

# Mellom tradisjonen og weben: katalogisering, metadata og bibliotekarutdanning

Tor Arne Dahl, Unni Knutsen og Kim Tallerås

## Innledning

Vi bruker *metadata* i en rekke sammenhenger uten at vi nødvendigvis tenker over hva det er eller hvordan de produseres. Digitale musikkbutikker og strømmings-tjenester tilbyr enorme musikkbibliotek til sine brukere, og mange eier private musikkksamlinger som er langt større enn bibliotekenes. Disse samlingene organiseres og gjøres søkbare ved hjelp av metadata. Produksjon av metadata har alltid vært en kjernevirksomhet i biblioteket. Med digitalisering og webens utbredelse har behovet for og bruken av metadata økt betraktelig, også utenfor bibliotekene.

Til tross for denne utvikling har bibliotekarutdanningene vært mest opptatt av den tradisjonelle metadataproduksjonen innenfor biblioteksektoren. I praksis innebærer det at undervisningen har hatt en håndverksmessig tilnærming til *katalogisering* i henhold gjeldende katalogiseringsregler og MARC-formatet<sup>1</sup>, først og fremst for å legge til rette for gjenfinning i en bibliotek katalog.

Weben er ikke bare et nytt anvendelsesområde for metadata utenfor bibliotek-katalogen, men stiller i tillegg nye krav til funksjonalitet og kvalitet. Metadata blir for eksempel i større grad enn tidligere distribuert og brukt på tvers av ulike produksjonsmiljøer. Det forutsetter at både mottakere og avsendere forstår hverandres metadata, også når mottaker og avsender ikke er mennesker, men data-maskiner.

Mange har kritisert bibliotekenes maskintilpassede MARC-format for å være for tett knyttet til sitt opphav og fysiske forgjenger – katalogkortet – og dermed også for å være dårlig tilpasset moderne krav og behov for datautveksling. Med de siste årenes webutvikling har denne kritikken eskalert (se for eksempel Coyle & Hillmann, 2007; Tennant, 2002).

På weben har altså konkurrerende aktører og tjenester dukket opp innenfor det som har vært bibliotekenes kompetansemonopol. Bibliotekene og utdanningene har i liten grad latt seg påvirke eller «forstyrre» av denne utviklingen. Både i norske og internasjonale bibliotekmiljøer har flere omtalt konkurrerende aktører og metadata på en nærmest nedlatende måte (se for eksempel Gorman, 2002; Hegna,

---

<sup>1</sup> MACHine-Readable Cataloging

udatert). Samtidig viser arbeidet med nye modeller og standarder som FRBR<sup>2</sup>, RDA<sup>3</sup> og lenkede data, at også sentrale institusjoner innenfor det internasjonale bibliotekmiljøet ser behovet for tilpasninger til de teknologiske endringene. De tre amerikanske nasjonalbibliotekene<sup>4</sup> konstaterer at MARC-formatet ikke er tilpasset webens infrastruktur, og har satt i gang et arbeid med å utvikle et nytt og mer fleksibelt metadatasystem som bedre støtter datautveksling med andre produsenter og brukere (U.S. RDA Test Coordinating Committee, 2011). En arbeidsgruppe hos Library of Congress understreker dessuten behovet for mer forskning på metadatakvalitet og økt samarbeid på tvers av utdanninger, bibliotek og andre relaterte metadatamiljøer (Library of Congress Working Group on the Future of Bibliographic Control, 2008).

Hvordan skal så bibliotekarutdanningene tilpasse sin undervisning innenfor et digitalt og webbasert metadatalandskap? Hvilken metadatakompentanse trenger den fremtidige bibliotekaren? Trenger bibliotekarene i det hele tatt tradisjonelle katalogiseringskunnskaper?

Vi vil svare på disse spørsmålene gjennom to tilnærminger til temaet. Først trekker vi opp de viktigste posisjonene i den internasjonale debatten om katalogiseringsfagets innhold og rolle i bibliotekarutdanningene. Her presenterer vi også katalogiseringsundervisningen slik den tilbys ved bachelorutdanningen vår i dag. For å finne ut hvordan metadata produseres og distribueres utenfor bibliotekene, undersøker vi et metadataproduksjonsmiljø som har oppstått digitalt. Vi har valgt ut MusicBrainz, som tilbyr metadata for digitale musikkfiler, og sammenlikner denne tjenesten med bibliotekenes tradisjoner. På bakgrunn av undersøkelsen diskuterer vi avslutningsvis en vei framover for katalogiseringsundervisningen.

## Metadata

Metadatabegrepet kan historisk knyttes til utviklingen av informasjonsteknologi, men kom først i allmenn bruk på 1990-tallet i forbindelse med gjenfinning på weben og massespredning av digitale ressurser. Metadata defineres ofte som «data om data». Dette er imidlertid ikke en helt entydig definisjon da begrepet brukes om litt ulike fenomener. Det finnes for eksempel administrative metadata, tekniske metadata eller strukturelle metadata. Her er vi opptatt av det som kalles *deskriptive* metadata, det vil si strukturert informasjon som beskriver en informasjonsressurs i den hensikt å identifisere og gjøre ressursen tilgjengelig (Hodge, 2001). Slike metadata inneholder både en beskrivelse av formelle og innholdsmessige egenskaper ved

---

<sup>2</sup> Functional Requirements for Bibliographic Records, på norsk Funksjonskrav til bibliografiske poster

<sup>3</sup> Resource Description and Access

<sup>4</sup> Library of Congress, National Library of Medicine, og National Agricultural Library

en ressurs. I denne artikkelen konsentrerer vi oss om selve beskrivelsene, og går ikke inn på klassifikasjonprosesser og analyser av ressursenes innhold.

I bibliotekene kalles gjerne produksjon av deskriptive metadata for *katalogisering*, ettersom sluttproduktet tradisjonelt samles i en katalog. Selve sluttproduktet kalles ofte *data* eller *poster* sammenstilt i uttrykk som «katalogdata», «katalogposter» eller «bibliografiske data». Sammenliknet med slike begreper er ordet metadata mer generisk. Det sier verken noe om hva som beskrives eller hvordan selve dataene er modellert.

Metadatastrukturer består av et sett *elementer* som angir hvilke aspekter ved ressursen som kan eller skal beskrives. Elementene inngår videre i et *metadataskjema*<sup>5</sup>. Dublin Core er et eksempel på et slikt skjema, og *creator* et eksempel på et definert element som beskriver et ansvarsforhold til en informasjonsressurs. Selv om MARC gjerne omtales som et utvekslingsformat, tolker vi i denne sammenhengen også MARC som et eksempel på et metadataskjema. Her brukes betegnelsen *felt* i vår betydning av begrepet element.

I tillegg kan selve metadataproduksjonen basere seg på en *registreringsstandard*, eller retningslinjer for registreringen knyttet til *hva* og *hvordan*. Dublin Core opererer ikke med særskilte retningslinjer for hva som skal beskrives og hvordan beskrivelsene skal utformes. MARC og en del andre metadataskjemaer baserer seg imidlertid på slike retningslinjer, som vil innvirke på selve produksjonen av metadata.

Alle som skal formidle informasjonsressurser, enten gjennom spesialiserte nettstedet og tjenester eller gjennom de generelle søkemotorenes trefflistet, er avhengige av metadata. Webens opphavsmann Tim-Berners Lee har med sine visjoner om en *semantisk web* tatt til orde for å styrke metadataenes rolle ytterligere. I den semantiske weben skal metadata skape nye sammenhenger i informasjonsmylderet – en «smartere» web – gjennom standardiserte metadatastrukturer som kan fortolkes av resonnerende maskiner. Til tross for enkelte motstemmer (se for eksempel Weinberger, 2007), er semantisk web i dag et viktig tverrfaglig forsknings- og utviklingsområde. Muligheten for å publisere bibliotekenes metadata som lenkede data vies særlig interesse (Kelley, 2011; U.S. RDA Test Coordinating Committee, 2011). Brukere trenger gode metadata for å kunne dra full nytte av digitale tjenester og produkter, og fagfolk diskuterer hva som karakteriserer slike gode metadata.

---

<sup>5</sup> I en semantisk web-kontekst inngår gjerne elementene i ulike «vokabularer» eller «ontologier».

## I: Katalogiseringsundervisning

Internasjonalt har det over lengre tid pågått en diskusjon om bibliotekarutdanningenes undervisning i katalogisering. Flere har argumentert for at faget må fjernes, omformes eller reduseres. Hovedbegrunnelsen er at det kun er ansatte hos de sentrale metadataprodusentene som trenger denne fagkunnskapen, mens andre bare skal gjenbruke dataene som disse sentrale miljøene skaper. Mange bibliotekarutdanninger har derfor sluttet å undervise i faget eller kuttet kraftig ned på timeantallet (Davis, 2008). Blant 47 ALA<sup>6</sup>-akkrediterte utdanningsinstitusjoner i USA er tidligere obligatoriske grunnkurs i praktisk katalogisering blitt erstattet med generelle innføringskurs i bibliografisk kontroll<sup>7</sup>. Noe mer enn halvparten gir avanserte kurs, men disse tilbys ikke regelmessig. Det er også utviklet få kurs som omhandler internett, digitale ressurser og alternative metadataskjemaer. Davis oppsummerer sin undersøkelse slik:

Many librarians and library educators feel that cataloging is a skill that is not as vital as it once was. The shift away from hands-on, detailed bibliographic control courses in library schools reflects the perceived changing needs of the profession. It may be possible that while cataloging as we currently understand it becomes less necessary, the basic skills and concepts of bibliographic control will remain vital (s. 197).

Som et innlegg i utdanningsdebatten skrev Michael Gorman artikkelen *Why teach cataloguing and classification?* Han svarer selv:

That obvious answer is: «We should teach cataloguing and classification because it is essential that those who wish to be librarians (not just those who wish to be cataloguers) understand the way in which recorded knowledge and information is organized for retrieval». (2002, s. 2)

Dette utdyper han ved å vise til at svært mange av bibliotekets aktiviteter er knyttet til katalogen. Han legger videre vekt på at katalogisatorer sikrer konsistente lokale kataloger som i sin tur bidrar inn mot nasjonale og internasjonale bibliografiske nettverk. Katalogisering er derfor ikke bare en aktivitet som er viktig i sentrale, bibliografiske miljøer. Han konkluderer med at katalogiseringsfaget må være en sentral del av undervisningstilbudet.

Både Gorman og Davis er mest bekymret for konsekvensene av en utilstrekkelig katalogiseringsekspertise får for aktivitetene i biblioteket. Rapporten *On the Record*

---

<sup>6</sup> American Library Association

<sup>7</sup> Bibliografisk kontroll omfatter bibliotekenes bestrebelser på å identifisere utgivelser, systematisk samle inn disse for innlemmelse i samlingene, bekjentgjørelse av utgivelsene gjennom registrering i kataloger eller publikasjoner samt å gi brukerne fysisk tilgang til dokumentene. Begrepet brukes ofte også i en snevrere betydning knyttet til selve registreringsprosessen.

peker på det samme, men løfter samtidig blikket mot et større informasjonslandskap:

There is an impending and critical shortage of catalogers, indexers, etc. as these positions are affected by retirements, resource reductions, and a dearth of qualified faculty to teach them. For almost three decades, it has been assumed that the demand for professionals in these positions will decline as more libraries rely on acquiring bibliographic control data from others. In actuality, there has been a shift in demand for these skills from libraries to the information industry, but LIS programs tend to focus on the former, rather than the latter. (Library of Congress Working Group on the Future of Bibliographic Control, 2008, s. 38)

Bibliotekarutdanningene anbefales deretter å gjøre følgende:

Require core levels of knowledge for all information professionals in the fundamentals of knowledge organization theory and practice, including application not only in libraries, but also in the broader range of related communities and information activities. (2008, s. 39)

Bibliotekarutdanningsinstitusjonene må altså ta inn over seg at kompetanse og problemstillinger knyttet til metadata i dag har et langt bredere utgangspunkt enn biblioteket som institusjon. Det samme gjelder dermed også nedslagsfeltet i arbeidsmarkedet for bibliotekarer. For utdanningen betyr dette at den må dekke temaer som strekker seg utover bibliotekenes metadatasfære. Mange vil hevde at dette er en nødvendig prosess for å modernisere metadatakompetansen også innenfor bibliotekene. Gorman (2002) er derimot skeptisk til denne typen argumentasjon. Han vil fortsatt ha fokus på bibliotekenes tradisjonelle kjerneoppgaver og verktøy. Library of Congress-rapporten viser en noe annen holdning.

Som ledd i Bologna-prosessen for å harmonisere europeisk høyere utdanning, ble det utarbeidet en rapport om innholdet i bibliotekarutdanningene i europeiske land (Kajberg & Lørring, 2005). Rapporten identifiserer 11 kunnskapsområder som til sammen utgjør bibliotek- og informasjonsvitenskap. Ett av dem er *kunnskapsorganisasjon*, som også omfatter katalogisering ved Oslo-utdanningen. 66 prosent av de europeiske bibliotekarutdanningene anser kunnskapsorganisasjon som et kjerneemne, viser svarene fra en spørreundersøkelse gjengitt i rapporten<sup>8</sup>. I rapporten trekkes i tillegg fram den sentrale og økende betydningen som metadata får innenfor det nye kunnskapsområdet *digitalisering av kulturarven*. Dette omtales som en transformering av tradisjonell katalogiseringskunnskap fra bibliografisk beskrivelse til metadatastandarder (s. 38).

---

<sup>8</sup> Det var imidlertid noe ulik oppfatning blant deltakerne om katalogisering hører inn under *kunnskapsorganisasjon*, så tallene må tolkes med forsiktighet. Det framgår likevel av rapporten at katalogisering er et fag det undervises i og som tillegges vekt.

Debatt om katalogiseringsundervisning har vi også i Norge. Argumentasjonen for å kutte ned på undervisningen har vært den samme som i USA, altså at denne kompetansen først og fremst anvendes i de store, bibliografiske miljøene som Nasjonalbiblioteket og Biblioteksentralen. Folkebibliotekene skal hovedsakelig importere poster fra de sentrale leverandørene, mens spesialiserte fagbibliotek har noe behov for grunnkatalogisering, hevdes det.

En undersøkelse blant folkebibliotek viser derimot at også disse bibliotekene grunnkatalogiserer visse typer litteratur (Knutsen, 2007). Dette gjelder i første rekke lokallitteratur, men også medietyper der tilfanget på katalogposter fra sentrale leverandører er for lite, som musikk og film. Folkebibliotekene understreker at de ønsker å bidra til den nasjonale katalogproduksjonen ved å registrere lokallitteratur. De oppsporer enklest og hurtigst denne typen litteratur selv, og har dessuten best forutsetninger for å tildele emneord til dokumentene. De opplever også at kunnskap om katalogisering gir en verdifull bestillerkompetanse, for eksempel ved innkjøp av biblioteksystemer og utvikling av digitale formidlingstjenester. Folkebibliotekene er dermed skeptiske til argumenter om at bibliografisk kompetanse bare skal finnes i de store produsentmiljøene, og flere anser katalogiseringskunnskap som kjernekompetanse.

En tilsvarende undersøkelse i fagbibliotekene (Knutsen, 2009) viser at det også her registreres mye materiale fra grunnen av, blant annet publikasjoner fra egen institusjon. Metadatapraksis i fagbibliotek er samtidig mer utfordrende enn i folkebibliotekene. Digitale vitenskapelige tidsskrifter forvaltes gjennom et mangfold av informasjonssystemer og tilgangsmetoder. Fagbibliotekene må derfor forholde seg til mange forskjellige metadataskjemaer. Biblioteksystemene baserer seg for eksempel gjerne på MARC-format, mens institusjonelle arkiv typisk bruker Dublin Core. En kartlegging av interoperabilitet mellom institusjonelle arkiv og bibliotekataloger i britiske fagbibliotek finner at samvirket mellom systemene er svært liten, selv om anvendelsesområdene i stor grad er overlappende (Birrell, Dunsire, & Menzies, 2009). Dette rammer effektiviteten i både metadataproduksjon og i sluttbrukertjenester. Birrell et al. anbefaler derfor at bibliotekarer på bakgrunn av sin metadatakompetanse deltar i utviklingen av verktøy som fremmer interoperabilitet. De heterogene systemene i fagbibliotekene krever med andre ord at bibliotekarene som skal forvalte dem har en bred og sammensatt kompetanse.

Som vist innledningsvis, har bruken av metadata de siste tiårene eskalert også utenfor bibliotekene. Dette har ført til et økende kompetansebehov, og at bibliotekarutdanningene ikke lenger er alene om utdanningstilbud på dette feltet.

For eksempel tilbyr norske utdanninger innen digitale mediefag, informatikk og arkivistikk kurs i metadata<sup>9</sup>.

I denne sammenhengen er det også verdt å nevne den såkalte iSchool-bevegelsen som opererer med en utvidet forståelse av informasjonsbegrepet:

The study of information is interdisciplinary, fed by multiple diverse fields. Librarianship and computer science have historically been the primary feeders of the field, but information studies is also fed by fields such as education, psychology, anthropology, business, journalism—indeed, the range of social sciences. (iSchools, udatert)

iSchools-bevegelsen består av utdanningsinstitusjoner med ulik forankring. Noen har utspring i rene bibliotek- og informasjonsvitenskapelige institusjoner, mens andre har røtter i informatikk, medie- og utdanningsfag. Det er vanskelig å få noen enhetlig oversikt over studietilbudene i disse utdanningene, særlig ettersom ikke alle er ALA-akkreditert og dermed heller ikke inngår i undersøkelser av bibliotekarutdanninger som den nevnt over. Ett eksempel på en iSchool uten slik akkreditering, UC Berkeley School of Information, viser likevel at et utvidet informasjonsbegrep også påvirker perspektivet på metadata. Her tilbys blant annet generelle kurs innen webteknologier som omhandler metadatamodellering uten særlige tilknytning til bibliotekenes tradisjoner<sup>10</sup>.

## **Katalogiseringsundervisningen ved bachelorutdanningen i Oslo**

Ved bachelorstudiet i bibliotek- og informasjonsvitenskap i Oslo er den obligatoriske undervisningen i katalogisering lagt til de tre første semestrene og inngår i emnekretsen *Kunnskapsorganisasjon og gjenfinning*. Tredje studieår tilbys et valgfritt fordypningskurs i kunnskapsorganisering som blant annet omhandler historikk og sentrale prinsipper innenfor bibliografisk virksomhet. De obligatoriske emnene inneholder katalogisering etter norske katalogregler (AACR2) og NORMARC-format. Det gis også innføringer i det norske «kataloglandskapet», sentrale produksjonsmiljøer og eierskap til bibliografiske data, samt en generell introduksjon til nyere metadataskjemaer. I tillegg til katalogisering får studentene undervisning i beslektede emner som relasjonsdatabaseteori, XML og datautveksling. For å skape en noe tettere sammenheng mellom disse emnene og katalogisering, gis det fellesforelesninger: Første år knyttet til FRBR og andre år til digitale bibliografiske tjenester.

---

<sup>9</sup> Se for eksempel Høgskolen i Gjøvik, NTNU og arkivutdanningen ved Høgskolen i Oslo og Akershus.

<sup>10</sup> Kurs ved UC Berkeley School of Information: <http://www.ischool.berkeley.edu/courses>

Antallet timer innenfor katalogisering er blitt redusert de siste årene i tråd med argumentene nevnt tidligere. Samtidig har en håndverksmessig tilnærming til katalogisering blitt tonet noe ned til fordel for et bredere perspektiv på bruk av deskriptive metadata. Likevel gjenstår en del på å integrere de ulike bitene i et konsistent, relevant og utvidet fagområde.

## **II: Et digitalt født metadataproduksjonsmiljø: MusicBrainz**

Vi vil videre i artikkelen undersøke hvordan metadata skapes og distribueres i et produksjonsmiljø uten bibliotekenes tunge historiske arv. Formålet er å finne ut hva som kjennetegner disse metadataene og deres produksjonsmåter, og vurdere hvordan vi kan dra nytte av disse innsiktene i utdanningen.

Vi bruker MusicBrainz som et eksempel på metadata som har oppstått digitalt og er dypt forankret i webens teknologiske og sosiale infrastruktur. Musikkmediet er valgt fordi det har vært spydspissen for endringene i digital kulturkonsumpsjon. I biblioteksektoren er registrering av musikk en spesialisert form for katalogisering. Dette er også et av de områdene innenfor biblioteksektoren hvor det per dags dato er et stort behov både for katalogiseringskompetanse og gode metadata som grunnlag for gjenbruk (Knutsen, 2007).

### ***MusicBrainz***

MusicBrainz ble etablert i 2003, men røttene ligger i musikkmetadatatenesten CDDB (Compact Disc Database) som startet opp i 1993, tre år etter at Tim Berners-Lee introduserte den tekniske infrastrukturen som weben hviler på. På denne tiden var mange personlige datamaskiner utstyrt med CD-lesere, og det fantes programvare som kunne spille av innholdet på musikk-CD-er. Fysiske CD-er inneholder ikke metadata som artistnavn, album- og låttitler og utgivelsesår, så denne informasjonen kunne derfor ikke vises i musikkspillernes display. To programmerere i California opprettet den webbaserte databasen CDDB for å tilby slike metadata til ulike mediespillere (Fry, 2001). Dette foregikk ved at mediespilleren beregnet en identifikator<sup>11</sup> for CD-en som ble spilt og sendte den over nettet til CDDB som søkte etter match i databasen. Ved treff ble metadata returnert til mediespilleren. Hvis CD-ens identifikator var ukjent, kunne brukeren taste inn informasjonen og registrere den hos CDDB. Slik ble CDDB en tidlig nettdugnad med bidrag fra musikkinteresserte over hele verden.

Etter utviklingen av MP3-formatet rundt midten av 1990-tallet, økte bruken av CDDB. Programvare som brukes til å rippe CD-er, dvs. å komprimere musikken på en CD til MP3-filer, tok i bruk CDDB for å legge inn metadata i disse filene. Selve MP3-spesifikasjonen sier ingenting om hvordan tekstlige metadata skal lagres. ID3

---

<sup>11</sup> Identifikatoren beregnes på grunnlag av fysiske egenskaper ved CD-en, som antall spor og deres lengde. Se [http://musicbrainz.org/doc/Disc\\_ID\\_Calculation](http://musicbrainz.org/doc/Disc_ID_Calculation)



(«Identify an MP3») ble utviklet som et tillegg i 1996, og er nå en de facto standard<sup>12</sup>. Metadataene som ligger i ID3-taggene blir brukt av programvare til å organisere musikkbiblioteker på de fleste datamaskiner og MP3-spillere.

CDDB ble kjøpt opp og relansert under navnet Gracenote i 2000, og dette varemerket eies i dag av Sony. Gracenote er den dominerende metadatakilden for digitale musikkspillere og -tjenester, og er tett integrert med for eksempel musikk-tjenester som iTunes og Spotify. Selve databasen vedlikeholdes med informasjon fra plateselskaper og distributører samt vanlige brukere som i CDDBs barndom.

Kommersiell utnyttelse av de brukerskapte dataene i CDDB ble sett på som et tillitsbrudd i fri programvaremiljøer (se for eksempel Van Buskirk, 2006; Dean, 2004). MusicBrainz<sup>13</sup> ble grunnlagt av Robert Kaye og lansert som et nettdugnadsinitiativ i 2003, delvis som en reaksjon på kommersialiseringen og lukkingen av CDDB-data. Ambisjonen var også å tilby bedre og mer konsistente metadata enn det Gracenote og andre musikkmetadatatenester gjør. MusicBrainz finansieres gjennom donasjoner fra brukere og øvrige sponsorer gjennom stiftelsen MetaBrainz<sup>14</sup>. Google har hittil vært den største bidragsyteren (MusicBrainz, 2010; mayhem, 2010).

MusicBrainz' hovedprodukt er selve databasen med informasjon om musikk-utgivelser. Per 22. mars 2011 består databasen av rundt 910.000 utgivelser og 10,8 millioner spor med 617.000 artister<sup>15</sup>. Til sammenlikning oppga Gracenote at de hadde rundt 8 millioner utgivelser og 100 millioner spor i sin database høsten 2010 (Gracenote, 2010). MusicBrainz hadde 22. mars 2011 drøyt 537.000 registrerte brukere som kan gjøre endringer i databasen, og av disse har rundt 1.400 gjort endringer den siste uken.

På nettstedet til MusicBrainz kan hvem som helst søke og bla i databasen. Der tilbys også klientprogramvare til gratis nedlasting. MusicBrainz Picard er den mest brukte klienten. En typisk bruker av MusicBrainz Picard eier en stor samling av MP3-filer, enten rippet fra CD eller lastet ned fra digitale musikkbutikker<sup>16</sup>. Eieren er ikke fornøyd med kvaliteten på metadataene i disse filene, og vil «vaske» samlingen mot MusicBrainz. MusicBrainz Picard gjør denne jobben ved å lese metadata fra filsamlingen, gruppere filene i album, hente metadata fra MusicBrainz, og skrive over og supplere ID3-tagger i filene med disse.

MusicBrainz er organisert som en nettdugnad, og ulike ordninger er innført for å sikre gode data og hindre vandalisme. Endringer i databasen blir ikke aktivisert

---

<sup>12</sup> Les mer om ID3 på <http://id3.org/>

<sup>13</sup> MusicBrainz: <http://musicbrainz.org/>

<sup>14</sup> MetaBrainz: <http://metabrainz.org/>

<sup>15</sup> Oppdatert statistikk kan hentes fra <http://musicbrainz.org/statistics/>

<sup>16</sup> Eller kanskje enda mer typisk: fra ulovlige fildelingsnettverk.

umiddelbart, men går gjennom en ganske kompleks avstemmingsprosess som er mer restriktiv enn den som brukes i Wikipedia. Bare registrerte brukere kan gjøre endringer i databasen, i motsetning til Wikipedia der det er åpent for alle. Etter å ha gjort et visst antall endringer, vil brukerne også kunne stemme på andre brukeres endringer.

Mindre endringer oppdaterer databasen direkte. Eksempler på slike er registrering og endring av sporenlengde, og bytte av store og små bokstaver i sangtitler. De fleste endringer må derimot kvalitetssikres gjennom avstemming. Hovedregelen er at endringer aktiviseres i databasen etter fjorten dager dersom ingen andre brukere stemmer nei til endringen. Dersom nei-stemmer blir registrert, vil det kreves en overvekt av tre ja-stemmer for at endringen skal tre i kraft. Hvis det etter fjorten dager ikke er en overvekt på tre ja-stemmer, blir endringen avvist.

Brukere kan legge inn notater til endringer de gjør, og det regnes som god skikk både å dokumentere med kildehenvisninger og å begrunne eventuelle nei-stemmer. Slik blir viktige prinsipielle diskusjoner om stilregler og redigeringspraksis knyttet til endringer i databasen, og disse bevares på samme måten som i Wikipedia. Slike diskusjoner er også viktig for opplæring og «disiplinering» av ferske deltakere i nettsamfunnet MusicBrainz.

### III: Undersøkelse

For å bli kjent med MusicBrainz' registreringspraksis og metadata, og kunne sammenlikne disse med tradisjonell bibliotekatalogisering, startet vi med å grunnregistrere to CD-er i henholdsvis MusicBrainz og NORMARC-formatet så detaljert som mulig. Vi ønsket å dekke bredden av musikkmetadata, og valgte én CD med populærmusikk og én med klassisk. Klassiske innspillinger er mer komplekse å beskrive enn populærmusikk, så det var spesielt viktig for oss å se hvordan de håndteres av MusicBrainz. Det andre kriteriet vi brukte, var at CD-ene ikke skulle være registrert i MusicBrainz fra før.

Følgende to CD-er, begge med norsk tilknytning, ble valgt:

- Populærmusikk: *Christianssands String Swing Ensemble* med Christianssands String Swing Ensemble, utgitt på CSSE Records i 2009
- Klassisk: *Concerto for Violin and Orchestra / Symphony No. 4 «The Inextinguishable»* med Royal Philharmonic Orchestra dirigert av Sir Yehudi Menuhin og Arve Tellefsen på fiolin, utgitt på Norsk plateproduksjon i 1989

#### ***Registreringspraksis***

Registrering i MusicBrainz startet for vår del med oppslag om CD-informasjon ved hjelp av klientprogramvaren Picard. MusicBrainz gjenkjente ikke CD-en, og genererte automatisk et webskjema for registrering basert på antall spor og med sporenlengde forhåndsutfyllt. Registreringsskjemaet består i stor grad av

nedtrekkslister med standardverdier. Dette sikrer konsistente data og virker tidsbesparende. Ved registrering av navn på artist, album, plateselskap eller liknende, må disse først søkes opp. Dersom søket resulterer i treff, blir brukeren presentert for en treffliste og kan knytte aktuell term eller navneform til egen registrering. I motsatt fall må informasjonen legges inn. Etter at aktuell artist (eller liknende entitet) er registrert, kan artisten søkes fram igjen og knyttes til den pågående registreringen.

MusicBrainz opererer med egne registreringstandarder, som kalles Official Style Guideline. De består av 25 enkeltregler inndelt i gruppene Utgivelsestitler, Sportitler, Artister, Generelle retningslinjer (herunder et separat dokument for klassiske utgivelser) og Utgivelsesattributter. De offisielle stilretningslinjene utvikles av The MusicBrainz Style Council, som er en e-postliste som alle interesserte MusicBrainz-medlemmer kan melde seg på. StyleCouncil er flatt organisert, men to personer har ansvaret for å styre diskusjonene og ta avgjørelser dersom diskusjonen på e-postlisten ikke når konsensus.

Stilreglene forklarer hvordan data i MusicBrainz skal formateres og organiseres. MusicBrainz understreker at stilreglene kun er retningslinjer, ikke et fastspikret regelverk som skal følges blindt ved registrering av data. Prinsippene om å følge artistenes intensjoner, om minst mulig bruk av forkortelser og liknende, kan overstyre stilreglene.

Registreringspraksis kjennetegnes ved at det blir lagt stor vekt på dokumentasjon ved hjelp av kildehenvisninger. Wikipedia, Amazon og Discogs er eksempler på kilder med troverdighet i miljøet. Endringsnotatene gjør det enklere for andre MusicBrainz-medlemmer å verifisere under avstemmingsprosessen. At data-registreringen overvåkes, fikk vi erfare ved registrering av CD-en fra Christianssand String Swing Ensemble. Det viste seg at CD-en anga feil komponist/tekstforfatter på ett av kuttene. Vi la inn Sholom Secunda som komponist og Jacob Jacobs som tekstforfatter til sporet *Bei mir bist du schön* og dokumenterte med en lenke til Wikipedia-artikkelen om sangen. To timer etterpå fikk vi advarsel fra en annen bruker om at våre opplysninger kunne være feil, fordi mange swing-versjoner av komposisjonen bygger på en engelskspråklig innspilling av The Andrew Sisters for det amerikanske markedet. Vedkommende bruker lyttet senere til Christianssand String Swing Ensembles versjon, og stemte deretter ned vår registrering. Vi endret våre data etter råd fra denne brukeren<sup>17</sup>.

Tradisjonelt har også bibliotekene lagt stor vekt på kvalitetskontroll gjennom nøye korrekturlesing og oppretting av feilregistreringer. Den økte mengden dokumenter og omfordeling av ressurser har ført til at det utføres langt mindre kvalitetskontroll

---

<sup>17</sup> Korrekte MusicBrainz-ansvarsangivelser for sporet er Sholom Secunda (komponist), Saul Chaplin (medkomponist) og Sammy Cahn (tekstforfatter).

enn tidligere. Det forventes at sentrale produksjonsmiljøer holder en god nok kvalitet til at poster kan gjenbrukes uten behov for korrigeringer.

Bibliotekenes registreringspraksis kjennetegnes ved at man baserer katalogiseringen på informasjon i det fysiske dokumentet og eventuell emballasje. Skrivefeil i dokumentet blir tydelig markert før man legger inn riktig form som «søkbart element» i MARC-posten. Dette står i motsetning til praksisen i MusicBrainz, hvor man legger inn riktig informasjon uten nærmere presisering annet enn i endringskommentarer. I bibliotekene inngår navn på personer eller korporasjoner ofte i autoritetsregistre, som er separate filer med foretrukket navneform og eventuelle aliaser. Funksjonsbetegnelse som angir en persons tilknytning til verket kan legges til. I slike tilfeller hentes verdien fra en relativt kortfattet liste i katalogiseringsreglene. Knappheten på valg av betegnelser gjør det ofte vanskelig å avgjøre hvilken tilknytning en person har til en gitt utgivelse. Dette vil kun framkomme i notefelt, som ren *presentasjon*. Tilsvarende vil sammenhengene mellom ulike verk, eller eksempelvis mellom en gruppe og medlemmer i denne, ofte bare framkomme som tekstlig beskrivelse i notefelt.

Katalogisering i henhold til katalogiseringsregler og MARC-format blander på denne måten maskinlesbare dataelementer med presentasjonen av innholdet i dem. Dette er en arv fra katalogkortene som var laget for lesing av mennesker, ikke maskiner. I MusicBrainz vil derimot alle slike sammenhenger være realisert som navngitte og datamaskin-prosesserbare relasjoner mellom entiteter i en relasjons-database (eksempel: **Bei mir bist du schön** *lyrics were written by Sammy Cahn*). De viktigste entitetstypene er artist, utgivelsesgruppe, utgivelse, utgivelseshendelse, spor og plateselskap. Alle entiteter innenfor disse typene får tildelt en unik identifikator (MBID). De eksplisitte identifikatorene og maskinlesbare relasjonene lagt inn manuelt av MusicBrainz-brukerne, brukes i dag av flere eksterne tjenesteleverandører.

Enkelte dataelementer registreres dobbelt i MARC-poster. Kuttene på et album vil for eksempel både listes opp i en innholdsnote og som autoriserte søkeinn ganger. Det forekommer også noe dobbeltregistrering ved at data kodes for maskinell tolkning, samtidig som de registreres for å presenteres som tekst for leseren. Kodingen er et uttrykk for automatiseringsmulighetene i MARC-formatet, mens tekstrepresentasjonen gjenspeiler katalogkortets funksjon. I databaseterminologi kalles dobbeltregistrering gjerne *redundans*, og anses som dårlig praksis.

Katalogiseringsregelverkets nær sagt altomfattende tilnærming gjør at få avgjørelser er overlatt til katalogisators skjønn. Dette sikrer konsistente data, men bidrar også til at katalogisering framstår som spesialisert arbeid forbeholdt en «ekspertgruppe». Dette står i motsetning til MusicBrainz, der enkle prinsipper styrer registreringen. Terskelen for å legge inn og endre data blir derfor lavere, og musikkinteresserte amatører kan delta i dugnaden.

Den største forskjellen i registreringspraksis og datakvalitet fant vi innen klassisk musikk. I bibliotekene får klassiske verk tilføyd standardtittel for å samle og presentere variasjoner av samme verk. Ved registrering av klassisk musikk i MusicBrainz legges det vekt på at artistnavnet skal være komponist (ikke utøvere) og at det i feltet *release title* legges inn en del tekstelementer for å skille ulike framføringer fra hverandre. Vi tolket stilreglene slik at vår klassiske CD fikk denne tittelen: *Concerto for Violin and Orchestra / Symphony No. 4 «The Inextinguishable» (Royal Philharmonic Orchestra feat. conductor: Sir Yehudi Menuhin, violin: Arve Tellefsen)*.

Samtidig som dirigent og fiolinist er kodet inn i tekststrengen, la vi inn relasjoner mellom de to utøverne og plateinnspillingen. Dette er et eksempel på redundante data i MusicBrainz. MusicBrainz erkjenner at dagens registreringspraksis og datamodell ikke har god støtte for klassisk musikk, og dette er også vår erfaring. Begge deler er derfor for tiden under revisjon<sup>18</sup>.

Ettersom det er brukerne som registrerer data, er MusicBrainz avhengig av å knytte til seg gode bidragsytere. MusicBrainz-grunnlegger Robert Kaye ser i så måte på samarbeidsavtalen som har blitt inngått med BBC som svært viktig for musikk-metadatatjenesten (Becky, 2009). MusicBrainz' pressemelding i sakens anledning<sup>19</sup> understreker at BBC ville tilføre en type kompetanse som har manglet i nettsamfunnet. BBC har svart til disse forventningene og blant annet vært med på å utvikle det nye databaseskjemaet hos MusicBrainz som er bedre tilpasset klassisk musikk. I en slik sammenheng er også naturlig å tenke seg at bibliotekarer kan inneha metadatakompentanse som vil være nyttig i MusicBrainz.

### ***Interoperabilitet***

I tillegg til registreringpraksis har vi sett på graden av *interoperabilitet*, som i vår sammenheng innebærer metadataskjemaenes evne til samvirke uten tap av data og funksjonalitet. Vi har konsentrert oss om en *semantisk* interoperabilitet, altså hvordan skjemaene understøtter forståelse for hverandres elementer (hva er en artist, hva er en tittel osv.) (Aalberg & Hegna, 2000).

For å sammenlikne de enkelte dataelementene systematisk, laget vi en konverteringstabell mellom MusicBrainz-registreringene og NORMARC-postene<sup>20</sup>. Vi hentet ut MusicBrainz-dataene fra deres XML-baserte programmeringsgrensesnitt, og oversatte hvert XML-element og -attributt til felter og delfelter i NORMARC. Vi fant at data stort sett lar seg konvertere til NORMARC-formatet, men noe informasjon vil gå tapt i konverteringen. Det gjelder for eksempel den utstrakte

---

<sup>18</sup> MusicBrainz ble oppdatert til NGS (Next Generation Schema), 18. mai 2011. Artikkelen ble skrevet før lanseringen, så vi har ikke hatt anledning til å vurdere støtten for klassisk musikk i den nye versjonen.

<sup>19</sup> <http://wiki.musicbrainz.org/BBCTeamsEditingMusicBrainz>

<sup>20</sup> Interesserte kan få tilgang til konverteringstabellen ved henvendelse til forfatterne.

bruken av identifikatorer og relasjoner, som ikke er like godt utviklet i NORMARC. Ulik registreringspraksis gjør også at ikke all informasjon som bibliotekene etterspør vil være representert i postene fra MusicBrainz. Noen av de manglende opplysningene kan autogenereres ved import, mens andre kan legges til manuelt i ettertid.

### ***Inskripsjoner***

Metadata fra MusicBrainz er altså i de fleste henseender kompatible med katalogdata. De største forskjellene ligger i tilpasningen til dagens teknologiske plattformer. Der har MusicBrainz løsninger som bibliotekene mangler. Dette skyldes først og fremst historie, mål og ønsker som er skrevet inn i metadata og praksiser, og som vi kaller *inskrripsjoner*.

Dette begrepet har vi lånt fra aktør-nettverkteori (ANT), hvor det brukes til å analysere og forstå vitenskapspraksis og teknologiutvikling (se for eksempel Bijker & Law, 1992; Hanseth, Aanestad, & Berg, 2004; Latour, 1987). I ANT-analyser betraktes teknologiutvikling og -bruk som et likeverdig samspill mellom mennesker og tekniske verktøy for å oppnå et bestemt mål (Monteiro, 2000). Både verktøyene, menneskene som konstruerer dem og de som bruker dem, inngår som aktører i et nettverk. Målsettingen med innføring av teknologi kan oppnås på forskjellige måter. Menneskenes rolle kan tillegges mest vekt ved utarbeidelse av regler og retningslinjer for bruk. Motsatt kan målet preges direkte inn i de tekniske verktøyene ved å utforme dem på en måte som fremmer spesifikke bruksmåter. Slike sterke føringer på hvordan en teknologi skal brukes, kalles inskripsjoner i ANT, og det skilles altså mellom ikke-tekniske (regelverk) og tekniske inskripsjoner (implementert i verktøy) (Hanseth & Monteiro, 1997). Inskripsjoner blir etter hvert del av en tung infrastruktur av mennesker, rutiner og teknologier. Slike historiske inskripsjoner kan være kostbare og vanskelige å endre.

De angloamerikanske katalogiseringsreglene (AACR2), som også benyttes i Norge, ble utviklet i en tid da automatiseringen av kortkatalogene var påbegynt. Likevel er all eksemplifisering – og dermed i praksis også bruken – av reglene i samsvar med ISBD (International Standard Bibliographic Description). ISBD er en internasjonal standard for form og innhold i bibliografisk beskrivelse, som ble utviklet for å forenkle menneskelig tolkning av fysiske katalogkort. En del data registreres på denne måten altså kun av presentasjonshensyn, samtidig som de inngår som en vesentlig del av det maskinleselige MARC-formatet. Det er svært vanskelig å programmere datamaskiner til å fortolke disse dataene, og de har derfor begrenset funksjon i digitale tjenester som maskinelt prøver å foredle data til kunnskap. ISBD, og dermed også AACR2 og MARC-formatet, framstår som anakronismer i en slik sammenheng.

At nye registreringspraksiser bevarte kortkatalogens fysiske egenskaper, skyldtes nok delvis at relasjonsdatabasene og deres muligheter til å uttrykke sammenhenger maskinelt, ble utviklet på et senere tidspunkt enn bibliotekenes standarder

(Thomale, 2010). Katalogkortet inngår rett og slett i bibliotekenes datasystemer som inskripsjoner i registreringsstandarder og metadataskjemaer.

På begynnelsen av 1990-tallet nedsatte IFLA en arbeidsgruppe som analyserte funksjonskrav til bibliografiske poster i et nytt digitalt landskap. Arbeidsgruppen avla rapporten *Functional Requirements for Bibliographic Records* (FRBR) i 1998. Gjennom en ER-analyse<sup>21</sup> av bibliografiske poster fikk bibliotekene et teoretisk rammeverk og et begrepsapparat tilpasset databasemodellering. FRBR er i liten grad implementert i systemene som benyttes i bibliotekene. Dette skyldes hovedsakelig arven fra tidligere standarder og manglende tilpasninger til nye teknologier (Westrum, 2011).

I 2004 startet et revisjonsarbeid av AACR2. Det nye regelverket, RDA, ble lansert i 2010 og er organisert i tråd med FRBR-rammeverket. I motsetning til sin forgjenger, er RDA en rendyrket registreringsstandard uten eksemplifisering med verken ISBD eller MARC-format. Likevel har RDA blitt kritisert for å ligge for tett opp til sin forgjenger, ettersom det blant annet fortsatt forutsetter en feltvis sammenblanding av potensielt maskinleselige data og tekst myntet på menneskelig fortolkning (se for eksempel Coyle & Hillmann, 2007). Det omfattende arbeidet med FRBR og RDA viser på ulike måter hvordan både tekniske og ikke-tekniske inskripsjoner kan forgrene seg inn i arbeidet med nye teknologier, og komplisere nødvendige endringsprosesser.

MusicBrainz har en helt annen opprinnelse enn bibliotekenes metadata og standarder, og er et ektefødt barn av weben. For eksempel bidrar unike identifikatorer på ulike deskriptive nivåer til at maskiner kan gjenkjenne artister og utgivelser, og relatere dem effektivt både til hverandre og til forekomster i andre webbaserte datasett hvor de også er unikt identifiserbare. Bruken av standardiserte formater som XML og RDF letter en slik utveksling på tvers av metadatamiljø. MusicBrainz inneholder også URL-lenker til beslektede tjenester som Discogs og Wikipedia, og alternative identifikatorer til de interne primærnøklerne. Identifikatorene er hensiktsmessig å bruke for å koble sammen data fra MusicBrainz med data fra andre leverandører. Disse egenskapene ved MusicBrainz er blant årsakene til at grunnleggeren Robert Kaye ble invitert av Tim Berners-Lee som en representant for webens framtid til webens 20-årsjubileum avholdt ved CERN i 2009 (Hemerly, 2010).

Bibliotekene har ikke utviklet et tilsvarende identifikatorsystem som MusicBrainz. Med mindre denne situasjonen endrer seg, er det lite sannsynlig at katalogdata vil kunne være en egnet metadatakilde for webbaserte tjenester. Forsøk på å konvertere katalogdata til FRBR-modellen viser også hvordan manglende

---

<sup>21</sup> ER står for entity-relationship, og er en konseptuell modelleringsteknikk som brukes mest innenfor databasutvikling.

identifikatorer er et problem (Pisanski & Žumer, 2010; Westrum, 2011). Bibliotekene deltar aktivt i internasjonale fora for å utvikle relevante identifikatorer, men foreløpig er ikke disse utbredt i bibliotekenes eller andres registreringspraksiser.

Oppsummert vil vi hevde at katalogkortet er videreført som en ikke-teknisk inskripsjon i bibliotekenes registreringsstandard (AACR2) og metadataskjema (MARC). Eksempler er registrering av data i notefelt, redundante data og språkavhengig informasjon<sup>22</sup> samt ISBD-skilletegn i tekststrenger. Alle disse praksisene er unødvendige i dag fordi dataprogrammer ved behov kan generere slike data og tekststrenger automatisk ved eksport til MARC. I tillegg er det sterke tekniske inskripsjoner for *registrering* av data i MARC, som opprinnelig var et dataformat for *datautveksling*. Derfor videreføres katalogkortene også inn i data-systemene, og i bibliotekarenes kompetanse og praksis.

På den andre siden har MusicBrainz' registreringsskjema ekstremt sterke, nærmest rigide, inskripsjoner som tvinger brukerne til en bestemt oppførsel: å registrere maskinlesbare data. Dette er et eksempel på tekniske inskripsjoner i et dataverktøy. Alle data som kommer inn i MusicBrainz, blir automatisk en del av den semantiske weben. Dette er et eksplisitt mål for grunnleggeren Robert Kaye, som til og med ser ut til å bli inskribert i brukerne. Her fra en undersøkelse av brukernes motivasjoner for å registrere data i MusicBrainz:

One editor remarked that he stopped thinking about MusicBrainz as a tagging solution and started to think about it as a Semantic Music Encyclopedia: «Something like a structured Wikipedia – that is, something that can be processed by machines and helps building various tools on top of it». (Hemerly, 2011, s. 56)

Vi skal ikke gjøre dette til en krig mellom bibliotekene og MusicBrainz, men for oss er det åpenbart at metadataskjema, registreringsstandarder og biblioteksystemer ville sett annerledes ut om de hadde blitt utviklet på 2000-tallet. Bibliotekenes historie eksisterer gjennom katalogkortene, og disse dataene må bevares gjennom alle teknologiske oppgraderinger. Tilsvarende vil muligens MusicBrainz' historiske inskripsjoner bli en klump om foten for tjenesten ved teknologiske endringer i fremtiden. Likevel vil vi konkludere med at MusicBrainz har egenskaper som er langt bedre tilpasset dagens teknologiske infrastruktur enn bibliotekene. En bibliotekar som ikke forstår dette, vil ikke ha den metadatakompetansen som er nødvendig i dag eller i årene som kommer.

### ***Åpne og fritt tilgjengelige data***

MusicBrainz er et eksempel på et sofistikert metadatamiljø bygd rundt åpne og brukerskapte data. Åpne data betyr i denne sammenhengen å legge færrest mulig

---

<sup>22</sup> For eksempel sidetall, som registreres som *s.* på norsk og *p.* på engelsk.



restriksjoner på kommersiell og ikke-kommersiell utnyttelse av dataene. Dette er et prinsipp som ligger tett opp til bibliotekenes samfunnsoppdrag. Selv om produksjon av bibliotekenes metadata direkte og indirekte er offentlig finansiert, er de i motsetning til data fra MusicBrainz i praksis en kommersiell vare i Norge (se for eksempel Knutsen, 2007; Massey, Knutsen, & Tallerås, 2010; Søndersrød, 2009). Bruerskapte data er ikke bare idealisme for MusicBrainz, men også motivert ut fra ønsket om å sikre bedre kvalitet enn dagens kommersielle og profesjonelle aktører kan tilby. Mens enkelte skribenter stempler amatørernes innholdsproduksjon som forflatning og forfall (Gorman, 2007; Keen, 2007; Siegel, 2008), ser altså MusicBrainz på distribuert brukerdeltakelse som en styrke.

Vi erfarte at nettdugnaden og kvalitetskontrollen fungerer godt. MusicBrainz-brukerne er kanskje katalogiseringsamatører, men til gjengjeld er de eksperter på utvalgte musikkformer. Denne ekspertisen kan være spredt globalt, men til sammen utgjør den et effektivt registrerings- og kontrollregime som sikrer gode metadata. Registrerte brukere i MusicBrainz kan abonnere på endringer for artister eller utgivelser på plateselskap, og har derfor muligheten til å overvåke sine spesialistområder. Disse brukernes ekspertkunnskaper vil også kunne være nyttige for bibliotekene.

#### **IV: Konsekvenser for utdanningen**

Vi spurte innledningsvis om hva slags metadatakompetanse den framtidige bibliotekaren trenger, og hvordan katalogiseringsundervisningen ved bibliotekarutdanningen i Oslo bør utformes.

Den internasjonale utdanningsdebatten innen biblioteksektoren er – noe satt på spissen – preget av to ulike posisjoner: på den ene siden nedbygging eller avskaffelse av det tradisjonelle katalogiseringsfaget, på den andre siden opprettholdelsen av det samme.

Hovedargumentet for å kutte ut eller skjære ned på katalogiseringsundervisningen er at kunnskapen er irrelevant for den generelle bibliotekaren i dagens teknologiske og bibliografiske landskap der sentrale leverandører dominerer. Dette er ikke en helt korrekt virkelighetsbeskrivelse ettersom grunnkatalogisering av visse typer dokumenter og tilpasning av importerte metadata fortsatt er viktige arbeidsoppgaver i folke- og fagbibliotekene (Knutsen, 2007, 2009). Dessuten er det grunn til å frykte at bibliotekarene i et slikt scenario blir redusert til passive konsumenter av sentrale leverandørers tjenester, ute av stand til å stille konkrete krav til utviklingsprosessene. Det er i denne sammenhengen også grunn til å spørre om hvem andre enn bibliotekarutdanningene som skal utdanne kandidater med tilstrekkelig kompetanse til å kunne virke i de sentrale katalogiseringsmiljøene.

Samtidig er det slik at bibliotekene trenger metadata som kan brukes til noe mer enn lokal logistikk. Standardene som brukes i dag baserer seg på logikk knyttet til

fysisk gjenfinning og formidling. Det er rett og slett vanskelig å håndtere bibliotekenes «atomære» katalogposter i webens infrastrukturer av lenker og relasjoner. Nyutdannede bibliotekarer må være i stand til å hjelpe til med å fornye bibliotekenes metadatapraxis. De som hevder at bibliotekarene må få massiv opplæring i tradisjonell katalogisering slik den har blitt praktisert i bibliotekene opp gjennom årene, har dermed etter vårt syn et for snevert perspektiv på bibliotekarenes metadatakompetanse. Et institusjonsinternt blikk er dessuten altfor lite offensivt i dagens ekspansive metadatalandskap. Vi har sett at metadata er blitt pensum i andre utdanninger enn vår egen. Dette er positivt og viser at én av bibliotekarprofesjons kjernekompetanser er blitt allmenngjort og etterspurt. Likevel er dette metadata forstått i en annen sammenheng enn i bibliotekene. Vår utdanning bør hente inspirasjon fra disse miljøene for å videreutvikle en relevant metadataundervisning. Det vil være en fallitterklæring å overlate metadatakompetansen, som alltid har vært sentralt i profesjonens kunnskapssystem, til andre.

MusicBrainz er ett av mange eksempler på at metadata i dag produseres i miljøer utenfor bibliotekene. Utbredt bruk i digitale musikkspillere og -tjenester viser at denne typen metadata er etterspurt. Flere egenskaper ved MusicBrainz skiller dette initiativet fra tradisjonelle metadataproduksjonsmiljøer: Metadataene som produkt er knyttet til digitale objekter, de utvikles desentralisert og av amatører, og de er fritt tilgjengelige for bruk. Disse egenskapene gjør på den annen side MusicBrainz til et typisk eksempel på samtidens webtjenester.

Vår undersøkelse av MusicBrainz-data viser at en bredt anlagt metadatakompetanse er sammensatt av flere teknologiske komponenter som henger tett sammen med selve produksjonen. I visjonen om en semantisk web uttrykkes koblingen mellom metadata og teknologi eksplisitt: «In the near future, these developments will usher in significant new functionality as machines become much better able to process and «understand» the data that they merely display at present. [...] The Semantic Web will enable machines to COMPREHEND semantic documents and data [...]» (Berners-Lee, Hendler, & Lassila, 2001). Det forutsettes med andre ord at metadata organiseres og lenkes sammen slik at maskiner kan forstå strukturene og resonnerer på bakgrunn av dem. Dette krever at vi som metadataprodusenter har innsikt i hvordan en datamaskin «tenker», og kjenner til teknologi som bidrar til at den «tenker» best mulig. Slike prinsipper om «machine-processability» (Nilsson, 2010) eller «computable semantic interoperability» (Mead, 2006) er etter hvert også godt representert i litteraturen som omhandler metadatakvalitet. Dette er en litteratur som gradvis har vokst frem på 2000-tallet, og som søker å etablere generelle kvalitetskriterier uavhengige av spesifikke metadataskjema, -tradisjoner og -praksiser (se for eksempel Bruce & Hillmann, 2004; Duval, Hodgins, Sutton, & Weibel, 2002; Haslhofer & Klas, 2010; Hillmann & Phipps, 2007; Nilsson, 2010).

Katalogiseringsundervisningen ved bibliotekarutdanningen i Oslo har tradisjonelt vært knyttet opp til MARC og katalogiseringsregler. Samtidig har vi i utdannings-

forløpet inkludert flere dataorienterte emner som XML, programmering og relasjonsdatabaseteori. Undervisningen innenfor disse områdene har allikevel i stor grad foregått parallelt med katalogiseringsundervisningen. Denne splittelsen kan spores tilbake til studieomleggingen i 1982. Da ble studietilbudet redusert fra sju til fire hovedområder (Henriksen, 1990):

- Metodefag
- Litteratur og samfunn
- Informasjonskunnskap
- Administrasjon/organisasjon

Metodefagenes innhold beskrives slik i studiehandboken for studieåret 1999-2000:

Metodefagene omfatter delfag som samfunnsvitenskapelige metoder, statistikk, EDB, matematikk, lingvistikk og pedagogikk. Det karakteristiske ved metodefagene er at de skal være til støtte ved tilegnelse av de mer spesifikke bibliotekfag, i tillegg til at de skal være til støtte ved senere yrkesutøvelse. (Høgskolen i Oslo. Avdeling for journalistikk, bibliotek- og informasjonsfag, 1999, s. 41)

Et eksempel på metodefag dette studieåret var WWW-publisering med PERL. Emnet Anvendt EDB, som inneholdt «datamodellering og øvelser i bruk av databaseverktøy i nettmiljø», var derimot et valgfag innen hovedområdet Informasjonskunnskap.

Faginndelingen fra 1982 ble opprettholdt fram til studieåret 2000-2001. Da ble hovedområdene omdøpt og redusert til tre *emnekretser*. De tre emnekretsene heter det samme som i dag: *Bibliotek og samfunn*, *Litteratur og bruker*, og *Kunnskapsorganisasjon og gjenfinning*. Metodefag ble fjernet som en selvstendig bolk, men videreføres i studiehandboken. Betegnelsen for emnekretsen *Kunnskapsorganisasjon og gjenfinning*, som er mest relevant i denne sammenhengen, etterfølges av teksten «med metoder og redskaper fra matematikk og informatikk». Skillet mellom bibliotekfag og metodefag er også tydelig i fagbeskrivelsene, der «de mer bibliotekspesifikke delene av IKT» blir omtalt som redskaps- og hjelpefag til *Kunnskapsorganisasjon og gjenfinning*. Delemnet som dekker disse fagene blir de nærmeste årene kalt for *Gjenfinningssystemer og verktøy* første og andre studieår. Her inngår databaseteori, markeringsspråk og programmering. (Høgskolen i Oslo. Avdeling for journalistikk, bibliotek- og informasjonsfag, 2000)

Vi mener at dette skillet mellom metadatakompetanse og datafag tydeligere må utvikles for å gjøre undervisningen relevant. Dersom vi ikke integrerer teknologiaspektene, risikerer vi å redusere metadatakompetansen til håndverksmessige ferdigheter knyttet til spesialiserte og etter hvert isolerte registreringspraksiser.

I tillegg til bibliotekenes tradisjonelle metadataproduksjon, må studentene gjøres kjent med andre metadatafamiljer, deres skjemaer og registreringsstandarder. Det er

samtidig viktig å overkomme et rent håndverksmessig perspektiv, både i henhold til bruk av konkrete skjema og spesialiserte teknologier. Utdanningen må gi studentene et teoretisk grunnlag som gir forståelse og overføringsverdi på et generelt nivå. Dette innebærer blant annet at historikk og teoretisk forankring må introduseres tidligere i studieløpet.

Universiteter og høyskoler skal «tilby høyere utdanning som er basert på det fremste innen forskning, faglig og kunstnerisk utviklingsarbeid og erfaringskunnskap» (Universitets- og høyskoleloven, 2005). Bibliotekarutdanningen vil legge mer vekt på teori, men samtidig er det verken ønskelig eller mulig å gjøre katalogisering til en ren akademisk disiplin. De senere årenes innsikter og erfaringer med webbruk av metadata generelt og MARC-data spesielt, må inkluderes i undervisningsopplegget. En slik bredde vil nødvendigvis bety at den tradisjonelle katalogiseringsundervisningen reduseres. Grunnprinsippene og den praktiske håndteringen av visse typer materiale må med, men kandidatene må i sterkere grad enn i dag tilegne seg praktiske ferdigheter på arbeidsplassen etter endt utdanning.

Vi konkluderer med at våre kandidaters framtidige metadatakompetanse må være bredere enn i dag. Dette gjelder både de som jobber i og utenfor bibliotekene. Metadata er motoren for de fleste store webtjenester, og for å kunne operere effektivt og innovativt må bibliotekaren forstå hele kretsløpet for produksjon og distribusjon. Vi må fortsatt legge særskilt vekt på kunnskap og ferdigheter innenfor bibliotekenes metadatapraksis, men uten en bredere forståelse av omgivelsene, vil bibliotekskatalogene bli irrelevante datasiloer i webens periferi. MARC-data brukes knapt nok utenfor bibliotekene. For å bevare og styrke sin posisjon som metadataekspert, må bibliotekarer utvide horisonten gjennom innsikt i ulike metadatapraksiser og kompetanse til å vurdere og dele metadata på tvers av skjemaer, formater og systemer.

Ved utdanningen i bibliotek- og informasjonsfag kan vi kontekstualisere både katalogiseringsfagene og de øvrige fagene som inngår i emneområdet *Kunnskapsorganisasjon og gjenfinning* ved å bygge ned skillene mellom datafag og bibliotekfag. Weben er en plattform som gjør metadatakompetanse mer aktuell enn noen gang. Vi har sett at historien videreføres som inskripsjoner i teknologi og metadata. Vi må unngå at bibliotekarenes kompetanse får tilsvarende inskripsjoner fra en svunnen tid. Det er viktig at kandidatene kjenner bibliotekenes metadatahistorie, men de må samtidig være i stand til å se utover egne tradisjoner. For å få til dette vil vi ruste opp undervisningen innenfor metadata ved å fylle den med en mer grunnleggende og abstrakt kompetanse tilpasset nye og bredere bruksområder.

## Litteratur

- Aalberg, T., & Hegna, K. (2000). *Arkitektur for digitale bibliotek*. Trondheim: Bibsys.
- Becky. (2009, 27. november). Featured project: MusicBrainz [Blogginnlegg]. Hentet fra <http://blog.okfn.org/2009/11/27/featured-project-musicbrainz/>
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37.
- Bijker, W. E., & Law, J. (1992). *Shaping technology/ building society: Studies in sociotechnical change*. Cambridge, MA: MIT Press.
- Birrell, D., Dunsire, G., & Menzies, K. (2009). *Online catalogue and repository interoperability study (OCRIS): Final report*. Hentet fra <http://eprints.rclis.org/bitstream/10760/13187/1/strathprints014037.pdf>
- Bruce, T. R., & Hillmann, D. I. (2004). The continuum of metadata quality: Defining, expressing, exploiting. I D. I. Hillmann & E. L. Westbrook (Red.), *Metadata in practice* (s. 203-222). Chicago: American Library Association.
- Coyle, K., & Hillmann, D. I. (2007). Resource Description and Access (RDA). *D-Lib Magazine*, 13(1/2). doi:10.1045/january2007-coyle
- Davis, J. M. (2008). A survey of cataloging education: Are library schools listening? *Cataloging & Classification Quarterly*, 46(2), 182-200.
- Dean, K. (2004, 7. juli). *The house that music fans built*. Hentet fra <http://www.wired.com/entertainment/music/news/2004/07/64033>
- Duval, E., Hodgins, W., Sutton, S., & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8(4). Hentet fra <http://www.dlib.org/dlib/april02/weibel/04weibel.html>
- Fry, J. (2001, 31. desember). Three veterans advise the next tech wave: It's all about business. *The Wall Street Journal*. Hentet fra [http://www.ibiblio.org/tkan/software/cddb\\_wsj\\_12.31.01.pdf](http://www.ibiblio.org/tkan/software/cddb_wsj_12.31.01.pdf)
- Gorman, M. (2002). Why teach cataloguing and classification? I J. S. Hill (Red.), *Education for cataloging and the organization of information: Pitfalls and the pendulum* (s. 1-13). Binghamton, NY: Haworth Information Press.
- Gorman, M. (2007, 11. juni). Web 2.0: The sleep of reason, part I [Blogginnlegg]. Hentet fra <http://www.britannica.com/blogs/2007/06/web-20-the-sleep-of-reason-part-i/>
- Gracenote. (2010, 9. september). *Gracenote reaches global media database milestone*. Hentet fra <http://www.gracenote.com/press/09/09/2010/>
- Hanseth, O., & Monteiro, E. (1997). Inscribing behaviour in information infrastructure standards. *Accounting, Management and Information Technologies*, 7(4), 183-211. doi:10.1016/S0959-8022(97)00008-8
- Hanseth, O., Aanestad, M., & Berg, M. (2004). *Actor-network theory and information systems*. Bradford: Emerald.
- Haslhofer, B., & Klas, W. (2010). A survey of techniques for achieving metadata interoperability. *ACM Computing Surveys (CSUR)*, 42(2), 7.

- Hegna, K. (udatert). *Dublin Core: Keiserens nye klær?* Hentet fra <http://heim.ifi.uio.no/~knuthe/dok/sesam.pdf>
- Hemerly, J. (2010, 3. mai). Interview [Blogginlegg]. Hentet fra <http://jhemerly.wordpress.com/robert-kaye-interview/>
- Hemerly, J. (2011, mai). *Making metadata: The case of MusicBrainz* (Masteroppgave, UC Berkeley School of Information). Hentet fra [http://www.ischool.berkeley.edu/files/student\\_projects/jhemerly\\_mims\\_finalproject.pdf](http://www.ischool.berkeley.edu/files/student_projects/jhemerly_mims_finalproject.pdf)
- Henriksen, T. (1990). Bibliotekutdannelsen i Norge. I H. E. Aarek (Red.), *Mer enn bøker: SBIH 50 år: Jubileumsskrift* (s. 9-16). Oslo: Statens bibliotek- og informasjonshøgskole.
- Hillmann, D. I., & Phipps, J. (2007). Application profiles: Exposing and enforcing metadata quality. I *Proceedings of the 2007 International Conference on Dublin Core and Metadata Applications: Application profiles: theory and practice* (s. 53–62). Hentet fra <http://portal.acm.org/citation.cfm?id=1344591.1344601>
- Hodge, G. M. (2001). *Metadata made simpler*. Bethesda, MD: NISO Press.
- Høgskolen i Oslo. Avdeling for journalistikk, bibliotek- og informasjonsfag. (1999). *Studiehåndbok 1999-2000*. Oslo: Høgskolen i Oslo.
- Høgskolen i Oslo. Avdeling for journalistikk, bibliotek- og informasjonsfag. (2000). *Studiehåndbok 2000-2001*. Oslo: Høgskolen i Oslo.
- iSchools. (udatert). *About the iSchools*. Hentet 20. september 2011 fra <http://www.ischools.org/site/about/>
- Kajberg, L., & Lørring, L. (Red.). (2005). *European curriculum reflections on library and information science education*. København: Royal School of Library and Information Science.
- Keen, A. (2007). *The cult of the amateur: How today's Internet is killing our culture*. New York: Doubleday/Currency.
- Kelley, M. (2011, 31. august). How the W3C has come to love library linked data. *Library Journal*. Hentet fra [http://www.libraryjournal.com/lj/home/891826-264/how\\_the\\_w3c\\_has\\_come.html.csp](http://www.libraryjournal.com/lj/home/891826-264/how_the_w3c_has_come.html.csp)
- Knutsen, U. (2007). *Ut og stjele poster? : Effektivitet og rasjonalitet i folkebibliotekenes katalogarbeid* (Masteroppgave, Høgskolen i Oslo). Hentet fra <https://oda.hio.no/jspui/handle/10642/315>
- Knutsen, U. (2009). *Fragmentering eller fellesløsning?: Organisering av norske bibliografiske produksjon*. Hentet fra <http://www.abm-utvikling.no/publisert/abm-skrift/abm-skrift-60>
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Milton Keynes: Open University Press.
- Library of Congress Working Group on the Future of Bibliographic Control. (2008, 9. januar). *On the record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Hentet fra <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>
- Massey, D., Knutsen, U., & Tallerås, K. (2010). Frigjør data - Hva fanden nøler I efter? *Bok og bibliotek*, (3), 60-62.

- mayhem. (2010, 22. februar). Google ups its sponsorship of MusicBrainz in 2010! [Blogginnlegg]. Hentet fra <http://blog.musicbrainz.org/?p=456>
- Mead, C. N. (2006). Data interchange standards in healthcare IT: Computable semantic interoperability: Now possible but still difficult, do we really need a better mousetrap? *Journal of Healthcare Information Management*, 20(1), 71-78.
- Monteiro, E. (2000). Monsters: From systems to actor-network. I K. Braa, C. Sørensen, & B. Dahlbom (Red.), *Planet Internet* (s. 239-249). Lund: Studentlitteratur.
- MusicBrainz. (2010, 15. september). *MusicBrainz sponsors*. Hentet fra [http://musicbrainz.org/doc/MusicBrainz\\_Sponsors](http://musicbrainz.org/doc/MusicBrainz_Sponsors)
- Nilsson, M. (2010). *From interoperability to harmonization in metadata standardization: Designing an evolvable framework for metadata harmonization* (Doktoravhandling, Kungliga tekniska högskolan). Hentet fra <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-26057>
- Pisanski, J., & Žumer, M. (2010, 15. juni). *Identifiers: Bridging language barriers*. Hentet fra [www.ifla.org/files/hq/papers/ifla76/93-pisanski-en.pdf](http://www.ifla.org/files/hq/papers/ifla76/93-pisanski-en.pdf)
- Siegel, L. (2008). *Against the machine: Being human in the era of the electronic mob*. London: Profile.
- Søndersrød, V. (2009). *Katalogdata: Juridiske problemstillinger knyttet til eierskap og produksjon*. Oslo: ABM-utvikling.
- Tennant, R. (2002, 15. oktober). *MARC must die*. Hentet fra <http://www.libraryjournal.com/article/CA250046.html>
- Thomale, J. (2010). Interpreting MARC: Where's the bibliographic data? *Code4Lib Journal*, 2010(11). Hentet fra <http://journal.code4lib.org/articles/3832>
- U.S. RDA Test Coordinating Committee. (2011, 20. juni). *Report and recommendations of the U.S. RDA Test Coordinating Committee*. Hentet fra <http://www.loc.gov/bibliographic-future/rda/rdatesting-finalreport-20june2011.pdf>
- Universitets- og høyskoleloven. (2005). *Lov om universiteter og høyskoler (universitets- og høyskoleloven)*. Hentet fra <http://www.lovdatab.no/all/hl-20050401-015.html>
- Van Buskirk, E. (2006, 13. november). *Gracenote defends its evolution*. Hentet fra <http://www.wired.com/entertainment/music/commentary/listeningpost/2006/11/72105>
- Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York: Times Books.
- Westrum, A.-L. (2011, 22. februar). Tre år med PODE [Blogginnlegg]. Hentet fra <http://www.bibpode.no/blogg/?p=1787>