatford reduced Ranganathan's five categories to four.<sup>24</sup> This resulted in a 3 × 4 matrix for the classification of image descriptions (see Table 8).

The Panofsky/Shatford-model we use corresponds to categories of subject headings presented in the rules given by Hjortsæter,<sup>25</sup> where syntax rules are based on categories like units, actions, space, and time. Due to this correspondence, we believe that the model can be meaningful when categorizing subject terms primarily formulated to describe the aboutness of books. The inclusion of four facets makes the model interesting to apply to books, as both the facets of a thesaurus and the syntactic rules of a synthetic language use categories originating from Ranganathan's fundamental categories.<sup>26</sup>

The term "facet" is widely used when dealing with subject descriptions.<sup>27</sup> In our categorization, we use only four facets already identified in the Panofsky/Shatford-model. The distinction in the model between specifics, generics, and abstracts (levels of meaning) gives the model a potential to reveal additional differences between the two agencies' indexing practices and underlying vocabularies, and potential gaps in the subject access for Norwegian media in general.

## 3. The agencies and their datasets used in this research

### 3.1. Brief history of the agencies

Historically there has been no common subject vocabulary in Norway. Biblioteksentralen's subject headings list, used by the majority of Norwegian public libraries, has been a de-facto standard in public and school library catalogs.<sup>28</sup> This list has its origins in the late 1950s, and it was first published in 1963 and consists of pre-coordinated strings.

Biblioteksentralen is owned by municipalities and county municipalities in Norway. They offer books, metadata, and other services to libraries.

Bokbasen was established in its initial form in 1984 by Forlagsentralen.<sup>29</sup> In 2007, it was separated from Forlagssentralen as its own company, and is now owned by a number of Norway's leading publishing groups. Bokbasen provides metadata and digital services to virtually all Norwegian publishers, book retailers, and some libraries. In the 1980s, Bokbasen started to develop a hierarchical thesaurus with controlled subject terms, and its cataloging department maintained it.

Both agencies provide bibliographic records for practically all publications published in Norway. Terms from their controlled vocabularies are applied to these records.

Before 2016, each public library decided whether to purchase centrally cataloged records and from where they would purchase them. Most libraries used Biblioteksentralen as their record vendor, some used records from Bokbasen, and a minority did not purchase records at all. In 2016 the National Library, acting as a directorate under the Ministry of Culture, changed the distribution of bibliographic records in Norway,<sup>30</sup> and entered a cooperation with Bokbasen, for the purchase of centrally cataloged records of books published by Norwegian publishers.<sup>31</sup> However, Biblioteksentralen continued to deliver records as well, and many libraries continued to use them as a record supplier.

### **3.2. Datasets**

The project uses three datasets:

- Bibliographic records created by Biblioteksentralen and Bokbasen for the same publications published between 2012 and 2019 inclusive.
- Biblioteksentralen's vocabulary.
- Bokbasen's vocabulary.

#### **3.2.1. The bibliographic records**

The 2017–2019 records for both agencies were available online using REST services which allowed us to search for records more precisely and enabled a more exhaustive download of records for media published in a certain year. This is also the case for earlier Biblioteksentralen records, from the period 2012 to 2016, but not for Bokbasen records from this earlier period. Here we were granted access to the Bokbasen API, which does not have a similar search facility. This meant that all records changed or registered since January 1, 2012, had to be downloaded and then filtered for the

applicable publication years. This may have resulted in some missing records for this period.

For the entire period, 2012–2019, we have downloaded a total of 185,804 records, 79,717 from Bokbasen and 106,087 from Biblioteksentralen. We identified 51,075 parallel publications (matched by ISBN). Of these, 47,235 have assigned subject terms (MARC fields 600, 610, 611, 630, 640, 650, 651, 653, and 656) from at least one of the agencies, and thus comprise our subset as presented in [Table 1](#).

### 3.2.2. *The vocabularies*

Biblioteksentralen's list consists of pre-coordinated strings, whereas Bokbasen's thesaurus is hierarchic and contains five main categories: topic, form, genre, time, and place. In addition to Bokbasen's own vocabulary, the agency makes extensive use of a Norwegian-English dictionary of basic education maintained by *Udir* (Utdanningsdirektoratet—The Norwegian Directorate for Education and Training) for cataloging education-related textbooks. We do not study this dictionary as a vocabulary, but as it is a part of Bokbasen's indexing policy, we study the usage of Udir terms in the downloaded bibliographic records.

As vocabularies are used as a source for terminology for subject terms, we have obtained downloads of the vocabularies used by the agencies. Each of the vocabularies features both of the official Norwegian written languages, Bokmål and Nynorsk, of which we only regard the Bokmål part.

When it comes to fields that are normally assigned from name authority files (like personal names), both agencies had their own proprietary name-authority files before 2017. These have, after 2017, been used as the basis for contributing to the common authority file held by the national library.<sup>32</sup>

### 3.2.3. *Technical layout of the imported data*

The bibliographic records were modeled in a relational database structure that facilitates detailed scrutiny and comparison of records.



**Table 1.** Our subset, number of record pairs by the year.

Year	Record-pairs
2012	5,151
2013	6,110
2014	5,846
2015	5,977
2016	6,132
2017	6,271
2018	6,094
2019	5,654
Total	47,235

Both the Biblioteksentralen subject headings<sup>33</sup> and the Bokbasen thesaurus<sup>34</sup> were supplied to us modeled as RDF (Resource Description Framework) files conforming to the SKOS (Simple Knowledge Organization System) ontology and are available via the Skosmos system developed by the National Library of Finland.<sup>35</sup> The Udir dictionary is available for download as an XML-file. After download, the files were adapted to our database model and imported into our database for further use.

Biblioteksentralen's vocabulary consists of strings. When a subject heading includes subdivisions (terms), they are delimited in the string by a hyphen with a blank to each side (" - "). Sometimes a qualifier is appended at the end, to state a discipline of the subject. The qualifier is delimited by a colon with a blank to each side (" : "). An example is the string **Farlig gods—Norge—Transport: lov og rett** (*Dangerous goods—Norway: Transportation: Legislation*) (see Figure 1). To facilitate the analyses, the terms were extracted from the vocabulary and stored in the database, each term pointing to the string it is a part of (strings were also stored in the database as separate entities). Thus, we do not study the syntax or the strings, only their components (terms), such that each of these terms is compared separately. In this example, the member **Lov og rett** (*legislation*) (subfield \$0) is omitted from the comparison, as in Bokbasen's records it typically goes into the Genre denotation, which is not part of our analysis.<sup>36</sup>

The Bokbasen thesaurus is hierarchical, and complex subjects, such as Philosophy, have one or more subordinate levels (see Figures 2 and 3). We do not include the hierarchy as such in this study, but subordinate terms are modeled as see-references for the Panofsky/Shatford analysis (Section 5).

ANBEFALT TERM	Farlig gods - Norge - Transport : lov og rett 	
TYPE	bs:ComplexType Generelt emne	
DDKS-NR.	343.094	
WEBDEWEY-NR. (GODKJENT)	343.48109/4	
IDENTIFIKATOR	1144167	
PÅ ANDRE SPRÅK	Farlig gods - Noreg - Transport : lov og rett	norsk nynorsk
URI	<a href="https://id.bs.no/bibbi/1144167">https://id.bs.no/bibbi/1144167</a> 	
Last ned dette begrepet:	RDF/XML TURTLE JSON-LD	Opprettet 16.05.2017, sist endret 25.01.2023
NÆRLIGGENDE BEGREPER	343.481094 Veitransport--lov og rett--Norge, ...	

**Figure 1.** An example entry from the Biblioteksentralen vocabulary as displayed in the Skosmos interface (<https://vokabular.bs.no/bibbi/nb/page/1144167?clang=nb>).

The screenshot shows the Skosmos interface for the term 'Moderne filosofi'. On the left is a navigation menu with 'Alfabetisk' and 'Hierarki' tabs. The 'Hierarki' tab is active, showing a tree structure: Emne > Databehandling og informasjon > Filosofi > Erkjennelsesteori > Estetik > Etik > Filosofer > Filosofiske retninger > Antikk filosofi > Idehistorie > Moderne filosofi > Dialektisk materialisme > Idealisme > Kritisk filosofi > Lingvistisk filosofi > Motkulturer > Positivisme > Postmodernisme > Østlig filosofi. The main content area shows the breadcrumb trail: Emne > Filosofi > Filosofiske retninger > Moderne filosofi. Below this is a table with two columns: 'ANBEFALT TERM' and 'Moderne filosofi'. The table lists 'OVERORDNEDE BEGREPER' (Filosofiske retninger) and 'UNDERORDNEDE BEGREPER' (Dialektisk materialisme, Idealisme, Kritisk filosofi, Lingvistisk filosofi, Motkulturer, Positivisme, Postmodernisme). Below the table is a section 'PÅ ANDRE SPRÅK' with 'Moderne filosofi' and 'norsk nynorsk'. The 'URI' is 'http://ns.boknett.no/thesaurus/emne\_modernefilosofi' and 'Last ned dette begrepet:' is 'RDF/XML TURTLE JSON-LD'.

Figure 2. An example entry from the Bokbasen vocabulary as displayed in the Skosmos interface. The figure includes the hierarchy for the term.

```

<skos:Concept rdf:about="http://ns.boknett.no/thesaurus/emne_modernefilosofi">
  <skos:inScheme rdf:resource="http://ns.boknett.no/thesaurus"/>
  <skos:prefLabel xml:lang="nb">Moderne filosofi</skos:prefLabel>
  <skos:prefLabel xml:lang="nn">Moderne filosofi</skos:prefLabel>
  <skos:broader rdf:resource="http://ns.boknett.no/thesaurus/emne_filosofiskeretninger"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_dialektiskmaterialisme"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_idealisme"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_kritiskfilosofi"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_lingvistiskfilosofi"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_motkulturer"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_positivisme"/>
  <skos:narrower rdf:resource="http://ns.boknett.no/thesaurus/emne_postmodernisme"/>
</skos:Concept>

```

Figure 3. RDF/XML-version of the example in Figure 2.

### 3.2.4. See-references

Whereas the Biblioteksentralen records employ *see*-type reference fields explicitly (using field tag 950), the Bokbasen records lack these fields. The reason for this may be that Bokbasen terms are drawn from a thesaurus (see Section 3.2.2). Nearly half of the preferred terms in the Bokbasen thesaurus have alternative labels which are used as *see* references for the terms with which they are associated. One example is the term "Moderne filosofi" (modern philosophy, see Figures 2 and 3), which among its alternative labels has *Positivisme* (positivism) and *Postmodernisme* (post-modernism). Bokbasen seems to assume that subscribing libraries, having access to this thesaurus, can use the thesaurus for facilitating *see* references. For these reasons, terms from the *see* references were not used in the statistical occurrence analysis and comparisons, but we do include them in the Panofsky/Shatford analysis. To facilitate that, we artificially remodeled the Bokbasen bibliographic records, automatically introducing 950-field entries with *See* references (alternative labels) to each of any record's existing 650-tagged field (general subject term). This process sometimes resulted in records featuring tens of 950 entries.



**Table 2.** Number of subject terms taken from the bibliographic dataset along with unique vocabulary terms in use.

Vocabulary	# 6XX-fields	Unique terms used
Biblioteksentralen	73,913	8,606
Bokbasen	101,945	2,431
Udir	4,579	468

### 3.2.5. Statistics of vocabulary usage

Table 2 shows the vocabularies' (unique terms) usage in the subject terms fields in our bibliographic dataset. As indicated in Section 3.2.2, the Bokmål only versions are counted<sup>37</sup>

## 4. Statistical analysis of subject term occurrences in bibliographic dataset

In this section, we statistically describe occurrences of subject terms in our records. We start by comparing occurrences between the two agencies in the entire dataset and proceed to compare subdivisions of the material.

### 4.1. Types and principles of comparison

We analyzed occurrences of terms found in the bibliographic records as well as for subsets of those, based on:

- years of publication (chronological)
- domain of publication represented by the first digit of the Dewey classification code, i.e., main classes (where records from both vendors share these)

Two comparison principles were used:

- *term-wise*, aggregating terms across subsets of record-pairs for either agency into term-sets and comparing the sets.
- *record-wise*, aggregating and comparing the sets of occurring terms across records pairs, counting record pairs where term-sets are equal, where term-sets intersect, and where term-sets are disjoint (see examples in Table 3).

For the sake of these analyses, we extracted subfields \$a, \$x, and \$z from the subject fields (MARC fields 600, 610, 611, 630, 640, 650, 651, 653, and 656).<sup>38,39</sup> When it comes to fields like 600, 610, 611, and 630 that are mostly updated from authority files, the authority files of the agencies, though originally proprietary, have been converging in recent years, including post-editing of older records.<sup>40</sup> This means that we do not expect that name-forms will be different, and when including these fields in our analysis, we actually compare the agencies' interpretation of

**Table 3.** Examples of record-pairs with equal, intersecting, and disjoint terms.

Relation	Record-pair
Equal	ISBN: 9788253040882 Title: 1989 Bibsent.: 1980–1989, Verdenshistorie Bokbasen: 1980–1989, Verdenshistorie
Intersecting	ISBN: 9788202630966 Title: Antikkens globale verden Bibsent.: Historie, Oldtiden Bokbasen: Afrika, Antikken, Asia, Europa, Historie
Disjoint	ISBN: 9788243012486 Title: Atlas for nysgjerrige sjeler Bibsent.: Atlas, Kuriositeter Bokbasen: Geografi, Historie, Kultursosiologi, Samfunnsfag

the work as having (or not having) the named person, organization, etc. as a subject.

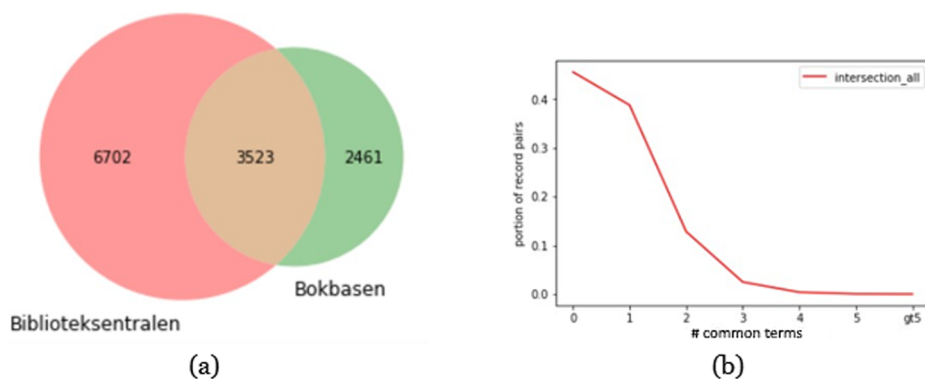
The agencies use different vocabularies, and while there are subject indexing rules for controlling permissible word-forms,<sup>41</sup> different forms (inclinations, prefixes, suffixes, etc.) of the same word do account for some of the differences.<sup>42</sup> Early thoughts about harmonizing word forms against Ordvev (the Norwegian version of the Wordnet lexical resource)<sup>43</sup> or applying lemmatization, were not pursued, because it was assumed that this would introduce its own noise into the analysis, offsetting any benefits. Moreover, in the analysis of subject terms using the Panofsky/Shatford categories (Section 5), we compare different grammatical forms of words and count them in different categories. Thus, a lemmatization would not benefit that analysis and the association between the analyses.

#### 4.2. Common and different terms in the entire set

In Figure 4(a), we show the intersection and differences of unique terms across all the records in our dataset. Figure 4(b) depicts *how many common terms* ( $x \in \{1, 2, 3, 4, 5, 6+\}$ ) are shared by different proportions ( $y \in [0.0, 1.0]$ ) of record pairs. We see that almost half of the record-pairs share no common terms whereas very few share four terms or more.

#### 4.3. Comparing subject term assignment over time

In Table 4, we show the intersection and differences of unique terms across all records-pairs belonging to each year since 2012. Looking at the percent columns to the right, there is a marked increase in the percentage of common subject terms after 2016. Table 5 lists the number and percentages of *record pairs* for which terms used are equal, intersecting, or disjoint. Also along this dimension, we see assignment practices coming closer. In Figure 5, we repeat the analysis of Figure 4(b) for subsets representing the year of publication, showing the rate of *record pairs* that share a number ( $x$ ) of identical subject terms.



**Figure 4.** The whole dataset. (a) Number of unique terms across sets (b) rate of parallel record pairs (y-axis) sharing  $n$  terms (x axis).

**Table 4.** Annual usage of unique terms across agencies.

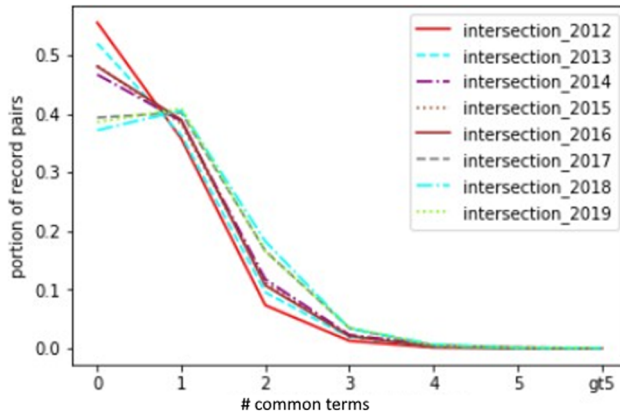
Year	Biblioteksentralen #	Intersection #	Bokbasen #	Biblioteksentralen %	Intersection %	Bokbasen %
2012	1,616	901	1,071	45.04%	25.11%	29.85%
2013	1,841	1,070	1,040	46.60%	27.08%	26.32%
2014	1,888	1,030	990	48.31%	26.36%	25.33%
2015	2,012	1,021	1,004	49.84%	25.29%	24.87%
2016	2,053	1,066	1,019	49.61%	25.76%	24.63%
2017	1,975	1,283	1,061	45.73%	29.71%	24.57%
2018	1,948	1,300	1,035	45.48%	30.35%	24.17%
2019	1,896	1,198	962	46.75%	29.54%	23.72%

**Table 5.** Comparing record pairs per year: How many record pairs (in a specific year) use entirely the same terms, how many intersect, and how many are disjoint?

Year	Equal	Intersecting	Disjoint	Total	Equal %	Intersecting %	Disjoint %	Total %
2012	152	2,138	2,861	5,151	2.95	41.51	55.54	100.00
2013	220	2,712	3,178	6,110	3.60	44.39	52.01	100.00
2014	368	2,749	2,729	5,846	6.29	47.02	46.68	100.00
2015	418	2,679	2,880	5,977	6.99	44.82	48.18	100.00
2016	421	2,765	2,946	6,132	6.87	45.09	48.04	100.00
2017	527	3,277	2,467	6,271	8.40	52.26	39.34	100.00
2018	539	3,290	2,265	6,094	8.84	53.99	37.17	100.00
2019	482	2,993	2,179	5,654	8.52	52.94	38.54	100.00

Numbers and percentages across domains.

For 2017–2019, we see a decrease in the rate of the record-pairs having no term in common ( $n = 0$ ), and a visible increase in the rate of pairs sharing two subject terms. Both analyses indicate a closer practice of subject assignment between the agencies toward the end of the time period.



**Figure 5.** Year-wise rates of record pairs (y-axis) sharing  $x$  terms (x axis).

**Table 6.** Usage of unique terms across the agencies by Dewey main classes.

Biblioteksentralen Bokbasen

Biblioteksentralen Bokbasen

Dewey class	#			%		
000–099	137	53	175	37.5%	14.5%	47.9%
100–199	244	159	227	38.7%	25.2%	36.0%
200–299	336	215	295	39.7%	25.4%	34.9%
300–399	1,561	864	950	46.3%	25.6%	28.1%
400–499	140	49	261	31.1%	10.9%	58.0%
500–599	203	127	222	36.8%	23.0%	40.2%
600–699	857	426	592	45.7%	22.7%	31.6%
700–799	804	616	590	40.0%	30.6%	29.4%
800–899	157	171	212	29.1%	31.7%	39.3%
900–999	867	655	494	43.0%	32.5%	24.5%

#### 4.4. Comparing subject term assignment across domains represented by Dewey main classes

Unlike years of publication, Dewey classes do not represent a linear development along an obvious dimension. Wishing to examine how the class of the book affects the assignment of subject terms, we counted occurrences of unique terms for either of the agencies in all records from the respective agency having the first digit of the main Dewey classification code of the record (Table 6). We also counted the usage of the terms across record pairs within those classification groups (Table 7).

For the 900–999 classes, History and Geography, the share of disjoint record pairs is relatively small, which can be explained by the extensive usage of geographical names. The share of common unique terms is also higher here, but not as markedly different as for the record-pair similarity.

**Table 7.** Comparing numbers and percentages of record pairs within class-code groups, for which term usage equals, intersects, or is disjoint.

Classes	Equal	Intersecting	Disjoint	Total	Equal %	Intersecting %	Disjoint %	Total %
000–099	47	87	227	361	13.0	24.10	62.88	100.00
100–199	115	300	718	1,133	10.1	26.48	63.37	100.00
200–299	52	337	626	1,015	5.12	33.20	61.67	100.00
300–399	422	2,825	2,362	5,609	7.52	50.37	42.11	100.00
400–499	29	267	307	603	4.81	44.28	50.91	100.00
500–599	236	217	335	788	29.9	27.54	42.51	100.00
600–699	532	944	1,687	3,163	16.8	29.85	53.34	100.00
700–799	183	1,348	847	2,378	7.70	56.69	35.62	100.00
800–899	5	319	303	627	0.80	50.88	48.33	100.00
900–999	55	2,460	371	2,886	1.91	85.24	12.86	100.00

This can be explained by the lack of lemmatization explained in [Section 4.1](#). Likewise, the high share of equal sets of terms for the books classified as natural sciences (500–599) may indicate that practices of assignment (selection from vocabulary) are more similar as the subjects of these books are more well-defined

In [Figures 6–8](#), we show, for different subsets of the material (not classified, classified, and classified 3XX,<sup>44</sup> respectively), occurrences/co-occurrences of unique main terms in the subsets ((a)-sub-figures), as well as the rate of the parallel records sharing one, two, three, etc. terms ((b)-sub-figures).

We have not fully analyzed the details here, but do see that there are interesting variations.

#### 4.5. Summary of data presentation

There are indications that the practices of subject assignment were more similar in 2017–2019 than they were in previous years, probably due to the change in the distribution of bibliographic records from the National Library of Norway. Their cooperation with Bokbasen from 2016, delivering data to potentially more public libraries from January 2017, appears to have changed their indexing practice. The cooperation demanded changes from Bokbasen. But it is also possible that Biblioteksentralen, risking a loss of customers, changed their records as well.

### 5. An analysis of subject terms using the Panofsky/Shatford categories

To compare subject term assignment by the two agencies, we categorized the subject terms of 490 randomly chosen nonfiction books published in 2019 into Panofsky/Shatford categories as described in [Section 2.2](#). We chose to analyze a sample of the most recently published nonfiction books in our dataset, to get an updated view of the indexing practice. With the selection of a single year, we also hoped to find records from a stable indexing practice not influenced by change of policy.

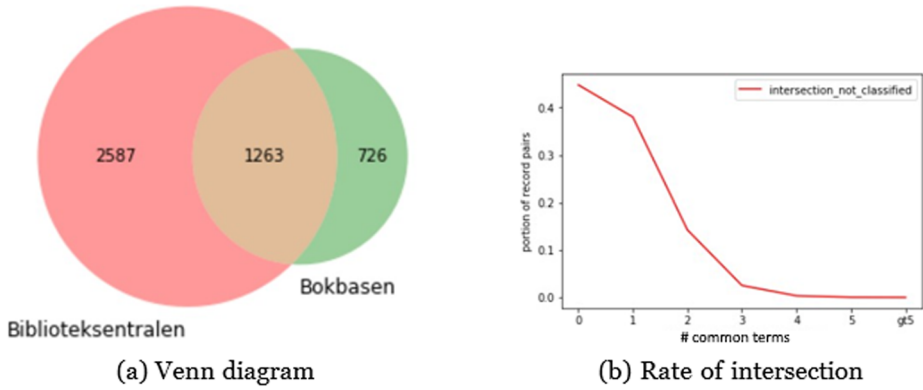


Figure 6. Unique terms and rates of intersection for the subset not classified.

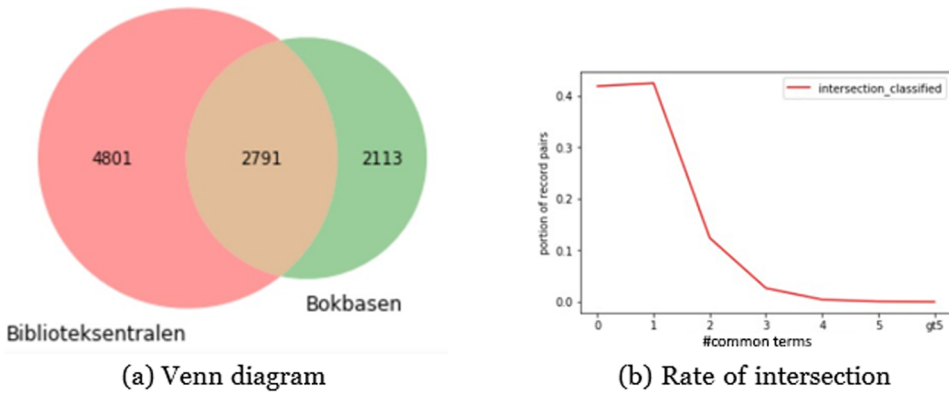


Figure 7. Unique terms and rates of intersection for the subset classified.

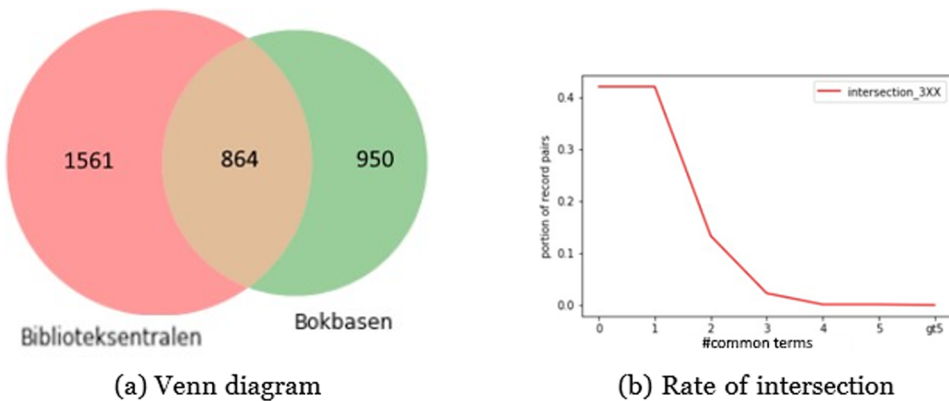


Figure 8. Unique terms and rates of intersection for the subset classified 300–399.

As our statistical analysis above indicates, the practices in 2019 were otherwise the most comparable.

Book			BokBasen															
ISBN	Title	Author	S6XX			Michae L_6	Mi/Ka _6	David_ 6	MK_6	S9XX			Michae L_9	Mi/Ka _9	David_ 9	MK_9		
97882055153 76	Dagbok	N. Jøe, Haddy Jatou	11	600	01	1966-												
			12	600	01	ND												
			20	650	01	Familie												
			40	650	01	Medier												
			60	650	01	Seksuelt trakassering												
			80	650	01	Politikere												
			100	650	01	2010-2019												
			120	651	01	Norge												

**Figure 9.** An example of annotating the Bokbasen terms assigned to one book. Two of the annotators assigned categories (representing the cells in Table 8) to each of the terms. The pink frames encircle the annotations, where, e.g. G1 corresponds to “Generics/Who.” The “Biblioteksentralen” section of the same book is hidden to save space.

Four researchers annotated the subject terms from our selected record pairs. The annotation was carried out in an Excel spreadsheet with columns for titles, authors, and terms, with separate columns for the annotations (see excerpt in Figure 9).

Bayerl et al.<sup>45</sup> provide an overview of the factors that influence inter-coder agreement in manual annotations of this nature. Subsequently, the following description is based on those factors and aims to elucidate the circumstances under which the terms were annotated. Our annotation process solely focused on subject terms, and the potential subject matters were extensive and could cover any topic discussed within a nonfiction book. All annotators were metadata experts who work with library metadata on a daily basis. However, none of us are experts in all possible subjects that could be discussed within the published books. The annotators are fluent in Norwegian, and all subject terms were written in Norwegian. The study employed four annotators, with one annotating 130 books, two annotating 250 books, and the remaining one annotating 270 books. Each book, or record pair, was annotated by two researchers. The annotators had an initial training period working with the Panofsky/Shatford categories and annotating a random sample of subject terms. Any divergent opinions were discussed, and a list of examples from the random sample of books was compiled to serve as a reference for the annotators when in doubt.

The annotation process involved twelve categories, with some categories geared toward visual culture objects that infrequently occurred in the material. Among the remaining categories, the selection process was challenging. The presence of more categories further complicates the process of achieving agreement between annotators.

**Table 8.** Examples of Panofsky/Shatford categories.<sup>a</sup>

	Iconography/specifics	Pre-iconography/generics	Iconology/abstracts
Who	Individually named person, group, thing, e.g. Northug, Petter	Kind of person or thing, e.g. Bakverk (Baked goods), Lesebøker (Books), Plast (Plastic)	Mythical or fictitious being, e.g. Nissar (Goblins), Olsenbanden
What	Individually named event, action, e.g. MeToo-kampanjen (The MeToo movement)	Kind of event, action, condition, e.g. Baking (Baking), Forureining (Pollution/polluting), Lesing (Reading)	Emotion or abstraction, e.g. Psykologi (Psychology), Historie (History)
Where	Individually named geographical location, e.g. Trondheim	Kind of place, e.g. Fjellomrøder (Mountain ranges), Elvar (Rivers)	Place symbolized, e.g. Fosterheimar (Foster homes)
When	Linear time: date or period, e.g. 2017	Cyclical time: season, time of day, e.g. Advent	Emotion, abstraction symbolized by time

<sup>a</sup>Adapted from Brian Stewart, "Pictures into Words," *Indexer* 33, no. 1 (2015): 8–25.

**Table 9.** Summary of category distributions by the agencies.<sup>a</sup>

	Biblioteksentralen		Bokbasen		Difference
	Subject terms	%	Subject terms	%	%-points
Specifics					
Who	212	9.75	256	6.08	<b>3.67</b>
What	75	3.45	131	3.11	0.34
Where	332	15.27	516	12.25	<b>3.02</b>
When	50	2.30	205	4.87	<b>-2.57</b>
Generics					
Who	511	23.51	1,267	30.09	<b>-6.58</b>
What	354	16.28	938	22.27	-5.99
Where	20	0.92	54	1.28	-0.36
When	6	0.28	9	0.21	0.06
Abstracts					
Who	9	0.41	8	0.19	0.22
What	599	27.55	825	19.59	<b>7.96</b>
Where	6	0.28	2	0.05	0.23
When	0	0.00	0	0.00	0.00
Sum	2,174	100.00	4,211	100.00	

Boldfaced differences indicate significance at  $p=0.05$  using the proportion Z statistic.

<sup>a</sup>Robert A. Hanneman, Augustine J. Kposowa, and Mark D. Riddle, *Basic Statistics for Social Research*, Research Methods for the Social Sciences (Jossey-Bass, 2012), 290–2.

We acknowledge that including more annotators may have increased the probability of inter-annotator disagreement. Additionally, an excessive number of annotators could have made it challenging to achieve agreement on categories.

### 5.1. Panofsky/Shatford categorization

In Table 8, we present the main categories as columns and the facets as rows, with examples of terms labeled by the sub-categories in the table cells.

A summary of the distribution of all Panofsky/Shatford categories by agencies can be found in Table 9. Tables 10 and 11 are summarizations of the categorization across broad categories and facets, respectively. We will concentrate our analysis on the categories where Biblioteksentralen and Bokbasen differ most.



**Table 10.** Distribution of broad categories by the agencies in our sample.<sup>a</sup>

	Biblioteksentralen		Bokbasen		Difference
	Subject terms	%	Subject terms	%	%-points
Specifics	669	30.77	1,108	26.31	<b>4.46</b>
Generics	891	40.98	2,268	53.86	<b>-12.87</b>
Abstracts	614	28.24	835	19.83	<b>8.41</b>
	2,174	100.00	4,211	100.00	

Boldfaced differences indicate significance at  $p=0.05$  using the proportion Z statistic.

<sup>a</sup>Hanneman, Kposowa, and Riddle, *Basic Statistics for Social Research*, 290–2.

**Table 11.** Distribution of broad facets by the agencies in our sample.<sup>a</sup>

	Biblioteksentralen		Bokbasen		Difference
	Subject terms	%	Subject terms	%	%-points
Who	732	33.67	1,531	36.36	<b>-2.69</b>
What	1028	47.29	1,894	44.98	2.31
Where	358	16.47	572	13.58	<b>2.88</b>
When	56	2.58	214	5.08	<b>-2.51</b>
	2,174	100.00	4,211	100.00	

Boldfaced differences indicate significance at  $p=0.05$  using the proportion Z statistic.

<sup>a</sup>Hanneman, Kposowa, and Riddle, *Basic Statistics for Social Research*, 290–2.

**Table 12.** Distribution of broad categories for biographies by the agencies in our sample.<sup>a</sup>

	Biblioteksentralen		Bokbasen		Difference
	Subject terms	%	Subject terms	%	%-points
Specifics	186	46.0	355	43.4	2.64
Generics	138	34.16	366	44.74	<b>-9.88</b>
Abstracts	80	19.80	97	11.86	<b>7.94</b>
	404	100.00	818	100.00	

Boldfaced differences indicate significance at  $p=0.05$  using the proportion Z statistic.

<sup>a</sup>Hanneman, Kposowa, and Riddle, *Basic Statistics for Social Research*, 290–2.

If we look at the broad categories, we find substantial differences between the agencies.

Bokbasen uses generic subject terms relatively more often than Biblioteksentralen (54 vs. 41%, see Table 10). Within Specifics and Abstracts categories, it is the opposite. Biblioteksentralen tends to use a higher percentage of subject terms compared to Bokbasen (Specifics: 26 vs. 31%, Abstracts: 20 vs. 28%, see Table 10).

When comparing the facets, the subject terms from Biblioteksentralen and Bokbasen are quite similar, all categories show differences smaller than three percentage points (see Table 11).

Biographies may be used to illustrate the differences between the agencies. Are they about the person only, or also about a subject? This depends on the specific book, but it can also be the result of the subject analysis. Out of the 490 books in our sample, 35 have metadata that indicate they are biographies. One example is the autobiography *Min historie* (My story), by and about cross-country skier Petter Northug. Biblioteksentralen uses only his name to describe the subject, while Bokbasen also uses the terms *Langrenn* (Cross-country skiing) and *Idrettsutøvere* (Athletes).

**Table 13.** Distribution of broad facets for biographies by the agencies in our sample.<sup>a</sup>

	Biblioteksentralen		Bokbasen		Difference
	Subject terms	%	Subject terms	%	%-points
Who	156	38.62	386	47.19	<b>-8.57</b>
What	157	38.86	238	29.10	<b>9.77</b>
Where	73	18.07	114	13.94	4.13
When	18	4.45	80	9.78	-5.32
	404	100.00	818	100.00	

Boldfaced differences indicate significance at  $p=0.05$  using the proportion Z statistic.

<sup>a</sup>Hanneman, Kposowa, and Riddle, *Basic Statistics for Social Research*, 290–2.

While we disagreed on whether *Langrenn* (Cross-country skiing) is a generic or abstract term in our categorization, *Idrettsutøvere* (athletes) is undoubtedly a generic term. Thus, this is one of the books where Bokbasen applied a generic term, while Biblioteksentralen did not.

Bokbasen has included subject terms that explain the role of the persons described in the biography, such as *Idrettsutøvere* (Athletes) in the previous example. This may be a useful subject term, but on the other hand, we may also see it as a violation of the rule that subject terms should only describe the specific subject of the book. *Min historie* is not about athletes in general, but about one specific athlete, named Petter Northug. Thus, according to the rule of specificity,<sup>46</sup> this term would be too broad.

In Table 12, we include a category distribution where only the works that are biographies are included. We can see that Bokbasen has a larger share of subject terms categorized as generic and abstract, compared to Biblioteksentralen. Biblioteksentralen also has applied more specific terms than generic, while Bokbasen has the opposite pattern: more generic terms than specific. This confirms our impression that the book *Min historie* (My story) is a typical example of how Bokbasen and Biblioteksentralen differ when it comes to biographies. The facet distribution for biographies (Table 13) resembles that for the whole material (Table 11), with larger differences for the Who and What facets.

Biblioteksentralen has a larger share of subject terms categorized as Abstract-What, Bokbasen has more subject terms categorized as Generics-What (see Table 9). These numbers are uncertain because categorizing Generics-What and Abstracts-What is difficult. On the other hand, all subject terms applied to one specific book, from both Biblioteksentralen and Bokbasen were always categorized by the same person. Thus, the distinction between Abstract-What and Generics-What for subject terms applied to the same book is considered by the annotators. All that said, Biblioteksentralen tends to use more abstract versions of words when assigning subject headings. The reasons for that may lie in the practices and traditions of the agencies, and this is something that might be further investigated qualitatively.

## 5.2. Specific subject terms

Biblioteksentralen tends to apply more subject terms categorized with Specifics-Who, Specifics-What, and Specifics-Where, Individually named persons, groups, things, events, actions, and geographical locations. Many of the Specifics-Who-terms are names of persons. We have not detected any difference when it comes to personal names. Most biographies have a personal name applied as a subject, from both vendors. For books that are not clear biographies, but include substantial biographical information, we find no systematic pattern: Sometimes one of them includes a personal name as a subject, sometimes the other does, and sometimes none or both. But all together Bokbasen applies a higher number of subjects to biographies compared to Biblioteksentralen, as they do with the other books as well.

Bokbasen rarely uses names of laws as subjects, even when a specific law is the topic of the book. Laws are also rare as related terms. Instead, Bokbasen uses words to describe what the law is about, like criminality or kindergartens. Biblioteksentralen uses the name of laws and thus does not always include words to describe what the law is about.

This is also the case for books about some other named entities, like *Grotten* (a state-owned residence lent out to merited artists for the remainder of their lives), *Apollo 11*, *Apex legends* (video game), or *Olsenbanden* (film).

## 5.3. Specifics-When

Bokbasen tends to have more terms that name specific time periods. They also have more standardized subject terms about time and use them regardless of the time period covered in the topic of a book. Examples are *1,500-tallet*, and *2000–2009*, which designate a century and a decade, respectively. Biblioteksentralen also has established time-periods as subject terms, but they are not as systematic. Thus it seems like time needs to be a more explicit part of the topic for Biblioteksentralen to apply a time-related subject heading.

## 5.4. Generic subject terms, Generics-Who and Generics-What

Bokbasen uses more Generics-Who and Generics-What-categorized subject terms compared to Biblioteksentralen. One reason can be their tendency to apply broader index terms. One example is the book *Informerte borgere?* (Informed citizens?). Here Biblioteksentralen applied one term: *Borgerdeltagelse* (citizen participation). Bokbasen applied three different terms: *Medier*, *Demokrati*, and *Sosiologi* (respectively Media, Democracy,

and Sociology). Together, these terms encircle the topic of the book but do not directly express the specific topic. Biblioteksentralen on the other hand, matches the term to the scope of the book. The differences in the number of Generics-Who-terms and Generics-What-terms here, are a result of Bokbasen's general tendency to apply more broad terms, rather than what categories the terms belong to. Another contribution to Bokbasen's higher number of Generics-Who and Generics-What terms, originates from Bokbasen's tendency to apply more terms to biographies.

### **5.5. Abstract subject terms, Abstracts-What**

Biblioteksentralen has more Abstracts-What-categorized subject terms than Bokbasen. We have so far not identified systematic differences between the agencies that account for such a large difference. It often seems like simply different wording, where Biblioteksentralen tends to end up with Abstract-What terms more often than Bokbasen. This corresponds to the fact that Bokbasen has more Generics-Who- and Generics-What-categorized terms. Many subjects can be named with words that are either Generics-Who (*bakverk/baked goods, sykkel/bicycle*), Generics-What (*baking/baking, sykling/biking*), or Abstracts-What (*bakerfag/bakery as a domain, sykkelfaget/bicycles as a trade*). In those cases, both Bokbasen and Biblioteksentralen use only one of the words, but we have not observed a systematic pattern for when either uses which word category. But altogether, Biblioteksentralen has a tendency to choose Abstracts-What-terms more often than Bokbasen.

For the remaining categories, such as Generics-Where, Generics-When, and Abstract-Where, there are only minor differences between the agencies when it comes to differences observable through our categorization.

### **5.6. Udir terms**

Bokbasen uses a combination of terms from their own thesaurus and *Udir* terms. This is mainly the case for books intended for use in schools. If we leave out Udir terms, the distribution of Panofsky/Shatford categories changes slightly. The changes affect three of the Panofsky/Shatford categories: Abstract-What, Generic-Who, and Generic-What all include Udir terms. This corresponds with Udir terms containing terms that name school subjects, like physics or Norwegian.

The Udir terms also raise questions about what can be a subject. Some of the terms that Bokbasen applies express the intended use of the book more than its aboutness. One example is the book *Kjemien stemmer* where Biblioteksentralen simply applied the term *Kjemi* (*Chemistry*). Bokbasen on the other hand, applied five terms: *Studiespesialisering*,<sup>47</sup> *Realfag vg3* (*Sciences for 3rd high school year*),

*Kjemi 2 (Chemistry 2)*, *VG3 (3rd high school year)*, and *Grunnbøker (basic level textbooks)*. None of the terms expresses the aboutness directly, instead, they all express aspects of the intended use of the book. However, the term *Kjemi 2 (chemistry 2)* includes the word *Kjemi (Chemistry)* that expresses aboutness, although the formulation strictly points to the level of chemistry knowledge you are supposed to achieve during your second year of reading chemistry. As a result, the aboutness of the book, chemistry, is searchable, but only indirectly expressed in the subject term.

Using the Udir terms, Bokbasen supposedly sees them as useful, especially for school libraries, and they probably are. But many of them do not express a book's aboutness. As there is no room for intended use or relation to discipline elsewhere in the record, Bokbasen has included those aspects as subject terms.

We do not know how Bokbasens' subject terms would be if they did not use the Udir terms at all. But the combination of the thesaurus and the Udir terms constitutes which terms Bokbasen's catalogers can use when they apply subject terms. Without Udir terms Bokbasen would probably apply fewer Abstract-What, Generic-Who, and Generic-What terms. But they could also have found a way to include such terms in their own thesaurus.

## 6. Discussion and conclusion

In the statistical comparison, we have found that records from Bokbasen and Biblioteksentralen were more similar after 2016. The two vendors have more subject terms in common during the years 2017–2019, compared to the years before. This corresponds to the change in policy by the National Library of Norway that happened in 2016. The imposed change in the distribution of bibliographic records appears to have had a harmonizing effect on the subject description practices of the two agencies (as prescribed by the tender mentioned in Preminger et al.<sup>48</sup>).

When examining the subject terms themselves, we found many similarities between the agencies. They more or less follow the Norwegian rules for subject term assignment. But they also have some practices that differ. Sometimes the agencies simply chose different words for their subject descriptions. These can be different synonyms, with similar meanings. It could also be because their subject analysis of the book differs slightly.

When looking at the Panofsky/Shatford categorization, some differences between the agencies are more interesting. Bokbasen sometimes applies more subject terms that we have categorized as generic, and Biblioteksentralen sometimes applies more abstract terms. One example is the book *Dybdeløring i naturfag*, where Biblioteksentralen uses the term *Undervisning*

(*Teaching*), while Bokbasen uses *Pedagogikk* (*Pedagogy*). We can see this in the number of terms categorized as abstract (Abstracts-Who) and generic (Generics-Who and Generics-What). But when looking at the books, it also seems that Biblioteksentralen's many abstract (Abstracts-Who) terms are a result of a tendency to choose the abstract version of a concept more often. Bokbasen's relatively more generic terms (Generics-Who and Generics-What) may also be a result of the same mechanism, where they choose the more concrete version of a concept more often. But our analysis also shows that Bokbasen quite often applies terms that violate the rule of applying the most specific term possible. This is visible in Bokbasen's relatively fewer number of terms categorized as specific (Specifics-Who, Specifics-What, Specifics-Where), but also within categories. One example of the latter is the book *Supertorsken*, where Biblioteksentralen has the term *Torsk* (*Codfish*) and Bokbasen the term *Fisk* (*Fish*), both categorized as Generics-Who.

We have stated that Biblioteksentralen and Bokbasen share a similar literary and use warrant, and we have observed many similarities. But some of the differences can be a result of differences in use warrant between the two agencies. Bokbasen's subject terms could be influenced by their slightly different view of the users of their data, where they have put emphasis on subject descriptions aimed at school libraries. Biblioteksentralen on the other hand, has a longer tradition as a vendor for public libraries.

When subject terms are too general, one can imagine consequences for precision and recall when searching. If users search for a specific topic, they may get zero hits even though there is a book about the topic in the collection. To find it, users must search with a slightly more general term. On the other hand, if users search with a more general term, they may find what they search for, and topics close to that. But if the collection is large, the hit list may be too long to look through. The usefulness of specific terms thus depends on how users behave and the size of the collection.

Before the advent of universal bibliographic control, every library would produce their own bibliographic records and decide what level of specificity was appropriate for each subject. If the number of documents within a certain subject was low, libraries would apply more general subject terms, thus helping users find what little they had. If the number of documents was high, they would apply more specific terms to help users find a reasonable number of hits. It seems Bokbasen has a practice that gives a similar result. We can see this as an indication of a collection warrant, or a literary warrant where the level of specificity is tuned according to the number of documents in the collection.

In this paper, we have identified several differences between subject vocabularies and their use. These changes are owed to differences in

vocabulary as well as differences in the practices and policies of the agencies. It would take a more qualitative research design to try and isolate the effects of any of these factors. Another path for further research is to compare the assignment of subject descriptions to subject searches taken from libraries' search logs.

## Acknowledgments

The authors wish to thank Biblioteksentralen and Bokbasen AS for help, support availability, and timely response to our enquiries during the writing of this paper.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## ORCID

Michael Preminger  <http://orcid.org/0000-0003-3524-4374>

## Notes

1. "Forside," Bokbasen, accessed December 6, 2023, <https://www.bokbasen.no/>.
2. "Metadata til bibliotek," Biblioteksentralen, accessed December 6, 2023, <https://www.bibsent.no/metadata/metadata-til-bibliotek>.
3. Michael Preminger et al., "The Public Library Metadata Landscape, the Case of Norway 2017–2018," *Cataloging & Classification Quarterly* 58, no. 2 (2020): 127–48.
4. F. Wilfrid Lancaster, *Indexing and Abstracting in Theory and Practice*, 3rd ed. (London: Facet Publishing, 2003).
5. Elaine Svenonius and William Y. Arms, *The Intellectual Foundation of Information Organization* (Cambridge, MA: MIT Press, 2009).
6. Jacques Savoy, "Bibliographic Database Access Using Free-text and Controlled Vocabulary: An Evaluation," *Information Processing & Management* 41, no. 4 (2005): 873–90, <https://doi.org/10.1016/j.ipm.2004.01.004>.
7. Tina Gross, Arlene G. Taylor, and Daniel N. Joudrey, "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching," *Cataloging & Classification Quarterly* 53, no. 1 (2015): 1–39; Karen M. Drabenstott, Schelle Simcox, and Marie Williams, "Do Librarians Understand the Subject Headings in Library Catalogs?," *Reference and User Services Quarterly* 38, no. 4 (1999): 369–87.
8. Philip Hider, "The Search Value Added by Professional Indexing to a Bibliographic Database," *Knowledge Organization* 45, no. 1 (2018): 23–32, <https://doi.org/10.5771/0943-7444-2018-1-23>; Philip Hider, Pru Mitchell, and Robert Parkes, "Measuring the Value of Professional Indexing," *Information Research – An International Electronic Journal* 24, no. 3 (September 2019), <https://informationr.net/ir/24-3/rails/rails1808.html>.
9. Ying-Hsang Liu and Nina Wacholder, "Evaluating the Impact of MeSH (Medical Subject Headings) Terms on Different Types of Searchers," *Information Processing & Management* 53, no. 4 (July 2017): 851–70, <https://doi.org/10.1016/j.ipm.2017.03.004>.

10. Dolf Trieschnigg et al., "MeSH Up: Effective MeSH Text Classification for Improved Document Retrieval," *Bioinformatics* 25, no. 11 (2009): 1412–8; Isabel Segura Bedmar, Paloma Martínez, and Adrián Carruana Martín, "Search and Graph Database Technologies for Biomedical Semantic Indexing: Experimental Analysis," *JMIR Medical Informatics* 5, no. 4 (2017): e48, <https://doi.org/10.2196/medinform.7059>.
11. Mario Barite, "Literary Warrant," *Knowledge Organization* 45, no. 6 (2018): 517–36.
12. Svenonius and Arms, *The Intellectual Foundation of Information Organization*.
13. Hjortsæters rules for subject headings was first published in 1990. They correspond to *Documentation: Methods for Examining Documents, Determining their Subjects and Selecting Indexing Terms* = *Documentation: m'ethodes pour l'analyse des documents, la d'etermination de leur contenu et la s'election des termes d'indexation*, International Organization for Standardization, Geneva, 1985 and partly to *Documentation: Guidelines for the Establishment and Development of Monolingual Thesauri* = *Documentation: principes directeurs pour l'establissement et la d'veloppement de thesaurus monolingues*, International Organization for Standardization, Genève, 1986. Parts of the rules were published in: Maria Inês Lopes, Julianna Beall, and Working Group on Principles Underlying Subject Heading Languages, *Principles Underlying Subject Heading Languages (SHLs)* (Berlin: K. G. Saur, 1999).
14. *Sears List of Subject Headings*, 22nd ed. (Amenia, NY: H. W. Wilson: Grey House Publishing, 2018).
15. Udir, Utdanningsdirektoratet is the Norwegian Directorate for Education and Training.
16. "Ordbok – for begreper i grunnopplæringen, norsk-engelsk/engelsk-norsk," UDIR, accessed December 6, 2023, <https://www.udir.no/verktoy/ordbok/>.
17. Ellen Hjortsæter, *Emneordskatalogisering: innholdsanalyse, emnerepresentasjon og lagring*, 3. utg. (Oslo: ABM-media, 2009).
18. Anne-Stine Ruud Husevåg, "Categorization of Known-Item Search Terms in a TV Archive," in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, CHIIR'17 (Oslo: Association for Computing Machinery, 2017), 321–24, <https://doi.org/10.1145/3020165.3022143>; Allen C. Benson, "Image Descriptions and their Relational Expressions: A Review of the Literature and the Issues," *Journal of Documentation* 71, no. 1 (2015): 143–64; L. Hollink et al., "Classification of User Image Descriptions," *International Journal of Human-Computer Studies* 61, no. 5 (2004): 601–26.
19. Erwin Panofsky, *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*, vol. 25, Icon editions (New York, NY: Icon Editions, 1972).
20. Karen Markey, "Computer-assisted Construction of a Thematic Catalog of Primary and Secondary Subject Matter," *Library Trends* 3, no. 1 (1983): 16–49; Karen Markey, "Access to Iconographical Research Collections," *Library Trends* 37, no. 2 (1988): 154–74.
21. Sara Shatford, "Analyzing the Subject of a Picture: A Theoretical Approach," *Cataloging & Classification Quarterly* 6, no. 3 (1986): 39–62.
22. Karen Collins, "Providing Subject Access to Images: A Study of User Queries," *The American Archivist* 61, no. 1 (1998): 36–55.
23. Shatford, "Analyzing the Subject of a Picture."
24. S.R. Ranganathan, *Prolegomena to Library Classification*, 3rd ed., vol. 20, Ranganathan Series in Library Science (Bombay: Asia Publishing House, 1967).
25. Hjortsæter, *Emneordskatalogisering: innholdsanalyse, emnerepresentasjon og lagring*.
26. Ranganathan, *Prolegomena to Library Classification*.
27. Michèle Hudon, "Facet," *Knowledge Organization* 47, no. 4 (2020): 320–33.



28. Unni Knutsen, *Fragmentering eller fellesløsning?: organisering av norsk bibliografisk produksjon*, vol. 60, ABM-skrift (trykt utg.) (Oslo: ABM-utvikling, 2009), 9.2.
29. Forlagsentralen ANS (<https://forlagsentralen.no>) is a distribution center historically owned by two of the largest publishers in Norway, Aschehaug, and Gyldendal, since 2021 wholly owned by Gyldendal.
30. Preminger et al., “The Public Library Metadata Landscape, the Case of Norway 2017–2018.”
31. Odd Letnes, “Konkurransen om metadata,” *Bok og bibliotek* (2017), 50–1.
32. Fride Fosseng, e-mail to author, November 3, 2023.
33. “Biblioteksentralens vokabulartjeneste: Bibbi autoriteter,” Biblioteksentralen, accessed December 6, 2023, <https://vokabular.bs.no/bibbi/nb/>.
34. “Bokbasen Tesaurus,” Bokbasen, accessed December 6, 2023, <https://support.bokbasen.no/hc/no/articles/115001692553-Bokbasen-Tesaurus>.
35. Osmo Suominen et al., *Publishing SKOS Vocabularies with Skosmos* (2015), <https://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf>.
36. As described later, the vocabularies were used differently for the occurrence analysis and PS-analysis. Subfields \$a, \$x, and \$z were taken into account for both agencies, whereas in the PS analysis, terms from the entire string (subfields \$q, \$x and \$z of Biblioteksentralen) were also considered.
37. Biblioteksentralen uses the language-value “nor” when the Nynorsk (“nno”) and Bokmål (“nob”) – versions are equal, so the “nor” versions are also counted in.
38. Biblioteksentralen uses these three subfield for pre-coordination, \$a, \$x, and \$z, whereas Bokbasen uses the \$a-subfield of repeating 6XX-fields.
39. Note that the *see* reference field, 950, was not included in the statistical comparisons, but taken into account in the Panofsky/Shatford analysis (Section 5).
40. Fride Fosseng, e-mail to author, November 17, 2023.
41. Hjortsæter, *Emneordskatalogisering: innholdsanalyse, emnerepresentasjon og lagring*.
42. An example of this is Jul (christmas inf. form) and Julen (Christmas fin. form) used by Biblioteksentralen resp. Bokbasen when cataloging Ashly Elston’s *10 Blinddates*.
43. “Norsk ordvev – bokmål,” Språkbanken, accessed December 6, 2023, <https://www.nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-27/>.
44. As the 3XX-class range is by far the largest group, we thought it will be interesting to see the distribution of only record pairs belonging to this class range to see if the relative tendencies manifest in this group as well.
45. Petra Saskia Bayerl and Karsten Ingmar Paul, “What Determines Inter-Coder Agreement in Manual Annotations? A Meta-Analytic Investigation,” *Computational Linguistics* 37, no. 4 (2011): 699–725.
46. Hjortsæter, *Emneordskatalogisering: innholdsanalyse, emnerepresentasjon og lagring*.
47. The Norwegian high school system offers two main specialization directions, where the “studiespesialisering” direction is meant to prepare the students for higher education.
48. Preminger et al., “The Public Library Metadata Landscape, the Case of Norway 2017–2018.”