

Ida Lovise Rustvåg

Dubletter – støy eller nytte?

Ein casestudie om metadata på e-bøker

**Masteroppgåve 2022
Master i bibliotek og informasjonsvitenskap**

Samandrag:

I denne casestudien blir metadata på e-bøker frå to kjelder samanlikna for å sjå om dei oppfyller IFLA LRM user task. Målet med samanlikninga er å sjå om det er grunnlag for å redusere dublettane brukarane blir presentert for i søkeverktøyet utan at det går utover atffinningsmoglegheitene for ein brukar. Funna visar at det er forskjellar i kvaliteten på metadata frå dei to kjelde som utfordrar brukaroppgåvene frå IFLA LRM user task for atffinning. Metadata frå den eine kjelda vil gi brukaren fleire søkeinngangar, sidan denne kjelda tek betre vare på brukaroppgåvene frå IFLA LRM user task.

Abstract:

In this case study, metadata on e-books from two sources are compared to see if they meet the IFLA LRM user task. The aim of the comparison is to see if there is a basis for reducing the duplicates in the search tool without going beyond the possibilities of discovery for a user. The findings show that there are differences in the quality of metadata from the two sources that challenge the user tasks from IFLA LRM user task for discovery. Metadata from one of the sources will give the user more access points since this source takes better care of the user tasks from the IFLA LRM user task.

Innhald

1	Innleiing	7
1.1	Problemstilling	7
2	Bakgrunn	9
2.1	Metadata	9
2.1.1	Metadastandarar i bibliotek.....	10
2.1.2	Metadatakvalitet	11
2.2	Funksjonen til ein bibliotekskatalog - frå Cutter til IFLA LRM user tasks	13
2.2.1	IFLA LRM user tasks	15
2.2.1.1	Finne.....	16
2.2.1.2	Identifisere	16
2.2.1.3	Velje.....	16
2.2.1.4	Skaffe.....	17
2.2.1.5	Utforske.....	17
2.3	Katalogisering og gjenbruk av postar	17
2.3.1	Katalogisering av e-bøker.....	18
2.4	Discovery-verktøy.....	20
2.5	Tidlegare forskning	20
3	Metode.....	25
3.1	Presentasjon av casen	25
3.1.1	Alma og BIBSYS-konsortiet.....	27
3.1.1.1	Katalogiseringspraksis i CZ.....	29
3.1.1.2	Katalogisering av e-bøker i BIBSYS-konsortiet.....	30
3.1.1.3	CDI.....	31
3.1.2	Oria.....	32
3.1.2.1	Dubletter i Oria	34
3.1.2.2	PNX.....	34
3.2	Utval	37
3.3	Avgrensingar	38
3.4	Datainnsamling.....	39
3.5	Svakheiter med undersøkinga	40
4	Funn	42
4.1	Gjennomgang av ein tittel	42
4.2	Presentasjon av funn	45

4.2.1	Forfattar	45
4.2.2	Tittel.....	48
4.2.3	Emneord.....	48
4.2.4	År.....	49
4.2.5	Utgåve	51
4.2.6	Beskriving	52
5	Diskusjon	55
5.1	Finne	55
5.1.1	Forfattar	55
5.1.2	Tittel.....	56
5.1.3	Emneord.....	57
5.2	Identifisere og velje.....	59
5.2.1	Forfattar	59
5.2.2	Tittel.....	60
5.2.3	Emneord.....	60
5.2.4	År.....	61
5.2.5	Utgåve	62
5.2.6	Beskriving	63
5.3	Utforske	64
5.3.1	Forfattar	64
5.3.2	Emneord.....	65
5.4	Samanliknande tabell over metadata frå CZ og CDI	66
6	Konklusjon	68
6.1	Vidare forskning.....	69
7	Referansar	71

Figurliste

Figur 3.1.1.3-1 Skjermdump av ei trefflista i Oria der ein har søkt på ein kjent tittel der treff nr 1 er frå CZ og treff nr 2 er frå CDI.....	32
Figur 3.1.2.2-1 Skjermdump av Marc-visning av ein post frå CZ.....	35
Figur 3.1.2.2-2 Skjermdump av korleis delar av ei PNX fil ser ut	36
Figur 3.1.2.2-3 Skjermdump av innhaldet i eit Display felt i ei PNX-fil.....	37
Figur 3.2-1 Tabell over utvalet i datasamlinga.....	38
Figur 4.1-1 Skjermdump av mangelfull post frå CZ	42
Figur 4.1-2 Skjermdump av tittel nummer 13 frå CZ	43
Figur 4.1-3 Skjermdump av tittel nummer 13 frå CDI	44
Figur 4.2.1-1 Diagram over antal redigerte bøker	45
Figur 4.2.1-2 agram over postar med author/editor i parantes	45
Figur 4.2.1-3 Diagram over antal ansvarbetegnelser per post	46
Figur 4.2.1-4 Diagram over bruk av ansvarbetegnelser	47
Figur 4.2.2-1 Diagram over titlar og undertitlar.....	48
Figur 4.2.3-1 Diagram over antal emneord per tittel.....	48
Figur 4.2.3-2 Tabell over fordeling av emneord.....	49
Figur 4.2.4-1 Diagram over utgjevingsår	50
Figur 4.2.4-2 Diagram over år	50
Figur 4.2.5-1 Diagram over utgåve opplysningar frå e-boka	51
Figur 4.2.5-2 Diagram over utgåve opplysningar i Oria.....	52
Figur 4.2.6-1 Diagram over antal linjer i beskrivinga	53
Figur 4.2.6-2 Diagram over postar som har feltet beskriving.....	53
Figur 4.2.6-3 Diagram over postar som har kapittel informasjon i beskrivinga	54
Figur 4.2.6-4 Diagram over postar som har kapittel og forfattar informasjon i beskrivinga	54
Figur 5.4-1 Tabell som samanliknar funna mellom CZ og CDI.....	67

Forord

Takk for lærerike år med nyttig innputt til min arbeidskvardag! Takk til rettleiar Trond Aalberg for nyttige innspel, og ikkje minst faste møter som har sikra ein jamn framgang, så eg har unngått dei verste skippertaka. Takk til mine leiarar og kollegaer som har gjort det mogleg at eg kan studer samstundes som eg har jobba fulltid. Utan dykkar godvilje og tilrettelegging hadde dette aldri gått!

Sist, men ikkje minst, tusen takk til mannen min og dottera vår! Viktigheita av planlegging og effektiv jobbing hadde ikkje kome så lett utan å ha dykk som motivasjon!

Ida Rustvåg

Follebu, juni 2022

1 Innleiing

Å katalogisere er ein dyr og tidkrevjande prosess som i dagens digitale bibliotekskatalogar lett kan effektiviserast på ein enkel måte ved å nytte seg av kopikatalogisering. Det vil da berre være eit bibliotek som treng å katalogisere eit dokument når andre kan kopiere jobben som alt er gjort over i sin katalog. Ei utfordring med delte katalogpostar er at dei som er brukarane av katalogen må være einige om å katalogisere på same måte (Denton, 2007, s. 50). Dette er ei utfordring for universitet- og høgskolebiblioteka i Noreg, som brukar biblioteksystemet Alma, når dei skal aktivere e-bøker i ein delte katalog, der både leverandørar, utgivarar og bibliotek leverer metadata. Her kan ein for eksempel finne mange dublettpostar for same tittel sidan utgivarar og bibliotek har ulik oppfatning av kva år som er riktig utgivingsår, eller andre manglar i metadata. Det gjer at det blir vanskeleg å verifisere at posten som alt ligger i katalogen er den same som den e-boka ein har kjøpt inn, og dermed kan skape dublettar.

E-bøker blir ofte kjøpt i store pakker der ein ikkje har kontroll på kva titlar som er i pakka til ei kvar tid når postane blir aktivert i bibliotekskatalogen, utan at ein bibliotekar ser igjennom postane. Nokre titlar går ut, mens andre kjem inn i pakka ved oppdateringar. Ein kan derfor oppleve at ein tidlegare kjøpte enkelt titlar også vert tilgjengeleg via ein pakke ein abonnerer på, og brukaren vil da bli møtt med to lenker på same posten. Ifølge brukartestinga til Davis og Song (2020, s. 197) trykkjer brukaren da berre på den øvste lenka, utan å reflektere noko meir over at det er to lenker. Dette er ein måte brukarar kan oppleve dublettar på i eit søkeverktøy. Ein annan måten brukaren kan møte på dublettar er når ein søker i søkeverktøyet der ein får treff på to postar, eller fleire, for same tittel frå ulike kjelder. Det er denne formen for dublettar denne undersøkinga vil sjå nærare på.

1.1 Problemstilling

Problemstillinga eg vil jobbe etter i denne masteroppgåva er som følgande:

Kan vi sjå nokre forskjellar på metadata på e-bøker frå to ulike kjelder, og oppfyller metadata på e-bøkene IFLA LRM user tasks?

Til å hjelpe meg å løyse denne problemstillinga skal eg sjå nærare på:

- Kordan kan ein brukar ved hjelp av metadata finne, identifisere og velje e-bøker?
- Kordan kan ein brukar ved hjelp av metadata på e-bøker utforske anna relevant materiale?

2 Bakgrunn

2.1 Metadata

Ein vanleg definisjon av metadata er data om data, eller data som beskriv data (Haynes, 2018, s. 9). Ein definisjon på metadata er definisjonen til Richard Smiraglia som omhandlar atffinning og administrasjon av informasjons resursar: «METADATA are structure, encoded data that describe the characteristics of information-bearing entities to aid in the identification, discovery, assessment and management of the described entities» (Smiraglia, 2005, s. 4).

Donna Frederick (2015, s. 3) definera metadata på denne måten “Metadata is structured information which represents a resource or service. This information is used to store, discover, retrieve, use and/or manage that resource in the present and over time.” Frederick (2015, s. 2) hadde spurt bibliotekarar korleis dei ville definere metadata før eit internt seminar der ho jobba. Mange av bibliotekarane svart da katalogisering. Som definisjonen til Frederick visar handlar ikkje metadata berre om å beskrive ein ressurs med bibliografiske data når ein katalogisere, det dreier seg også om atffinning.

Gilliland (2016, s. 2) seier at via metadata er det tre eigenskapar om ein ressurs som bør bli uttrykt: innhald, kontekst og struktur. Innhald seier noko om kva ressursen inneheld, eller kva ressursen er om, mens kontekst er «who, what, why, where and how aspects associated with the object’s creation and subsequent life» (Gilliland, 2016, s. 2). Det siste punktet struktur «relates to the formal set of associations within or among individual information objects and can be *intrinsic*, *extrinsic* or both» (Gilliland, 2016, s. 2).

Ulike modellar forklarar hensikta med metadata. Roy Tennant (1998, s. 30) deler metadata inn i tre typar: deskriptive, administrative og strukturelle. I kategorien deskriptive metadata finn vi element som brukast for å søke og lokalisere materiale i ein bibliotekatalog, slik som informasjon om kven som er forfattar, tittel og emneord. Administrative metadata tar for seg “how the digital file was produced and its ownership” (Tennant, 1998, s. 30). Eit eksempel på strukturelle metadata på e-bøker

kan være om kvar side er scanna som eit bilete og må ved hjelp av strukturelle metadatane blir satt samane til ei logisk heilheit (Tennant, 1998, s. 30).

I motsetning til Tennants (1998, s. 30) tredeling av hensikta med metadata delar Gilliland (2016, s. 10) metadata inn i fem kategoriar: administrative, deskriptive, bevaring, tekniske og bruk. Eit eksempel på administrative metadata er innkjøpsinformasjon. Katalogpostar er eit eksempel på deskriptive metadata. Bevaring består av «Metadata related to the preservation management of the collections and information resources» (Gilliland, 2016, s. 10). Metadata om korleis eit system virkar eller oppfører seg kallas tekniske metadata. Det siste kategorien metadata i følgje Gilliland er bruk. Denne type metadata er knytta til korleis bruken av ei samling er, slik som for eksempel søkeloggar (Gilliland, 2016, s. 10).

2.1.1 Metadastandarar i bibliotek

I store boksamlingar i bibliotek er ein avhengig av metadata for å sikre god atffinding, slik ein kan finne den spesifikke tittelen ein ynskjer blant alle bibliotekets titlar (Haynes, 2018, 29). Metadata i bibliotek er knytta til det å få tilgang til det materialet eit bibliotek har i samlinga si. «Library metadata includes indexes, abstracts, and bibliographic records created according to cataloging rules [...]» (Gilliland, 2016, s. 3). Det finnes ulike katalogiseringsreglar slik som Anglo American Cataloging Rules (AACR) og Resource Description and Access (RDA). Dette er eksempel på deskriptive katalogiseringsstandarar. Disse standarane gir retningslinjer for korleis ein skal beskrive ein ressurs (Frederick, 2015, s. 2). Med deskriptiv katalogisering beskriver den som katalogiserer kjenneteikna til materialet i ein bibliografisk post og lagar autoriserte søkeinngangar (Lazarinis, 2014, s. 19).

Det finnast ulike metadataskjema som mange bibliotek brukar, slik som Dublin Core og Marc. Dublin Core var laga for å beskrive ressursar på nett og nettsider (Haynes, 2018, s. 74). Dublin Core er utbredt i bibliotek der standaren brukast «to allow exchange of basic information and as a basis for interoperability between systems» (Haynes, 2018, s. 51). I Dublin Core er det 15 felt som brukast for å beskrive eit dokument (Baeza-Yates & Riveiro-Neto, 2011, s. 205). Når MARC, Machine Readable cataloguing, kom på midten av 1960-tallet var det med hensikt å skape ein standar på formatet av data i ein katalogpost for alle material typar. Standarden var laga for å

kunne lage katalogkort raskare, men var etter kvart brukt for å enkelt dele metadata, for å spare kostnader frå å katalogisere frå botn, mellom bibliotek (Haynes, 2018, s. 5-6). Ulike MARC-felt gjer det mogleg å uttrykke ulike attributtar ved eit dokument slik som tittel og forfattar (Baeza-Yates & Ribeiro-Neto, 2011, s. 205) MARC gir detaljerte bibliografiske opplysningar om dokumenta i eit bibliotek, som eignar seg best til å søke etter kjente titlar. For å auke andelen av søkbare emner kan ein berike postar ved å legge inn innhaldsinformasjon og samandrag i MARC felt 505 (Baeza-Yates & Ribeiro-Neto, 2011, s. 689-691). Metadatataskjema, slik som MARC21 og Dublin Core, definerer kva metadata som er obligatoriske og kva metadata som er valfrie å registrere i ein bibliotekskatalog, og er da ein viktig standardar for strukturen på metadata (Frederick, 2015, s. 2).

I tillegg til katalogiseringsreglar og standardar for struktur på metadata, har ein standardar for emneord, slik som Library of congress Subject Headings, LCSH (Gilliland, 2016, s. 3). Fordelen med at det finnes standardar for metadata er at «By having a documented standard, it is possible for future users of the metadata to understand the conventions in place and the intentions behind the metadata. Declared standards are an important tool for controlling the quality of data about information resources» (Haynes, 2018, s. 49). Standardane er nødvendige for beskrive ressursar på ein nokon lunde same måte mellom ulike bibliotek. Ein kan da utveksle data mellom bibliotek på ein fornuftig måte når dei ulike felta beskriv det same, noko som vil spare bibliotek for tid og pengar når metadata kan gjenbrukas.

Sidan 1960-tallet har det i biblioteka blitt systematisk laga og delt bibliografiske metadata via «automated systems such as bibliographic utilities, online public access catalogs (OPACs), and commercially available databases» (Gilliland, 2016, s. 3). Ei endring sidan 60-talet er at i dag blir ikkje berre metadatane laga av mennesker, men dei kan også blir oppretta på ulike automatiske måtar slik som «metadata mining, metadata harvesting and web crawling» (Gilliland, 2016, s. 3.).

2.1.2 Metadatakvalitet

Informasjons attfinning kan definerast som dette:

[...] the representation, storage, organization of, and access to information items such as documents, Web pages, online catalogs, structures and semi-

structures records, multimedia objects. The representation and organization of the information items should be such as to provide the users with easy access to information of their interest (Baeza-Yates & Riberiro-Neto, 2011, p. 1).

Forskjellen mellom dataatfinning og informasjonsatfinning, ifølge Van Rijsbergen (1979, s. 1-2), er at i informasjonsatfinning brukar ein metadata for å finne dokument som matchar søket, mens i dataatfinning er målet å finne faktasvar på eit spørsmål, og ikkje sjølve dokumentet. Sidan metadata beskriv ein ressurs på ulike måtar er metadata veldig viktig for atfinning av informasjons ressursar (Haynes, 2018, s. 96). For brukaren vil metadata med god kvalitet være viktig for å få løyst sitt informasjonsbehov. Bruce og Hillmann (2004, s. 243) seier at "completeness, accuracy, provenance, conformance to expectations, logical consistency and coherence, timeliness and accessibility" er dei viktigaste kjenneteikna på metadatakvalitet.

«The different items of information, such as author, title or subject, used for locating specific bibliographic records are called access points, known also as entry points.» (Lazarinis, 2014, s. 6). Når ein ressurs blir beskriven med metadata får ein brukar ulike søkeingangane som brukarane er avhengige av for å løyse informasjonsbehovet dei har. Det er da viktig at desse metadata, søkjeingane, har god kvalitet.

Wiersma og Tovstiadi (2017, s. 624) beskriv godt viktigheita av gode metadata på e-bøker i dette sitatet: "Many libraries report that one of the benefits of e-books is that, unlike physical materials, they cannot be misshelved, lost, or stolen. But an e-book without sufficient metadata is just as lost as its misshelved print counterpart." Brukaren kan ikkje vandre langs hyllene i eit biblioteket for å finne e-bøker på slump, bøkene må søkast etter når dei er digitale for å kunne oppfylle informasjonsbehovet til brukaren. Utan gode søke inngangar blir det vanskeleg, om ikkje umogleg å finne riktig e-bok når den trengs. Dette er noko David og Thomas også understrekar:

Metadata quality is particularly important in digital formats, such as ebooks, which cannot be found on the shelf. One may say that they exist only in so far as their metadata exists, as their presence is only determined by discovery services—whether local or across the Internet. If metadata is of low quality or incomplete, users will not have sufficient information to determine whether the

content is relevant to their research needs. On most ebook platforms, users make their decision to read the work or abort the search based on the first page of metadata (David & Thomas, 2015, s. 802).

David og Thomas (2015, s, 802) peikar på manglande metadata og metadata med lav kvalitet, som hindrar attfinning. I tillegg til desse utfordringane er det som Frederick (2015, s. 11) seier, ikkje uvanleg at utfordringar som omhandlar e-bøker, slik som metadatakvalitet i MARC postar, og forvirringar knytta til metadata i leverandørernes plattformer for kjøp, blir eit samtaleemne når bibliotekarar møtast.

2.2 Funksjonen til ein bibliotekskatalog - frå Cutter til IFLA LRM user tasks

Charles Cutter var i 1904 den første som beskrev målet med ein bibliotekskatalog (Svenonius, 2000, s. 15). I Cutters (1904, s. 12) *Rule for a Dictionary Catalog*, som vist i sitatet under, kan ein lese at måla til ein bibliotekskatalog er:

1. To enable a person to find a book of which either
 - (A) the author
 - (B) the title
 - (C) the subjectis known.
 2. To show what a library has
 - (D) by a given author
 - (E) on a given subject
 - (F) in a given kind of literature.
 3. To assist in the choice of a book
 - (G) as to its edition (bibliographically).
 - (H) as to its character (literary or topical).
- (Cutter, 1904, s. 12).

Den fyrste målet til Cutter for ein katalog, finne, baserar seg på at brukaren ser etter ei bok det tittel, forfattar eller emne alt er kjent for brukaren. I det andre målet katalogen skal dekke er nå brukaren ynskjer ei oversikt over alle bøker i bibliotekskatalogen som er skrevet av ein forfattar, eller handlar om eit emne. Det tredje målet går ut på at

brukaren skal kunne velje bland mange liknande dokument, som for eksempel blant bøker som har ulike utgåver eller er på ulike språk (Svenonius, 2000, s. 15-16).

I måla for ein bibliotekskatalog frå Charles Cutter snakka han om attfinning av ei bok, mens Seymour Lubetzky innfører begrepet verk (Denton, 2007, s. 45). Hos Lubetzky finn ein eit todelt brukaroppgåver der den fyrste går ut på at det skal være lett for ein brukar å identifisere om biblioteket har den boka brukaren ynskjer i samlinga eller ei. Den andre brukaroppgåva katalogen burde oppfylle hos Lubetzky var at brukaren skal kunne få informasjon om kva verk som finnes av ein bestemt forfattar, og kva utgåve og oversetningar eit bibliotek har av bestemte verk. (Svenonius & McGarry, s. 113, sitert i Denton, 2007, s. 46).

Arbeidet til Seymour Lubetzky, med endring frå bok til verk, satte sitt preg på ordlyden til Paris prinsippa frå International Conference on Cataloging Principles i 1961 (Denton, 2007, s. 46-47). Målet med møte var å få ein enighet mellom dei 53 landa og 12 organisasjonane som var med for ein felles standar, slik at katalogpostar enkelt kunne delast på tvers av landegrenser (Denton, 2007, s. 47) I Paris prinsippa er funksjonen til bibliotekskatalogen definert slik:

The catalogue should be an efficient instrument for ascertaining

2.1 whether the library contains a particular book specified by

(a) its author and title, or

(b) if the author is not named in the book, its title alone, or

(c) if author and title are inappropriate or insufficient for identification, a suitable substitute for the title; and

2.2 (a) which works by a particular author and

(b) which editions of a particular work are in the library.

(The International Conference on Cataloguing Principles, 1961, s. 1)

Functional Requirments for Bibliographic Records, FRBR, er ein conceptual model laga av IFLA (Lazarinis, 2014, s. 4) som byggar vidare på Paris Prinsippa (Denton, 2007, s. 51). I modellen finn ein brukaroppgåvene *find*, *identify*, *select* og *obtain* som ein ser i samheng med ein entitet (Lazarinis, 2014, s. 4). I FRBR er det tre grupper med entiteter. Gruppe 1 verk, utrykk, manifestsjon og eksemplar, består av «the

product of intellectual or artistic endeavor that are named or described in bibliographic record” (Lazarinis, 2014, s. 94). I gruppe 2 finn ein den som er ansvarleg for innhaldet i gruppe 1, slik som ein person eller korporasjon. Gruppe 3 er entitetar som blir presentert som emner for begrep, gjenstandar, hendingar og stadar (Lazarinis, 2014, s. 94).

2.2.1 IFLA LRM user tasks

The Library Reference Model (LRM) erstattar dei tre entitets relasjons modellane Functional Requirements for Bibliographic Records (FRBR), Functional Requirements for Authority Data (FRAD) og Functional Requirements for Subject Authority Data (FRSAD). Desse modellane har alle ulike, men overlappende, brukaroppgåver metadata skal oppfylle i ein bibliotekskatalog. Desse behova er slått saman og justert i IFLA LRM user tasks (Strander, 2021, s. 443). Det som er likt for alle desse modellane er brukaroppgåvene *find* og *identify*. Dei brukaroppgåvene i FRAD som var retta mot tilsette i bibliotek er ikkje tatt med i LRM, elles er brukaroppgåvene overlappende mellom fleire av modellane (Strander, 2021, s. 444).

Modellen IFLA LRM user task har fokus på brukarens informasjons behov. Ein brukar kan være både en sluttbrukar og ein bibliotekar som brukar same metadata for å utføre sine oppgåver. Denne modellen dekker ikkje administrative metadata og metadata for rettigheter (Riva et al., 2017, s. 15).

IFLA LRM user tasks presenterer fem brukaroppgåver som metadata på eit dokument i ein bibliotekskatalog skal dekke: *find*, *identify*, *select*, *obtain* og *explore*. (Riva et al., 2017, s. 15). Desse brukaroppgåvene er i denne masteroppgåva oversett til: finne, identifisere, velje, skaffe og utforske.

Som det poengterast av Riva et al. (2017, s. 15) følgjer ikkje prosessen med innhenting av informasjon nødvendigvis i den rekkefølga som er satt opp i brukaroppgåvene i IFLA LRM user tasks med finne, identifisere, velje, skaffe og utforske.

Although the tasks are listed here in a particular order, there is no intention to imply that these are all obligatory steps in an ideal information seeking process. In reality information seeking is iterative and may move in a tangent at any stage. Some user tasks may happen essentially simultaneously in the user's

mind (identify and select, for example). In particular, explore is a separate dimension from the other tasks: in some cases providing starting points for further information seeking processes, and in others allowing browsing without any particular information goal. (Riva, Bøeuf & Žumer, 2017, s. 15).

2.2.1.1 *Finne*

Den fyrste brukaroppgåva, *finne*, går ut på at det skal være mogleg å finne ein eller fleire ressursar ved å søke på relevante kriteria slik som tittel, forfattar og emneord. Dette er attributta, og relasjonen mellom entitetane, ein brukar søkar på som skal være mogleg å søke på ved hjelp av metadata (Riva et al., 2017, s. 15-16). Dette er da beskrivinga av ressursen og søkeingangane, metadata, som har blitt laga når materialet vart katalogisert, som brukarane kan søke på i eit søkeverktøy (Strader, 2021, s. 445). Eksempel på brukaroppgåva finne, som skal kunne løysast med metadata, er at det skal være mogleg å finne alle utgåver av eit verk som er på eit bestemt språk, å finne alle verk som er skrevet av ein forfatta, og finne verk når ein søker på emneord (Riva et al., 2017, s. 97).

2.2.1.2 *Identifisere*

Med å *identifisere* meinast det at det skal være mogleg for ein brukar å skilje mellom ressursar som liknar på kvarandre ved å få tilstrekkeleg informasjon om dokumentet (Riva et al., 2017, s. 15). For ein brukar skal det være mogleg å skilje mellom titlar som har same tittel, men er av ulike forfattarar i ei treffliste. Av metadata skal det komme klart fram kva slags materialtype ressursen har, slik ein kan skilje mellom ulike ressursar slik som ei bok, kapittel i ei bok eller ein bokomtale ((Riva et al., 2017, s. 97). Bøker som har ulik utgåve og bøker med same tittel, er eit eksempel der ein treng nøyaktig beskriving av ein resurs for å kunne skilje dei frå kvarandre (Strader, 2021, s. 445).

2.2.1.3 *Velje*

Den tredje brukaroppgåva, *velje*, går ut på at ved hjelp av metadata skal brukaren kunne velje eller avvise dei ulike resursane. Denne brukaroppgåva skjer ofte samtiding som når ein brukar identifiserer ein ressurs. (Riva et al., 2017, s. 15). Ein brukar skal kunne velje ut i frå metadata om ressursen er mogeleg å låne med ein gong eller om den er utlånt. Metadata skal gi brukaren informasjon om kva materialtype ressursen

har slik som om det er ei bok eller eit bokkapittel. Brukaren skal kunne velgje ut i frå utgivingstidspunktet om boka er aktuell (Riva et al., 2017, 2017, s. 98).

2.2.1.4 Skaffe

Punktet om å *skaffe* vil seie at ein brukar skal få nok informasjon om kvar det er mogleg å finne resursar, enten ved lenke til e-ressursar, eller ved å få informasjon om hylleoppstilling (Riva et al., 2017, s. 15). For e-bøker vil det da være veldig viktig at lenka til e-ressursen faktisk virkar frå søkeverktøyet for at brukaren skal få fullført brukaroppgåva.

2.2.1.5 Utforske

Det siste punktet i IFLA LRM user tasks, *utforske*, vil sei at brukaren skal kunne oppdage andre relevante resursar ved å kunne navigere vidare frå eit treff til andre, som for eksempel har same emneord eller forfattar. Denne brukaroppgåva visar viktigheita av serendipitet i informasjonssøkinga til ein brukar (Riva et al., 2017, s. 16). Ved å utforske kan brukarane komme over ressursar dei ikkje var klar over fantes.

2.3 Katalogisering og gjenbruk av postar

Katalogisering, eller bibliografisk kontroll, består av to delar; deskriptiv katalogisering og emne katalogisering. Ved hjelp av disse to delane beskrivast «an information object in terms of what it is ('is-ness') and the representing what it is about ('aboutness') through an analysis of the intellectual content" (Miksa, 2021, s. 98). Ein får da ein katalogpost som representerer objektet eller ressursen, slik som ei bok, som er registrert ved hjelp av metadataskjema (Miksa, 2021, s. 97).

Autorisering av søkeinngangar, som namn, tittel og emne er også ein viktig del av det å katalogisere i ein bibliotekatalog. For at ein bibliotekatalog skal kunne vedlikehaldas trengs det reglar som er basert på prinsipp og mål. Noko som er enda viktigare i dag når kopikatalogisering er vanleg, som gjer at ein kan dele postar med bibliotek over heile verden (Miksa, 2021, s. 98).

IFLA oppretta ei gruppe for Universal Bibliographic Control (UBC) på 1970-tallet. Tanken med universell bibliografisk kontroll var at dei ulike landas nasjonalbibliotek hadde i ansvar å lage katalogpostar og autoritetar for materialet som var utgitt i deira

land. Disse postane kunne da delast og gjenbrukas over heile verda (IFLA, 2012). Det ville da berre være nødvendig å katalogisere materiale ein gong.

Det har skjedd ei stor utvikling frå dei fyrste bibliotekatalogane som var skreve for hand, til kortkatalogar og til dagens digitale bibliotekatalogar. Bibliotekatalogane som var skrevet for hand, og seinare vart trykte, var vanskelege å organisere og var gjerne utdatert før dei var utgitt (Denton, 2007, s. 37). For å dekke behovet for søkeinnangar i ein kortkatalog måtte ein ha fleire kort, gjerne seks til åtte, som beskrev eit dokument slik som eit kort for forfattar, eit kort for tittelen og eit kort for kvart emne som var brukt for å beskrive dokumentet (Lazarinis, 2014, s. 6-8). Dette er i stor kontrast til dagens digitale katalogar det ein berre trenger ein bibliografisk post per tittel for å dekke dei søkeinnangane brukarane våre treng (Lazarinis, 2014, s. 9).

I følgje Frederick (2016, s. 120) er det å katalogisere frå botn er noko ein prøvar å unngå, og ein gjer det berre om det ikkje finnes noko post som ein kan kopiere, sidan dette er ein dyr og tidkrevjande prosess. Derfor vart det tidleg oppretta tenesta der biblioteka kunne kjøpe ferdige kortkatalogpostar frå for eksempel Library of Congress eller kommersielle bedrifter for å unngå å katalogisere frå botn. I dag kan bibliotek få tak i digitale MARC21 postar blant anna frå OCLC der dei som er medlemmer kan lasta opp og ned postar. Gratis postar er tilgjengelege frå British Library via Z39.50 protokollen (Lazarinis, 2014, s. 13-14). Med teknologi slik som Z39.50 kan postar delast og gjenbrukas over heile verda og ei bok treng da berre å bli katalogisert ein gang (Frederick, 2016, s. 118). Ein anna ting som er viktig for at det kun skal være behov for å katalogisere ei bok ein gong er at postane skal være leverandør nøytrale (Culbertson, Mandelstam & Prager, 2011, sitert i Barderheier, 2013, s. 62). Katalogpostar som er leverandør nøytrale kan da gjenbrukas av bibliotek over heile verda.

2.3.1 Katalogisering av e-bøker

I 1997 vart det mogleg for akademiske bibliotek å kjøpe e-bøker via ebrary og i 1999 via NetLibrary. I åra som fulgte kom det fleire leverandørar inn i markete slik som Safari, SpringerLink og Ebook Library (Ward et al., 2015, s. 2).

Med dette nye formatet kom det nye utfordringar for bibliotekas behov for gode metadata i bibliotekatalogane. Dei første e-bok postane som vart laga vart berre konvertert frå den trykte boka til ein e-bok post. Dette var ein effektiv måte for å enkelt få sikra søkeinngangar til forfattar, tittel og emne (Frederick, 2016, s. 108). Men ikkje så heldig med tanke på å skilje dei ulike dokumenttypene og for atffining. Både Minčić-Obradović (2011, s. 89) og Walters (2013, s. 100) seier at leverandørar ofte gjenbrakar postar for trykte dokumenter utan å fjerne informasjon som er unikt slik som ISBN når dei leverer postar for e-bøker. For trykte bøker er det unike ISBN nummer for paperback og hardcover, men for e-bøker er det ikkje alltid unike ISBN som skiller ulike versjonar av ei e-bok. Både forlag, agregatorer og leverandører av e-bøker kan gi ut ISBN (Walters, 2013, s. 98). Noko som kan skape usikkerheit om det er same boka som blir tilbutt frå ulike stadar.

I tillegg til at ISBN er feil eller manglar i postar frå leverandør for e-bøkere, er andre feil som er rapportert frå slike postar, at det er manglande eller ufullstendige opplysningar om utgjevingsinformasjon og fysisk beskriving, problem med emneorda som enten manglar eller er emneord som ikkje lenger brukast frå Library of Congress, at informasjon er registrert på andre språk enn engelsk, at namn ikkje stemmer med den autoriserte namneformen (Walters, 2013, s. 100).

Til skilnad frå katalogisering av trykte bøker, der ein enten katalogiserer frå botn eller kopikatalogiserer, er det for e-bøker også ein tredje måte ein kan katalogisere e-bøker ved at ein laste opp store pakkar til bibliotekatalogen. Denne måten brukast når eit bibliotek kjøpe pakkar med e-bøker, som kan innehalde med mange tusen e-bøker, som det hadde vore alt for omfattande å katalogisere ein og ein (Frederick, 2016, s. 117). Opplasting av pakkar brukast også når biblioteket gjer innkjøp via innkjøpsmodellar der brukaren er involvert i kva som kjøpast, i for eksempel ein PDA, patron driven acquisiton, der ein lastar opp eit utval av titlar til bibliotekatalogen innan for eit emneområdet der kjøpet først blir gjennomført når brukaren les boka (Ward et al, 2015, s. 11). Ved å lasta opp store pakkar med e-bøker mistar biblioteka kontrollen over kva metadata som er registrert i sine katalogar og korleis kvaliteten på metadata er, til skilnad frå når ein katalogisere materialet i bibliotekatalogen frå botn sjølv.

2.4 Discovery-verktøy

I dei siste åra har det blitt meir og meir vanleg med Discovery-verktøy i akademiske bibliotek som lar brukarane søke etter både dei trykte og digitale ressursane eit bibliotek har i same søkeverktøy. Brukaren kan da få treff i søkeverktøyet på både det biblioteket har i si fysiske samling, open access materiale og det biblioteket har av e-ressursar, slik som abonnement på tidsskrift og e-bøker (Hoepfner, 2012, s. 7-8). Vanlege funksjonar i eit discovery-verktøy, i tillegg til ein søkeinngang til ein central index, er rask respons tid, trefflister som er rangert etter relevans, ulike moglegheiter for å sorterte og bruker fasettar på trefflista, lenker til fulltekst og brukaren har ein eigen konto for å handtere lån, fornyingar og bestillingar (Hoepfner, 2012, s. 8).

Grunnen for at så mange bibliotek har gått over til discovery-verktøy er for å gi brukarane ein enklare måte å søke på som liknar Google, som dei alt er vant med å søke i, og dermed kan tru at bruken av bibliotekets elektroniske ressursar vil auke når dei blir lettare å finne enn via spesial databasar (Rose-Wiles & Hofman, 2013, s. 151).

Ein brukar kan lett tru at ein vil finne alle dei elektroniske ressursane eit bibliotek har tilgang til i eit discovery-verktøy, men sanninga er at sidan ulike leverandørar har sine egne søkeverktøy og ikkje ynskjer å leverer frå seg metadata til konkurrentar kan ein som brukar gå glipp av relevante treff som ikkje er indeksert i systemets central index. Brukaren er da avhengig av at kjelden er indeksert via andre kjelder biblioteket har tilgang til for å få treff på så mykje som mogleg som kan være relevant i discovery-verktøyet (Breeding, 2014, s. 14).

2.5 Tidlegare forskning

Dette finnes mykje forskning og undersøkingar om e-bøker og metadata, men i liten grad forskning som har samanlikna metadatane på e-bøker i Oria/Primo frå CZ og CDI, og heller ikkje nokon som har sett på metadatane frå CDI.

Wiersma og Tovstiadi (2017) har i artikkelen *Inconsistencies between Academic E-book Platforms: A comparison of Metadata and Search Results* sett på korleis den same e-bok tittelen blir presentert hos ulike leverandørar sine plattformer. I den komparative analysen til Wiersma og Tovstiadi (2017, s. 624) viste funna at det var manglar i grunnleggande metadata slik som forfattar, tittel, utgivar og utgivingsår. «Common discrepancies included listing editors as authors, using different dates,

such as copyright or publication date, and not including subtitles» (Wiersma & Tovstiadi, 2017 s. 624). I undersøkinga såg Wiersma og Tovstiadi (2017, s. 628) at emneord for ei og same bok varierte i så stor grad på tre ulike plattformer, at dei kunne ha beskrive tre ulike bøker.

Funna til Wiersma og Tovstiadi (2017, s. 624) stemmer overeins med dei manglane Knutsen (2017) presenterte på BIBSYS-konferansen i 2017. I undersøkinga Knutsen (2017) hadde gjort i 2014 vart metadata frå BIBSYS konsortiet samanlikna med andre e-bok leverandørar sine metadata. Manglane dei såg gjekk igjen i undersøkinga frå 2014 var blant anna at året i MARC felt 260/264 ikkje stemte med utgivingsåret, at det var feil bruk av funksjonskodar slik som \$4 red ikkje vart brukt på redigerte bøker, men heller koden \$4 aut for forfatar, og at namnet på forfattaren eller redaktøren mangla i 245 \$c. Postane mangla også autorisering av forfattarnamn (Knutsen, 2017 s. 14). I Knutsen (2017, s. 13) såg ein at det kunne være mange emneord frå ulike kjeldar på ein CZ post der ein hadde MESH-termar frå ulike nivå frå emneordshierarkiet.

Knutsen (2017, s. 2) sin presentasjon konkluderte med at katalogiserings nivået var lågare enn det vi i BIBSYS konsortiet har bestemt at vi skal ha. Noko som førte til lågare atffinning og fare for dublettar i CZ (Knutsen, 2017. s. 15).

Irene Refsland presenterte også ein undersøking på metadatakvalitet på BIBSYS-konferansen 2017. Ho presenterte ein undersøking frå 2016-2017 der UiO hadde følgt med på metadata i CZ over tid for å sjå kva som skjedde om dei endra postane. (Knutsen, 2017, s. 16-17). Det dei såg var at det tok lang tid før små MARC-postar vart utvida i CZ, og om ein hadde gjort endringar på desse kunne ein risikere at dei var skrivne over. Om ein gjorde endringar på MARC postar med betre kvalitet vart dei ikkje endra (Knutsen, 2017, s. 33).

Eit anna interessant prosjekt med fokus på atffinning er artikkelen *Missing the MARC: Utilization of MARC fields in the search process* av Woolcott et al. (2021, s. 49) som undersøkte samanhengen mellom termane ein brukar søker på og kor vellykka dei ulike MARC felte var. Dei mest vellykka MARC felte for atffinninga var 245 med tittel, 505 for innhaldsnote, 650 for generelle emneord og 520 for samandrag. (Woolcott et al., 2021, s. 41). Funna til Woolcoot et al. (2021, s. 51) viste at å ha eit MARC felt 505 for innhaldsnote er meir effektive for å auke atffinninga enn emneord, sidan orda

brukarane brukte ikkje samsvarar med emneorda brukt i katalogen. Sassen og Harker (2014) såg også i ei anna undersøking at bruken av e-bøkene auka etter å ha lagt til innhald i MARC 505 og samandrag i 520 i MARC-posten. Ein tredje som også så auke i bruken av ei e-bok pakke var Minčić-Obradović (2011, s. 89) når kvaliteten på metadata vart forbetra.

Elzi og Clair (2020, s. 43) hadde ein undersøking med brukartesting der dei fekk brukarane til å kommentere kva metadata felt dei syns var nyttige ved søk der beskriving og emneord var mest nemnt. Mange av postane i denne undersøkinga hadde få eller ingen emneord, noko som gjorde at brukarane syns det var vanskeleg å forstå kva posten handla om (Elzi & Clair, 2020, s. 43). Undersøkinga foregikk i ArchivesSpace, men har resultert i at biblioteket endra praksisen sin i korleis metadata skulle bli presentert i deira Primo. Dei hadde ikkje forventa at brukarane brukte MARC 520 så masse som dei gjorde for å identifisere materiale dei ynskja. Biblioteket har derfor ordna rutiner for å forbetre dette feltet årleg (Elzi & Clair, 2020, s. 46).

David og Thomas (2015, s. 807) undersøkte om MARC felt som tittel, forfattar, utgivingsår stemte overeins med sjølv PDF'en av e-boka i artikkelen *Assessing Metadata and Controlling Quality in Scholarly Ebooks*. I deira prosjekt fekk biblioteket tilsendt MARC-filer frå levernadørane på e-bøkene dei kjøpe som dei lasta opp i ein lokal database der brukarane kunne søker etter e-bøkene. Hensikta for prosjektet var å sjå om det var nokre grep dei kunne gjere for å bedte metadatane (David & Thomas, 2015, s. 801). Samanlikninga mellom metadatane og sjølv PDF'en av e-boka viste at i denne undersøkinga stemte metadata i stor grad (David & Thomas, 2015, s. 813). Ein ting dei såg i undersøkinga som kan være relevant for mitt prosjekt var at emne, tittel og forfattar var MARC felt som var viktigast for søking og atffinding, sidan desse felte var dei som gav flest treff på søkeorda brukarane brukte (David & Thomas, 2015, s. 810).

Dabin og Preminger (2018, s. 37) så i sin artikkel *Discovery systems as a n Alternative to stand-Alone Databases. The Example of Primo at BI Norwegian Business School* på korleis metadata på artklar frå Oria var samanlikna med metadata i ein spesialisert databasen. Sjølv om undersøkinga til Dabin og Preminger (2018) ta for seg korleis det er å søke etter artklar vil eg seie den har element som er relevant i mitt prosjekt

sidan metadatane til artiklar også kjem frå CDI, slik som postane på e-bøker. Eit av funna frå denne undersøkinga er at dei fekk treff på emneord som ikkje var synlege for ein brukar under detaljer på ein post vist i fullvisning, men når dei undersøkte PNX-fila fant dei at emneorda var i «search cection» delen av fila, og dermed ikkje synleg for brukaren (Dabin & Preminger, 2018, s. 49).

Artikkelen til Proctor (2013) *How Old is That Ebook: A Call for Standardization in Publisher-Provided Ebook* tar for seg utfordringa om opplysningar om utgjevingsår som fortsatt er gjeldande. I artikkelen ser ein på korleis åtte forlag presenterar utgjevingsåret på e-bøkene på nettsidene og tittelsider. Funna viser at «Publishers often use terminology that does not accurately describe the dates they are providing, and sometimes they provide only the date of electronic publication, which can be misleading when the print version was published many years prior.» (Proctor, 2013, s. 100).

“The definition of publication date is the day the book, in that format, becomes available to consumers. For publishers, this means the day we digitized the book.” (Branch et al, 2018, s, 128). For å gjere det tydelegare kva utgjevingsår som er gjeldane har forlaget *Taylor and Francis* lagt til eit felt for “First Published on” som viser kva år som er det fyrste ein tittel har blitt utgitt på. Dette for at det skal blir klarare for andre kva år boka er i frå (Branch et al, 2018, s, 128).

Bardenheier (2013, s. 60) såg i sin casestudie på kvaliteten på MARC postar frå fire ulike kjelder. For alle kjeldene var det feil som gjekk igjen i MARC postane slik som «the General Matrial Designation (GMD), punctuation, date and indicators» (Bardenheier, 2013, s. 67). Bardenheier (2013, s, 69) såg at det var ulike år registrert i MARC 008 og 260 feltet som skapar forvirring for brukarane. Feil i indikatorar i MARC 245 fører til at titlar som skal være søkbare ikkje er det, og berre kan søkast etter med emneord (Bardenheier, 2013, s. 69-70).

I casestudien til Kleeck et al. (2017, s. 560) ville dei undersøke om metadata på elektroniske ressursar oppfyller brukaroppgåvene finne, velje og det å få tilgang til frå to kjelder. Samanlikninga viste at den eine kjelda støtta brukaroppgåvene betre, sidan

brukarane fekk fleire søkeinngangar når det var fleire postar som hadde samandrag og innhaldsforteikning.

3 Metode

For å svare på denne masteroppgåvas problemstilling:

Kan vi sjå nokre forskjellar på metadata på e-bøker frå to ulike kjelder, og oppfyller metadata på e-bøkene IFLA LRM user tasks?

og spørsmåla som skal hjelpe meg å løyse denne problemstillinga:

- Kordan kan ein brukar ved hjelp av metadata finne, identifisere og velje e-bøker?
- Kordan kan ein brukar ved hjelp av metadata på e-bøker utforske anna relevant materiale?

vil det være hensiktsmessig å velgje ein casestudie for denne undersøkinga.

Ein casestudie kan definerast som «en inngående studie av en eller noen får undersøkelseenheter» (Jacobsen, 2015, s. 97). Ved å velje casestudie som metode for denne masteroppgåva kan ein da undersøke eit begrensa utval av postar meir inngående, enn om ein skulle ha sett på alle postane som finnes i ein bibliotekatalog. I ein casestudie er målet å generere kunnskap om casen (Tjora, 2020, s. 41). Ved å undersøke eit utval av postar for e-bøker vil ein kunne skape kunnskap ved å samanlikne korleis metadata er for dei to kjeldene, for å sjå om det er noko forskjell mellom dei, og om dei oppfyller brukaroppgåvene ein finn i IFLA LRM user tasks.

For å svare på oppgåvas problemstilling treng eg eit datasett som kan analyserast for å samanlikne metadata frå dei to kjeldene som skal undersøkast. Datamaterialet vil bli sortert og systematisert for å sjå om det er fellestrekk og ulikheiter i materialet frå dei to kjeldene.

3.1 Presentasjon av casen

Det er eit kjent problem for dei som jobbar med e-bøker i BIBSYS-konsortiet at når ein søker i Oria etter ein kjent e-bok tittel får ein treff to postar. Dette er fordi ein i Oria får metadata frå både biblioteksystemet Alma og Central discovery index (CDI), som er ein database over artiklar, e-bøker og andre e-ressursar (Unit, 2020). I CDI blir artiklar som det finnes fleire postar av slått samane til ein berika post, men dette skjer ikkje

med e-bøker når dei er aktivert frå ulike indeksar, Alma og CDI, som ikkje kan slåast saman (Unit, 2020b). Derfor opplever vi når vi søker etter e-bøker i Oria at vi får treff på to postar for ein tittel, ein frå CDI (eller fleire), og ein for den lokale posten frå Community Zone (CZ) i Alma. Om ein får fleire postar for same ressurs frå CDI, har prosessen med å slå saman postane feila slik at postane ikkje blir til det som kallast *Merged records*, også kaldt *Composite records* (Unit, u.å.).

I eit webinar heldt av Laura Jacobs frå Exlibris (ExLibrisLtd, 2021) om CDI blir det presentert to ulike måtar å løyse dublett problematikken på e-bøker i Oria. Den eine måten går ut på at ein kan skjule postane som kjem frå CDI, og berre visar dei lokale postane frå CZ. Om ein gjer dette vil ein miste informasjon om at postar som er indeksert som bokkapittel er tilgjengelege. Den andre måten er å skjule dei postane som er frå CZ. Fordelen med å skjule postane frå CZ er at ein da får treff på bokkapittel. Ulempa er at dette berre fungerer på collection som brukar "link in records". (ExLibrisLtd, 2021, 24:27). Ved å skjule postane frå CZ vil ein miste eventuelle lokale emneord som ein kan legge til så det blir mogeleg å søke på norsk på dei engelske e-bøkene om ein ynskjer denne moglegheita.

I webinarret om CDI frå Exlibris får ein også informasjon om at det vil komme ei forbetring i siste halvdel av 2021, som vil gå ut på at ein kan bruke kapittel informasjonen frå CDI i kombinasjon med den lokale posten frå CZ (ExLibrisLtd, 2021, 24:27). I november 2021 kom denne oppdatering for Oria der ein kan skru av postane frå CDI, men behalde kapittel informasjon slik at det i Oria vil bli kapittel informasjon frå CDI og den lokale posten frå CZ som er synleg (Exlibris, 2021).

Fredrik Solvang Pettersen (personlig kommunikasjon, 19. november 2021) sendte ut ein e-post til e-post-lista oria@unit.no 19. november 2021 der han åtvarar mot dei uheldige konsekvensane det vil føre til ved å aktivere denne løysninga, sidan ALLE postar frå CDI vil da forsvinne frå Oria. Noko som vil sei at om vi skrur av postar frå CDI, vil vi ikkje få treff på andre norske universitet- og høgskolebibliotek sine e-bøker når ein søker i «Norske fagbibliotek». Ein vil da gå glipp av titlar som kan være interessante å vite om, slik brukaren enten kan bestille inn eit kapittel frå boka eller be biblioteket sitt om å eventuelt kjøpe inn eit eksemplar når vi søker i "norske

fagbibliotek". Fredrik Solvang Pettersen (personleg kommunikasjon, 19. november 2021) seier også i e-posten at oppdateringa som kom i november 2021 var ikkje slik den var lova i NERS #6702, og at Exlibris seier dei jobbar vidare med denne foreløpige løysninga, sidan det var meir komplisert enn dei trudde, men har ikkje kome med noko tidsplan for når dette skal løysast.

I tillegg til dublett problematikken på e-bøker i Oria er ei anna kjent utfordring med e-bøker at kvaliteten på metadata kan variere veldig, og at det kan ta lang tid før postar frå CZ i Alma blir oppdatert. Noko som fører til at det kan ta lang tid før at mangelfulle postar får ein tilfredsstillande kvalitet. Denne problematikken presenterte Unni Knutsen på BIBSYS-konferanse i 2016 under foredraget *Metadata i Almas Community Zone (CZ)* der undersøkingar frå Universitet i Oslo var bakgrunnen (Knutsen, 2016, s. 2).

Med utfordringane med metadatakvalitet og dublett problematikken som bakteppe, vil det være interessant å samanlikne metadatane på e-bøker frå CDI og CZ i Oria. Dette for å betre kunne ta ei avgjerd om kva tiltak høgskolebiblioteket eg jobbar ved, skal gjere for å redusere dublett problematikken på e-bøker i Oria. Samstundes må ein være sikker på at vi har tilstrekkeleg gode metadata for å sikre ei god atffinning for våre brukarar. Alternativa er:

- behalde dublettane i Oria slik dei er i dag
- skru av postane frå CDI, og berre behalde kapittel informasjonen frå CDI
- skru av postane frå CZ

I denne undersøkinga vil data bli samla inn frå Oria instansen til biblioteket ved Høgskolen i Innlandet. Denne Oria instansen er opent tilgjengeleg for alle, datane i denne masteroppgåva, er dermed tilgjengeleg for andre som ynskjer å undersøke det same.

3.1.1 Alma og BIBSYS-konsortiet

BIBSYS-konsortiet består av omlag 80 fag- og forskingsbibliotek som nyttar det internasjonale biblioteksystemet Alma og discovery vektøyet Oria, den norske versjonen av Primo, frå Exlibris (Unit, 2020a).

Alma består av tre katalogar: Institution Zone (IZ), Network Zone (NZ) og Community Zone (CZ). Institution Zone er den lokale katalogen til eit bibliotek med oversikt over materialet bibliotek har tilgang til. I IZ finn ein kopi av postane eit bibliotek eig frå NZ, som ein samarbeidar med resten av BIBSYS-konsortiet for trykte ressursar, og CZ for dei e-ressursane (Faggruppe for Resource Management, 2021a).

Network Zone er ein felles base med bibliografiske postar som skal være til nytte for alle biblioteka i BIBSYS-konsortiet. Alle i konsortiet samarbeidar om dei bibliografiske postane, men det er Nasjonalbiblioteket som har ansvaret for den norske nasjonalbibliografien, og opprettar bibliografiske postar for alle norske titlar i NZ. Resource Description and Access (RDA) har sidan 1. oktober 2020 vore det gjeldane katalogiseringsregelverket i BIBSYS-konsortiet. Sidan postane i NZ er oppretta på ulike tidspunkt, der ulike regelverk har vore gjeldane, kan ein dermed sjå postar med ulik kvalitet og standarardar (Faggruppe for Resource Management, 2021a).

BIBSYS-konsortiet samarbeider ikkje berre om postane i NZ, men også det felles autoritetsregisterer BARE som opprettar ei lenke mellom den bibliografiske posten og autoritetsposten for namnet (Faggruppe for Resource Management, 2021a). Noko som gjer at brukarane får autorisertesøkeinnangar på forfattarnamn, slik dei kan finne alt ein forfattar har skrevet.

BIBSYS-konsortiet har tilgang til ein stor base med postar for elektroniske ressursar i CZ i Alma, derfor har BIBSYS-konsortiet som ein hovedregel at ein ikkje skal opprette bibliografiske postar for e-ressursar i NZ, men heller kople seg på posten som er i CZ. Det er ein stor fordel for oss i BIBSYS-konsortiet at vi har tilgang til delte postar både i NZ og CZ, sidan biblioteka kan sparar mykje tid og ressursar når det berre er nødvendig at eit bibliotek katalogiserar ein ressurs. For e-ressursar er dette ekstra nyttig med tanke på at når ein kjøper pakkar med e-bøker blir det gjerne aktivert tusenvis av titlar samstundes, desse e-bøkene hadde vore uoverkommeleg å registrere manuelt (Faggruppe for Resource Management, 2021a).

«For fysisk materiale registreres beholdningen i form av eksemplarposter (Items) knyttet til beholdningsposter (Holdings). Beholdningspostene forteller oss hvor

eksemplaret befinner seg. Beholdning av e-ressurser beskrives i form av portefolier» (Faggruppe for Resource Management, 2021a). For dei fleste nettressursar, slik som tekst, lyd, bilde eller video som er tilgjengeleg via nett, vil ein finne ein bibliografisk post i CZ som ein kan aktivere portefolien på. Ein kopi av posten vil da bli oppretta i både IZ og NZ når ein aktiverar ein portfolio i CZ (Faggruppe for Resource Management, 2020).

3.1.1.1 Katalogiseringspraksis i CZ

CZ er eit delt katalog over e-ressursar for alle bibliotek rundt omkring i verda som brukar Alma, ikkje berre for oss i BIBSYS-konsortiet. Katalogen består av bibliografiske metadata, autoritetsregister for blant anna namn og emner, og ein Central Knowledge base for elektroniske ressursar frå leverandørane (Kortick, u.å. s. 4-5).

Fordelen med ein delt katalog er at biblioteka som brukar Alma sparar tid ved at ei bok berre treng å bli katalogiser ein gong, og ein får automatiske oppdateringar om andre bibliotek, eller leverandøren, gjer endringar i metadatane (Kortick, u.å, s. 15). Eit viktig poeng med metadata i CZ er at «Changes made to Community Catalog records should focus on improving and increasing the quality of the records, with a special focus on improving discoverability» (ExLibris, 2014, s. 7) og “For any changes that would not benefit the broader community, users may copy the record to the local Institution Zone catalog” (ExLibris, 2014, s. 8). Dette vil seie at endringar som gjerast i CZ skal være til nytte for fleire enn det biblioteket ein tilhøyrar slik at atfinninga skal være bestmogleg, for flest moglege.

Standaren for postane i CZ er at det berre skal vare ein post for kvar elektroniske tittel. Brukarar kan endre og legge til data i postane, men det er ikkje lov å opprette dubletter for same tittel (ExLibris, 2014, s. 6). Eit problem med denne praksisen er «Since we have a one record policy in the CZ, the issue of languages has been the most challenging part of the solution. The current decision is that the language of the item dictates the language for cataloging» (ExLibris, 2014, s. 7). Noko som vil seie at for eksempel emneord på ei engelsk bok skal være på engelsk, og ikkje på det språket den som ynskjer å forbetrar posten brukar.

Postane som er i CZ skal være leverandørnøytrale. Informasjon om leverandør skal heller registrerast på behaldningsnivå i Alma. Postane i CZ hentar informasjon frå mange ulike kjelder og er dermed ikkje identisk til «the source record». RDA er det foretrukne katalogiseringspråket og det er oppfordra til at om ein skal endre noko i ein post at ein da brukar RDA, men om ein berre skal legge til ekstra informasjon brukar katalogiseringspråket som alt er brukt (ExLibris, 2014, s. 6).

I dokumentet *Alma Community Catalog: Cataloging Standards, Policies, Rights, and Responsibilities* (ExLibris, 2014) finn ein ExLibris sine retningslinjer for korleis metadata i CZ skal være. Disse retningslinjene for CZ godtar fleire ISBN i 020 \$a for e-bøker. Det er ikkje lov å fjerne ISBN som alt ligger inne. Ein er oppfordra til å legge inn ISBN som skiller dei ulike tilbyderane i 020 \$q. Det er ønskeleg med ISBN for den trykte boka i 776 \$z . ExLibris oppfordra ein til å legge til Marc-felt 505 og 520 for innhald og samandrag når det er mogleg (ExLibris, 2014, s. 7).

Mangelfulle postar er ikkje uvanleg i CZ, noko som hindrar attfinninga. Alle som har behaldning på ein post i CZ kan berike posten. Om ein gjer endringar vil ein i posten i CZ sjå kva institusjon som har gjort endringa ved at ein ser teksten «Contributed by xx». Med ujamne mellomrom blir endringar ein gjer i CZ skreve over når EX Libris oppdatera postane i CZ med ein import der gamle postar blir kobla med nye postar frå ein av ExLibris sine leverandørar. Derfor anbefaler ein ikkje i BIBSYS-konsortiet å bruke tid å fullkatalogisere postar i CZ. Det er berre lokale felt, slik som lokale emneord, som ikkje blir overskreve i ein slik import (Faggruppe for Resource Management, 2020).

3.1.1.2 Katalogisering av e-bøker i BIBSYS-konsortiet

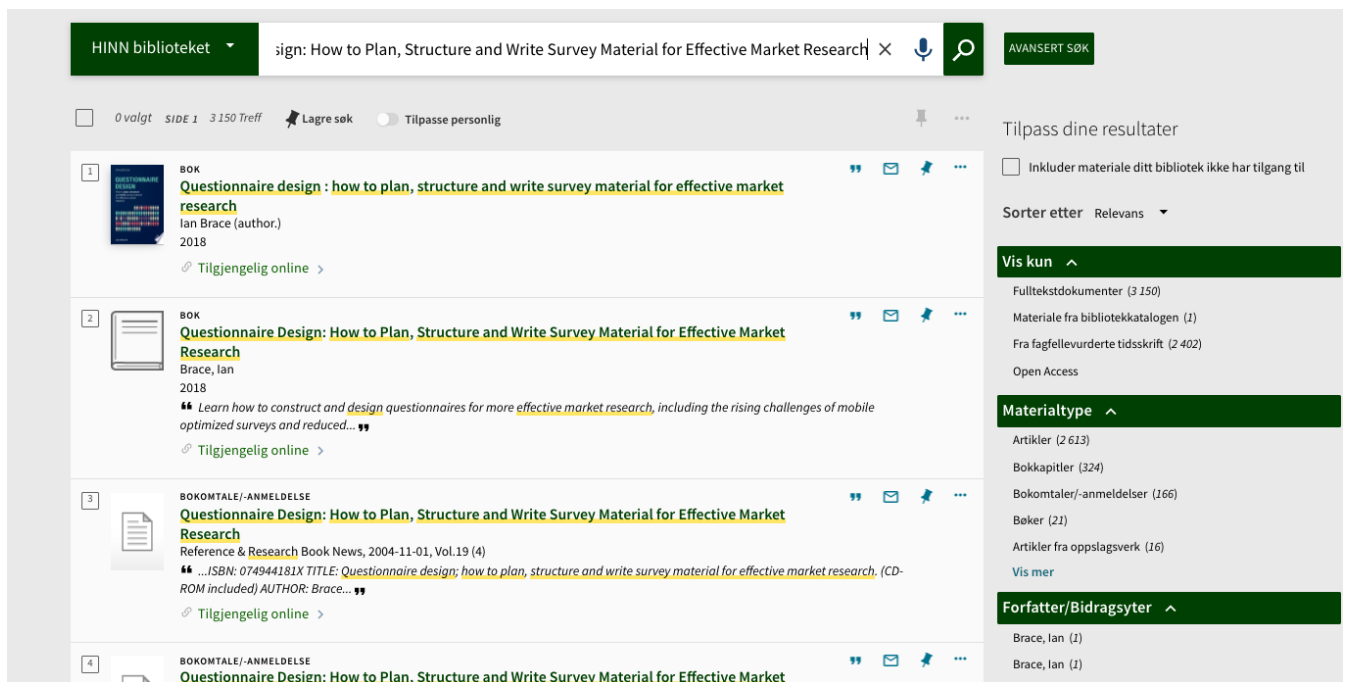
BIBSYS-konsortiet følger eksempla *Faggruppe for Resource Management* har laga for korleis ein skal katalogisere ulike dokumenttypar etter RDA. I rettleiinga for *E-bok med personlig forfatter* får ein informasjon om kva MARC-felt ein skal bruke og kor ein i RDA finn informasjon om dei ulike felta (Faggruppe for Resource Management, 2021b).

Nokre ting ein kan trekke fram frå denne retningslinja er at tidspunktet for utgivinga alltid skal registrerast. «Om kilden bare har året for copyright, registreres tidspunkt for

utgivelse i skarpe klammer og copyrightåret registreres i en egen instans av 264 \$c med andreindikator 4» (Faggruppe for Resource Management, 2021b). Ein skal berre registrere dei ISBN nummera som gjeld for den e-ressursen ein lagar post for. I BIBSYS-konsortiet skal ein oppgi i \$4 kva rolle personen har når den skal autoriserast mot BARE, for slik som personar i marcfelt 100 og 700 (Faggruppe for Resource Management, 2021b). Koden \$4 aut er for forfattar og \$4 edt er for redaktør (Library of Congress, 2013).

3.1.1.3 CDI

Når ein søker i Oria får ein metadata frå biblioteksystemet Alma og Central discovery index (CDI), som er ein database over artiklar, e-bøker og andre e-ressursar (Unit, 2020). CDI består av 4,5 milliardar postar der 755 millionar er bøker og e-bøker, mens 28 millionar er bokkapittel (ExLibris, u.å.c). I CDI finn ein innhald frå over 2000 kjelder. Derfor har ein behov for å slå saman postar som er for den same tittelen, men frå ulike kjelder for å gjere trefflista meir brukarvennleg. Om postar deler dei same identifikatorer, slik som ISBN eller EISBN, kan dei bli slått saman (ExLibris, u.å.e). Denne *match and merge* prosessen med å slå samane postar skal være leverandør nøytral, slik at postar frå enkele ten utgivarar ikkje skal bli prioritert (ExLibris, u.å.e). I CDI blir artiklar som det finnes fleire postar av slått samane til ein berika post, dette skjer ikkje med e-bøker sidan postane er aktivert frå ulike indeksar, CZ og CDI, som ikkje kan slåast saman (Unit, 2020b). Derfor opplever vi ofte at når vi søker i Oria at vi får treff både frå CDI, og den lokale posten frå CZ i Alma når vi søker etter e-bøker, slik ein kan sjå eit eksempel på i skjermdupen i *Figur 3.1.1.3-1*.



Figur 3.1.1.3-1 Skjermdump av ei trefflista i Oria der ein har søkt på ein kjent tittel der treff nr 1 er frå CZ og treff nr 2 er frå CDI

Det tar opptil 72 timar før postar for e-bøker frå CDI er på plass i Oria etter at tittelen er aktivert Alma (Unit, u.å.). Noko som vil sei at det kan ta opp mot tre dagar før ein får treff på metadata frå CDI posten i Oria. Posten frå CZ kjem normalt 24 timar etter e-boka er aktivert i Alma.

For at ei pakke med e-ressursar skal bli markert som at den er i CDI må minst 80% av innhaldet være indeksert. For nokre leverandørar slik som ProQuest og Gale er dekninga 100%. (ExLibris, u.å.d). EBSCO er eit eksempel på ein leverandør der ein ikkje har tilgang til metadata, utan om via andre leverandørar pakkar, som er indeksert i CDI. I følgje Breeding (2014, s. 14) ynskjer ikkje EBSCO å dele sine metadata med konkurrerande leverandørar sitt discovery-verktøy, sidan dei også er leverandør av sitt eige søkeverktøy.

3.1.2 Oria

Oria er den norske versjonen av discovery-verktøyet Primo frå ExLibris, der ein kan søke etter både trykte og elektroniske ressursar i same system. Dette utdypast i sitatet under:

Oria er en felles portal til det samlede materialet som finnes ved de fleste norske fag og forskningsbibliotek. Supplert med en mengde elektronisk materiale fra åpne kilder, gir Oria en enhetlig tilgang til materiale som bøker, elektroniske bøker, tidsskrifter, elektroniske tidsskrifter, dokumenter, artikler, musikk og filmer (Unit, 2021).

På eit amerikansk universitetsbibliotek opplevde Klatt og Meeks (2020) ei markant auke i bruken av bibliotekets e-bøker når dei fekk samla alle ressursane biblioteket hadde i eit søkegrensesnitt, etter at dei hadde gått over til Alma og discovery verktøyet Primo. Ein kan tenke seg til at grunnen for auka er at mange brukarar ikkje har kjennskap til dei spesialdatabasane biblioteket har tilgang til, som er ein stand brukarane kunne ha søkt etter e-bøker.

I Oria kan brukarane søke via enkelt søk eller avansert søkt. Ved å bruke enkelt søk søkar brukaren i alle felt og kan søke på forfattar, tittel og emneord. Brukaren kan avgrense søket i etterkant ved hjelp av fasettar, for å få ei meir oversikteleg og relevant treffliste. Ein kan blant anna avgrense på emne, forfattar, år, dewey, språk og materialtype slik som tidsskrift eller e-bøker (Unit, 2019). Når Niu et al. (2014, s. 427) såg på bruken av fasetter i Primo såg dei at det berre var 9,7% av søka som vart avgrensa ved å bruke fasettar.

Når ein brukar søkar via avansert søk kan ein avgrense søket når ein setter det opp med dei same kriteria som finnes i fasettane, ved å for eksempel velje materialtype og utgivingsdato for eit emne før ein søker (Unit, 2019). Søket kan så avgrensast ytlegare med fasettane om det er behov for det.

Det fine med å være ein del av eit konsortium er at brukarane i Oria lett kan få tilgang til andre norske fag- og forskningsbibliotek sine bøker om dei velje «Norske fagbibliotek», eller «Alle bibliotek» som nokre bibliotek har kalla instansen, i nedtreksmenyen ved sidan av søkeboksen i Oria. (Unit, 2019). Brukarane kan da få ei oversikt over materiale biblioteket ein tilhøyrar ikkje har, og kan komme med fjernlånsynskjer.

I Oria har brukarane tilgang til mange nyttige funksjonar slik som at ein kan sende referansen til ulike referanseverktøy, eller få forslag til korleis referere ved å trykke på

«Sitér», ein kan nytte seg av ein browse funksjon for trykte bøker ved å klikke på «Bla i liste» og velje om ein vil bla i emne, forfattar, tittel eller hylleplassering. Brukarane kan også lagre søk til seinare bruk. Ved å logge seg på «Min konto» får brukaren oversikt over sine lån, reserveringa og bestillingar, og kan fornye låna sine sjølv (Unit, 2019).

På fullvisinga av posten i Oria kan brukaren søke vidare på fleire bøker av same forfattar, eller bøker med same emnerord, ved å trykke på orda som er lenker til vidare søk. Det er ikkje mogleg å bruke dewey-nummer for vidare utforsking i fulltekstvisninga av ein post, berre via bruk av fasettar når ein har søkt via enkelt søk, eller avgrense til dewey-nummer når ein søker i avansert søk.

3.1.2.1 Dubletter i Oria

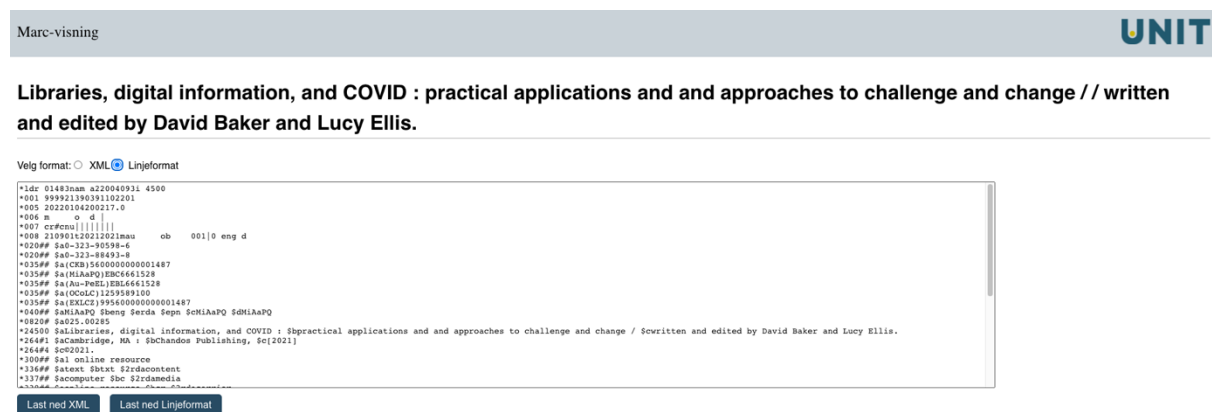
For å unngå dubletter i Oria kjøres det ein DEDUP prosess. «The goal of this process is to avoid duplicated results for the same book in different libraries within the same institution, or consortium» (ExLibris, u.å.f, s. 19). Når det er titlar som samsvarar blir den eine posten valt ut som den føretrekte posten, og ein post som bli kalla «dedup-merge» blir oppretta. Denne posten inneheld data frå den føretrekte posten i tillegg til data frå andre posten som ikkje var valt. Det er berre dedup-merge postane som er søkbare i Oria (ExLibris, u.å.f, s. 19).

I tillegg til ein DEDUP prosess er det også ein FRBR-prosess i Oria der ikkje ulike versjonar av ein post blir fjerna, men heller samla under ein samlepost. Ved å klikke på «Klikk for å se alle utgaver» vil ein kunne sjå alle variantane av ein tittel med for eksempel ulike utgåver, ulike språk eller format (ExLibris, u.å.f, s. 20).

3.1.2.2 PNX

Postar som er synlige i Oria har gått igjennom ein prosess der data fra ulike metadataformat som MARC, DublinCore, XML og Web Archive blir konvertert og normaliset til Primo Normalized XML, også kjent som PNX. På denne måten kan ein presentere metadata frå mange ulike format og kjelder i same søkeverktøy (ExLibris, u.å a, s. 4).

Postar frå CZ har ein moglegheita for sjå kva MARC-felt som er brukt ved å trykke på «Bibliografisk beskrivelse (MARC21)», som ein finn under *Andre ressurser* når ein har trykt seg inn på fullvisninga av ein post. Ein kan sjå ein eksempel av korleis MARC-visnigna er i *figur 3.1.2.2-1*. Denne mogleigheita for å sjå MARC-visning av ein post finnes ikkje for postane frå CDI.



Figur 3.1.2.2-1 Skjermdump av Marc-visning av ein post frå CZ

Det ein kan gjere, for både postane frå CZ og CDI, er å legge til `showPnx=true` i slutten av URL'en når ein er inne på fullvisninga av ein post i Oria for å sjå kva metadata som ligg bak fullvisninga (ExLibris, u.å.a, s. 5). Ein får da opp ein PNX fil som har gått igjennom ein normaliseringsprosses der det er oppretta 13 ulike deler i fila, der datane frå original postane er grunnlaget for PNX-fila, eksempel på ei slik visning ser ein i *figur 3.1.2.2-2*. Dei 13 delane er: *Control, display, links, search, sort, facets, dedup, FRBR, delivery, enrichment, ranking, addata, browse*. Kvar av disse 13 delane har ulike funksjonar slik som at *display* er for å presentere ei post for ein brukar i Oria. (ExLibris, u.å a, s.15). Fordelen med at alle postane har fått same format er at ein da kan søke samstundes i Oria på alle typar ressursar og at ein da har ein standard måte å vise fram disse søkeresultata på. «It also allows us to organize and group records in different ways to create search scopes, facets, and enable virtual shelf browsing» (ExLibris, u.å.f, s. 8).

```

<?xml version="1.0" encoding="UTF-8"?>
<record xmlns="http://www.exlibrisgroup.com/xsd/primos/primos_nm_bib" xmlns:sear="http://www.exlibrisgroup.com/xsd/jaguar/search">
  <control>
    <sourceid>71617578720002201</sourceid>
    <sourceid>BIBSYS_ILS</sourceid>
    <recordid>BIBSYS_ILS71617578720002201</recordid>
    <addsrcid>999921390391102201</addsrcid>
    <sourceformat>MARC21</sourceformat>
    <sourcesystem>Alma</sourcesystem>
    <almaid>47BIBSYS_UBIS:51120765770002208</almaid>
    <almaid>47BIBSYS_HIOA:51148266460002212</almaid>
    <almaid>47BIBSYS_HH:5185585200002214</almaid>
    <almaid>47BIBSYS_NETWORK:71617578720002201</almaid>
  </control>
  <display>
    <type>book</type>
    <title>Libraries, digital information, and COVID : practical applications and approaches to challenge and change</title>
    <contributor>David Baker</contributor>
    <contributor>Lucy Ellis</contributor>
    <publisher>Chandos Publishing</publisher>
    <creationdate>2021</creationdate>
    <format>1 online resource</format>
    <identifier>9781108849386</identifier>
    <subject>Libraries -- Information technology; Electronic information resources; Digital libraries; COVID-19 Pandemic, 2020 -- Social aspects; Libraries and public health</subject>
    <language>eng</language>
    <source>Bibliotekatalogen</source>
    <ds03>Includes bibliographical references and index.</ds03>
    <ds07>
    <ds09>71617578720002201</ds09>
    <ds10>025.00285</ds10>
    <ds12>Libraries, digital information, and COVID</ds12>
    <ds22>written and edited by David Baker and Lucy Ellis.</ds22>
    <ds28>Engelsk</ds28>
    <ds32>Chandos digital information review series. Chandos digital information review series.</ds32>
    <ds38>
    <ds150>Includes bibliographical references and index.</ds150>
    <ds190>Libraries -- Information technology</ds190>
    <ds190>Electronic information resources</ds190>
    <ds190>Digital libraries</ds190>
    <ds190>COVID-19 Pandemic, 2020 -- Social aspects</ds190>
    <ds190>Libraries and public health</ds190>
    <ds190>Electronic books</ds190>
    <ds191>Libraries -- Information technology ; Electronic information resources ; Digital libraries ; COVID-19 Pandemic, 2020 -- Social aspects ; Libraries and public health ; Electronic books</ds191>
    <ds06>Cambridge, MA</ds06>
    <ds46>Electronic books.</ds46>
  </display>
  <links>
    <thumbnail>
    <ln12>
    <ln12>
    <ln16>
  </links>
  <search>
    <creatorcontrib>written and edited by David Baker and Lucy Ellis.</creatorcontrib>
    <creatorcontrib>Baker, David.</creatorcontrib>
    <creatorcontrib>David. Baker</creatorcontrib>
    <creatorcontrib>Ellis, Lucy.</creatorcontrib>
    <creatorcontrib>Lucy. Ellis</creatorcontrib>
    <creatorcontrib>Baker, D</creatorcontrib>
    <creatorcontrib>Ellis, L</creatorcontrib>
  </search>

```

Figur 3.1.2.2-2 Skjermdump av korleis delar av ei PNX fil ser ut

Metadata som er i kort og fullvisning av ein post i Oria hentar informasjonen sin frå den delen i PNX fila som heiter *Display*, sjå Figur 3.1.2.2- 3. Data elementa i PNX fila har tatt utgangspunkt i Dublin Core sine *element set* sidan denne metadatastandaren kan dekke mange ulike ressursar og formål. Nokre av Dublin Core felta har blitt endra på, og nokre felt har blitt lagt til (ExLibris, u.å b).

```

<display>
<type>book</type>
<title>Libraries, digital information, and COVID : practical applications and and approaches to challenge and change</title>
<contributor>David Baker$$QDavid Baker; Lucy Ellis$$QLucy Ellis</contributor>
<publisher>Chandos Publishing</publisher>
<creationdate>2021</creationdate>
<format>1 online resource</format>
<identifier>$$CISBN$$V0-323-90598-6; $$CISBN$$V0-323-88493-8</identifier>
<subject>Libraries-- Information technology;Electronic information resources;Digital libraries;COVID-19 Pandemic, 2020--- Social aspects;Libraries and public health</subject>
<language>eng</language>
<source>Bibliotekatalogen</source>
<lds03>Includes bibliographical references and index.</lds03>
<lds07>$$INB</lds07>
<lds09>7161757872002201</lds09>
<lds10>025.00285</lds10>
<lds12>Libraries, digital information, and COVID :</lds12>
<lds22>written and edited by David Baker and Lucy Ellis.</lds22>
<lds28>Engelsk</lds28>
<lds32>Chandos digital information review series.$$QChandos digital information review series.</lds32>
<lds38>$$INB</lds38>
<lds150>Includes bibliographical references and index.</lds150>
<lds190>Libraries -- Information technology</lds190>
<lds190>Electronic information resources</lds190>
<lds190>Digital libraries</lds190>
<lds190>COVID-19 Pandemic, 2020- -- Social aspects</lds190>
<lds190>Libraries and public health</lds190>
<lds190>Electronic books</lds190>
<lds191>Libraries -- Information technology ; Electronic information resources ; Digital libraries ; COVID-19 Pandemic, 2020- -- Social aspects ; Libraries and public health ; Electronic books</lds191>
<lds06>Cambridge, MA</lds06>
<lds46>Electronic books.</lds46>
</display>

```

Figur 3.1.2.2-3 Skjermdump av innholdet i eit Display felt i ei PNX-fil

Feltet *publisher* under *display* i ei PNX-fil er for eksempel henta frå MARC21 felt 260 \$a og \$b, mens feltet *subject* er frå MARC21 6XX felt. *Title* i PNX fila er frå 245 \$a og \$b. *Creationdate* står for det året ressursen var laga eller publisert og er henta frå MARC21 008/07-10 og 260 \$\$c. *Creator* hentes fra 246 \$c, og i dei tilfella det ikkje er noko 245 \$c hentas informasjonen om forfattar frå 100 felte. Mens *Description* er frå MARC felt 502, 505 og 520 (ExLibris, u.å.b).

3.2 Utval

E-bøker kan kjøpast på mange ulike måtar for eit bibliotek, slik som via kjøp av e-bok pakkar som kan ha hundretusen vis av titlar, Evidence based acquisition (EBA) der ein får eit utval av titlar som blir kjøpt inn til biblioteket etter ein har sett på bruken etter eit år, eller kjøp av enkelt titlar. Denne undersøkinga har valt å sjå på nyinnkjøpte enkelt titlar, sidan bøker aktivert i pakker eller via EBA ikkje blir aktivert ein og ein, men i bolker. Enkelttitlane blir manuelt aktivert ein og ein. Om ein skal gjerne endringar på metadata på e-bøkene vi kjøper inn, er det berre realistisk å tenke at det er enkelttitlane ein kan ha moglegheit for å endre og forbetre metadata på.

For å få ei oversikt over kva enkelt titlar som er kjøpt inn har eg i denne undersøking brukt ein rapport i Alma Analytics over e-bøker som er aktivert i oktober 2021. Ved å bruke denne rapporten får ein fange opp e-bøker som er kjøpt inn på alle dei fem studiestadsbiblioteka ved Høgskolen i Innlandet, uavhengig av kven som har kjøpt dei inn og kva fagområde titlane tilhøyrar. Dermed kan ein få variasjon i

fagområda bøkene tilhøyrar. Sidan høgskolebiblioteket ikkje berre kjøpe inn e-bøker utgitt i det løypane året vil eg få eit utval av e-bøker som er publisert i ulike år.

Tabellen i *figur 3.2-1* visar korleis utvalet i denne undersøkinga vart redusert frå 74 til 42 titlar. I oktober 2021 var det aktivert 74 titlar ved høgskolebiblioteket på Høgskolen i Innlandet. Av desse 74 titlane er det 42 postar som vil bli undersøkt. Dei andre 32 postane var enten feil aktivert, at posten frå CDI ikkje hadde komme inn i Oria enda, sjølv om dei 72 timane som er lova hadde gått, eller at tittelen hadde tre eller fleire postar i Oria frå CDI, noko som gjer det vanskeleg å samanlikne på ein lett måte, og ein post viste seg å være eit tidsskrift. Derfor har eg valt å berre undersøke dei 42 titlane som det er mogeleg å samanlikne på ein enkelt måte, sidan ein i Oria finn ein post frå CZ og ein post frå CDI for desse. Ved å berre undersøke 42 titlar vil ein få eit inntrykk av korleis metadata kvaliteten er, men det er ikkje ein fasit for alle titlane som finnes i basen.

	Aktiver t titlar	Titlar som enten var feilaktivert, ikkje å finne i Oria, eller hadde tre eller fleire postar	Titlar med ein post frå CZ og ein frå CDI
Oktober 2021	74	32	42

Figur 3.2-1 Tabell over utvalet i datasamlinga

3.3 Avgrensingar

Brukaroppgåvene i IFLA LRM user tasks som metadata skal oppfylle er finne, identifisere, velje, skaffe og utforsk. Det er metadatafelt i Oria for forfattar, år, tittel, beskriving med innhaldsforteikning og samandrag, utgåve og emneord som vil bli undersøkt i denne masteroppgåva for å sjå om brukaroppgåvene er oppfylt. Desse er valt fordi dei samsvarar med felt andre liknande undersøkingar har sett på.

Brukaroppgåva skaffe vil ikkje bli undersøkt nærare sidan alle lenkene til e-bøkene i datasamlinga førte til riktig bok i datainnsamlinga. I datainnsamlinga vart postane frå dei to kjeldene søkt opp via tittel og lenka på posten førte til riktig e-bok hos leverandøren.

I denne undersøkinga har det ikkje blitt sett på bruk av ISBN og dewey nummer. Avanserte brukarar kan nok bruke dette som søkeinngangar, men sidan dei fleste brukarar nok ikkje har noko bevist forhold til denne måten å søke på, er ikkje desse metadatafelta undersøkt.

I casestudien er det dei metadata ein brukar finn i fullvisninga av ein post i Oria som vil bli undersøkt. Denne visninga har sine manglar, men sidan det ikkje er mogleg å samanlikne MARC-felt for postane frå CZ og CDI, er fullvisninga av posten mest hensiktsmessig å bruke for samanlikninga. Ein kan ikkje sjå i Oria om eit forfattarnamn er autorisert. Nokre autoritetar har ein parentes med fødsel, og eventuelt dødsår, i Oria, men ikkje alle, så ein kan ikkje sjå om ein namn er autorisert i fullvisninga. I fullvisninga av ein post kan ein heller ikkje sjå kva emneordssystem som er brukt. Undersøkinga vil derfor ta for seg talet på emneord. Kvaliteten på emneorda blir ikkje sett på, og det blir heller ikkje samanlikna kva emneord dei to kjeldene brukar på ein tittel. Ved å sjå på antall emneord vil ein få eit inntrykk av om ein brukar har mange eller få søkeinngangar på postane.

3.4 Datainnsamling

Denne undersøkinga tar utgangspunkt i metadata som ein brukar blir presentert for i discovery-verktøyet Oria. Oria er nok denne inngangen studentar og dei tilsette oftast brukar for å finne e-bøkene høgskolebiblioteket har kjøpt inn frå ulike leverandørar, og ikkje e-bok leverandørens si søkeside i spesial databasar. I datainnsamlinga er det søkt i Oria på e-bøkenes tittel slik dei framkjem av ein rapport frå Alma Analytics som vart henta ut over titlar som var aktivert i oktober 2021. Ved berre 1 treff i Oria har det blitt søkt ved å ta vekk undertittelen.

For postane frå CZ i Oria kan ein endre visningsformat av metadata til ei MARC visning, denne moglegheita finnes ikkje for postane frå CDI. For å kunne samanlikne metadata må ein da finne andre måtar som vil være hensiktsmessige. Ein kan hente ut metadata i PNX-format via eit java-tillegg i nettlesaren for postane både frå CZ og CDI, som ein da kan køyre spørjingar mot, men dette vil ikkje være hensiktsmessig når kvar tittel, med ein post frå CZ og ein post frå CDI, må kontrollerast mot tittel og kolofonsida i sjølve e-boka. Derfor vil informasjonen ein

brukar blir presentert for under detaljar i fullvisninga av ein post i Oria bli bruk for å kunne samanlikne metadata frå dei to kjeldene.

Data er henta manuelt ut for kvar tittel som er aktivert mellom 1. oktober og 29. oktober 2021 med ein post frå CZ og ein for CDI i Oria i form av ein skjermdump. Dette er ein tidkrevjande prosess som begrensar utvalet. Skjermdump er valt fordi data da blir i eit lett lesbart format til manuell registrering i eit førehandsdefinert kodeskjemaet i Excel, der data skal kontrollerast mot tittel og kolofon sida på sjølve e-boka. Metadata kan også endre seg overtid, så ein kan ikkje sjekke om ein har registrert data korrekt ved å søke i Oria seinare, data må derfor kontrollerast mot skjermdumpane. Undersøkinga blir da eit augneblinksbilete av korleis metadata var når data var henta ut frå Oria 3.-5. november 2021.

Analyse og grafiske framstillingar av data i denne casestudien er utført i Excel.

3.5 Svakheiter med undersøkinga

Ved berre å undersøke 42 e-bok titlar vil ein ikkje få eit fasit svar på korleis metadata i Oria er, men ein vil få eit inntrykk av korleis dei kan være.

Denne undersøkinga tar berre for seg dei metadata ein brukar ser under fullvisninga av ein post i Oria. Interessante metadata som finnes i dei 13 delane av PNX-fila slik som *fasett* og *browse*, som kan vere interessante for attfinning, er ikkje undersøkt i denne studien. Det er berre nokre av dei metadata som ein finn i *display* delen av PNX-fila som er sett på, forfattar, tittel, år, utgåve, emneord og beskriving. Det finnes andre felt i denne visninga som ikkje er sett på i denne masteroppgåva slik som bruk av dewey nummer og ISBN. Forfattar, tittel, år, utgåve, emneord og beskriving er valt sidan dette er felt som tidlegare undersøkingar har sett er viktige for søking og attfinning for ein brukar.

Ein annan svakheit ved undersøkinga er når data skal registrerast i Excel i eit førehandsdefinert skjema er at ein må være konsekvent i registreringa. Det krevjar at ein har gode, målbare førehandsbestemte svaralternativ å fylle ut så det går ann å samanlikne data. Ein kan ikkje berre ha ja og nei spørsmål, må ofte ha eit alternativ til

for at alle aspekt skal bli dekket. Slik som at ein kan ikkje berre spørje om året er riktig eller feil, når det i nokre tilfelle ikkje er registrert noko år.

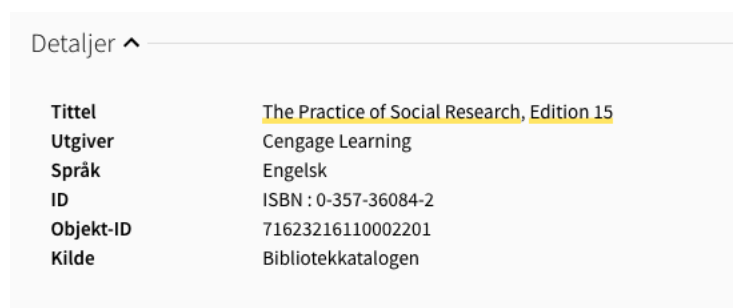
4 Funn

Biblioteket ved Høgskolen i Innlandet aktiverte 74 e-bøker i oktober 2021 der berre 42 e-bøker av desse vert analysert i denne dataanalysen, sidan det var mange av dei 74 e-bok titlane som blant anna hadde fleire enn to postar i Oria, og dermed ikkje kunne samanliknast på ein lett måte.

For å få eit betre innblikk i korleis metadata kan være for ein tittel frå CZ og CDI vil eg fyrst i dette kapittelet gå igjennom ein e-bok tittel for å samanlikne metadata frå CZ og CDI i Oria. Dette for å gi innsikt i korleis heilheita i ein post i søkeverktøyet Oria kan sjå ut. Deretter vil funna i undersøkinga blir presentert i kapitla som følgjer. Fokuset i samanlikninga og i analysen vil være på år, forfattar, tittel, beskriving, utgåve og emneord.

4.1 Gjennomgang av ein tittel

Lengda og kvaliteten på postane i datasamlinga varierer veldig. Nokre er så ufullstendige at det berre er tittel ein vanlig brukar hadde fått treff på. Slik som denne posten som ein ser i *figur 4.1-1* frå CZ som både manglar år, forfattar emneord og beskriving.



Detaljer ^

Tittel	The Practice of Social Research, Edition 15
Utgiver	Cengage Learning
Språk	Engelsk
ID	ISBN : 0-357-36084-2
Objekt-ID	71623216110002201
Kilde	Bibliotekatalogen

Figur 4.1-1 Skjermdump av mangelfull post frå CZ

Mens andre postar i datasamlinga er lengre, men har likevel manglar, slik som tittel nummer 13 frå datasamlinga som ein ser i *figur 4.1-2* med posten frå CZ og *figur 4.1-3* frå CDI, som eg no skal gå nærare inn på. Tittel nummer 13 kan med eit raskt blikk sjå like ute i Oria frå CZ og CDI, men når ein ser litt nærar på dei ser ein fleire ulikeheiter.

Tittel	EcoDesign and sustainability I : products, services, and business models
Medforfatter / Bidragsyter	Yusuke Kishita (editor.) >
Ansvarsangivelse	Yusuke Kishita [and three others] editor.
Utgiver	Springer
Utgiveststed	Gateway East, Singapore
Årstall	2021
Utgave	1st ed. 2021.
Format	1 online resource (X, 532 p. 212 illus., 153 illus. in color.)
Serie	Sustainable Production, Life Cycle Engineering and Management, >
Språk	Engelsk
Dewey	338.927
ID	ISBN : 981-15-6779-4
Beskrivelse	Innhold: Product & Service Design -- Industrial Designers Towards Design Concepts Based on the Water and Fire Themes-- A Review and Comparison of Sustainability Considerations -- Time Axis Design as an EcoDesign Method -- Design of Household Appliances considering Remanufacturing: A Case Study -- Persuasive Design for Improving Battery Swap Service Systems of Electric Scooters -- Service Design of Rehabilitative Exoskeleton for Sustainable Value Creation: A Case Study of Exoskeleton for Stroke Rehabilitation in China -- State-of-the-Art on Product-Service Systems and Digital Technologies -- Material- Service Systems for Sustainable Resource Management -- Designing for Vehicle Recyclability from the Perspectives of Material and Joining Choices -- Eco-Innovation by Integrating Emerging Technologies with ARIZ Method Video Networks of Sustainable Design: the Doughnut Perspective. This book highlights cutting-edge ecodesign research, covering product and service design, smart manufacturing, and social perspectives in ecodesign . Featuring selected papers presented at EcoDesign 2019: 11th International Symposium on Environmentally Conscious Design and Inverse Manufacturing , it also includes diverse, interdisciplinary approaches to foster ecodesign research and activities. In the context of Sustainable Development Goals (SDGs), it addresses the need for the manufacturing industry to design innovations for sustainable value creation, taking into account technological developments, legislation, and consumer lifestyles. Further, the book discusses the concept of circular economy, which originated in Europe and aims to increase resource efficiency by shifting away from the linear economy. Focusing on product life cycle design and management, smart manufacturing, circular economy, and business strategies, and providing useful approaches and solutions to these emerging concepts, this book is intended for both researchers and practitioners working in the broad field of ecodesign and sustainability .
Emner	Sustainable development > Manufacturing, Machines, Tools, Processes > Industrial design >
Objekt-ID	71604329410002201
Kilde	Bibliotekatalogen

Figur 4.1-2 Skjermdump av tittel nummer 13 frå CZ

Tittelen på boka er registrert på same måte med undertittel i begge postane. I posten frå CZ er det nemnt ein redaktør, som ansvarsangivelse saman med tre andre som ikkje er nemnt med namn i parentes. Posten frå CZ har også eit felt der det same namnet er registrert under Medforfatter/Bidragsyter med ein parentes med editor i. I posten frå CDI er det fire namn under feltet forfattar. Denne posten får ikkje fram at boka er redigert.

Ein annan forskjell er at året ikkje er likt for dei to postane. I posten frå CZ er utgjevningsåret 2021, mens frå CDI er det 2020 som er registrert som utgjevningsår. Når ein skjekkar kolofonsida i e-boka er det 2021 som er det riktige året. Posten frå

CZ har også med at dette er 1. utgåve av boka. Dette har ikkje posten frå CDI med. Dette er heller ikkje noko som står i kolonsida i e-boka.

Begge postane frå CZ og CDI har i dette eksempelet eit samandrag av kva boka handlar om. Posten frå CZ har, i tillegg til ei beskriving av innhaldet av boka, med innhaldsinformasjon over kapitela i boka. Ingen av postane har med kven som har skreve dei ulike kapitla.

Tittel	EcoDesign and Sustainability I: Products, Services, and Business Models
Forfatter	Kishita, Yusuke > Matsumoto, Mitsutaka > Inoue, Masato > Fukushige, Shinichi >
Relaterte titler	Sustainable Production, Life Cycle Engineering and Management
Utgiver	Singapore: Springer Singapore Pte. Limited
Årstall	2020
Språk	Engelsk
ID	ISBN: 9789811567780 ISBN: 9811567786 EISBN: 9811567794 EISBN: 9789811567797 OCLC: 1204135226
Beskrivelse	This book highlights cutting-edge ecodesign research, covering product and service design, smart manufacturing, and social perspectives in ecodesign . Featuring selected papers presented at EcoDesign 2019: 11th International Symposium on Environmentally Conscious Design and Inverse Manufacturing , it also includes diverse, interdisciplinary approaches to foster ecodesign research and activities. In the context of Sustainable Development Goals (SDGs), it addresses the need for the manufacturing industry to design innovations for sustainable value creation, taking into account technological developments, legislation, and consumer lifestyles. Further, the book discusses the concept of circular economy, which originated in Europe and aims to increase resource efficiency by shifting away from the linear economy. Focusing on product life cycle design and management, smart manufacturing, circular economy, and business strategies, and providing useful approaches and solutions to these emerging concepts, this book is intended for both researchers and practitioners working in the broad field of ecodesign and sustainability .
Emner	Industrial design > Manufacturing, Machines, Tools, Processes > Sustainable development >
Kilde	Ebook Central Perpetual and DDA

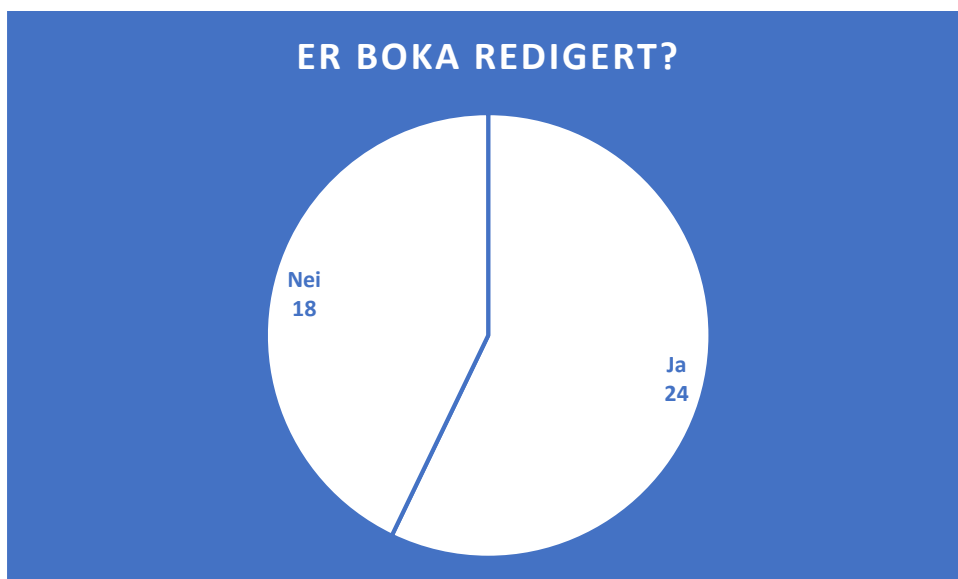
Figur 4.1-3 Skjermdump av tittel nummer 13 frå CDI

Begge postane har tre emneord, for å beskrive ressursen, som er identiske, berre i ulik rekkefølge.

4.2 Presentasjon av funn

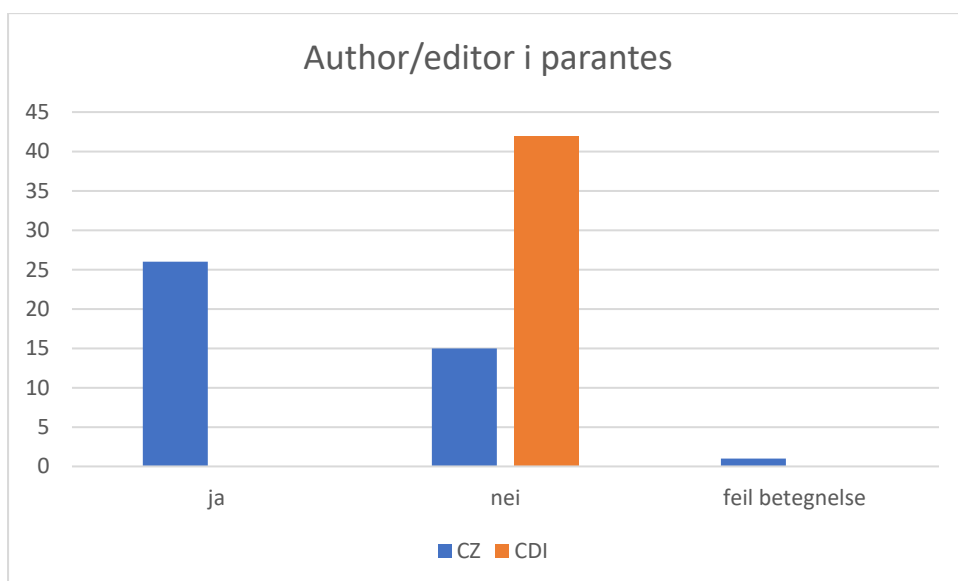
4.2.1 Forfattar

Av dei 42 titlane i datasamlinga var 24 av dei redigert, mens 18 av dei ikkje var det som ein ser i *figur 4.2.1-1*.



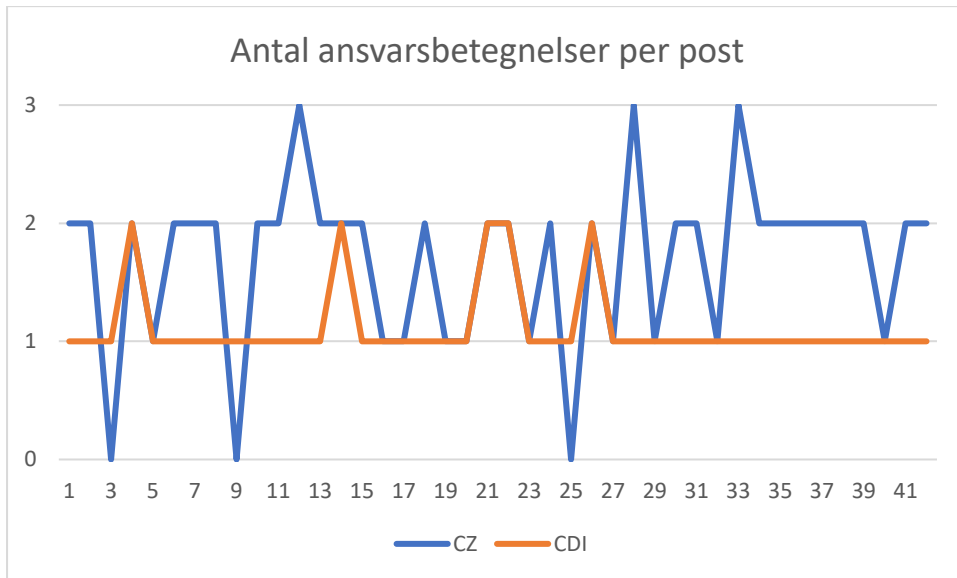
Figur 4.2.1-1 Diagram over antal redigerte bøker

I *figur 4.2.1-2* ser ein at 26 av postane frå CZ har ei markering av kva rolle namnet har med enten author, for forfattar, eller editor, for redaktør, i parentes bak namnet. 15 postar frå CZ har ikkje denne parentesen, mens 1 post har feil betegnelse. Ingen av postane frå CDI har ein parentes med author/editor..



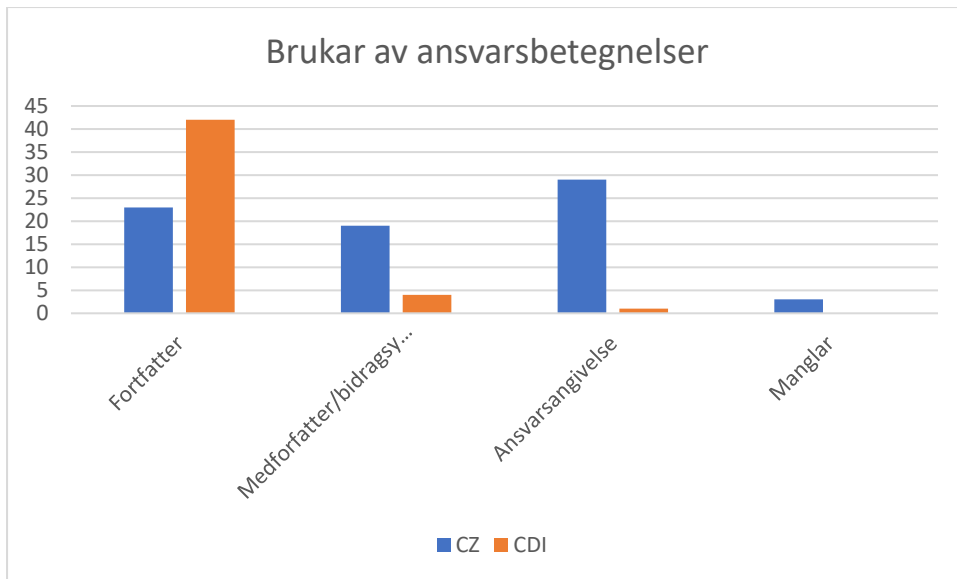
Figur 4.2.1-2 agram over postar med author/editor i parantes

I Oria er det i tillegg til funksjonen author og editor i parentes bak eit namn ulike ansvarbetegnelsar som er brukt slik som forfatter, medforfatter/bidragstyrer og ansvarangivelse. Fleire betegnelsar kan være brukt på same tittel slik ein ser av diagrammet i figur 4.2.1-3.



Figur 4.2.1-3 Diagram over antal ansvarsbetegnelser per post

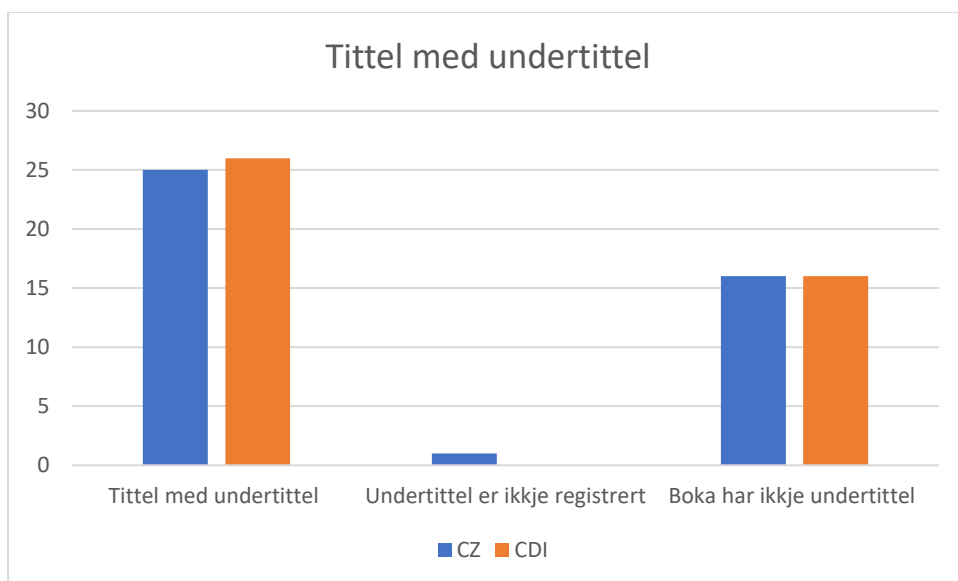
Av funna ser ein at det er ein forskjell i bruk av disse betegningane frå postane frå CZ og CDI. Postane frå CZ, vist med blått i diagrammet, har mange fleire betegnelsar på same post, opp mot tre, mens postane frå CDI, har ofte berre ei betegning som er forfattar. Det er tre postar frå CZ som manglar informasjon om kven som er ansvarleg for innhaldet av tittelen i posten.



Figur 4.2.1-4 Diagram over bruk av ansvarbetegnelser

I diagrammet som er vist i *figur 4.2.1-4* ser ein kva betegnelse som er brukt mest i datasamlinga og ein ser at det er tre postar frå CZ som manglar informasjon om namnet på forfattaren av tittelen. I figur 4.2.1-4 ser ein at postane frå CDI har ein stor overvekt av betegnelsen forfattar. Det er fire postar frå CDI som har både forfattar og medforfatter/bidragssyter og ein post som har forfattar og ansvarsangivelse, men elles er det kun forfattar som er brukt. Det kjem derfor ikkje fram av postane frå CZ om boka er redigert, sidan postane frå CDI stort sett berre brukar betegninga forfattar sjølv om boka er redigert, og ikkje har noko parentes med author eller editor bak eit namn. Av dei 42 titlane er det 24 titlar som er redigerte i denne datasamlinga., får ikkje postane frå CDI fram at boka er redigert eller ei.

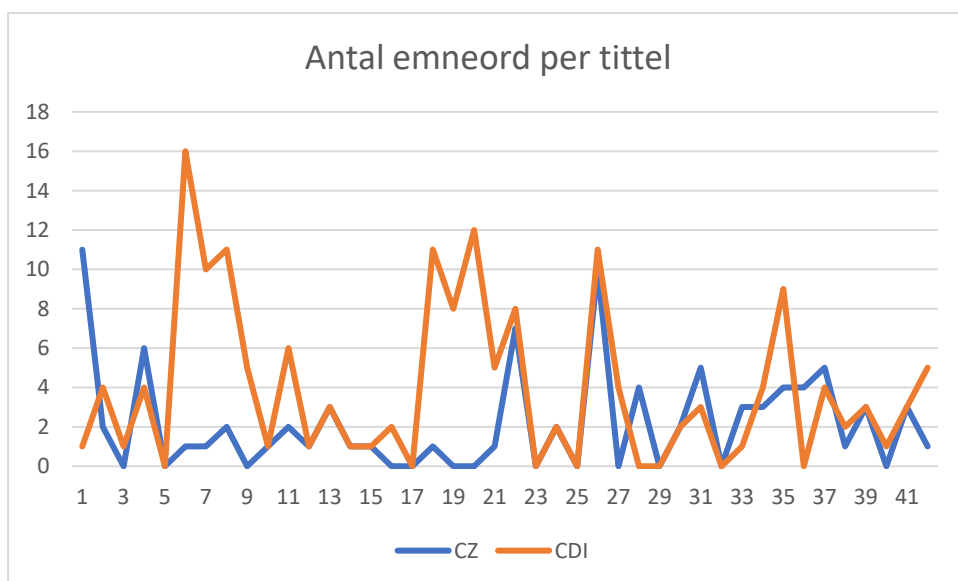
4.2.2 Tittel



Figur 4.2.2-1 Diagram over titlar og undertitlar

I datasamlinga er det 16 titlar av 42 som ikkje har ein undertittel. Dei 26 andre titlane i datasamlinga har ein tittelen med undertittel. Av titlane med undertittel er det berre ei post frå CZ der det finnes ein undertittel som ikkje er registrert som ein ser i diagrammet i figur 4.2.2-1.

4.2.3 Emneord



Figur 4.2.3-1 Diagram over antal emneord per tittel

Det er stor spreiding og lite samsvar i talet på emneord i postane frå CZ og CDI som ein ser i *figur 4.2.3-1*. Den posten som har flest emneord er på ei tittel frå CDI som har heile 16 emneord. Det er fleire postar som har ingen emneord. Frå CZ er det 13 postar som har ingen emneord, mens for CDI er det 8 som har ingen. I *figur 4.2.3-2* ser ein at samanlagt har postane frå CZ 91 emneord, mens postane frå CDI har 165, noko som gir eit gjennomsnitt med nesten 4 emneord per post frå CDI mens det berre er 2 for postane frå CZ.

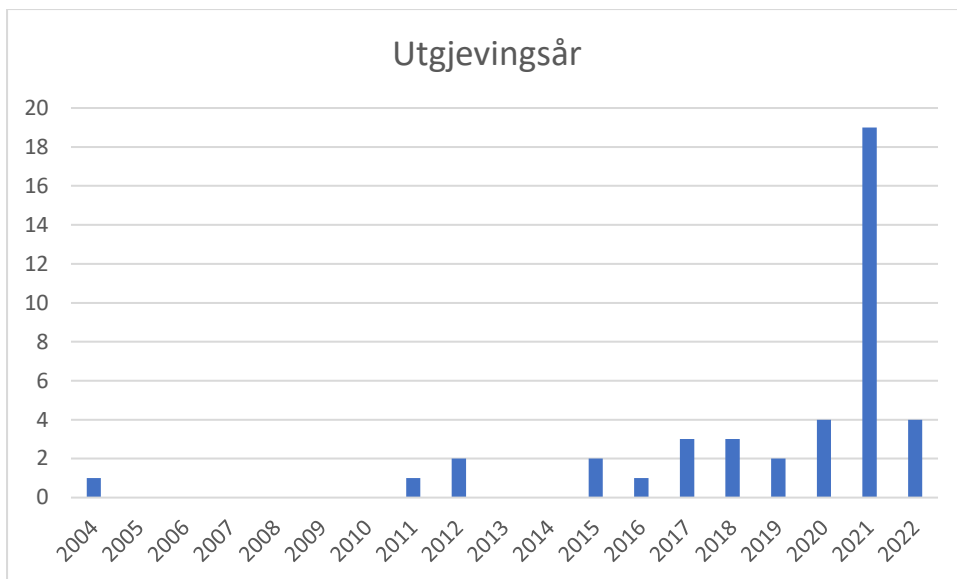
	Postar med emneord	Postar utan emneord	Alle emneord	Gjennomsnitt emneord per post
CZ	29	13	91	2,16
CDI	34	8	165	3,9

Figur 4.2.3-2 Tabell over fordeling av emneord

Alle titlane i datasamlinga er på engelsk. Språket på emneorda i datasamlinga var engelsk for alle emneord utanom på ein post frå CDI der språket var tysk, og ein anna post frå CDI der det var eit emneord på svensk i tillegg til engelske.

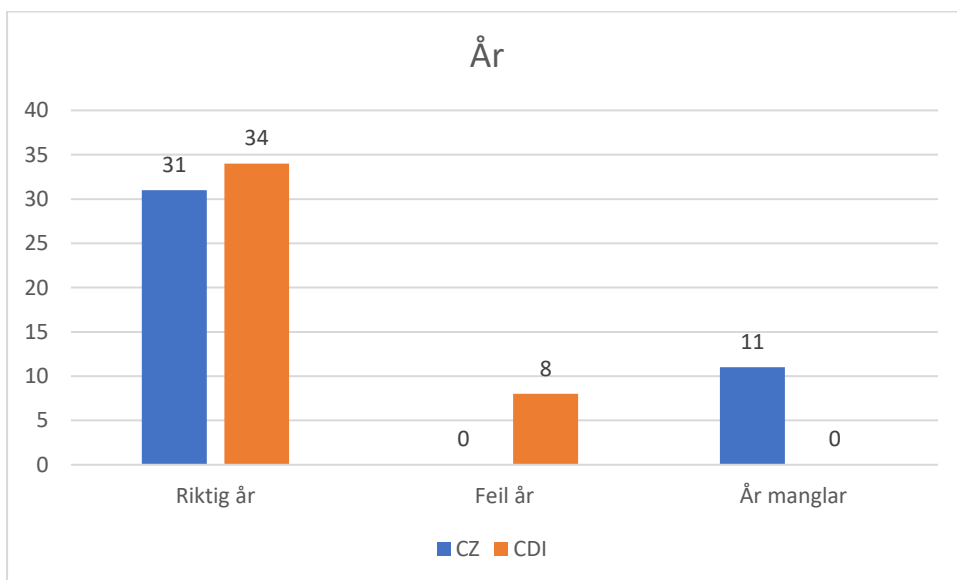
4.2.4 År

Utgjevingsåret på bøkene i denne datasamlinga fordeler seg over fleire år som ein ser i *figur 4.2.4-1*, der dei fleste, 19 av 42, er utgjeven i 2021. Som ein kan sjå av diagrammet i *figur 4.2.4-1* er det fire bøker som det i kolofonsida av boka står at utgjevingsåret er 2022, sjølv om vi har kjøpt og aktivert boka i oktober 2021.



Figur 4.2.4-1 Diagram over utgjevingsår

Når ein samanliknar utgjevingsåret i postane frå CZ og CDI med kolofonsida i e-boka ser ein at det er postane frå CDI som oftast har riktig utgjevingsår registrert vist i figur 4.2.4-2. Det er 31 postar frå CZ som har riktig år, mens det er 34 av 42 frå CDI. Det er heile 11 postar frå CZ som manglar år, mens det er 8 postar frå CDI som har feil år registrert.

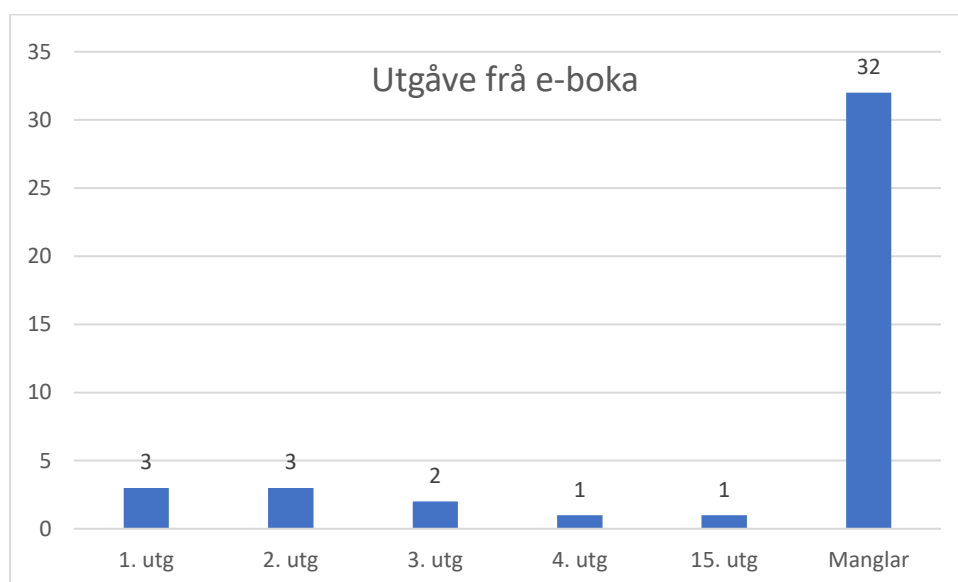


Figur 4.2.4-2 Diagram over år

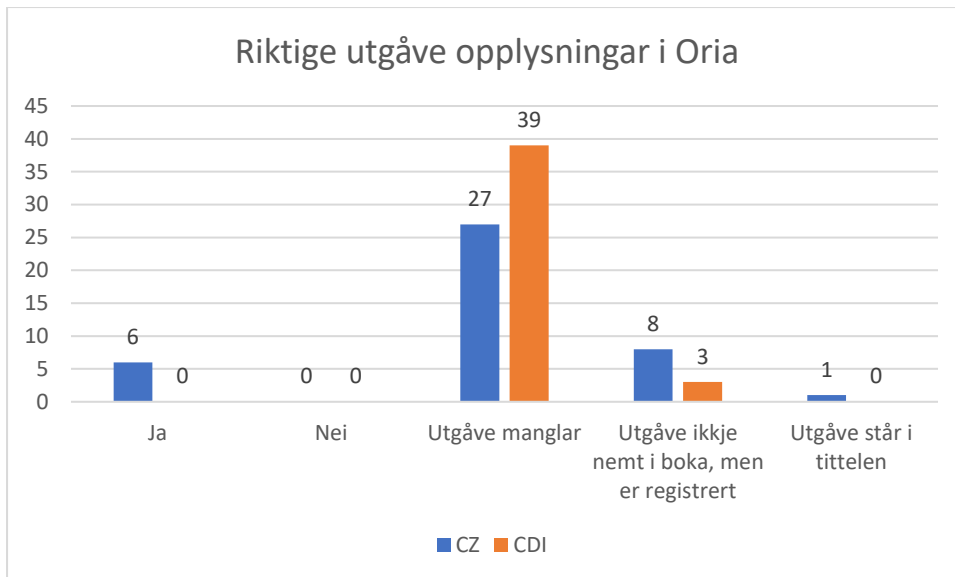
4.2.5 Utgåve

Det var heile 35 titlar av dei 42 bøkene som var første utgåver, dette var markert i boka ved at det var oppgitt at det var stod *1. utgåve*, at det ikkje stod noko om kva utgåve det var, som da betyr at det er fyrste gong boka er gitt ut, eller at det stod *first published* som visar at det er første gong tittelen var utgitt.

Som ein ser i *figur 4.2.5-1* var det berre 3 titlar i datasamlinga der det i boka tydeleg stod at tittelen var 1. utgåve. I datasamlinga var det også nokre titlar som var 2. utgåve, 3. utgåve, 4. utgåve og ein tittel som var 15. utgåve. Det var 32 titlar som ikkje hadde noko informasjon om utgåve eller at det stod *first published* på kolofonsida i e-boka.



Figur 4.2.5-1 Diagram over utgåve opplysningar frå e-boka



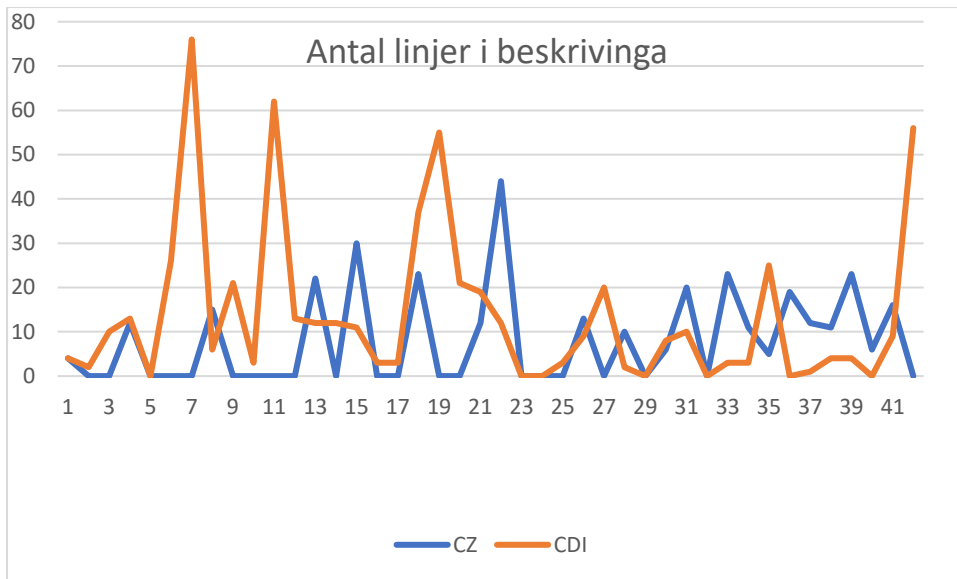
Figur 4.2.5-2 Diagram over utgåve opplysningar i Oria

Det er berre postane frå CZ som har metadata som gir oss riktige opplysningar om kva utgåve tittelen er på 6 titlar som ein ser i *figur 4.2.5-2*. I tillegg er det 8 postar frå CZ som har oppgitt utgåve, sjølv om dette ikkje står i kolofonsida på boka.

For postane frå CDI er det berre registrert opplysningar om utgåve for bøker der utgåve ikkje er nemnt i kolofonsida av boka. Det er 27 postar frå CZ og 39 postar frå CDI som manglar opplysningar om utgåve i posten i Oria.

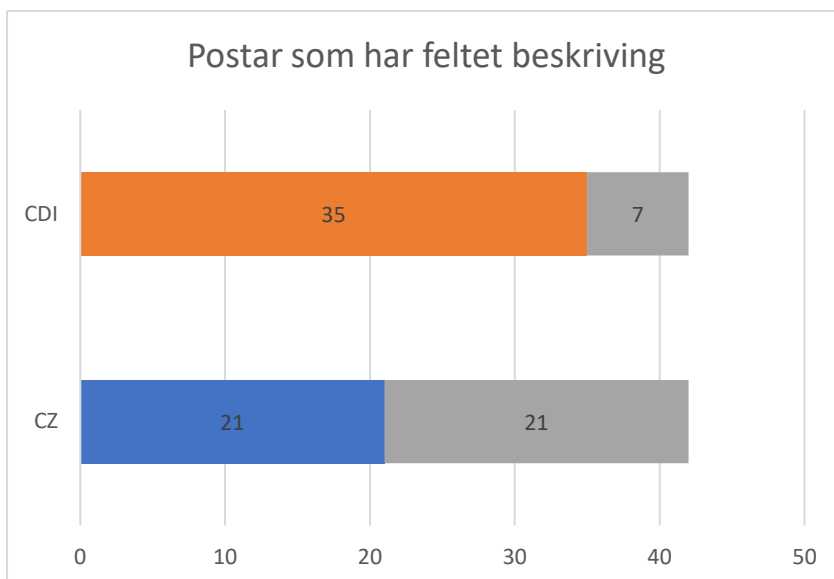
4.2.6 Beskriving

Postane i Oria har eit felt for beskriving der omtale, innhald og samandrag av boka kan bli registrert. Dette er informasjon som er henta frå MARC felt 502, 505 og 520 (ExLibris, u.å.b). Nokre beskrivingar er fylldige med både oppsummering av kva boka handlar om, namn på kapittel, og kapittelforfattar, mens andre postar har ingen informasjon i dette feltet.



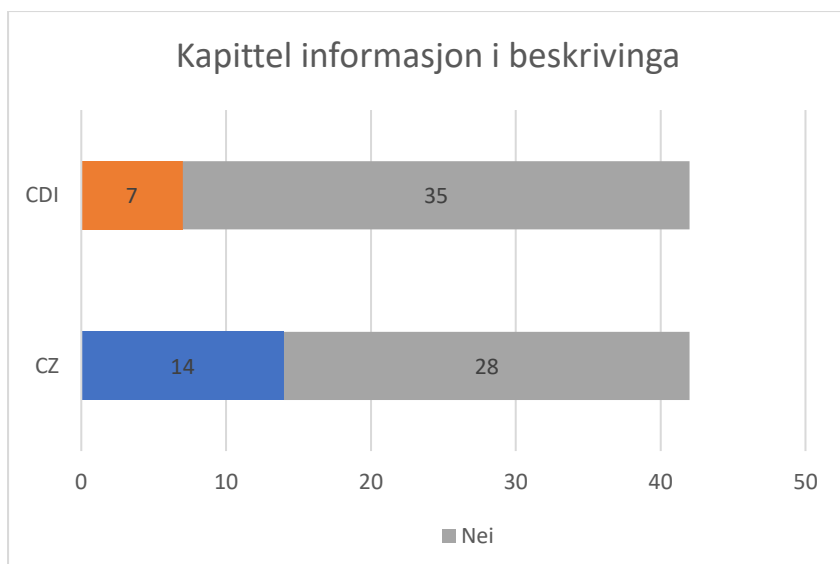
Figur 4.2.6-1 Diagram over antal linjer i beskrivinga

Av diagrammet i figur 4.2.6-1 ser ein at antal linjer i beskrivinga varierer veldig frå 0 til 76 som er det høgaste antalet som er frå ein post frå CDI. Det er lite samanfall mellom dei to kjeldene på antal linjer i beskrivinga.



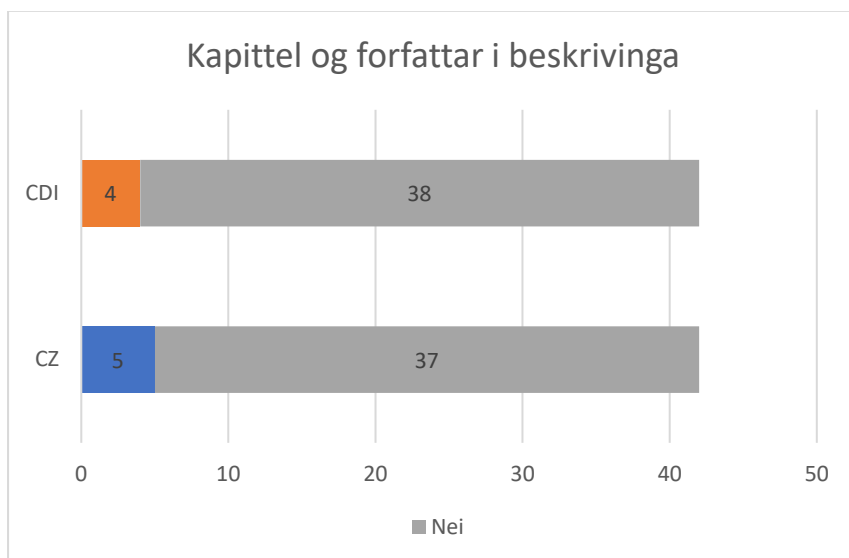
Figur 4.2.6-2 Diagram over postar som har feltet beskriving

Av dei 42 titlane er det for CZ halvparten som har noko i beskrivings feltet, mens frå CDI er det 35 av 42 som har beskriving som ein ser i figur 4.2.6-2.



Figur 4.2.6-3 Diagram over postar som har kapittel informasjon i beskrivinga

I figur 4.2.6-3 ser ein at det er 7 titlar frå CDI som har kapittel informasjon i beskrivingsfeltet, men det er 14 frå CZ.



Figur 4.2.6-4 Diagram over postar som har kapittel og forfattar informasjon i beskrivinga

Det er få av postane frå dei to kjeldene som har både informasjon om kven som har skreve dei ulike kapitla og har med kapittelittelen i beskrivinga som figur 4.2.6-4 visar. Det er 4 postar frå CDI, mens det er 5 frå CZ som har med både informasjon om tittel på kapittelet og kven som har skrevet det.

5 Diskusjon

For at ein brukar skal oppleve at treffa i Oria er relevante for sine informasjonsbehov må tilstrekkelege metadata være registrert. I følgje IFLA LRM user tasks er det fem brukaroppgåver som metadata på eit dokument skal dekke: finne, identifisere, velje, skaffe og utforske (Riva et al., 2017, s. 15). Det ein har sett i funna av denne datasamlinga er at det er fleire moment som kan utfordre desse brukaroppgåvene.

5.1 Finne

Den fyrste brukaroppgåva, finne, går ut på at det skal være mogleg å finne ein eller fleire ressursar ved å søke på relevante kriterier slik som tittel, forfattar og emneord (Riva et al., 2017, s. 15). Brukarar skal kunne søke etter ein kjent tittel, brukarane skal kunne finne noko som er skreve av ein bestemt forfattar, og brukarane skal kunne å finne ressursar som handlar om eit bestemt emne.

Eit aspekt med brukaroppgåva finne frå IFLA LRM user task som ein også må huske på, er at postane frå CZ og CDI kjem inn i Oria etter ulike tider. Postane frå CZ skal være i Oria etter 24 timar, mens postane frå CDI kjem ikkje inn før etter 72 timar (Unit, u.å.). Dette har konsekvensar for atffinninga for ein brukar som da får mindre og dårlegare metadata å søke på, mens ein ventar på CDI postane. Om ein hadde skrudd av postane frå CZ for å redusere dublettane vil det seie at ein brukar ikkje vil kunne få treff på relevante titlar før 3 dagar etter at boka er kjøpt og aktivert av biblioteket. Biblioteka får dermed ikkje utnytta sine boksamlingar til det fulle når det tar så lang tid å få metadata inn i Oria, som nok er den fyrste plassen våre brukarar søker etter e-bøker, og ikkje på sjølve plattformen frå leverandøren for e-bøker som biblioteket tilbyr.

5.1.1 Forfattar

I datasamlinga hadde dei fleste postane i Oria med informasjon om forfattar. Det var tre postar frå CZ som ikkje hadde denne informasjonen. Brukarar som da hadde søkt på forfattarane av desse titlane ville ikkje ha fått treff på dei bøkene om det berre var postar frå CZ i Oria. Men sidan det i Oria er dublettar frå CDI vil brukaren få treff så lenge alle postane frå CDI har med feltet for forfattarnamn, slik som postane frå CDI hadde i denne datasamlinga.

Når ein ikkje kan søke på namnet til ein forfattaren av ei bok, slik vi som dei tre postane som manglar forfattar frå CZ i denne datasamlinga, mistar brukaren ein viktig søkeinnang til bibliotekets materialet. Ifølge David og Thomas (2015, s. 810) er MARC feltet for forfattar eit av dei viktigaste felta for søking og atffinning sidan det er eit av dei felta som gjev mest treff på dei søkeorda brukarane brukte i undersøkinga dei hadde.

Når forfattarnamn manglar kan ein brukar heller ikkje være sikker på at ein får alle relevante treff om ein avgrensar eit søk ved hjelp av fasetten for eit forfattarnamn. Dei titlane som manglar forfattarnamn vil ikkje være med i slike søk og brukarane kan da gå glipp av relevante treff når dei ynskjer å få ei oversikt over kva biblioteket har tilgang til av ein spesifikk forfattar.

5.1.2 Tittel

Tittel var eit av dei viktigaste MARC felta for søking og atffinning i undersøkingane til både Woolcott et al. (2021, s. 41) og David og Thomas (2015, s. 810). Ingen av postane i denne datasamlinga mangla tittel. I datasamlinga var det 26 titlar der tittelen hadde ein undertittel. Det var berre ein post frå CZ av desse som mangla undertittel. I undersøkinga til Wiersma og Tovstiadi (2017, s. 624) var det eit problem at undertittelen mangla. Når ein brukar søker i Oria på ein fulltittel vil ein brukar truleg få treff sjølv om undertittelen ikkje er registrert. Men om undertittelen manglar på ein tittel som denne *Philosophy : the classics* vil det være vanskeleg å finne denne att om berre hovudtittelen er registrert i posten, og ein ikkje søker på forfattarnamn i tillegg.

I undersøkinga til Barderheier (2013, s. 69-70) såg ein eksempel på feil i bruk av indikatorar i MARC 245 som fører til at titlar som skal være søkbare ikkje var det. Sidan alle titlane i denne datasamlinga vart henta inn frå Oria ved å søke på tittelen under datainnsamlinga kan ein seie at dette ikkje er eit problem i denne datasamlinga, sjølv om ein ikkje har sett på bruken av MARC felt, sidan ein har fått treff på tittelen. Brukaroppgåva finne ved å søke på tittel støttast i denne datasamlinga for begge kjeldene.

5.1.3 Emneord

I datasamlinga var det emneord på 34 postar frå CDI og 29 postar frå CZ. Det vil sei at det er mange av dei 42 titlane som ikkje har emneord å søke på for ein brukar. Feltet for emneord var eit av dei mest vellykka MARC felta for atffinning i fleire undersøkingar (Woolcott, et al., 2021, s. 41 ; David & Thomas, 2015, s. 810).

I denne undersøkinga er det ikkje samanlikna kva emneord som er brukt for ein tittel frå CZ og CDI, men heller sett på kor mange emneord kvar post har. Ved å sjå på talet på emneord kan ein få eit inntrykk av om brukarane har mange eller få moglegheiter for å søke etter relevante titlar som har emneord. Talet på emneord per post varierte i datasamlinga frå 0 til 16. Gjennomsnittet for CZ postane var to emneord, mens det for postane frå CDI var nesten fire. Ein brukar som søker på emneord vil ha større sjanse å finne ein post frå CDI enn frå CZ, sidan postane frå CDI oftast har emneord og har i tillegg fleire emneord per post.

I undersøking til Wiersma og Tovstiadi (2017, s. 628) såg dei at emneorda for den same tittelen varierte så masse på tre ulike e-bok plattformar at dei kunne ha beskrive tre ulike bøker. Dette viser at det kan være ein stor variasjon i kvaliteten på emneorda på e-bøker. Noko som ein kan tenke seg også kjem til uttrykk når nokre postar har få og andre mange emneord. Ein får ikkje fram dei same aspekta ved eit dokument når mengda emneord variera frå 0 til 16 på postane.

Knutsen (2017, s. 13) såg i si undersøking at det kunne være CZ postar med heile 27 emneord, der ein hadde mange MESH-termar frå ulike nivå frå emneordshierarkiet. Å dumpe alle termar frå eit emneordsregisteret vil ikkje være hensiktsmessig. Ein brukar vil da kunne få treff på titlar som ikkje er relevante, sidan emneorda kan være meir spesifikk enn det innhaldet i boka faktisk er. I denne datasamlinga har tittelen med flest emneord heile 16 emner, men denne undersøkinga har ikkje sett på bruken av systematiske emneordsregister. Så ein kan berre lure på om alle disse er nødvendige for å sikre god atffinning for ein brukar, og om mange av desse emna berre er dumpa frå eit hierarkiet i ei emneordsliste.

Behovet for mange emneord vil variere ut i frå ulike titlar og kva slags bibliotek tittelen er ein del av. Ei handbok i for eksempel psykologi vil ha behov for mange fleire

emneord enn ei bok som berre handlar om ein liten del av psykologien. Ei bok om IT vil ha behov for meir spesifikke emneord på eit universitet- og høgskolebibliotek som har IT-studentar enn dei biblioteka som har studentar som studere sjukepleie. Dette viser at ulike brukarar har ulike behov for kva slags emneord som er registrert i posten for å utføre brukaroppgåva finne frå IFLA LRM user task. Sidan det i denne undersøkinga ikkje er sett på om emneorda er «riktige» i forhold til boka kan ein berre spekulere i om boka som hadde flest emneord i denne datasamlinga kanskje er ei handbok som treng mange emneord for å få uttrykt flest moglege aspekt ved boka, eller om emneorda er dumpa frå eit emneordshierarki.

Språket på emneorda i datasamlinga var engelsk utanom for to postar frå CDI der den eine hadde eit emneordet på tysk, og den andre hadde eit svensk emneord i tillegg til engelske på ein engelsk tittel. Ifølge retningslinjene for katalogisering i community zone så er det språket på tittelen som bestemmer kva språk metadata som beskriver eit dokument skal ha (ExLibris, 2014, s. 7). I tillegg skal forbetringane av ein post i CZ være til nytte for alle brukarane (ExLibris, 2014, s. 7) . Dette er prinsipp som også burde gjelde postane frå CDI, sidan CDI også er felles for alle som brukar produkta til ExLibris. At det finnes tyske og svenske emneord i datasamlinga er eit eksempel på at dette vil berre være nyttig for brukarar som søker på tysk og svensk. Ein brukar som hadde søkt etter emneord på engelsk hadde gått glipp av denne tittelen, som ikkje har engelske emneord. Brukaroppgåva finne vil da være vanskeleg å utføre om språket ikkje matchar brukarens.

I CZ har biblioteka moglegheit for å legge til lokale felt, slik som emneord, som ikkje blir skreve over ved neste import (Faggruppe for Resource Management, 2020). Ved å legge til lokale emneord på norsk vil ein kunne gi brukarane våre fleire søkeinnangar, og da moglegvis auka mengda relevante treff, om brukaren er interessert i å lese e-bøker på engelsk når dei har søkt på norsk. Dette blir da eit tillegg til det som alt er felles for alle, og berre tilgjengeleg for eit bibliotekets sin lokale katalog. Om ein hadde lagt til lokale emneord på norsk kan det tenkast at dette hadde vore nyttig for andre bibliotek i BIBSYS-konsortiet også. Da er dumt at dette da berre vil nære nyttig for eit bibliotek og at andre bibliotek som har same boka da må gjere den same jobben ein gong til. Dette bryter med tanken og retningslinjene om at ei bok berre skal trengast å bli katalogisert ein gong.

Eit av funna til Dabin og Preminger (2018, s. 49) var at dei fekk treff på emneord som ikkje var synlege for ein brukar under detaljer på ein post vist i fullvisning i Oria. Når dei undersøkte PNX-fila fant dei at grunnen for at dei fekk treff på dei emneorda var at dei var registrert i «search cection» delen av PNX-fila. Ein kan tenke seg at nokre brukarar vil synes at det er forvirrande å få treff i trefflista som ein ikkje kan sjå kvifor er relevante for det søket ein har utført. Dette er noko som ikkje har blitt sett på i denne undersøkinga, men er viktig å være klar over at det finnes søkeinngangar i Oria som ikkje er synlege for ein vanleg brukar.

5.2 Identifisere og velje

Med å identifisere meinast det at det skal være mogleg for ein brukar å skilje mellom materiale som liknar på kvarandre ved å få tilstrekkeleg informasjon om dokumentet (Riva, Bøeuf & Žumer, 2017, s. 15). Brukaroppgåva å velje går ut på at ved hjelp av metadata skal ein brukar kunne velje eller avvise dei ulike resursane (Riva, Bøeuf & Žumer, 2017, s. 15). Desse to brukaroppgåvene skjer ofte samtidig, eller like etter at ein har identifisert materialet, og blir derfor behandla samla i denne teksten.

Ein kan tenke seg at etter ein brukar har identifisert og valt ein ressurs vil dei bruke denne informasjonen til å lage referansar til ei litteraturliste, om dei driver med ei skriveprosjekt. Om det da er feil eller manglar i metadata dei har brukt for å identifisere og velje vil også skape feil og manglar i litteraturlista for ein brukar som nyttar seg av moglegheita for å laste ned referansar enten frå Sitèr funksjonen som nokre bibliotek brukar i Oria, eller ved å laste ned til EndNote, eller andre refereringsverktøy, frå Oria.

5.2.1 Forfattar

At funksjonen forfattar er brukt i staden for redaktør i metadata er noko tidlegare forskning (Wiersma & Tovstiadi, 2017, s. 264 ; Knutsen, 2017, s. 14) har sett er ein mangel som går igjen i metadata på e-bøker. Dette er noko ein ser i denne datasamlinga også, spesielt for postane frå CDI som stort sett berre brukar funksjonen forfattar sjølv om boka er redigert, og ikkje har noko parentes med author eller editor bak eit namn. Av dette får ein ikkje fram i postane frå CDI at av dei 42 titlane er 24 redigerte. For ein brukar vil det nok være lettare å identifisere ei bok om ein er sikkert på at funksjonen til eit namn er riktig. Om ein brukar har ein referanse der det står at

boka er redigert, kan ein tru at brukaren som oftast leite etter informasjon om dette i trefflista for å identifisere at ein har funne riktig bok. Når brukaren da berre finner titlar med ein forfattar når ein leitar etter ei redigert bok kan søket feilaktig ende i null relevante treff for ein brukar.

For dei tre titlane frå CZ som manglar forfattarnamn i denne datasamling vil det være vanskeleg for ein brukar å identifisere og velje at dei har funne riktig tittel. Når forfattar manglar forutsett det at brukaren har søkt på tittel eller emneord for så å skulle identifisere og velje, men utan forfattarnamn kan dette være utfordrande.

5.2.2 Tittel

Å identifisere og velje ein tittel skjer i samband med dei andre metadata som er i posten. Det må være nok andre metadata rundt ein tittel slik det vil være mogleg å identifisere og velje to bøker som har same tittel frå kvarandre. Forfattar og år er eksempel på metadata som er nyttige når ein skal identifisere og velje ein tittel.

I undersøkinga til Wiersma og Tovstiadi (2017, s. 624) var det eit problem at undertittelen mangla. I denne datasamling er det berre ei bok frå CZ der det finnes ein undertittel der den ikkje er registrert. Av dette kan ein ikkje sei at det er eit stort problem i denne datasamlinga av undertittel manglar, og dermed kan ein sei at brukaroppgåva identifisere og velje er tatt vare på når det søkjast etter ein tittel.

Om tittelen ikkje er fullstending med undertittel kan der føre til usikkerheit for ein brukar som søker på den heile tittelen og ikkje får noko treff som er identisk med det ein søker på. Om metadata om forfattar manglar vil det være viktig at undertittelen er registrert så det blir lettare å identifisere og velje riktig bok om det skulle være treff på fleire liknande titlar.

5.2.3 Emneord

Brukarane i brukartestinga til Elzi og Clair (2020, s. 43) syns det var vanskeleg å forstå kva ein post handla om når det var mange postar som mangla, eller hadde få, emneord. Ein kan anta at våre brukarar i Oria også vil syns det er vanskeleg å identifisere og velje bøker som har få, eller ingen, emneord. Noko som ein såg i denne datasamlinga der 13 postar frå CZ og 8 frå CDI mangla emneord. Om ein brukar ikkje

forstår av metadata kva dokumentet handlar om vi dei ha vanskar med å kunne identifisere og velje bøker som passar deira behov.

Ein såg i denne datasamlinga at det var postar med emneord på tysk og svensk. Dette kan gjere det vanskeleg for ein brukar som ikkje beherskar desse språka å forstå kva ressursen handlar om, og dermed kanskje veljer bort interessante treff der tittelen ikkje beskriv godt nok kva boka handlar om.

5.2.4 År

Tidlegare forskning har sett at år ofte er feil eller manglar i metadata på e-bøker (Wiersma & Tovstiadi, 2017, s. 624 ; Knutsen, 2017, s. 14). At året er feil eller manglar er noko ein også ser i denne datasamlinga, men dei flest titlane har riktig år. Postane frå CDI har flest titlar med riktige år, der 34 av 42 er riktig. Mens for postane frå CZ har 31 av 42 riktig utgjevingsår.

For postane frå CDI er det åtte postar som har feil år, for CZ er dette talet 0, men det er eleve postar frå CZ som manglar år. Når utgjevingsåret er feil eller manglar vil det vær vanskeleg for ein brukar å identifisere og velje riktig tittel. Bruk av fasettar der ein avgrensar søket i Oria på utgjevingsåret vil heller ikkje være mogleg, eller gi feil resultat. Treff som hadde vore relevante hadde blitt utelat frå trefflista om ein for eksempel hadde søkt på eit emne og avgrensa med fasetten for dei fem siste åra som utgjevingsår.

Utgjevingsåret på bøkene i denne datasamlinga fordeler seg over fleire år, der dei fleste er utgjeven i 2021. Det er fire bøker som det i kolofonsida av boka står at utgjevingsåret er 2022, sjølv om vi har kjøpt og aktivert boka i oktober 2021. Hadde boka vore katalogisert i 2022 hadde året vore riktig, men når ein veit at denne var mogleg å kjøpe alt i 2021 kan ein lure på kva som er riktig. I følge RDA skal ein skrive det ein ser i kjeldene ein katalogiserer. Av dette kan ein lure på om dei 8 bøkene som har feil år frå CDI er slike som har blitt gitt ut året før, der ein har registrert det året boka faktisk var gitt ut, og ikkje det året som står i boka. Og dermed gir dette feil i metadata som kan forvirre brukarane.

Forlaga har ei litt anna oppfatning av utgjevingsår enn bibliotek, der forlaga meiner at det er datoen boka blir lansert i eit spesielt format som er den riktige (Branch et al, 2018, s. 128 ; Proctor, 2013, s. 100). Dette vil sei at ei bok vart utgitt trykt i 2021, som blir utgitt året etter som e-bok, får eit nytt utgjevingsår, sjølv om det er akkurat det same innhaldet. Når året er feil vil det skape mykje forvirring for ein brukar både for attfinning og identifisering av ein tittel som kan skape spørsmål som «er dette ein ny utgåve?». Dette kan føre til unødvendig mykje tid for ein bibliotekar i samband med innkjøp og fjernlån for å undersøke om den aktuelle tittelen er den same som er gitt ut tidlegare, eller om det faktisk har kome ei ny utgåve. Faren vil derfor være stor for dobbelt innkjøp og unødvendige fjernlånsbestillingar når året er feil eller manglar i metadata.

5.2.5 Utgåve

Informasjon om utgåve er viktig for brukaroppgåvene identifisere og velje for å finne riktig bok. Om ein er på jakt etter den nyaste utgåva av ein tittel må ein ha denne informasjon for å være sikker på at ein har funne riktig bok. Det er berre seks postar frå CZ som har riktig informasjon om utgåve i postane i Oria.

I datasamlinga er det 27 postar frå CZ og 39 frå CDI som manglar informasjon om utgåve. Når så mange postar manglar denne informasjonen vil det være vanskeleg for ein brukar å kunne identifisere og velje riktig bok. For ein brukar vil det være ekstra vanskeleg å velje riktig bok om fleire viktige metadata element manglar, slik som år og forfattarnamn, for å kunne identifisere boka.

I denne datasamlinga var det berre ti titlar som hadde informasjon om utgåve i kolofonsida av e-boka. Det er dermed ikkje så rart at det er mange postar manglar denne informasjonen i fullvisninga av postane i Oria, sidan ein ifølge 2.5.1.4 i RDA skal «Gjengi en utgaveangivelse slik den forekommer i informasjonskilden» (RDA toolkit, 2018). Dette vil sei at sjølv om ein bok er 1. utgåve skal dette registrerast i posten om det i kjelden står at det er 1. utgåve, om det ikkje står, skal det ikkje registrerast. I datasamlinga såg ein at åtte postar frå CZ og tre postar frå CDI hadde registrert informasjon om utgåve sjølv om dette ikkje står i informasjonskjeldene.

5.2.6 Beskriving

Fleire undersøkinga har sett at MARC felt 505 for innhaldsnote og 520 for samandrag er meir effektive for å auke atfinninga enn emneord, sidan orda brukarane brukte ikkje samsvare med emneorda brukt i katalogen (Woolcott et al., 2021, s. 51 ; Sassen & Harker, 2014 ; Elzi & Clair, 2020, s. 43 ; Kleeck et al., 2017, s. 560). Informasjonen i metadatafeltet beskriving kan gi brukarane nytting informasjon som hjelper dei å identifisere og velje. Sidan dei da kan lese kva boka handlar om, og kanskje får informasjon om kva kapitla heiter, og kven som har skrevet dei.

I denne datasamlinga var det 35 postar frå CDI med ei beskriving av tittelen i Oria, frå CZ er det berre halvparten av postane, altså 21 postar, som har med feltet for beskriving. I dokumenter *Alma Community Catalog: Cataloging Standards, Policies, Rights, and Responsibilities* med anbefalingar frå ExLibris om korleis ein skal katalogserie i CZ oppfordrar dei til å legge til 505 og 520 for innhald og samandrag når det er mogleg (ExLibris, 2014, s. 7). Av funna i denne datasamlinga ser det ut som at det er få som går inn i CZ og legger til desse felte når dei manglar.

Lengde på beskrivingane varierer frå 0 til 76 linjer. Nokre beskrivingar har berre med litt informasjon om kva boka handlar om, men andre har med kva kapitla heiter, og nokre får har i tillegg med kven som har skreve kapitla. Lettare for ein brukar å velje og identifisere ei bok om samandraget er representert i posten. Brukaren vil da få informasjon om kva boka handlar om og kanskje til og med kva kapitla heiter og kven som har skreve dei.

I datasamlinga var det 14 av dei 21 postane frå CZ som hadde med kapittel informasjon i beskrivinga, mens CDI hadde kapittel informasjon på 7 av dei 35 postar som hadde beskriving. Beskriving med både kapitteltittel og forfattar av kapitla var den berre 4 frå CDI og 5 frå CZ som hadde. Brukarane får fleire søkeinnningar når ein kan få treff på kapittel og forfattarnamn som ellers ikkje er nemnt i posten. Om det er få eller ingen emneord på ein post kan tittelen på eit kapittel være dekkande for det brukaren er på jakt etter.

Ei god beskriving kan dermed hjelpe til å få oppfylt fleire av brukaroppgåvene i IFLA LRM user tasks. Når beskrivinga inneheld både kapittel tittel og forfattere av kapitlet

vil det være mogleg å finne og identifisere bokkapittel skreve av ein bestemt forfattar, som ikkje elles er nemnt i posten. Ved å få denne informasjonen har brukaren moglegheit til å velje, og velje bort, bøker som har ein kapittelittel eller forfattar som passar til brukarens behov.

5.3 Utforske

Den siste brukaroppgåva i IFLA LRM user task er at ein brukar skal kunne utforske ved hjelp av metadata for å kunne oppdage andre relevante dokument ved å kunne navigere vidare frå eit treff til andre, som for eksempel har same emneord eller forfattar (Riva, Bœuf & Žumer, 2017, s. 16). I Oria kan ein brukar berre navigerer vidare til andre relevant materiale ved å trykke på forfattarnamn og emneord i fullvisninga av ein post. Ein brukar kan da få ei oversikt over alle ressursar ein forfattar har skreve og andre ressursar som har same emneord. I andre søkeverktøy er ofte dewey-nummer også ein felt brukarane kan trykke vidare på for å finne anna materiale som handlar om det same. Dette er ikkje mogleg i fullvisninga i ein post i Oria, så det er berre forfattarnamn og emneord som er mogleg å utforske vidare på.

5.3.1 Forfattar

I datasamlinga er det 3 av 42 titlar frå CZ som manglar informasjon om forfattar. Dette skapar utfordringar for ein brukar som ynskjer å utforske etter meir materialet skrevet av same person. Brukaren kan ikkje være sikker på at ein finn alt forfattere har skrevet når ein trykkar på forfattarnamnet i fulltekstvisninga i ein post når ein veit det finnes postar som manglar informasjon om forfattar. Ein brukar som utforskar vidare frå ein post vil da kunne gå glipp av relevante treff når metadata manglar.

I Knutsen (2017, s. 14) si undersøking såg ein at forfattarnamn mangla autorisering. Autorisering er viktig for å få samla alt materiale ein forfattar har skrevet og få skilt ressursar ut frå andre forfattare sitt arbeid, som har same eller liknande namn. I nokre postar i Oria ser ein at posten er autorisert ved at det er ein parentes med fødselsår, og eventuelt dødsår, men ikkje alle autoritetar har fødselsår så ein brukar kan ikkje ut i frå fullvisninga av ein post i Oria vite om namnet er autorisert eller ei. Ved å utforske på namn er det viktig at namna er autoriserte så ein får fram titlar som handlar noko lunde om det same, sidan ein forfattare stort sett skriv om dei same temaa, slik får ein brukar utforska vidare til ressursar som er relevante for brukarens behov.

Sidan det berre var 3 av 42 titlar som mangla metadata om forfattar i postane frå CZ, og ingen postar frå CDI som manglar informasjon om forfattar, vil det for ein brukar vere mogleg i dei fleste tilfelle å utforske vidare til anna materialet av same forfattar i Oria når det finnes postar både frå CZ og CDI. Men brukaren vil da få dubletter postar når namn på forfattaren er registrert i posten både frå CZ og CDI. Det kan da sjå ut som at ein forfattar har skreve dobbelt så mange bøker enn ein faktisk har om ein berre ser på antall treff i trefflista og ikkje studerar alle titlane så nøye med ein gong.

Ei anna utfordring med å utforske på forfattarnamn kan være at forfattarar av eit kapittel i ei bok, som berre er nemnt i samandraget i posten, ikkje vil være med i vidare utforsking frå fullvisninga av ein post, der kanskje berre redaktøren av boka er registrert. Brukaren vil da kunne få fleire treff om ein søkar i søkefeltet på forfattarnamnet på nytt, slik ein da får treff både på metadata som er registrert under forfattar og beskrivingsfeltet i fullvisninga av Oria posten. Dette viser at brukaroppgåva utforske kan være utfordrande for det å utforske vidare på forfattarnamn av forfattarar av kapittel av ei redigert bok.

5.3.2 Emneord

Bruken av emneord varierte veldig i datasamlinga. Nokre postar hadde mange emneord, mens andre hadde ingen. I datasamlinga var det 13 postar frå CZ og 8 postar frå CDI som manglar emneord.

Når emneord manglar i fullvisninga av ein post i Oria mistar brukarane moglegheita til å utforske eit søk vidare til andre relevante titlar, ved å trykke på eit emneord i fullvisninga av ein post dei alt har funne. På denne måten kan brukarane finne anna materialet som handlar om det same, og kanskje til og med oppdage ressursar som er relevant som ein ikkje hadde forventa å finne.

I datasamlinga såg ein at det varierte kor mange emneord ein post hadde frå 0 til 16. Om det er mange postar i ein bibliotekatalog som har få, eller ingen emneord, kan ein gå glipp av moglege treff ved å utforske på emneord, sidan ein ikkje kan være sikkert på at alle relevante titlar har emneord og dermed mistar moglegheita for å utforske. I undersøkinga til Wiersma og Tovstiadi (2017, s. 628) såg dei når dei

samanlikna emneorda for den same tittelen frå tre ulike kjelder at emneorda kunne ha beskrive tre heilt ulike bøker. Om det er så stor variasjon i korleis ein setter emneord i ein katalog vil det ikkje være mogleg å utføre brukaroppgåva utforske vidare til liknande titlar. Brukarane vil da kunne gå glipp av interessante ressursar når det ikkje brukst dei saman emneorda for det same innhaldet i ulike titlar.

Hadde ein brukar som har utforskar eit søk vidare, ved å trykke på eit emneord i ein Oria post, heller søkt på det interessant emne som eit fritekstsøk, hadde brukaren kanskje fått fleire treff. Ved eit fritekstsøk hadde brukaren fått treff både i tittel, emneord og frå beskrivinga av ei bok, i staden for å berre utforske på emneord, som kanskje ikkje er registrert på alle relevante titlar og kanskje ikkje har gode beskrivande emneord for ressursen.

Woolcott et al. (2021, s. 51) fant i si undersøking at det var meir effektivt for attfinninga å ha MARC felt 505 for innhald og 520 for samandrag enn emneord, sidan det gav betre treff i katalogen med dei orda brukarane søkte med. Når ein les om funna til Woolcot et al. (2021) kan eg tenke meg at det er freistande å berre forbetre MARC 520 slik Elzi og Clair (2020, s. 46) har gjort ved å innføre rutinar for å forbetre dette feltet årleg. Om ein berre forbetrar 520 for samandrag vil brukarane ikkje få moglegheita til å utforske på fleire emneord, om desse manglar frå postane som blir oppdatert. Brukarane vil da kunne miste interessante emneord og utforske vidare på om ein berre har forbetra metadata for samandrag.

5.4 Samanliknande tabell over metadata frå CZ og CDI

I tabellen i *figur 5.4-1* visast det ei oversikt over kva metadata felt som oftast tek vare på brukaroppgåvene frå IFLA LRM user task i postane frå CZ og CDI, for denne datasamlinga. Feitskrift i tabellen markera konklusjonen for dei ulike områda år, forfattar, tittel, beskriving, utgåve og emneord, mens grønskrift visar kven av CZ og CDI som oftast har postar som er «best» i ei ulike kategoriane.

	CZ	CDI
År		x
Riktig år	31 postar	34 postar
Feil år	0 postar	8 postar
År manglar	11 postar	0 postar

Forfatter	x	x
Forfatter er registrert	39 postar	42 postar
Kjem fram i posten om det er forfatter eller redaktør	26 postar	0 postar
Tittel		x
Tittel og undertittel er registrert	41 postar	42 postar
Beskriving	x	
Postane har beskriving	21 postar	35 postar
Beskriving med kapittel	14 postar	7 postar
Beskriving med kapittel og forfatter	5 postar	4 postar
Utgåve	x	
Riktig utgåve er registrert	6 postar	0 Postar
Emneord		x
Tal på postar med emneord	29 postar	34 postar
Sum flest x	3	4

Figur 5.4-1 Tabell som samanliknar funna mellom CZ og CDI

Av tabellen i figur 5.4-1 ser ein at postane frå CDI så vidt tek betre vare på brukaroppgåvene i IFLA LRM user tasks enn metadata frå CZ. CDI har fire kryss for år, forfatter, tittel og emneord. Mens CZ har tre kryss, eit for forfatter (slik som CDI sidan begge kjeldene oppfyller eit av to kriterium kvar), beskriving og utgåve. Dette visar at brukarane kan ha nytte av både postane i Oria frå CZ og CDI for å være sikker på å ha nok metadata som tar vare på brukaroppgåvene i IFLA LRM user task.

6 Konklusjon

Problemstillinga for denne casestudien har vore:

Kan vi sjå nokre forskjellar på metadata på e-bøker frå to ulike kjelder, og oppfyller metadata på e-bøkene IFLA LRM user tasks?

Til å hjelpe meg å løyse denne problemstillinga har eg sett nærare på

- Kordan kan ein brukar ved hjelp av metadata finne, identifisere og velje e-bøker?
- Kordan kan ein brukar ved hjelp av metadata på e-bøker utforske anna relevant materiale?

Når ein post i Oria har dårlege eller manglane metadata er det ikkje berre vanleg søking der ein brukar skal finne ei bok etter kriteriar slik som forfattar, tittel og emne som er skadelidande, men også dei andre brukaroppgåvene i IFLA LRM user task blir utfordra. Er metadata dårlege eller manglar vil ein brukar ikkje kunne identifisere, velje eller velje bort, og heller ikkje utforske til anna relevant materialet. I tillegg vil det å avgrense eit søk ved hjelp av fasettar være problematisk når ein ikkje kan være sikkert på at ein får treff på for eksempel alle bøker som er utgitt i eit år når ein veit at nokre postar ikkje har med utgjevingsår, eller at året er feil.

Eit ynskje for denne casestudien var å hente inn kunnskap for å kunne ta ei avgjerd om tiltak for å redusere dublettane i Oria frå dei to kjeldene, slik som det vart moglegheiter for etter oppdateringa i november 2021, utan at avgjerda skal gå utover attfinningsmoglegheitene for brukaren. Alternativa er om vi skal behalde dublettane i Oria slik det er i dag, skru av postane frå CDI og berre behalde kapittel informasjonen, eller om vi skal skru av postane frå CZ. Av funna såg ein at postane frå CDI oppfyller så vidt oftare brukaroppgåvene i IFLA LRM user task enn dei frå CZ, men med tanke på at posane frå CDI kjem inn i Oria først etter 3 dagar, vil ein risikere at mange av brukaroppgåvene i IFLA LRM user task ikkje vil være moglege å oppfylle om ein hadde skrudd av postane frå CZ som kjem inn i Oria etter 24 timar.

Ei datasamling på berre 42 titlar er alt for få til å konkludere med at metadata er mykje betre i CZ eller CDI for alle postar i Oria. Men av denne dataanalysen kan ein få eit inntrykk av korleis metadatakvaliteten kan være på e-bøkene i Oria frå CZ og CDI. Tittelen på denne masteroppgåva har vore "Dublettar – støy eller nytte?". Gjennom kunnskapen ein har fått gjennom denne casestudie såg ein at metadata frå CDI så vidt oppfylte brukaroppgåvene i IFLA LRM user task oftare enn metadata frå postane frå CZ i Oria. Det kan sjå ut som av funna at metadata frå dei to kjeldene utfyller kvarandre, og dermed sikrar brukarane tilstrekkeleg med søkeinngangar. Om det hadde vore mogleg å slå saman postane frå CZ og CDI slik det blir gjort for dublettpostar for artiklar i Oria hadde dette nok vore meir hensiktsmessig for brukarane, som da hadde sloppe å forhalda seg til dublettar, og ville truleg ha fått ein post per e-bok tittel med meir fullstendige metadata. Meir kunnskap om korleis våre brukara søker, om dei nyttar seg av moglegheita for å utforske til anna materiale og kva metadata dei syns er viktig for attfinning hadde vore nyttig for å vite om dublettane oppfattast som støy, eller om dei faktisk er mest til nytte for våre brukarar, hadde vore nyttig før ein tar ei avgjere på kva ein skal gjere med dublettane på e-bøkene i Oria.

6.1 Vidare forskning

Mens eg jobba med denne masteroppgåva dukke det opp mange nye spørsmål som eg tenker hadde vore interessant å undersøke vidare. Nokre av desse spørsmåla er:

Korleis søker brukarane etter e-bøker i Oria? Er det berre kjente titlar dei er interessert i, eller er dei avhengig av å ha andre gode søke inngangar enn tittel og forfattar for å finne bøkene våre slik som emneord og beskriving? Avgrensar dei søket med fasettar og utforskar dei vidare på emneord eller forfattarnamn? Ein kunne da i tillegg ha sett nærare på dei 13 ulike delane av PNX-fila og korleis metadata påverkar moglegheitene for fasettar og utforskningsfunksjonalitet.

Har dei postane som hadde dårleg kvalitet i dette masterprosjektet endra seg til det betre eller er metadata framleis dei same? Ved tilfeldig kontroll av eit par av titlane i denne datasamlinga har eg sett endringar i både CZ og CDI postane i Oria lang tid etter at eg henta ut data til dette prosjektet, både til det betre og verre for begge kjeldene. Det hadde vore interessant med ei undersøking der ein såg på korleis postane frå CZ og CDI endrar seg over tid.

I dette prosjektet har eg ikkje set på *match and merge* prosessen for postane frå CDI, som skal slå saman postar for e-bøker som er frå CDI. Meir kunnskap om korleis dette fungerer hadde vore nyttig. I utvalge eg hadde for aktivert titlar i oktober var det 72 titlar, men berre 42 som kunne brukast i denne undersøkinga sidan det for mange titlar det var både to og tre postar i Oria frå CDI. Dette tenker eg viser at *match and merge* prosessen ikkje fungerer bra nok .

Eit anna prosjekt som hadde vore interessant å sjå på er om det er noko samanheng mellom metadatakvaliteten og kor mykje dei ulike e-bok titlane blir brukt. Vil bruken av titlar som knapt er brukt auke om vi legg til MARC felt 505 og 520, som andre undersøkingar har sett? Eller har vi rett og slett kjøpt feil titlar som brukarane ikkje er interessert i?

7 Referansar

- Baeza-Yates, R. & Ribeiro-Neto, B. (2011). *Modern information Retrieval: the concepts and technology behind search* (2. utg). Person.
- Bardenheier, P. (2014). The quality of e-book records: a case study. *New Zealand Library & Information Management Journal*, 53(3), 60-76.
- Branch, D., Aronoff, K., Elias, E. & Waecker, E. (2018). *The E-Book Story: The Key to a Happy Ending*. Proceedings of the Charleston Library Conference.
<http://dx.doi.org/https://doi.org/10.5703/1288284317025>
- Breeding, M. (2014). Library resource discovery products: context, library perspectives and vendor positions. *Library Technology Reports*, 50(1), 1.
- Bruce, T. R. & Hillmann, D. I. (2004). The Continuum of Metadata Quality: Defining, Expressing, Exploiting I D. I. Hillmann & E. L. Westbrook. *Metadata in practice*. American Library Association.
- Cutter, C. A. (1904). *Rules for a dictionary catalog*. (4 utg., omredigert). Washington: UNT Digital Library. <http://digital.library.unt.edu/ark:/67531/metadc1048/>
- Dabin, E. & Preminger, M. (2018). Discovery Systems as an Alternative to Stand-Alone Databases: The Example of Primo at BI Norwegian Business School. *Nordic Journal of Information Literacy in Higher Education*, 10(1), 37–53.
<https://doi.org/10.15845/noril.v10i1.274>
- Davis, R. C. & Song, X. (2020). Uncovering the Mystery of How Users Find and Use eBooks though Guerilla Usability testing. *Serials Review*, 46(3), 193-200.
<https://doi.org/10.1080/00987913.2020.1806648>
- David, R. & Thomas, D. (2015). Assessing Metadata and Controlling Quality in Scholarly Ebooks. *Cataloging & Classification Quarterly*, 53(7), 801-824.
<https://doi-org.ezproxy.inn.no/10.1080/01639374.2015.1018397>
- Denton, W. (2007). FRBR and the History of Cataloging. I A. G. Taylor (Red.), *Understanding FRBR: What It Is and How It Will Affect Our Retrieval* (s. 35 – 58). Westport: Libraries Unlimited.
- Elzi, E. & Clair, K. (2020). Measuring Metadata Quality: A User Research Approach. I S. Marien (Red.), *Library Technical Services : Adapting to a Changing Environemnt*. Purdue University Press.

- ExLibris (u.å.a). *Primo Administration: Primo Normalized XML*.
https://knowledge.exlibrisgroup.com/@api/deki/files/83928/05_PRIMO-Primo_Normalized_XML.pdf?revision=1
- ExLibris (u.å.b). *The Display Section*.
https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Technical_Guide/010The_PNX_Record/040The_Display_Section
- ExLibris (u.å.c). *CDI Record Summary and Sources*.
https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/CDI_Record_Summary_and_Sources
- ExLibris. (u.å.d). *CDI Collection List for Alma Customers*.
https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/CDI_Record_Summary_and_Sources
- ExLibris. (u.å.e). *Match and Merge in CDI*.
[https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_\(English\)/CDI_-_The_Central_Discovery_Index/060Match_and_Merge_in_CDI](https://knowledge.exlibrisgroup.com/Primo/Content_Corner/Central_Discovery_Index/Documentation_and_Training/Documentation_and_Training_(English)/CDI_-_The_Central_Discovery_Index/060Match_and_Merge_in_CDI)
- ExLibris. (u.å.f). *Primo Publishing Platform*.
https://knowledge.exlibrisgroup.com/@api/deki/files/83920/03_PRIMO-Publishing_Platform.pdf?revision=1
- ExLibris. (2014). *Alma Community Catalog: Cataloging Standards, Policies, Rights, and Responsibilities*.
<https://knowledge.exlibrisgroup.com/@api/deki/files/60631/AlmaCommunityCatalogCatalogingStandardsPolicies.pdf?revision=2>
- ExLibris. (2021). *Primo 2021 Release Notes*.
https://knowledge.exlibrisgroup.com/Primo/Release_Notes/Primo/2021/001Primo_2021_Release_Notes?mon=202111BASE
- ExLibrisLtd. (2021, 11. juni). *Central Discovery Index*. [Video]. YouTube.
<https://youtu.be/Y8vOFgG8lo0>
- Faggruppe for Resource Management. (2020, 17. november). *Om e-ressurser*.
<https://rdakatalogisering.unit.no/om-e-ressurser/>
- Faggruppe for Resource Management. (2021a, 17. juni). *Introduksjon til katalogisering i BIBSYS-konsortiet*.
<https://rdakatalogisering.unit.no/introduksjon-til-katalogisering-i-bibsyskonsortiet/>

- Faggruppe for Resource Management. (2021b, 16. april). *E-bok med personlig forfatter*. <https://rdakatalogisering.unit.no/e-bok-med-personlig-forfatter/>
- Frederick, D. E. (2015). *Managing EBook Metadata in Academic Libraries : Taming the Tiger*. Elsevier Science & Technology.
- Gilliland, A. J. (2016). Setting the Stage. I M. Baca (Red.), *Introduction to Metadata* (3. utg., s. 1-21). Getty Research Institute.
- Haynes, D. (2018). *Metadata for information management and retrieval : understanding metadata and its use* (2. utg.). Facet Publishing.
- Hoeppner, A. (2012). The Ins and outs of evaluating web-scale discovery services. *Computers in Libraries*, 32(3), 6-40.
- IFLA (2012). *IFLA Professional Statement on Universal Bibliographic Control*. <https://www.ifla.org/wp-content/uploads/2019/05/assets/bibliography/Documents/ifla-professional-statement-on-ubc-en.pdf>
- Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser? : Innføring i samfunnsvitenskapelig metode* (3. utg.). Cappelen Damm Akademisk.
- Klatt, C., & Meeks, K. (2020). *Increasing eBook usage: The importance of investing time and money into access points*. <https://dc.uthsc.edu/cgi/viewcontent.cgi?article=1074&context=scmla>
- Knutsen, U. (2016, 16. mars). *Metadata i Almas Community Zone (CZ)* [Paperpresentasjon]. BIBSYS Trondheim <http://2016.bibsyskonferansen.no/files/2016/03/CZ-problematikk.pdf>
- Knutsen, U. (2017). *Metadatakvalitet på e-bøker: Undersøkelse 1*. [Paperpresentasjon]. BIBSYS Trondheim http://2017.bibsyskonferansen.no/files/2017/03/Metadataunders%C3%B8kelseneBIBSYS_2017.pdf
- Kortick, Y. (u.å.). *Working with the Alma Community Zone and Electronic Resources*. https://knowledge.exlibrisgroup.com/@api/deki/files/49365/CZ_-_Introduction_to_working_with_the_Alma_Community_Zone_and_Electronic_Resources.pptx?revision=3
- Lazarinis, F. (2014) *Cataloguing and Classification : An Introduction to AACR2, RDA, DDC, LCC, LCSH and MARC 21 Standards*. Elsevier Science & Technology.
- Library of Congress (2013, 28. juni). *MARC Code List for Relators*. <https://www.loc.gov/marc/relators/relacode.html>

- Minčić-Obradović, K. (2011). *E-books in academic libraries*. Chandos Publishing.
- Miksa, S.D. (2021). Cataloging Principles and Objectives: History and Development. *Cataloging & Classification Quarterly*, 59(2-3), 97–128.
<https://doi.org/10.1080/01639374.2021.1883173>
- Niu, X., Zhang, T. & Chen, H-I. (2014) Study of User Search Activities With Two Discovery Tools at an Academic Library, *International Journal of Human-Computer Interaction*, 30(5), 422-433,
<https://doi.org/10.1080/10447318.2013.873281>
- Proctor, J. (2013). How Old is That Ebook: A Call for Standardization in Publisher-Provided Ebook Publication Dates. *Journal of Electronic Resources Librarianship*, 25(2), 100–114. <https://doi.org/10.1080/1941126X.2013.785288>
- RDA Toolkit (2018, 13. februar). *Registrering av utgaveangivelser*. Henta 05. Mai, 2022, frå <https://original.rdatoolkit.org/document.php?id=rdachp2-no&target=rda2-4894#rda2-4894>
- Riva, P., Bœuf, P. L. & Žumer, M. (2017). *IFLA Library Reference Model : A Conceptual Model for Bibliographic Information*.
https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf
- Rose-Wiles, L. M. & Hofmann, M. A. (2013). Still Desperately Seeking Citations: Undergraduate Research in the Age of Web-Scale Discovery. *Journal of Library Administration*, 53(2-3), 147–166.
<https://doi.org/10.1080/01930826.2013.853493>
- Sassen, C. & Harker, K. R. (2014, 17. mars). *Usage of E-books With Enriched Bibliographic Records* [Paperpresentasjon]. Electronic Resources and Libraries Conference. <https://digital.library.unt.edu/ark:/67531/metadc277292/>
- Smiraglia, R. (Red.) (2005). *Metadata: a cataloger's primer*. Haworth Information Press.
- Strader, C. R. (2021). Cataloging to Support Information Literacy: The IFLA Library Reference Model's User Tasks in the Context of the Framework for Information Literacy for Higher Education. *Cataloging & Classification Quarterly*, 59(5), 442–476. <https://doi.org/10.1080/01639374.2021.1939828>
- Svenonius, E. (2000). *The intellectual foundation of information organization*. MIT Press.

- Tennant, R. (1998). Digital Libraries: 21st-Century Cataloging. *Library Journal*, 123(7), 30-31.
<https://login.ezproxy.inn.no/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=507559&site=ehost-live>
- The International Conference on Cataloguing Principles. (1961). *Statement of Principles Adopted by The International Conference on Cataloguing Principles*. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/IMEICC/IMEICC1/statement_principles_paris_1961.pdf
- Tjora, A. (2020). *Kvalitative forskningsmetoder i praksis* (3. utg.). Gyldendal
- Unit. (2020a, 2. september). *Alma Biblioteksystem*.
<https://www.unit.no/tjenester/alma-biblioteksystem>
- Unit. (2020b, 19. mai). *Overgang til Central Discovery Index (CDI)*.
https://dok.unit.no/oria/overgang_til_cdi
- Unit. (u.å.). *Feilsøking i CDI*. <https://dok.unit.no/oria/feilsokcdi>
- Unit. (2021, 4. november). *Oria Søkjetjeneste*. <https://www.unit.no/en/node/469>
- Unit. (2019, 18. juli). *Søkveiledning nytt grensesnitt*.
<https://dok.unit.no/oria/sokeveiledning-nytt-gui>
- Van Rijsbergen, C. J. (1979). *Information Retrival* (2. utg.). Butterworth
- Walters. (2013). E-books in Academic Libraries: Challenges for Discovery and Access. *Serials Review*, 39(2), 97–104.
<https://doi.org/10.1080/00987913.2013.10765501>
- Ward, S., Freeman, R. S., & Nixon, J. M. (2015). *Academic E-Books : Publishers, Librarians, and Users*. Purdue University Press.
- Wiersma, G., & Tovstiadi, E. (2017). Inconsistencies between Academic E-book Platforms: A Comparison of Metadata and Search Results. *portal: Libraries and the Academy* 17(3), 617-648. doi:10.1353/pla.2017.0037
- Woolcott, L., Payant, A., Skeen, B., & Daybell, P. (2021). Missing the MARC: Utilization of MARC Fields in the Search Process. *Cataloging & Classification Quarterly*, 59(1), 28–52. <https://doi.org/10.1080/01639374.2021.1881010>