Bibliotheca Nova
4-2014

Kunnskapsorganisering

Nasjonalbiblioteket

Bibliotheca Nova
4-2014

Kunnskaps-organisering

Nasjonalbiblioteket 2014

Skriftserien Bibliotheca Nova
Skriftserien Bibliotheca Nova er en av Nasjonalbibliotekets formidlingskanaler for bibliotekutvikling. Ambisjonen er å publisere artikler som inspirerer og gir faglig støtte til innovasjon i bibliotek.

Gjennom tildeling av utviklingsmidler fra Nasjonalbiblioteket synliggjøres ulike innsatsområder på bibliotekfeltet. Disse er politisk definerte satsingsområder og prioriterte i samråd med Nasjonalbibliotekets rådgivende utvalg. Ved tildelingene til prosjekter har Nasjonalbiblioteket forutsatt at erfaringene fra prosjektene og innsatsområdene lett skal kunne deles med andre via artikler i Bibliotheca Nova. Vi skal ikke bare lære av hverandre, men også av andre. Faglige artikler innenfor områder som er viktige og relevante for bibliotekutvikling vil inngå i skriftserien Bibliotheca Nova. Det kan være foredrag holdt på konferanser eller artikler innenfor aktuelle temaer på bestilling av fagfolk.

Vi tar sikte på at Bibliotheca Nova skal komme ut med 4–5 utgaver i året.
Alle utgavene vil bli sendt til bibliotekene. Ekstra eksemplarer kan bestilles fra magasintjenesten@nb.no så langt opplaget rekker.

Prosjektoversikter:
http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Prosjektoversikter

Utgiver: Nasjonalbiblioteket
Nr. 4–2014

Redaktør: Gunhild H. Aalstad
Redaksjonskomitè: Irene Hole, Oddrun Pauline Ohren og Elise Conradi

ISSN 1894-1427
ISBN 978-82-7965-227-4 (trykt)
ISBN 978-82-7965-228-1 (digital)

Design: Melkeveien Designkontor AS
Trykk: Rolf Ottesen as

Opplag: 1200

www.nb.no

Bibliotheca Nova
4-2014

Innhold

FORORD

Helt fra de første bibliotekene oppsto, har en av de sentrale aktivitetene rundt samlingene vært kunnskapsorganisering. Samlingsforvalterne har arbeidet systematisk med å gjøre det enklere å finne fram i informasjonen, uavhengig av om mediene var leirtavler, papyrus eller pergament. Da – som nå – var det ofte tre sentrale elementer som gikk igjen når det skulle skapes orden i samlingen: forfatter, tittel og emne.

I dagens digitale virkelighet er behovet for god og presis gjenfinning sannsynligvis større enn det var for bare noen få år siden. Informasjonsmengdene som er tilgjengelig vil som regel langt overskride det som er mulig å få oversikt over. Et enkelt søk på Google returnerer en treffliste som man som regel ikke kan ha noe håp om å komme gjennom. En av farene med denne overveldende datamengden er at man nøyer seg med det man får. Gråsteinen som glimrer anses som god nok, mens gullet blir liggende begravd dypt nede i resten av fjellet.

Her har bibliotekene en viktig oppgave å utføre. I alle år har bibliotekansatte forsøkt å lage strukturer rundt dokumentmengdene. Materiale er blitt katalogisert, klassifisert og indeksert. Forfattere er identifisert og lagt inn i autoritetsregistre. Bibliografier er blitt utarbeidet. Emneordslister og klassifikasjonstermer brukes for å gi kunnskap om innholdet i publikasjonene. Standardnumre brukes for å skille mellom ulike publikasjoner. Denne kompetansen er ikke blitt overflødig i det digitale domenet, snarere tvert i mot. Strukturerte metadata er fortsatt et kraftig verktøy og et helt nødvendig supplement til fritekstsøking, og potensialet er på langt nær utnyttet fullt ut. Søkegrensesnitt som utnytter fordelene både med strukturerte metadata og fritekstsøk, vil omdanne den digitale støyen og gi oss muligheter ingen har hatt før oss.

Dette nummeret av Bibliotheca Nova handler om kunnskapsorganisering. Vi har forsøkt å trekke fram en del av de spennende tingene som skjer på

Foto: Evy Andersen

dette feltet her til lands. En fellesnevner ved artiklene er at de beskriver den evnen til nytenkning og nyorientering som finnes i bibliotekene og lignende institusjoner. Ulike aktører prøver hver på sitt felt å skape bedre systemer for gjenfinning i store informasjonsmengder. Noen av artiklene tar utgangspunkt i utviklingsprosjekter støttet av Nasjonalbiblioteket. Andre artikler er knyttet an til de oppgavene Nasjonalbiblioteket har som nasjonal koordinator på dette feltet. Andre artikler kommer fra bibliotekenes naboområder, museer og arkiv.

De fleste artiklene peker framover mot morgendagens informasjons- og kunnskapshåndtering. Gode digitale verktøy og nye muligheter til å skape sammenheng i kunnskapsuniverset, gjør at bibliotekansatte kan bli viktige bidragsytere i kampen mot «information overload». Dette krever imidlertid at sektoren tar inn over seg utviklingen på området de siste tiårene, at bibliotekene satser på tjenester som fungerer godt sammen med andres tjenester og at fokus legges på å utvikle verktøy som setter sluttbrukerne i stand til å finne fram i datamengdene. I tillegg er det ingen ulempe å forholde seg til globaliseringen, og sørge for at nasjonale trender spiller på lag med det som skjer av utvikling i resten av verden. Det er også viktig å ha i mente at mye av utviklingen som må skje på disse feltene er avhengig av et samspill mellom ulike fagfelt. Lingvister, informatikere og bibliotekarer har hver på sin side fagkunnskaper som satt sammen kan gi gode løsninger.

Foreliggende hefte vil forhåpentligvis kunne inspirere til økt aktivitet på kunnskapsorganiseringsfeltet. En videreutvikling av fagområdet er avhengig av gode fagmiljø, dedikerte aktører og en ikke ubetydelig grad av innovasjonsevne.

Jonny Edvardsen

DIREKTØR

TILVEKST OG KUNNSKAPSORGANISERING

Kunnskapsorganisasjon – «kjerneteknologi» for bibliotek- og informasjonsfag

TEKST: **RAGNAR NORDLIE**

«allerede de gamle egyptere –» er en parodisk innledningsfrase når et fagfelt eller problemområde skal beskrives. Når dette fagfeltet er kunnskapsorganisasjon, kan innledningen faktisk være relevant, selv om det muligens like gjerne kunne hett «allerede de gamle sumerere ...». Så lenge mennesker har vært i stand til å nedtegne og dermed lagre kunnskap, har de også vært opptatt av å finne en måte å ordne disse nedtegnelsene på. Det velbevarte tempelet i Edfu i Egypt inneholder et «bokhus» der inskripsjoner indikerer at papyrusruller har vært oppbevart i kister ordnet i henhold til tematikken de behandler, og i store funn av mer enn fire tusen år gamle samlinger av kileskriftplater fra Mesopotamia er det konstatert at disse også har hatt en tematisk organisering. Dokumentsamlingen i antikkens største og best kjente bibliotek, Alexandria-biblioteket, som av noen påstås å ha omfattet så mye som 700 000 bokruller, var beskrevet i en katalog som både var klassifisert etter fagområde og ordnet alfabetisk etter forfatter og verkenes førstelinjer. Allerede her var altså de grunnleggende hjelpemidlene for organisering av kunnskap etablert, nemlig faglig eller tematisk systematisering og alfabetisk ordning. Med disse to grunnleggende ordningsprinsippene som

utgangspunkt er det siden utviklet mer og mer sofistikerte kunnskapsorganisatoriske prinsipper og systemer, etter som voksende dokumentmengder og skiftende teknologier har stilt stadig større krav til ordning og samtidig åpnet for stadig bedre muligheter for ordning.

Ragnar Nordlie, professor ved Institutt for arkiv-, bibliotek- og informasjonsfag, Høgskolen i Oslo og Akershus
Foto: HiOA

Begrepet kunnskapsorganisasjon kan forstås i en snever og en bred forstand, hevder den danske bibliotekprofessoren Birger Hjørland. I bred forstand betegner begrepet aktiviteten eller prinsippene som ordner alle typer kunnskap, for eksempel innenfor vitenskapelige disipliner, undervisningsinstitusjoner, leksika og oppslagsverk eller, i videste forstand, innenfor et gitt teorisystem, et språksystem – eller i menneskenes sinn. I snevrere forstand hevder han at kunnskapsorganisasjon befatter seg med organiseringen av bibliografiske poster eller metadatarepresentasjoner av kunnskap, dvs. de elementene kunnskapsorganisatoren velger ut for å representere det dokumentet som er den faktiske kunnskapsbæreren (og her må dokument forstås i videste betydning, brukt om all representert kunnskap, i form av tekst, lyd, bilde –).

I *International encyclopedia of information and library science* defineres kunnskapsorganisasjon slik:

> «the description of documents, their contents, features and purposes, and the organization of these descriptions so as to make these documents and their parts accessible to persons seeking them or the messages they contain. Knowledge organization encompasses every type and method of indexing, abstracting, cataloguing, classification, records management, bibliography and the creation of textual or bibliographic databases for information retrieval»

Denne definisjonen understreker for det første at kunnskapsorganisasjon både dreier seg om selve *beskrivelsen* av dokumentene og *ordningen* av disse beskrivelsene, det vil si de systemene som er utviklet for å tilgjengeliggjøre beskrivelsene for publikum. Dessuten peker definisjonen på at organiseringen har som formål å tilgjengeliggjøre både selve dokumentene og kunnskapsinnholdet i dem («the messages they contain»), det vil si at både dokumentenes fysiske egenskaper og deres emneinnhold, egnethet for ulike formål osv. må beskrives. Endelig peker definisjonen implisitt på et

skille mellom *kunnskapsorganisasjon* som en aktivitet «sett fra dokumentets side», der dokumentet og dets potensial som kunnskapskilde står i fokus, og *informasjonsgjenfinning*, der brukerens behov for å bli informert, dvs. motta kunnskap tilpasset et mer eller mindre veldefinert informasjonsbehov, er sentralt.

Skillet mellom *kunnskap* og *informasjon* er verken absolutt eller veldefinert, og særlig termen «informasjon» kan dekke mange ulike betydninger. Den er blitt brukt mer eller mindre synonymt med «data», og refererer til en meddelelse, melding eller sett med innsamlede fakta, der betydningen av meldingen er underordnet. Den er av og til forstått som mer eller mindre parallell med «kunnskap»: data som er ordnet, gitt mening og brakt i kommuniserbar form. Den kan også referere til en melding som gir (eller har potensial for) mening i en spesifikk brukssammenheng, for en spesifikk mottager, og som derfor har potensial for å skape kunnskap hos denne mottageren. Denne siste forståelsen av begrepet informasjon gjør det meningsfullt å snakke om kunnskapsorganisasjon når dokumenter tilrettelegges for gjenfinning, og informasjonsgjenfinning når de skal hentes fram på basis av en brukerforespørsel. Slik er begrepene skilt fra hverandre blant annet i norsk bibliotekundervisnings-tradisjon, men i kunnskapsorganisatoriske tekster vil det forekomme at «knowledge organization» og «information organization» brukes om hverandre.

Selv om det kunnskapsorganisatoriske *behovet* er like gammelt som evnen til å nedtegne kunnskap, kan *termen* kunnskapsorganisasjon som betegnelse på en aktivitet og et fagfelt knyttet til beskrivelse av dokumenter bare spores tilbake til slutten av 1800-tallet. På det tidspunktet hadde mengden av dokumenter og dermed bibliotekenes samlinger vokst til et omfang som krevde ny gjennomtenking av systemer og standarder for ordning. Samtidig medførte folkeopplysningstanken og fremveksten av store folkebibliotek, særlig i USA, at disse samlingene skulle gjøres tilgjengelige for et helt nytt og sammensatt publikum, som skulle kunne finne dokumentene selv, på åpne hyller. En «teknologisk revolusjon» kom også omtrent samtidig, nemlig kortkatalogen. Den tillot en dynamisk og fleksibel ordning av dokumentbeskrivelser på kort i standardisert størrelse, som muliggjorde beskrivelse og gjenfinning etter flere kriterier enn det som tidligere hadde vært praktisk mulig, og som samtidig inviterte til samarbeid om utarbeidelse av disse dokumentbeskrivelsene og til deling av katalogdata. Dette økte igjen presset mot standardisering. Det samme gjorde ideer om et universelt bibliografisk system – en oversikt over

all dokumentert kunnskap – som på dette tidspunktet ble sett på som en reell mulighet, for første gang siden Alexandria-biblioteket.

Det første systematisk uttrykte utsagn om «the objectives of a bibliographic system» (som vi snevert kan forstå som en bibliotekkatalog, men som i videre forstand kan ses som et uttrykk for formålet med kunnskapsorganisasjon) sprang da også ut fra amerikansk bibliotekmiljø. Disse «objectives» ble uttrykt av Charles Ammi Cutter, bibliotekar ved The Athenaeum library i Boston, første gang i 1876, som sa at det bibliografiske systemets formål var

1. to enable a person to find a book of which either the author, the title or the subject is known

2. to show what the library has by a given author, on a given subject, or in a given kind of literature

3. to assist in the choice of a book as to its edition (biliographically) or as to its character (literary or topical).

Prinsippene sto uforandret i nærmere hundre år, og er i dag videreført av den internasjonale biblioteksammenslutningen IFLA, som i 2009 uttrykte at «the catalogue» skulle assistere brukere med å *finne* en bibliografisk ressurs, *identifisere* en ressurs, *velge* en ressurs tilpasset brukers behov, *anskaffe* eller skaffe tilgang til en bibliografisk enhet, og *navigere* mellom ressurser.

IFLAs «formålsparagrafer» ligger nær Cutters. De legger gjennom uttrykket «bibliografisk ressurs» inn et eksplisitt skille mellom «verk» (det intellektuelle produktet) og «manifestasjon» (for eksempel boka), et skille som bare var implisitt tilstede i «assist»-formålet hos Cutter, og legger inn et formål om å bistå til å skaffe adgang til dokumentet, samt et formål om å støtte navigasjon mellom ressurser eller ressurs-beskrivelser, aktualisert ved at både dokumenter og dokumentbeskrivelser er tilgjengelige primært i digital form.

IFLA kaller dette «cataloguing principles» og beholder dermed betegnelsen «katalog» på den samlingen av dokumentbeskrivelser eller metadata som disse prinsippene skal anvendes på. Begrepet «katalog» er imidlertid ikke lenger helt enkelt å definere. Katalogen har tradisjonelt vært forstått som det settet av metadata som beskriver et enkelt biblioteks eller en gruppe

bibliotekers dokumentsamling, eller en annen avgrenset samling av dokumenter. I dag er «egen samling» en mindre veldefinert størrelse, det er ikke nødvendigvis klart skille mellom ressurser en institusjon eier og de som den på andre vis gir tilgang til, og gjenfinning av dokumenter skjer i økende grad på tvers av og uavhengig av institusjonsskiller. Samtidig er det et stadig mindre tydelig skille mellom dokumentene selv og deres metadata. Men selv om katalog-begrepet er i ferd med å bli mindre fruktbart, er prinsippene fortsatt relevante som *kunnskapsorganisatoriske* retningslinjer dvs. som uttrykk for hva kunnskapsorganisatorisk praksis skal tjene publikum med.

Skal IFLA-formålene kunne tilfredsstilles, krever det at dokumentene beskrives både med henblikk på hvilke personer eller institusjoner som er intellektuelt og praktisk ansvarlig for dem (forfatter, utgiver, utøver, oversetter ...), hva dokumentene heter og hva de handler om. Det er nødvendig å beskrive de kriteriene som brukerne må kjenne til for å kunne velg et dokument framfor et annet, for eksempel dokumentets alder, omfang, intellektuelle nivå, målgruppe eller formål. For tilgangsformål må rettighetshavere, visningsformater osv. identifiseres, og for navigasjonsformål må relasjonene mellom ulike opphavspersoner, dokumentnavn, emner og dokumenttyper også være beskrevet. Tradisjonelt er denne beskrivelsesprosessen delt i på den ene siden formell, på den andre siden emnemessig eller innholdsmessig beskrivelse, med hvert sitt sett av systemer og standarder som hjelpemidler.

Den formelle dokumentbeskrivelsen retter seg i prinsippet mot å beskrive de delene av et dokument som ikke krever kjennskap til dokumentets innhold: opphavspersoner, dokumentnavn, omfang, presentasjonsform osv. Dette skillet er ikke udelt enkelt å opprettholde – er f.eks. genrebestemmelse, beskrivelse av intellektuelt nivå e.l. formell eller innholdsmessig beskrivelse?

Hvor grensen enn måtte settes, innebærer formell beskrivelse en rekke beslutninger: hvilke elementer skal beskrives, etter hvilke kriterier skal disse elementene avgrenses og defineres (hva kreves f.eks. for at en person skal kunne sies å være opphavsperson til et verk, hvilke typer av opphavspersoner finnes, hvor mange opphavspersoner skal registreres ...), hvordan skal disse elementene standardiseres og presenteres, hvordan skal de kodes og lagres? I bibliotekene har disse beslutningene vært avhjulpet av mer eller mindre internasjonalt anerkjente standarder på ulike nivå, gjennom en årrekker har f.eks. AACR (Anglo-American cataloging rules) eller deres nasjonale varianter styrt valg av elementer, ISBD (International standard

bibliographic description) for ulike materialtyper styrt beskrivelsen av elementene og MARC (Machine-readable cataloging) med lokale «dialekter» beskrevet kodesettet for overføring av disse elementene i maskinleselig form. For øyeblikket er alle disse standardene under revisjon og omforming, men den store utfordringen ligger i at det kunnskapsorganisatoriske feltet utvides til å gjelde dokumenttyper som tradisjonelt har hatt egne, og av og til manglende, standarder, for eksempel arkiv- og museumsmateriale. Et viktig ledd i revisjonsarbeidet er å gjøre standardene mer prinsipielle og mindre materialavhengige, slik at alle dokumenter som representerer interessant materiale for en bruker kan gjenfinnes under ett, uansett om materialet er organisert i et bibliotek, en avis, et arkiv eller et museum.

Å kalle den formelle dokumentbeskrivelsen «kunnskapsorganisasjon» kan muligens være en tilsnikelse – til en viss grad kan det sies at det er dokumentene heller enn kunnskapen de representerer som organiseres. Den innholdsmessige beskrivelsen av dokumentene har imidlertid utvilsomt som formål å organisere kunnskap, i betydningen stille emnemessig relevant innhold til rådighet for brukere. Tradisjonelt har denne organiseringen skjedd gjennom enten klassifikasjon eller verbal beskrivelse av dokumentene.

Systemer for organisering av kunnskap har som nevnt innledningsvis meget lang historie, men bestrebelser for å ordne og presentere kunnskap i systematisert form blomstret på 17–1800-tallet dels gjennom bestrebelser på å samle kjent kunnskap tematisk og presentere den alfabetisk gjennom store encyclopedi-prosjekter (først ut var den franske encyclopedien i 1751–72), dels gjennom å systematisere fenomener i naturen, samfunnet osv i taksonomier ordnet etter et eller annet inndelingskriterium. Linnés *Systema naturalis* (1735) var tidligst ute, og ordnet plante- og dyreverdenen gjennom systematisk nomenklatur – slekt, familie, art osv. – etter polygenetisk prinsipp, dvs. etter utviklingsfellesskap; det periodiske system ordnet grunnstoffene etter atomvekt (Dalton, 1808/ Mndelejev, 1863)).

I vitenskapelig sammenheng har klassifikasjon gjennom taksomomier som formål å organisere et fagfelt systematisk etter prinsipper som gir hvert fenomen en entydig og klart definert plass – klassifikasjonen skal ordne alt som er likt i en klasse og skille det entydig fra alt annet i systemet. Klassifikasjon av dokumenter i biblioteksammenheng har i prinsippet samme formål – å samle alt som emnemessig hører sammen, og skille det så tydelig som mulig fra det øvrige materialet i samlingen. Det er imidlertid forskjeller – klassifikasjon av

dokumenter i biblioteksammenheng er basert på hensiktmessighet i forhold til den aktuelle samlingen eller det aktuelle ordningsformålet, ikke på streng vitenskapelighet; og klassifikasjon av dokumentert kunnskap skal samle det som er tilstrekkelig likt heller enn det som er absolutt likt. Disse forskjellene påvirker både klassifikasjonssystemenes konstruksjon og deres anvendelse.

De første systemene for klassifikasjon i kunnskapsorganisatorisk sammenheng som kom i utbredt bruk og hadde universelle formål dukket, som katalogiseringsprinsippene, opp i USA på slutten av 1800-tallet – Cutter presenterte sin «Cutter expansive classification» i 1882, den ble senere videreutviklet til Library of Congress classification, og Melvil Dewey presenterte første utgave av sin Decimal Classification i 1876. Begge disse systemene er basert på hierarkiske prinsipp, der kunnskapsuniverset inndeles i overordnede grupper av disipliner eller fagområder (10 hos Dewey, derav Decimal Classification) som så gis mer og mer spesifikke hierarkiske underinndelinger. Fordelene med dette hierarkiske prinsippet er flere: systemet kan enkelt brukes både til ordning av fysiske objekter f.eks. sekvensielt på hyller og til ordning av dokumentrepresentanter i gjenfinningssystemer, og det er enkelt for systemene å tilby økt eller minsket spesifisitet under søking ved å bevege seg oppover eller nedover i hierarkiene.

Ulempen ved hierarkiske system er at emnebeskrivelsen i prinsippet låses til å presentere ett aspekt ved den kunnskapen som dokumentet representerer, mens kunnskap ofte er tverrdisiplinær eller presenterer en disiplins eller emneområdes påvirkning på et annet. Kunnskapsorganisatoriske teoretikere har forsøkt å imøtekomme dette ved å introdusere klassifikasjon basert på fasetteknikk, grunnet på den indiske biblioteketeoretikeren Ranganatnans teorier uttrykt gjennom hans Colon Classification (1933). Hans observasjon at de fleste emner ville kunne gis to eller flere plasseringer i Deweys system (*Indisk historie* kunne være enten *India* eller *Historie*), førte til ideen om at et klassifikasjonssystem burde konstrueres rundt de fasetter et komplekst emne var sammensatt av. Fasetterte klassifikasjonssystem har vist seg for kompliserte å vedlikeholde og bruke til at de har fått allmenn utbredelse, men elementer av fasetteknikk er introdusert i senere utgaver av i utgangspunktet hierarkiske systemer.

Fra katalogsalen i publikumsavdelingen ved Universitetsbiblioteket i Oslo, 1949. Per Kleppa(1903-1991), sjef for bibliotekets tekniske tjeneste. Foto: N.T.B.s Billedavdeling A/S

Alternativet til å uttrykke emneinnhold gjennom plassering i et klassifikasjonsskjema er å gi en verbal beskrivelse av dokumentets innhold, *emneindeksering*. Kunnskapsorganisasjonsteorien byr på flere muligheter for systematisk verbal beskrivelse. Prekoordinerte indekseringsprinsipper tilbyr regelverk for å sette sammen emnebeskrivende termer til semantisk meningsfylte uttrykk for emneinnhold i dokument (*India – historie – 1920–1945*), – prekoordinering innebærer da at kunnskapsorganisator bestemmer sammenhengen mellom og rekkefølgen av de beskrivende termene. Alternativt kan et postkoordinert indekseringsprinsipp legges til grunn, da baseres indekseringen på vokabularer av emnebeskrivende termer som kan uttrykke de grunnleggende semantiske enhetene emnet er bygd opp av, og sammensetningen av dem overlates til brukeren i søkeøyeblikket. Slik postkoordinert indeksering krever et vokabular, en tesaurus, som skal standardisere de termene som tillates brukt til å uttrykke disse semantiske enhetene, og som angir hvordan termene er relatert til hverandre, både hierarkisk og assosiativt. Tesauri skulle da ideelt være tilgjengelig både for kunnskapsorganisator under indekseringen og for brukeren under søking.

Det eksisterer standarder for tesauruskonstruksjon og det foreligger en rekke både allmenne og fagspesifikke tesauri.

Så langt er kunnskapsorganisasjon beskrevet som en manuell prosess der fagpersoner støttet av regelverk, standarder og vokabularer utarbeider de metadata – dokumentrepresentasjoner – som skal støtte et antall stipulerte gjenfinningsbehov og gi brukere tilgang til kunnskapsinnholdet i dokumenter. Denne aktiviteten utsettes nå for press fra flere kanter. Prosessen er kostbar ettersom den er arbeidskrevende, dokumentmengden som potensielt kunne trenge beskrivelse øker tilnærmet eksponensielt, og dokumenter som tidligere måtte beskrives manuelt fordi de i utgangspunktet forelå i fysisk form, er nå digitale og tilgjengelige for maskinell behandling. Allerede på 1960-tallet ble det foretatt eksperimenter som syntes å vise at gjennomsnittlig gjenfinningseffektivitet, i den grad den lot seg måle, ble like god når søking ble basert på maskinell analyse og utvelgelse av ord i dokumentteksten som når den ble basert på manuell indeksering av de samme dokumentene, uavhengig av hvilket indekseringsprinsipp som ble lagt til grunn. Siden den tid har maskinelle indekseringsmetoder blitt mer sofistikerte, selv om de fortsatt baserer seg på den grunnleggende antagelsen at frekvensen av forekomsten av ord i en tekst er en tilfredsstillende representasjon av semantikken i teksten. Tesauri er i ferd med å utvikle seg til *ontologier*, maskinleselige oversikter over termer og relasjoner mellom dem som vil kunne gjøre maskiner i stand til å resonnere over termers spesifikke semantiske rolle i en tekst og dermed raffinere den semantiske representasjonen av teksten. En kan si at dokumentet er blitt sin egen metadatarepresentasjon.

På samme tid er oppfatningen av hvem som er ansvarlig for kunnskapsorganiseringen under endring. I prinsippet er det tre parter som kan ha en mening om hva kunnskapsinnholdet i et dokument er – produsenten av dokumentet, den som har ansvaret for å formidle det, og den endelige brukeren av kunnskapen i dokumentet. Hittil har formidleren – en bibliotekar, arkivar – vært stort sett eneansvarlig for metadata-tilordningen. I større og større grad åpnes det nå for at produsenten – en forfatter, en fotograf – kan la teksten følge av supplerende informasjon som kan bidra til forståelse og tolking av innholdet. Samtidig åpnes det for mulighet for at brukeren både direkte og indirekte kan bidra med metadata – direkte gjennom kommentarer og omtale, vurderinger og karaktersetting, stikkord og annen supplerende tekst; indirekte ved at gjenfinningssystemer registrerer brukerdata som hvilke søkeord som førte til at et gitt dokument ble valgt og sett, eller

hvilke dokument som velges i sammenheng med hverandre. Det er altså god grunn til å tenke gjennom den tradisjonelle kunnskapsorganisatorens rolle i dokumentbeskrivelses-prosessen.

Høgskolen i Oslo og Akershus. Fotograf: John Anthony Hughes/HiOA

Undervisningen i fagkretsen som gjerne er betegnet «kunnskapsorganisasjon og gjenfinning» i norsk bibliotekutdanning reflekterer denne utviklingen. Faget har tradisjonelt vært delt i katalogisering, som stort sett har omfattet det som her er kalt den formelle beskrivelsen, klassifikasjon, som har omfattet alle former for emnebeskrivelse, samt «bruker og gjenfinning», som har omfattet kunnskap om brukere, om gjenfinningshjelpemidler (kataloger, indekser) og om konkrete informasjonskilder (fagdatabaser). Fagfeltet har opptatt inntil halvparten av studietiden, og det har alltid vært sterke innslag av kunnskapsorganisatorisk teori (konstruksjon av klassifikasjonssystem og tesauri, innholdsanalyse), men det har samtidig vært et ferdighetsfag der studentene har vært forventet å ha gode ferdigheter i katalogisering etter det til enhver tid dominerende norske regelverket og klassifikasjon og indeksering etter de mest brukte klassifikasjonsskjemaene og emneordssystemene.

Nå utvikler fagkretsen seg i retning av større vekt på prinsipper for beskrivelse, for klassifikasjon og indeksering, og studentene utfordres til å se på de konkrete systemene mer som eksempler enn som spesifikke ferdighetsområder, med sikte på at disse prinsippene skal kunne anvendes i et bredere spekter av kunnskapsorganisatoriske sammenhenger – basert på en antagelse om at der det generelle beskrivelsesbehovet kanskje i større grad ivaretas av maskinell indeksering eller sentralisert beskrivelse og datautveksling, vil det være spesifikke bruksområder og brukergrupper som vil ha økende behov for tilpasset kunnskapsorganisatorisk innsats. Det paradoksale er jo at aldri har så mange yrkesgrupper vært engasjert i arbeid som grenser mot kunnskaporganisasjon (utvikling av vevsider, nettsalg) og aldri har allmennheten vært eksponert for og avhengige av søkebaserte (og dermed kunnskapsorganisasjons-baserte) tjenester som i dag. Samtidig legges det i undervisningen vekt på å utvikle forståelse for de automatiserte systemenes oppbygging og funksjonalitet, for forståelse av brukeroppførsel og søkeadferd, for konstruksjon av vev-baserte tjenester, for forståelse av systemer for maskinell koding og lagring av informasjon.

Kunnskapsorganisasjon er trolig like relevant og nødvendig i dag som noensinne. Men kunnskapsorganisatoren må trolig i økende grad være i stand til å analysere og innrette seg mot spesielle brukerbehov, til å forstå og kunne utnytte de automatiserte systemenes muligheter og begrensninger, til å anvende kunnskapsorganisatoriske prinsipper på nye felter og nye datatyper, og til å fungere som brukerens talsperson mot systemer og kunnskapsprodusenter.

Formidling, kunnskaps-organisering og Dewey

TEKST: **ELISE CONRADI**

Elise Conradi, Prosjektleder norsk WebDewey, Nasjonalbiblioteket
Foto: Nasjonalbiblioteket

Deweys desimalklassifikasjon er det mest brukte bibliotek-klassifikasjonssystem på verdensbasis[1]. Det er oversatt til mer enn 30 språk og brukes i over 200 000 bibliotek i 135 land. Men det er også et av de mest kritiserte klassifikasjonssystemene i verden. Systemet ble utviklet på 1870-tallet av Melvil Dewey, en ambisiøs mann som på det tidspunktet var i begynnelsen av 20-årene, og som ifølge sin biograf, allerede da var besatt av effektivitet og viste en ubetvingelig trang til å reformere[2]. Kritikken av systemet inkluderer, men er ikke begrenset til, at det er vestlig-sentrert, overflatisk, tungrodd og utdatert. Den amerikanske filosofen og teknologen David Weinberger har kalt systemet «ureparerbart»[3], da det forutsetter at kunnskap kan kartlegges, et premiss han og mange andre er svært uenige i. Men i dag satses det likevel hundrevis av millioner kroner i nasjonalbibliotek verden over på Deweys desimalklassifikasjon. I denne artikkelen skal jeg forsøke å forsvare denne enorme satsningen ved å se på opphavet og utviklingen av klassifikasjonssystemet i lys av bibliotekenes rolle som formidler av kunnskap og i lys av det voksende kunnskapsunivers bibliotekene formidler ut ifra.

1 http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/9042
2 Wiegand, Wayne A. «Irrepressible reformer: a biography of Melvil Dewey». London, 1996.
3 Weinberger, David. «Everything is Miscellaneous». New York, 2007.

FORMIDLING AV KUNNSKAPSUNIVERSET

Formidling av kunnskap er et av bibliotekenes viktigste oppgaver. Kunnskap kan her være den personlige kunnskapen som åpenbarer seg gjennom debatt, foredrag og blogginnlegg, eller den samlede kunnskapen som tradisjonelt sett er blitt huset i bibliotekets samling, gjerne kalt «bibliotekets kunnskapsunivers». Bibliotekets kunnskapsunivers formidles gjennom utstillinger, personlig veiledning og ikke minst, gjennom fysiske oppstillinger og digitale verktøy som oppmuntrer til oppdaging av ny kunnskap på egenhånd. I bibliotekene kan begge former for kunnskap være tett knyttet opp mot hverandre, og ofte, jo mer de formidles sammen, jo mer vellykket kan formidlingen sies å være. En debatt i et folkebibliotek kan for eksempel motivere til bruken av bibliotekets samling for å tilegne seg mer kunnskap om et bestemt emne, mens elementer i selve samlingen kan være utgangspunktet for å invitere til foredrag. Formidlingen av begge former for kunnskap er en grunnleggende forutsetning for et demokratisk samfunn, og kan være en nøkkel til inspirasjon og innovasjon.

Mens formidlingen av personlig kunnskap avhenger av alt fra bibliotekets evne til å lage arrangementer til foredragsholderens retorikk, er formidlingen av bibliotekets kunnskapsunivers mest av alt bunnet i god kunnskapsorganisering. Her spiller selvfølgelig gjenfinning av dokumenter en stor rolle, men like viktig er tilretteleggingen av dokumenter slik at man lett kan vise frem samlingen eller deler av samlingen i interessante kontekster. Dette gjelder spesielt for formidlingen som foregår gjennom fysiske oppstillinger og digitale verktøy.

Praktiseringen av kunnskapsorganisering innen bibliotekvitenskap har inntil nylig vært stort sett begrenset til tillegging av deskriptive metadata (tittel, ansvarshaver, utgiver, osv) og emnemetadata (hva det handler om) til hvert dokument biblioteket besitter. Med førstnevnte kan biblioteket for eksempel vise frem alt som finnes i samlingen av en bestemt forfatter, i en bestemt form, eller av en viss lengde. Med emnemetadata, på den andre siden, kan man potensielt få en helhetlig modell av bibliotekets kunnskapsunivers fordelt på fagområder eller emner. Slik kan man for eksempel lage temabaserte utstillinger med alle dokumentene biblioteket besitter som handler om et bestemt emne, eller søkeverktøy som lar bibliotekbrukere traversere bibliotekets kunnskapsunivers på egen hånd ved å vise relasjoner mellom emnene (og dermed dokumentene) i samlingen. Nyere metoder innen kunnskapsorganisering inkluderer bruken av eksterne ressurser

(gjerne i form av Linked data) for å supplere beskrivelsen av dokumentene i bibliotekets samling, og bruken av språkverktøy til automatisk indeksering av digitaliserte dokumenter. Samspillet mellom tradisjonell praksis og nye metoder kan spille en stor rolle i hvordan bibliotekets ressurser blir formidlet i årene fremover.

Det finnes i hovedsak to forskjellige typer emnemetadata: de mange variantene som baserer seg på tillegging av ord, og de mange variantene som baserer seg på tillegging av klassifikasjonskoder. Som navnet tilsier, er Deweys desimalklassifikasjon en variant av sistnevnte.

DEWEYS DESIMALKLASSIFIKASJON

Deweys desimalklassifikasjon er det eldste bibliotekklassifikasjonssystem i Vesten som fremdeles brukes i dag[4]. Før det ble lansert i 1876, var det vanlig praksis å ordne bibliotekets samlinger i brede kategorier, der bøker og annet materiale ble gitt en fast, nummerert plass basert på innkjøpsdato. I jakten på en mer effektiv måte å arrangere biblioteksamlinger, utviklet Melvil Dewey et system der dokumenter får en relativ hylleplassering basert på fagområde og emne. Han beskriver ambisjonen sin på følgende måte i 1893:

Melvil Dewey.
Foto: Wikimedia Commons

> Economy and simplicity calld[5] [...] for some plan of consolidating the two sets of marks heretofore used ; the one telling of what subject a book treats, the other where the book was shelvd. By relativ location and decimal class numbers, the simple arabic numbers have been made to tell each book and pamflet, both *what* it is, and *where* it is[6].

Systemet er i dag så utbredt at det er lett å ta denne sammensmeltingen av emne og plassering for gitt, men på 1870-tallet var det revolusjonerende. For første gang kunne plasseringen av en bok formidle både dets innhold og dets kontekst innenfor bibliotekets kunnskapsunivers samtidig[7].

4 http://www.pitt.edu/~agtaylor/articles/ICC10DeweyChapter.pdf
5 Melvil Dewey kjempet også for mer effektiv staving av det engelske språket, men kom ingen vei med dette (annet enn endringen av fornavnet fra Melville til Melvil).
6 https://archive.org/stream/abridgeddecimalc00dewerich#page/8/mode/2up
7 Les «Hvordan fungerer Dewey i folkebibliotekene», side 29, for et innblikk i hvordan dette aspektet ved Deweys desimalklassifikasjon utarter seg i folkebibliotek i dag.

Relativ plassering av bøker. Foto: Wikimedia Commons

For å få til sammensmeltingen, trengte Dewey en modell av bibliotekets kunnskapsunivers og et kunstig språk til å beskrive den. Modellen han endte med er gjenstand for mesteparten av kritikken mot deweysystemet i dag; den gir nemlig en svært ujevn representasjon av kunnskapsuniverset, med uforholdsmessig stort fokus på enkelte fagområder og tilsynelatende forsømmelse av andre. Selv om det er veldig lett å si seg enig i denne kritikken, så er den også uberettiget. Dewey, ambisiøs som han var, hadde nemlig aldri ambisjoner om å få til en sann modell av all kunnskap. Han tok heller en svært pragmatisk tilnærming til arbeidet ved å utvikle en modell av *bibliotekets* kunnskapsunivers, det vil si, en modell basert på litteraturbelegg. Tilnærmingen beskriver han i innledningen til den første utgivelsen av klassifikasjonssystemet i 1876:

> In all work, philosophical theory and accuracy have been made to yield to practical usefulness[8]. The impossibility of making a satisfactory classification of all knowledge as preserved in books, has been appreciated

8 Første utgave ble utgitt før Dewey tok opp kampen for stavereform på ordentlig.

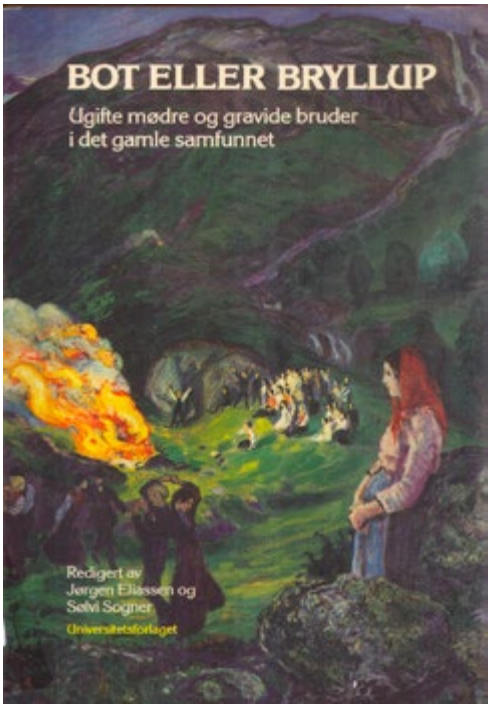
from the first, and nothing of the kind attempted. Theoretical harmony and exactness has been repeatedly sacrificed to the practical requirements of the library[9].

Modellen ble i hovedsak basert på pensumlister og litteraturbelegget ved Amherst College, hvor han på det tidspunktet var student. Amherst College var i 1870-årene et lite «liberal arts» universitet, med en sterk forankring i tradisjonelle og klassiske fag som latin og gresk, matematikk, retorikk, filosofi, kjemi, botanikk, astronomi, psykologi, geologi, bibelhistorie, logikk, statsvitenskap og historie[10]. Med dette utgangspunktet, laget Dewey en modell der kunnskapsuniverset ble delt inn i 9 hovedområder: filosofi, religion, samfunnsvitenskap, språkvitenskap, naturvitenskap, anvendt vitenskap, kunst, litteratur og historie. Logiske relasjoner ble ofret der det var nødvendig for at titallsystemet (det desimale tallsystemet) kunne brukes både som struktur for modellen og som representasjon av modellen i form av et kunstig språk skrevet i arabiske tall. En tiende kategori som omfattet dokumentform ble lagt til for å opprettholde denne strukturen, men flere av underkategoriene her (bibliografi, bibliotekvitenskap, journalistikk) er i senere tid blitt anerkjent som egne fagområder. Desimalstrukturen tilsa at hvert hovedområde skulle kunne deles inn i 10 underklasser, som igjen kunne deles inn i 10 underklasser, ad infinitum, slik at systemet i teorien skulle kunne utvides uten grenser.

Man kan i dag gremmes over noen av resultatene av denne pragmatiske tilnærmingen. Kristendommen i Deweys desimalklassifikasjon opptar åtte av de ti underklassene til hovedområdet religion. Greske, romerske og germanske religioner fra gamle dager (også kjent som mytologi) får plass mellom kristendommen og øvrige religioner som buddhismen, jødedommen og islam. Det finnes en egen klasse for ugifte mødre, og utenomekteskapelige barn plasseres i samme klasse som forlatte barn og misbrukte barn. Informatikk blir skviset inn i 004–006 mens parapsykologi og okkultisme får strekke seg ut mellom 130–139. Men faktum er at forvalterne av deweysystemet gjennom årene har beholdt Deweys pragmatiske tilnærming til kunnskapsuniverset, og utvidelser av systemet har alltid vært basert på litteraturbelegg (etter hvert ved Library of Congress og nylig også ved nasjonalbibliotek som forvalter oversettelser av systemet). Videre, til tross for utdaterte synspunkter innen diverse fagområder, har man så langt som mulig latt være å endre på

9 https://archive.org/stream/classificationan00dewerich#page/4/mode/2up

10 Wiegand, Wayne A. «Irrepressible reformer: a biography of Melvil Dewey". London, 1996, side 15



En bok i nasjonalbibliografien fra 1981 om ugifte mødre. Foto: Universitetsforlaget

deres plasseringer i deweysystemet slik at tidligere klassifisert litteratur forblir gjenfinnbart. Ugifte mødre er, for eksempel, et utdatert fenomen i dag, men det ble skrevet veldig mye om det tidligere, og det skrives fremdeles historisk-sosiologiske beretninger om emnet. Nyere fagområder som ikke fantes på 1870-tallet, som for eksempel informatikk og bioteknologi, får riktig nok mindre plass i de øverste nivåene av systemet, men det betyr ikke at de ikke får nødvendige videreinndelinger. Det betyr bare at det krever lengre numre å beskrive alle aspektene ved dem.

Gjennom årene er systemet blitt tatt i bruk ved flere og flere bibliotek i USA og i andre land. Selv om dette mest sannsynlig skyldes at systemet er relativt lett og effektivt å bruke, spesielt når man kan gjenbruke klassifikasjonsarbeid gjort ved andre bibliotek, så kan det også forklares med at modellen av kunnskapsuniverset stort sett har fungert bra i bibliotek verden over (spesielt i vestlige land). Dette kan tyde på at litteraturbelegget i USA grovt sett ikke har vært så annerledes enn litteraturbelegget i resten av Vesten. I den norske nasjonalbibliografien, for eksempel, har vi p.t. 14 776 dokumenter om diverse aspekter ved kristendommen, men bare 1410 dokumenter som handler om samtlige andre religioner. Oversettelser av deweysystemet inkluderer også flere inndelinger, såkalte utvidelser, basert på litteraturbelegget i de aktuelle landene. I norske oversettelser finnes det naturligvis langt flere klasser for norsk geografi, litteratur og historie enn det som finnes i originalen. Oversettelser i ikke-vestlige land (som Egypt og Vietnam) har tatt i bruk tillatte løsninger i deweysystemet for å kompensere for avvik mellom deres og vestlig litteraturbelegg. I den arabiske oversettelsen, for eksempel, er det islam som får åtte av de ti underklassene innen religion, og det er lagt inn mange utvidelser for islamsk rett. Sistnevnte er nylig blitt tatt inn i originalen, da mer og mer blir skrevet om dette også i Vesten.

Bøker organisert etter deweysystemet på et bibliotek i Hong Kong.

Foto: Wikimedia Commons

DEWEYREGISTERET

En større kritikk av deweysystemet, som for så vidt gjelder for alle klassifikasjonssystemer der hovedinndelingen baserer seg på fagområder, er at generelle emner blir spredt rundt innen systemet. Man kan, for eksempel, skrive om abort ut ifra et medisinsk perspektiv, et samfunnsvitenskapelig perspektiv eller et filosofisk perspektiv, og filmer fra et kunstnerisk perspektiv, et markedsføringsperspektiv og et sosiologisk perspektiv. Man kan teoretisk sett skrive om hvilket som helst generelt emne fra et hvilket som helst faglig perspektiv, og vi har tusenvis av doktoravhandlinger i nasjonalbibliografien som reflekterer nettopp dette. Oppstillingen av biblioteksamlingen etter deweysystemet vil aldri kunne formidle overfor brukeren alt biblioteket har om et bestemt emne.

For å kompensere for dette, utviklet Dewey et emneregister til klassifikasjonssystemet sitt. Emneregisteret er en alfabetisk liste over emner med henvisninger til fagområdene de kan klassifiseres med. I den første utgave av systemet, beskriver han dets funksjon slik (min uthevelse):

> SUBJECT INDEX. Find the subject in this Alphabetical INDEX; The number following it is its Class Number. The entire resources of the library on this subject will be found under this number either in the Subject Catalogue, the Shelf Catalogue, or *on the shelves*.

Deweyregisteret ble allerede i andre utgave omdøpt på engelsk fra «subject index» til «relative index» for å tydeliggjøre dets funksjon som en liste over emner og deres plasseringer *relativ til* fagområdene i deweysystemet.[11] Fra og med samme utgave ble emnene også angitt med kvalifikatorer, altså ord som beskrev hvilke fagområde emnene tilhørte. Det er tydelig fra sitatet over at registeret var ment som et hjelpeverktøy både for klassifikatorer og sluttbrukere helt fra starten. Til førstnevnte gruppe er bruken av registeret blitt en integrert del av klassifikasjonsarbeidet. Men bruken av registeret i formidlingen overfor sluttbrukere av hvilke emner som finnes i bibliotekets kunnskapsunivers, har i beste fall gitt varierende resultater.

I Norge har det vært vanlig praksis i bibliotek å tilgjengeliggjøre for publikum enten hele Deweys desimalklassifikasjon (i fysisk kopi) eller egenutviklede emneregistre (i fysiske eller digitale kopier) til dette formålet. Det største problemet med første variant er at den ikke gir sluttbrukeren noen indikasjon om emnet i det hele tatt finnes i det enkelte bibliotekets kunnskapsunivers. Videre finnes det ingen registertermer knyttet til bygde numre i de fysiske kopiene av deweyregisteret. Begge problemene er løst med lokale emneregistre, som viser til litteraturbelegget ved det aktuelle bibliotek. For hvert dokument i biblioteket som klassifiseres etter dewey, legges det et emneord med eller uten kvalifikator til emneregisteret og med koblinger til deweynummeret som ble brukt. Gjennom årene er det blitt utviklet utallige lokale emneregistre til deweysystemet, i utallige formater og former. Disse brukes stort sett i emnegjenfinning av dokumenter, men kan teoretisk sett også brukes for å få en oversikt over hvilke emner som finnes i bibliotekets kunnskapsunivers. Dessverre er det ingen samarbeid i vedlikeholdet av de lokale emneregistrene og de må i høyeste grad betraktes som lokale løsninger.

Begge metodene over er problematiske. Emneregistre viser et emnes plasseringer relativt til fagområdene i deweysystemet, men sier ingenting ellers om deweyklassene de tilhører. De kan derfor aldri brukes for å få en oversikt over hvor mye hvert bibliotek har om hvert emne. En sluttbruker som vil vite

11 I de fem norske oversettelser vi har av deweysystemet, er det kun i 3. utgave av *Arnesen* at det heter «Relativt register». Ellers har det bare hett «Register».

hva biblioteket har om radioer, for eksempel, vil bli henvist fra registeret til tre fagområder. Et av disse er en underklasse i fagområdet bilmekanikk, hvor dokumenter om radioer er plassert i samme klasse med dokumenter om bilers lydanlegg, hanskerom og askebegre. Sluttbrukeren vil potensielt måtte lete gjennom mye støy i hylla eller trefflisten for å finne dokumentene om nettopp bilradioer. Dette gjelder heldigvis mest for sære emner med lite litteraturbelegg.

Biblioteksentralen har forsøkt å løse alle de ovennevnte problemene ved å lage sitt eget emneregister til deweysystemet, som de har kalt BIBBI-emner og som de knytter til både deweynumre og til katalogposter. BIBBI-emner er basert på de dokumentene de selv klassifiserer, i praksis vil det si litteraturbelegget ved de aller fleste folkebibliotek i Norge. Den største forskjellen mellom BIBBI-emner og deweyregisteret er at førstnevnte er bygd opp som emneordstreng (etter Hjortsæters regler for emneordskatalogering) og viser dermed ikke emneordenes faglige plassering relativt til deweysystemet, men dette kan eventuelt tolkes ut ifra deweynummeret. Dessverre er det p.t. ikke mulig å få en full oversikt over emnene deres sammen med deweynumrene de er koblet til.

ENDRINGER I BIBLIOTEKETS KUNNSKAPSUNIVERS

Kunnskapsuniverset har ekspandert enormt siden 1876, og Deweys desimalklassifikasjon har holdt tritt med utviklingen. Mens første utgave av systemet inneholdt 921 klasser og 2765 registertermer, inneholder dagens utgave (den 23.) 40 155 klasser og 102 530 registertermer[12]. Systemet har også sakte, men sikkert blitt mer og mer syntetisk: man kan bygge nye klasser basert på eksisterende klasser ved hjelp av 12 194 fasetter som befinner seg i eksterne og interne tabeller for form, geografi, språk og m.fl.

De siste tiårene har bibliotekets kunnskapsunivers også undergått en drastisk formendring. Der man før kunne snakke om det som samlingen av fysiske dokumenter huset i biblioteket, består nå de fleste biblioteksamlinger av både fysiske og digitale dokumenter. Videre er det en forventning at bibliotek skal kunne formidle kunnskap som ikke befinner seg i deres samlinger. Bibliotekets kunnskapsunivers mangler grenser. Dette gjelder spesielt for Norge, hvor digitaliseringsarbeidet til Nasjonalbiblioteket har gjort mesteparten av norsk litteraturbelagt kunnskap til allemannseie. Blant Nasjonalbibliotekets innsatsområder for 2015 står blant annet dette (min uthevelse):

12 Tallene er hentet fra 1.09.2014: http://ddc.typepad.com/025431/2014/09/dewey-by-the-numbers.html

Avhengig av rettighetsstatus vil materialet i Nasjonalbiblioteket kunne gjenbrukes på ulike måter av andre bibliotek til utvikling av egne digitale tjenester og *formidlingstiltak*[13].

Selv om det finnes mange nye metoder innen kunnskapsorganisering for emnebeskrivelse av digitale dokumenter som man nå kan benytte seg av, består utfordringen i norske bibliotek i å beskrive digitale og fysiske dokumenter innenfor samme system, slik at man kan formidle begge i interessante kontekster sammen. Når plasseringen av fysiske dokumenter fremdeles spiller en stor rolle i formidling, og når de fleste bøkene i nasjonalbibliografien siden 1956 er tilkoblet deweykoder, er det naturlig å tenke seg at man kan dra nytte av deweysystemet i tilretteleggingen av begge typer dokumenter fremover. Når deweysystemet i tillegg har fått et ansiktsløft i form av WebDewey det siste tiåret, er det lett å se hvorfor Norge og andre land satser såpass mye penger på det.

WEBDEWEY

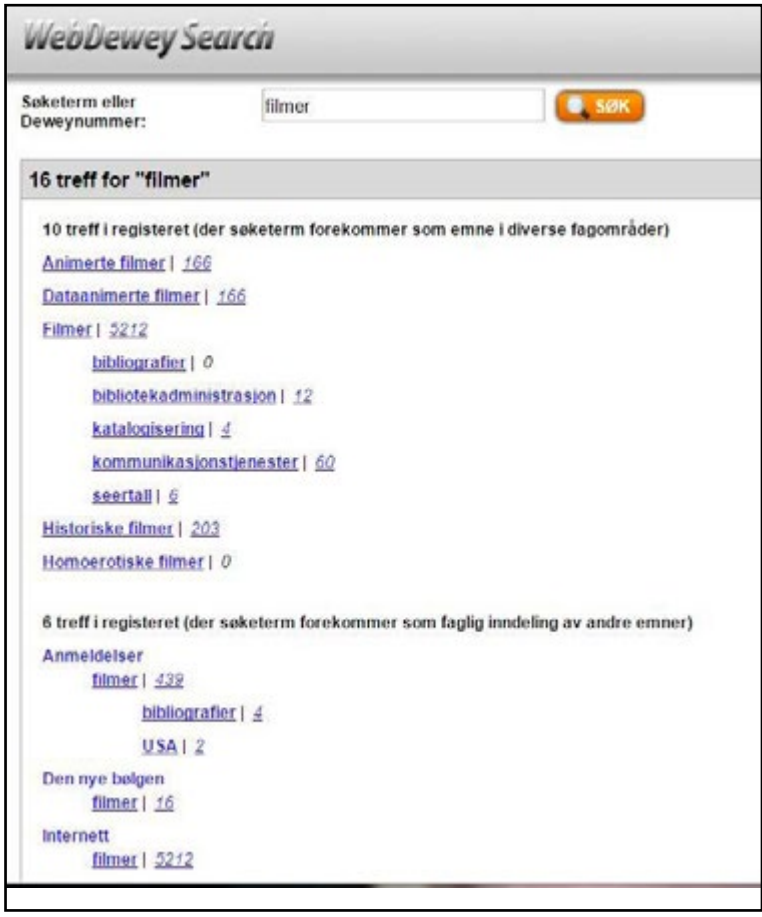
WebDewey ble lansert av OCLC[14] i 2003 og blir reklamert som den letteste måten å bruke deweysystemet på[15]. Fremfor å finne frem til riktig klassifikasjonsnummer ved å bla i fysiske bøker, kan man med webversjonen benytte seg av avanserte søkefunksjoner for å komme rett til beskrivelsen av klassen, samt enkelt sjekke hvordan klassenummeret er brukt på andre dokumenter. Videre kan man lagre bygde numre i WebDewey, slik at disse kan gjenbrukes av andre klassifikatorer. Dette fører naturligvis til mer effektivitet i klassifikasjonsarbeidet. Mens effektiviteten i klassifikasjonsarbeidet øker desto flere som bruker WebDewey, øker potensialet for verdien i formidlingsarbeidet parallelt. Dette er mest tydelig i felleskataloger som Biblioteksøk[16] og WorldCat[17], der man kan sammenstille deweyklassifiserte dokumenter i bestemte fagområder på tvers av samlinger og språk.

WebDewey danner også grunnlaget for å skape nye sluttbrukerverktøy basert på deweysystemet, da det inneholder maskinlesbar data som kan utnyttes på nye måter. Dette er mest tydelig i koblingen mellom ord og deweynumre. Klassebetegnelser i alle språk som deweysystemet er oversatt til, ligger til-

13 http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Omtale-av-innsatsomraader-2015
14 OCLC har eid deweysystemet siden 1988
15 https://oclc.org/dewey/features.en.html
16 http://www.nb.no/bibsok/start.jsf
17 http://www.worldcat.org/

gjengelige som Linked data og kan for eksempel brukes i sluttbrukersystemer til å belyse betydningen bak hvert deweynummer eller til å la sluttbrukere browse seg gjennom deweyhierarkiene på egen hånd. World Digital Library[18], for eksempel, benytter seg av Linked data for å formidle i sju språk deres digitale ressurser inndelt etter hovedkategoriene i deweysystemet.

Andre ord som er maskinelt koblet til deweynumre er registertermer. Dette betyr at man kan utvikle sluttbrukersystemer som på sikt erstatter de lokale emneregistrene som finnes i Norge. Dersom man i tillegg begynner praksisen som følges på Biblioteksentralen med å registrere registertermer i katalogposten[19] knyttet til hvert enkelt dokument, vil man løse problemet med for mye støy i trefflisten ved å øke granulariteten til deweynummeret. Norske registertermer i WebDewey er også maskinelt koblet til engelske registertermer, hvilket muliggjør utviklingen av søketjenester på både engelsk og norsk.



Registertermer i deweysystemet koblet til Biblioteksøk i beta-versjonen av WebDeweySearch. Foto: privat skjermbilde

Ord er ikke bare koblet til deweynumre, men de er i tillegg koblet til enkeltstående fasetter og til fasetter i sammensatte numre. Førstnevnte kan (og bør) registreres i katalogposten[20] når et deweynummer ikke klarer å uttrykke de viktigste emnene i dokumentet. Fasetter i sammensatte

18 http://www.wdl.org/en/topic/
19 I MARC 21 kan dette gjøres i emnefeltene, med «ddcri» gjengitt som kilde i $2
20 I MARC 21 kan dette gjøres i 083-feltet for ekstra deweynumre knyttet til dokumentet

numre blir automatisk lagret i WebDewey-databasen ved bygging av numre. Disse kan også registreres i katalogposten[21], slik at de kan utnyttes i sluttbrukersystemer ved for eksempel å sammenstille aspekter ved deweyklassifiserte dokumenter som ikke har kommet frem tidligere.

Med WebDewey har man også muligheten til å vise eksterne vokabularer som er knyttet til deweysystemet gjennom mapping. Amerikanske emneordsystemer som LCSH, BISAC og MeSH er mappet til det amerikanske systemet, mens tyske og svenske emneord er mappet til henholdsvis det tyske og svenske deweysystemet. I Norge er Universitetsbiblioteket i Oslo godt i gang med å undersøke muligheten for å mappe Humord til det norske deweysystemet[22]. I tillegg til den tydelige gevinsten med et økt tilfang av ord i WebDewey-verktøyet for klassifikatorer, muliggjør mapping en traversering av flere typer kunnskapsorganiseringssystemer via deweysystemet. En sluttbruker vil potensielt kunne bevege seg mellom samlinger som er organisert på forskjellige måter, og dermed se dokumentene i kunnskapsuniverset i nye kontekster.

Parallelt med Nasjonalbibliotekets lansering av den norske WebDewey høsten 2015, kommer sluttbrukerverktøyet WebDeweySearch. WebDeweySearch er et åpent søkeverktøy som vil gi sluttbrukere tilgang til deweyklassifiserte dokumenter i Norge via Biblioteksøk. Med verktøyet vil man kunne søke etter deweynumre eller registertermer, samt browse i deweyhierarkiet, for å få en oversikt over hva som finnes i Biblioteksøkets kunnskapsunivers. Det utvikles også en API som vil kunne benyttes av bibliotekssystemleverandører slik at de kan integrere denne funksjonaliteten i sine kataloger.

Aldri før har så mye skrevet kunnskap vært tilgjengelig for sluttbrukere i Norge. Utviklingene i WebDewey gjør forhåpentligvis deweysystemet godt rustet til å tilrettelegge for både fysiske og digitale dokumenter slik at norske bibliotek kan formidle kunnskapen i disse, i nye og interessante kontekster. Når den norske WebDewey endelig er på plass, vil det bli spennende å se hvordan deweydata kan kombineres med nyere metoder innen kunnskapsorganisering for å øke dets relevans ytterligere i formidlingsarbeidet. [23]

---

21 I MARC 21 kan dette gjøres i 085-feltet. Nasjonalbiblioteket undersøker muligheten for automatisk å registrere denne informasjonen i klassifiseringsarbeidet.

22 Les «På randen av mapping" på side 36 for å lære mer om dette arbeidet.

23 Les "Digitalisering og klassifikasjon", side 64, for et innblikk i noe av forskningen som foregår allerede på dette området ved Nasjonalbiblioteket.

Hvordan fungerer Dewey i folkebibliotekene?

Hvordan brukes Dewey i folkebibliotekene? Hvordan er samsvaret mellom det Dewey-systemet uttrykker og brukernes behov? Hvilken betydning har formelle kunnskaper om klassifikasjon i forhold til kunnskaper om brukerne? Hvordan er Dewey nedfelt i bibliotekets strukturer og rutiner?

TEKST: **INGEBJØRG RYPE**

Jeg holder nå på med en masteroppgave ved Høgskolen i Oslo og Akershus, Bibliotek- og informasjonsvitenskap, hvor målet er å få mer kunnskap om disse spørsmålene.

Svarene kan forhåpentligvis hjelpe Nasjonalbiblioteket i arbeidet med å gjøre den kommende norske WebDeweyen til et nyttig verktøy også for folkebibliotekene, og kanskje gi nyttige innspill til andre som utvikler tjenester på området.

Undersøkelsen er utført som kvalitative intervjuer med til sammen 16 bibliotekarer i ni små og mellomstore bibliotek (bibliotek i kommuner med under 20 000 innbyggere).

Ingebjørg Rype, hovedbibliotekar, Nasjonalbiblioteket
Foto: Nasjonalbiblioteket

Foto: Kathleen McIntyre

BAKGRUNN

Unni Knutsen skrev i 2007 en masteroppgave om katalogarbeid i norske folkebibliotek. (Knutsen, 2007). Det som der ble avdekket når det gjelder klassifikasjon, var behovet for å tilpasse klassifikasjonen fra Biblioteksentralen til den lokale praksisen. Det ble også pekt på at selv om man får poster fra Biblioteksentralen, vil det alltid være områder man må klassifisere selv, for eksempel lokallitteratur. (s.47-48).

I 2012 gjennomførte Nasjonalbiblioteket en spørreundersøkelse blant alle typer norske bibliotek om hvordan klassifikasjon og emneord ble brukt. Svarprosenten var svært lav, særlig fra små og mellomstore folkebibliotek. I etterkant av denne undersøkelsen utførte Unni Knutsen en uhøytidelig kvalitativ undersøkelse, hvor hun spurte et utvalg folkebibliotek hvordan de brukte og oppfattet Dewey. Undersøkelsene er oppsummert i en artikkel som ble publisert i Bok og Bibliotek i 2012 (Knutsen og Rype, 2012).

I min undersøkelse ønsket jeg å undersøke Deweys funksjon i de små og mellomstore folkebibliotekene nærmere.

FUNN FRA UNDERSØKELSEN

Hensikten med å bruke Dewey i folkebiblioteket?

Bibliotekarene jeg snakket med, jobbet i såpass små bibliotek at de i stor grad hadde oversikt over både brukere og samlinger. De fikk få referansespørsmål, og Dewey ble brukt lite i søk. Det ble også påpekt at det var enklere å bruke emneord enn Dewey i søk. Bibliotekarene hadde heller ikke inntrykk av at brukerne benyttet Dewey til emnesøk.

Det alle jeg snakket med poengterte, er at målet på om Dewey er hensiktsmessig i et folkebibliotek, er i hvilken grad det kan være en støtte for formidling. Systemet blir hensiktsmessig i den grad det klarer å bidra til å gjøre bøkene synlige og gjenfinnbare. Dermed blir hylleoppstillingen et viktig aspekt ved Dewey i et fysisk bibliotek.

Et annet viktig aspekt, er at Dewey utgjør en felles struktur for alle folkebibliotek. En felles struktur gjør det enklere å utveksle og gjenbruke klassifikasjon. En felles struktur gjør det enklere å finne fram og orientere deg hvis du begynner å jobbe på et annet bibliotek. En felles struktur er også viktig for brukerne. Det bidrar til at brukerne kan kjenne igjen en grunnstruktur når de oppsøker ulike bibliotek.

Betydningen av klassifikasjon fra Biblioteksentralen (BS)
Bibliotekene jeg snakket med, abonnerte på, eller hadde bestemt seg for å begynne å abonnere på poster fra Biblioteksentralen. På landsbasis abonnerer nå ca 87 % av folkebibliotekene på postene derfra.

Blant de jeg snakket med, ble det i liten grad gjort endringer i klassifikasjonsnummeret i postene fra BS. I den bibliografiske registreringen er klassifikasjonsnummer og hyllesignatur prinsipielt to forskjellige ting. Klassifikasjonsnummeret viser dokumentets faglige plassering i Dewey-hierarkiet. Hyllesignaturen viser til dokumentets fysiske plassering i biblioteket. Biblioteksentralens poster inneholder forslag til begge deler, men det er viktig å være klar over at hyllesignaturen egentlig kan bestemmes lokalt. Det er feltet for klassifikasjonsnummer det søkes i ved søk på Dewey-nummer, og det er det som er grunnlag for fellessøk.

Det at det gjøres få endringer i klassifikasjonsnumrene, samsvarer altså ikke helt med det Unni Knutsen fant ut i sin masteroppgave (s.47). Det kan bety at det har skjedd en holdningsendring, at bibliotekarene har blitt mer bevisst på at det er viktig å legge til rette for gjenbruk og utveksling. En annen grunn til at resultatene er forskjellige, kan være at det i Knutsens undersøkelse også var med bibliotekarer fra store bibliotek, at gjenfinning gjennom egen katalog betyr mer der, eller at de store bibliotekene i større grad er opptatt av å følge sin egen tidligere praksis.

Bortsett fra noe av lokallitteraturen, blir mesteparten av litteraturen klassifisert av BS. Flere av informantene er inne på at bøker som ikke blir klas-

sifisert av BS, er så sære og lite utlånt, at de heller blir lånt inn. Tidsbruken er en viktig faktor ved bruken av BS-poster. Tiden som blir spart på å bruke poster fra BS, frigjør tid til å gjøre andre ting, i praksis til «å være sammen med brukerne».

Det er litt forskjellig forståelse av og praksis angående forkortelse av nummer, og hvor nyttig det er med fininndelt klassifikasjon. Flere er opptatt av at det går an å forkorte i hyllesignaturen selv om man ikke forkorter i det søkbare Dewey-nummeret. Grunnene til at mange Dewey-numre har så mange sifre, er at de kan være satt sammen av ulike faglige fasetter. Denne informasjonen blir ikke godt nok nyttiggjort i dagens søkesystemer, men vil kunne utnyttes i framtida. Et eksempel på en bok med et langt og sammensatt klassifikasjonsnummer, er Per Fugellis «Døden skal vi danse». Dewey-nummeret er 362.1969940092. Dette kan virke unødvendig fininndelt i et lite bibliotek, men nummeret er satt sammen av numre som viser at boka er en biografi om en som har sykdommen kreft, sett fra en sosialmedisinsk vinkling. M.a.o. kan alle vinklingene eller fasettene gi nyttig informasjon og være interessant i ulike faglige sammenhengene. Og bibliotekene kan velge å forkorte i oppstillingssignaturen selv om de ikke gjør det i klassifikasjonsnummeret.

OPPSTILLING. DEWEY OG HENSYNET TIL BRUKERNE

Bibliotekene endrer altså ikke i klassifikasjonsnummeret i stor grad, mens hylleoppstillingen kan bli forkortet, eller den kan bli endret eller brutt på andre måter. Endringer i oppstillingssignaturen blir begrunnet i erfaring om hva bibliotekarene har sett fungerer best for brukerne. Kriteriene for å fungere for brukerne er at temaer skal være synlige, at det skal være enkelt å finne fram til temaer brukerne er interessert i. Med dette forstår jeg at hylleoppstillingen blir betraktet som en del av formidlingen. Det er flere av informantene som sier at de hele tida endrer på hvordan bøkene blir oppstilt, de ser hva som fungerer best for brukerne.

Et av områdene som oftest blir utskilt fra Dewey-oppstilling er biografier. Dette gjøres også i flere av bibliotekene som ellers holder seg til Dewey-oppstilling. Biografier skal etter Dewey klassifiseres med fagområdet den biograferte hører hjemme. Det at det er en biografi, blir i Dewey merket ved at -092 fra Hjelpetabell 1 blir lagt til på slutten av nummeret. Tidligere ble biografier klassifisert i 921, og stilt opp etter den biograferte. Mange bibliotek har valgt å fortsette praksisen med å stille opp bøkene alfabetisk etter den biografertes navn, da de har erfaring med at brukerne er interessert i å lese sjangeren biografier. Men

Asker bibliotek. Foto: Olav Dahl

dette betyr igjen at f.eks. «Døden skal vi danse», blir plassert sammen med andre biografier, og ikke blant de andre bøkene om kreft. I katalogen kan man tenke seg løsninger hvor det kan søkes på de ulike fasettene i et sammensatt Dewey-nummer. I et fysisk bibliotek velger bibliotekarene hva de mener er det som fungerer best for brukerne. Dette kan de gjøre uten å endre på Dewey-nummeret, de kan skille ut biografiene som en egen kategori og stille opp bøkene etter Dewey-nummer innenfor kategorien, eller de kan velge å sette boka sammen med andre bøker om kreft. Dette gjøres altså på ulike måter, alt ettersom hvilken erfaring de har med hva som fungerer best for brukerne.

Barnelitteraturen er et annet område de fleste av bibliotekarene jeg snakket med enten skiller ut deler av, eller hele samlingen fra ren Dewey-oppstilling.. Erfaringen er at det særlig for barn er vanskelig å forstå inndelingen etter Dewey. F.eks. vil bøker om «Hester» i Dewey avhengig av faglig vinkling klassifiseres både med sport i 798, med husdyr i 636, og på skjønnlitteratur. Brukerne er gjerne interessert i alle disse aspektene, og det kan være fornuftig å samle dem. Det samme gjelder sjangeren «Eventyr», som i Dewey blir skilt mellom folkeeventyr i 390-gruppen, og kunsteventyr, som plasseres sammen med skjønnlitteratur.

Kategorisering

Dewey er delt inn etter fag, ikke emner, derfor vil mange temaer bli spredt rundt i systemet, og dermed på hyllene. Dette er den korte begrunnelsen for at noen bibliotek velger å bryte opp Dewey-oppstillingen ved å sette opp hele eller deler av samlingen etter temaer eller kategorier. Det som er verdt å merke seg, er at kategorisering er en alternativ måte å sette opp bøkene på, det er ikke et alternativt klassifikasjonssystem. De jeg snakket med mente at Dewey var viktig for å beholde en kontroll med kategoriene, og for å sikre en gjenkjennelig struktur.

Noen av bibliotekene hadde kategorisert hele samlingen, noen hadde kategorisert deler av den, to av bibliotekene hadde begynt på eller gjennomført kategorisering og gått tilbake til ren Dewey-oppstilling. Noen kategoriserte ikke i det hele tatt. Det var vanskelig å finne fellestrekk ved de som var for, evt. i mot kategorisering. Det som var felles for alle bibliotekene, var at standpunktet til kategorisering var basert på hva de mente fungerte best for brukerne. De som kategoriserte, mener at det er mer brukervennlig å dele inn i kategorier som f.eks. «Helse», hvor bøker om både psykologi, sykdommer, kosthold og trening kan plasseres, enn å sette dem opp i Dewey-rekkefølge, hvor de vil komme langt fra hverandre på hyllene. De mener også at kategorisering er et mer fleksibelt system, kategorier kan lett fjernes og legges til etter brukernes behov. En forutsetning for å lykkes med kategorisering, i hvert fall når det gjelder bøker for voksne, virker å være at Dewey blir brukt innenfor kategoriene og i katalogen.

De som ikke kategoriserer

De bibliotekene som ikke kategoriserer begrunner dette delvis med at det skaper nye problemer, som hvem skal bestemme hvilke kategorier man skal ha. De mener også at det byr på problemer ved utskifting av personale. De to bibliotekene som i større eller mindre grad hadde kategorisert, men gått tilbake til Dewey-oppstilling, mente også at det ikke var noe enklere for brukerne å forholde seg til kategorier enn til oppstilling etter Dewey. En av dem mente at det gjorde at brukerne «gresset» mindre, at de ble for bundet til temaene de på forhånd var interessert i. Den andre hadde en interessant sammenligning med oppstillingen i et supermarked. Vi vet etter hvert hvor vi finner lettmelk, vi trenger ikke å skjønne systemet bak. Brukerne blir vant til at ting er plassert på en bestemt måte.

Hvilke kunnskaper er det nødvendig å ha om klassifikasjon når du jobber i et mindre folkebibliotek? Forholdet mellom erfaring og formelle kunnskaper

Alle informantene var enige om at det er viktig å ha formelle kunnskaper om Dewey som system hvis du jobber i et folkebibliotek. Det er viktig å forstå systemet, men det er ikke så viktig å kunne klassifisere selv. Hvordan ting bør plasseres og stilles opp, læres gjennom erfaring, men det er viktig å ha en formell kunnskap for å forstå systemet og vite hvordan det kan utnyttes. Det å kunne utføre selve teknikken det er å klassifisere selv ble ikke sett på som så viktig. En av bibliotekarene sier at her er det er veldig stor forskjell på å jobbe på et stort og et lite bibliotek. På et lite bibliotek, hvor man stort sett kjøper poster fra BS, trenger man ikke å kunne klassifisere selv, men det er viktig å forstå hvordan det fungerer.

De færreste av bibliotekene hadde noe samarbeid med andre bibliotek når det gjaldt klassifikasjonsspørsmål. Noen nevnte at de kunne ha uformelle diskusjoner med andre bibliotek om dette, men at det ikke var tema for formelle diskusjoner eller møter mellom bibliotek f.eks. på fylkesnivå.

KAN DISSE FUNNENE HA NOEN BETYDNING FOR FRAMTIDIG UTVIKLING AV TJENESTER?

Funnene mine tyder på at klassifikasjon fortsatt er viktig i det fysiske biblioteket. Det hjelper til med å formidle samlingen, og det fungerer som en felles standard. Det ses på som nødvendig for å sikre kontroll og kontinuitet både for samlinger, ansatte og brukere.

Det at bibliotekene i liten grad endrer i det søkbare klassifikasjonsnummeret, gjør at bibliotekene i stor grad har felles klassifikasjon, noe som kan bidra til god kvalitet på sluttbrukertjenester basert på Dewey. Det er i den forbindelsen viktig at alle bibliotek i praksis får anledning til å basere seg på felles katalogposter.

Bibliotekene vil antakeligvis fortsette å bruke Dewey slik de syns det fungerer best i oppstillingen. Dette er en del av formidlingen, og katalog- og gjenfinningssystemene bør bedre utnytte at Dewey inneholder mange bygde numre som uttrykker ulike fasetter av et emne.

Ett Dewey-nummer kan bestå av mange ulike fasetter, ett nummer kan uttrykke mange aspekter av et emne. Dette gjør at Dewey kan brukes fleksibelt. Både WebDewey og marcformatet legger nå bedre til rette for å utnytte de ulike fasettene i Dewey-numrene. Det betyr at fleksibliteten i Dewey-systemet kan utnyttes bedre, og Dewey kan bli et bedre verktøy også lokalt. Jeg mener at undersøkelsen min viser at Dewey også i framtida kan bli et godt verktøy også for de mindre folkebibliotekene hvis denne fleksibiliteten utnyttes.

LITTERATUR

Knutsen, Unni og Rype, Ingebjørg. (2012). *Bruk av Dewey i norske bibliotek*. Bok og bibliotek 2012(4), 36-37. Lokalisert på http://www.bokogbibliotek.no/images/stories/pdf_2012/Bob-4-2012_web.pdf

Knutsen, Unni. 2007. *Ut og stjæle poster. Effektivitet og rasjonalitet i folkebibliotekenes katalogarbeid.* Masteroppgave ved JBI. Lokalisert på http://hdl.handle.net/10642/315 6.desember 201

På randen av mapping

Universitetsbiblioteket i Oslo fikk i 2014 midler fra Nasjonalbiblioteket til å utvikle metodikk for mapping mellom tesaurusen Humord og norsk WebDewey. I artikkelen redegjør vi for tidligere mappingprosjekter ved UBO, Realfagsbiblioteket, og hvordan vi bygger på dette arbeidet ved utvikling av mappingverktøy. I tillegg til å beskrive hvilke tekniske standarder vi følger, viser vi også noen av de utfordringene som mapping mellom et emnevokabular og et klassifikasjonssystem byr på. Den underliggende drivkraften er å tilby sluttbrukere bedre emnemessig gjenfinning i katalogene våre.

TEKST: **UNNI KNUTSEN** OG **ARE GULBRANDSEN**

Høsten 2013 sendte Universitetsbiblioteket i Oslo (UBO) en søknad til Nasjonalbiblioteket (NB) om mapping av tesaurusen Humord samt emnevokabularet Realfagstermer til den kommende, fullstendige norske oversettelsen av Deweys desimalklassifikasjon.

Ved UBO har vi sakte, men sikkert opparbeidet oss innsikt i hva mapping innebærer gjennom to tidligere NB-støttede prosjekter. I det første prosjektet mellom Realfagsbiblioteket og NTNU (Kuldvere et al., 2013) ble det gjort sammenlikninger av terminologi mellom Realfagstermer og TEKORD. Ved hjelp av automatisk tekstsammenlikning fant man et fellesvokabular som utgjorde ca. 20 prosent av hvert av vokabularene. I en videreføring av prosjektet (Kuldvere et al., 2014) ble det utviklet metoder for mapping

mellom Realfagstermer og fellesvokabularet med TEKORD til aktuelle deler (500-gruppen og 600–640) av den norske oversettelsen av DDC.

Mappingsøknaden vi sendte i 2013 resulterte også i midler fra Nasjonalbiblioteket[1]. Prosjektet består av to atskilte deler. Den ene delen er et forprosjekt mellom NB og UBO som skal gi beslutningsgrunnlag i spørsmålet om det skal etableres en generell, norsk tesaurus med utgangspunkt i Humord. Den andre delen av prosjektet drives kun av UBO og skal resultere i at det utvikles metodikk for mapping mellom tesaurusen Humord og norsk WebDewey. Her lener vi oss selvsagt tungt på de erfaringene som er gjort i Realfagsbiblioteket og ved NTNU.

Unni Knutsen, seksjonssjef, Universitetsbiblioteket i Oslo. Foto: Tove Jareld

Are Gulbrandsen, senioringeniør, Universitetsbiblioteket i Oslo. Privat foto

HVA ER MAPPING – OG HVA SLAGS FUNKSJONALITET KAN DET GI?

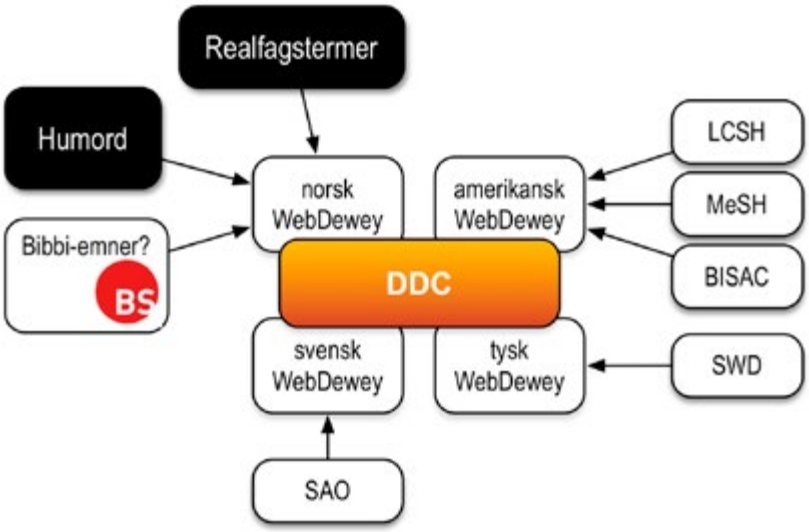
Mapping kan defineres som en aktivitet hvor det etableres relasjoner mellom begreper i ett vokabular med begrepene i et annet. Gjennom de tidligere omtalte prosjektene har UBO både laget relasjoner (mappinger) mellom emneordsvokabularer og mellom emneordsvokabularer og Deweys desimalklassifikasjonssystem.

Det teoretiske utgangspunktet for mappingarbeidet vårt er ISO-standard 25964 om tesauri (International Organization for Standardization, 2009, 2013). Det er særlig del to, *Interoperability with other vocabularies,* som er relevant for vårt arbeid. Standarden beskriver tre hovedmodeller for mapping mellom vokabularer. En av disse omtales ofte som nav-modellen[2]. Her mappes forskjellige vokabularer inn mot et sentralt vokabular som billedlig talt blir navet i hjulet og som distribuerer søkene videre.

Valget av modell var i vårt tilfelle mer eller mindre gitt. Andre vokabularer er allerede mappet til DDC. DDC framstår dermed som et sentralt koblingspunkt, noe figur 1 viser:

1 https://www.ub.uio.no/for-ansatte/om-ubo/prosjekter/tesaurus-mapping/delte-dokumenter/tildelingsbrev-tesaurus.pdf

2 Hub på engelsk



Figur 1: Dewey Decimal Classification (DDC) som nav. Humord og Realfagstermer er her mappet til norsk WebDewey. Andre vokabularer er mappet til DDC på tilsvarende måter.

Når Humord blir mappet til DDC gjennom norsk WebDewey, har vi lagt grunnlaget for at en bruker skal kunne søke på en term fra Humord i eksempelvis UiO biblioteksøk[3] og ledes inn i samlinger som har brukt helt andre emneordsvokabularer. DDC blir dermed det som ofte kalles «switching language» (se for eksempel Koch, 2006). Forutsetningen er at alle emnevokabularene er mappet til navet (DDC).

For UBO-brukere vil mappingen mot DDC også være nyttig for emnegjenfinning i egen samling. Mye av det elektroniske materialet som finnes i katalogen er tildelt engelskspråklige emneord fra kilder som MeSH (Medical Subject Headings), BISAC (Book Industry Subject and Category Subject Headings) og ikke minst LCSH (Library of Congress Subject Headings). Det er hensynet til økt funksjonalitet og bedre gjenfinning for brukeren som er drivkraften for mappingaktiviteten vår.

MAPPINGRELASJONER OG TEKNISKE STANDARDER

Når vi mapper et vokabular mot et annet opererer vi med begrepene kildevokabular og målvokabular. I vårt tilfelle er Humord og Realfagstermer

3 http://bibsys-primo.hosted.exlibrisgroup.com/primo_library/libweb/action/search.do?vid=UBO

kildevokabularene og målvokabularet er norsk WebDewey. Vi mapper altså én vei, mot DDC. Det er viktig å være klar over at vi mapper meningsinnholdet i ordene. Det betyr at vi må være bevisst på at sammenfallende ord ikke nødvendigvis bærer samme mening.

Før mapping kan starte opp må det avklares hvordan relasjonene mellom de to vokabularene skal uttrykkes. I det tyske CrissCross-prosjektet hvor man mappet Schlagwortnormdatei (SWD) mot den tyske WebDewey, opererte man med fire nivåer for overensstemmelse mellom emneord og Dewey-nummer.

I vårt prosjekt opererer vi med følgende relasjoner hentet fra ISO-standarden:

=EQ	(EQ = Equivalence). Likhetstegnet angir at mappingen er eksakt
~EQ	Tilden angir at mappingen ikke er eksakt. Det innebærer at begrepene kan være like i noen sammenhenger, men ikke i alle eller at begrepene kan være delvis overlappende eller avvike noe i betydningsinnhold
BM	Broader mapping. Termen i målvokabularet har en videre betydning enn termen i kildevokabularet
NM	Narrower mapping. Termen i målvokabularer har en mer spesifikk betydning enn termen i kildevokabularet
RM	Related mapping. Termen i målvokabularet assosieres med termen i kildevokabularet, men er ikke et synonym, et kvasisynonym eller en bredere eller smalere term

Ved mapping har man altså behov for å uttrykke ekvivalensrelasjoner, hierarkiske relasjoner og assosiative relasjoner. I følge ISO-standarden er det ekvivalensrelasjonen som vil være mest brukt.

Et av hovedmålene i alle våre utviklingsprosjekter har vært å publisere både vokabularene og mappingene våre med tanke på den semantiske weben. Av den grunn har vi helt fra starten av ønsket å bruke SKOS/RDF-standardene i arbeidet vårt.

SKOS-modellen opererer med *closeMatch, ExactMatch, narrowMatch, broadMatch* og *relatedMatch*. I tillegg har formatet noter som er nyttige

for definisjoner, redegjørelse for hvordan begrepet er brukt, historikk, kommentarer med videre. Vi har konkludert med at SKOS akkomoderer en tesaurus godt og uttrykker de relasjonene vi trenger å uttrykke. Det er selvsagt ikke tilfeldig at ISO-standarden og SKOS sammenfaller langt på vei[4]. De to utviklingsmiljøene har fulgt hverandre tett i dette arbeidet. I ISO-standardens del to (International Organization for Standardization, 2013, s. 43) er dette uttrykt slik:

> As yet there is no standard schema that fully complies with this part of ISO 25964, and the development of such a schema is not within scope. However, this is a rapidly evolving field and implementers of this part of ISO 25964 should be alert to developments among interested parties, e.g. the SKOS user community.
>
> The schemas developed for storage purposes may also be used or adapted to enable publication of the mappings. A SKOS-compliant format [...] is recommended if use in the Semantic Web is desired.

Dette understøtter valgene våre av SKOS og RDF.

UTVIKLING AV MAPPINGVERKTØY

I de fleste tilfeller bør mapping være et datastøttet intellektuelt arbeid. I forbindelse med mappingutprøvingen ved Realfagsbiblioteket ble det utviklet en prototype på et mappingverktøy, µmapper[5] for å gjøre mapping mellom Realfagstermer og DDC enklere og mer effektivt.

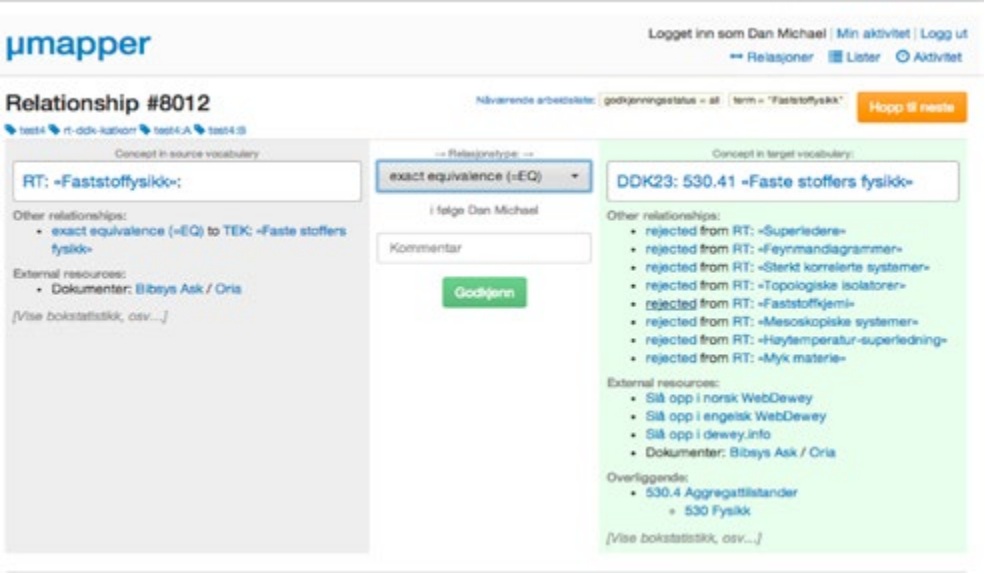
I eksempelet under er termen Faststoffysikk foreslått som eksakt ekvivalens (=EQ) til DDC-nummeret 530.41 med klassebetegnelse Faste stoffers fysikk. Verktøyet gir anledning til å gå til WebDewey for å sjekke eksempelvis hvilket faglig hierarki dette klassenummeret opptrer i og eventuelt foreta videre undersøkelser. Relasjonen til målvokabularet kan endres ved at mapperen velger et annet alternativ på nedtrekkslisten i midten.

Bak i kulissene i mappingverktøyet finnes Realfagstermer og fellesvokabularet fra det tidligere mappingprosjektet med TEKORD, alle uttrykt i SKOS. Deweydataene fra Nasjonalbiblioteket ble konvertert (svært pragmatisk)

4 http://www.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf

5 https://github.com/scriptotek/mumapper/

fra MARC 21 Classification til SKOS. En dump fra BIBSYS-basen med alle poster som er klassifisert etter DDC 23 ble avlevert i MARC 21-formatet. Både statistisk sammenlikning og ordlikhet er brukt som metodikk for å foreslå mappingrelasjoner. For ytterligere detaljer, henvises det til sluttrapporten for prosjektet (Kuldvere et al., 2014).



Figur 2: Brukergrensesnitt for prototypen µmapper.

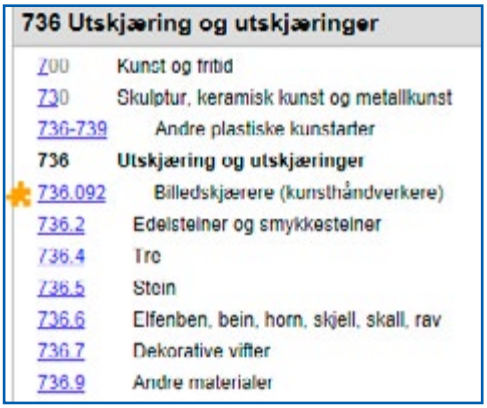
I det inneværende prosjektet vil vi bygge videre på det arbeidet som er gjort i µmapper. Det vil imidlertid også bli gjort en del endringer, både når det gjelder datakilder, infrastruktur, interaksjonsdesign og prototyping. Vi vil inkludere både Humord og UBOs emneregister til Dewey[6] som datakilder. De teknologiske valgene blir også annerledes; vi beveger oss mot en mer rendyrket RDF-løsning og vil utvikle en prototype som baserer seg på et trippellager (triplestore) fremfor en relasjonell database. Vi vil bruke rammeverket Callimachus til å utvikle selve brukergrensesnittet. For en mer detaljert systemskisse, se http://www.ub.uio.no/for-ansatte/om-ubo/prosjekter/tesaurus-mapping/prosjektplan/. I skrivende stund jobber vi med å diskutere arbeidsflyt, funksjonalitet og brukergrensesnitt for sluttbrukerverktøyet.

6 http://wgate.bibsys.no/search/pub?base=USVDEMNE

MAPPING SOM INTELLEKTUELL UTFORDRING

Det er mange problemer forbundet med å knytte to eller flere ulike vokabularer sammen. I vår sammenheng er hovedutfordringen at Deweys desimalklassifikasjon har inndeling etter fag mens en tesaurus som Humord gir emneinnganger til litteraturen. Vi vil dermed bli stilt ovenfor en situasjon hvor ett Humord kan tilsvare flere mulige Deweyplasseringer. Vi ser også tydelig at de faglige hierarkiene kan være ulike, Humords måte å dele inn verden på stemmer ikke alltid overens med Deweys inndelingskriterier. Videre kan emneord som er like eller nesten like med Dewey-termer ha ulikt betydningsinnhold.

Noen klassebetegnelser i DDC gir først full mening når de opptrer i sin faglige sammenheng:



Figur 3: Visualisering av kontekst for klasse 736 Utskjæring og utskjæringer og 736.5 Stein i norsk WebDewey.

Termen Stein i 736.5 må tolkes inn i rammen av at dette er stein brukt som materiale for utskjæring. Dette viser at man i mange sammenhenger må studere konteksten nøye før mapping kan foretas.

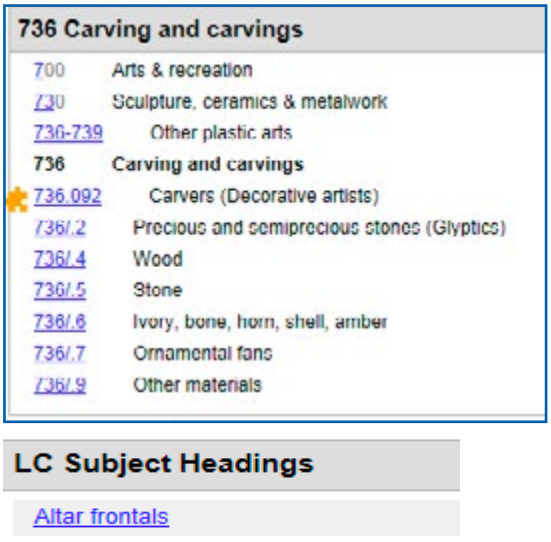
Å angi riktig relasjon mellom de to vokabularene er heller ingen lett oppgave. Ved eksakt mapping må man være obs på at identiske eller nesten identiske ord kan ha ulike betydninger. Forholdet mellom en ikke-eksakt mapping og en bredere/snevrere mapping, er heller ikke alltid klart. Dette fikk vi erfare på et seminar i regi av mappinggruppen i EDUG (European DDC Users Group) i mai 2014 hvor vi framla et konkret eksempel. Debatten gjorde det klart for oss at fokuset måtte flyttes fra operasjonen mapping til selve sluttbrukeropplevelsen. Hvilken (mer)verdi gir det for sluttbrukerne å bli henvist til et ikke helt eksakt treff eller til hierarkiske og assosiative relasjoner? Vi

ønsker å utforske dette gjennom å arrangere et seminar i tilknytning til neste møte i EDUG. Som et forarbeid til dette vil vi utarbeide ulike «mock-ups» for å synliggjøre hvordan ulike mappingvalg kan påvirke funksjonaliteten i sluttbrukerverktøyet.

MAPPING OG MERVERDI TIL NORSK WEBDEWEY

Vi har til nå mest fokusert på mappingens verdi for sluttbrukersøk inn mot ulike samlinger og for utviklermiljøer som vil bruke dataene våre i ulike sammenhenger. Mapping vil imidlertid også ha en verdi for indekserere. Ved mapping av Humord og Realfagstermer inn mot norsk WebDewey vil vi bygge numre i WebDewey-verktøyet med relevans for norske forhold. Disse numrene vil kunne gjenbrukes av andre bibliotek. At også termene fra Humord og Realfagstermer blir eksponert i WebDewey vil gi flere emnemessige innganger inn til Dewey og lette klassifikasjons- og emneordsarbeidet for andre indekserere.

Et eksempel på at Humord kan berike norsk WebDewey med terminologi: I Humord finnes termen Antemensaler (med synonymene Alterfronter (BF) og Alterfrontaler (BF)). I den engelskspråklige webutgaven er LCSH Altar frontals mappet til klassenummer 736 Carving and carvings (se figur 4). I norsk WebDewey ser det foreløpig ut som i figur 3. Plassen for Additional Terminology Sources står med andre ord tom. Målet vårt er å fylle ut den tomme plassen med termene fra Humord og Realfagstermer.



Figur 4: Visualisering av kontekst for klasse 736 Carving and carvings i engelskspråklig WebDewey og verdien av supplerende vokabular (LCSH)

SLUTTKOMMENTARER

Ved Universitetsbiblioteket i Oslo har vi over flere år jobbet med problematikk knyttet til mapping. Gjennom skrittvis tilnærmelse har vi opparbeidet stadig mer forståelse for hva arbeidet innebærer. Vi har erfart at mapping ikke er en eksakt vitenskap. Det er mange snubletråder, tolkningsmuligheter og fare for subjektivitet knyttet til dette arbeidet. Vi har tro på at risikoen for subjektive valg kan reduseres ved å utvikle mappingverktøy som ved hjelp av statistisk mapping og termlikhet kan guide indekserere til å foreta riktige mappingvalg. Vi har blitt overbevist om at det er sluttbrukerperspektivet som må være styrende for selve utførelsen av mappingarbeidet. Vi har en lang vei å gå ennå, men den mappingutprøvingen vi har foretatt i tidligere prosjekter gir oss tro på at vi vil klare å mappe vokabularene våre til WebDewey. Vi vil imidlertid også i fortsettelsen ha stor nytte av å knytte oss til miljøer som arbeider aktivt med tilsvarende prosjekter.

LITTERATUR

International Organization for Standardization. (2009). *Information and documentation: Thesauri and interoperability with other vocabularies: Part 1: Thesauri for information retrieval* (*ISO* 25964-1: 2011). Geneve: International Organization for Standardization.

International Organization for Standardization. (2013). *Information and documentation: Thesauri and interoperability with other vocabularies: Part 2: Interoperability with other vocabularies*. (*ISO* 25964-2: 2013). Geneve: International Organization for Standardization.

Koch, T. (2006, 10. januar). *Electronic thesis and dissertation services: Semantic interoperability, subject access, multilinguality: E-Thesis Workshop, Amsterdam 2006-01-19/20*. Hentet fra http://www.ukoln.ac.uk/ukoln/staff/t.koch/publ/e-thesis-200601.html

Kuldvere, V., Lundevall, M., Hegna, K., Konestabo, H. S., Låberg, K. T. , Flatby, E. S., & Greenall, R. (2013, 7. mai). *Realfagstermer og TEKORD: RDF som plattform for sammenlikning og sammenføyning av emnesystemer?: Rapport*. Hentet fra ub.uio.no/om/prosjekter/avsluttet/realfagstermer-tekord/realfagstermer-og-tekord- rapport.pdf

Humord: En tesaurus opprinnelig utviklet innen humaniora. Fra 2011 er vokabularet utvidet med samfunnsvitenskapelige begreper. Omfatter ca. 18 500 hovedtermer og 8 500 se-henvisninger. Humord er organisert som et indekseringssamarbeid mellom universitetsbibliotekene i Oslo, Bergen og Tromsø samt Senter for studier av Holocaust og livssynsminoriteter.

Realfagstermer: Et kontrollert pre-koordinert emneordsvokabular som hovedsakelig dekker fagområdene naturvitenskap, matematikk og informatikk. Omfatter ca. 15 000 hovedtermer og ca. 2000 se-henvisninger. I dette inngår også fellesvokabular fra prosjektet mellom Realfagstermer og

TEKORD (NTNU).

Norsk WebDewey: En pågående, fullstendig norsk oversettelse av Dewey Decimal Classification (DDC), publisert på nett.

ISO-standarden *Information and documentation:* *Thesauri and interoperability with other vocabularies* (International Organization for Standardization, 2009, 2013) danner det teoretiske grunnlaget for prosjektet som omtales i artikkelen. Det er spesielt del to*: Interoperability with other vocabularies* som er relevant for arbeidet vårt. SKOS (Simple Knowledge Organization System)[7]: En modell og et RDF-vokabular som er laget for å kunne modellere vanlige, men relativt enkle og semi-formelle kunnskapsorganisasjonssystemer.

RDF: (Resource Description Framework)[8]: En familie av W3C-standarder som er basis for SKOS, arbeidet med åpne lenkede data og semantisk web.

http://urn.nb.no/URN:NBN:no-44610

Artikkelen refererer til to prosjekter.

FAKTA

Prosjektets tittel: På vei mot en generell norsk tesaurus
Prosjektansvarlig: Universitetet i Oslo, Universitetsbiblioteket
Ansvarlig kontaktperson: Halvor Kongshavn
Varighet: 2014
Prosjektstøtte fra Nasjonalbiblioteket: 2014: kr 1 350 000

Prosjektet er todelt:

1. Utvikle metodikk for mapping fra tesaurus til WebDewey, med Humord som utprøvingsarena.
2. Forprosjekt som gir beslutningsgrunnlag i spørsmålet om det skal etableres en universell tesaurus med utgangspunkt i Humord. Forprosjektet gjennomføres av NB og UBO i fellesskap, med utgangspunkt i prosjektplan utarbeidet av NB.

Innsatsområder: Samarbeid og partnerskap
Nasjonalbibliotekets digitale tjenester som grunnlag for nye tilbud

7 w3.org/2004/02/skos/
8 w3.org/RDF/

FAKTA

Prosjektets tittel: Realfagstermer og TEKORD
Prosjektansvarlig: Universitetet i Oslo, Universitetsbiblioteket
Ansvarlig kontaktperson: Live Rasmussen
Prosjektstøtte fra Nasjonalbiblioteket: 2012: kr 200 000
Mål for prosjektet: En sammenføying av kontrollerte emneord brukt på realfagslitteratur ved UBO (Realfagstermer) og termer innen teknikk og naturvitenskap koplet til UDK-klassifikasjon fra UBiT (TEKORD) vil kunne gi:

- Erfaring og uttesting av sammenføying av ulike emneordssystemer som kan utvides til å gjelde andre norske emneordssystemer senere
- Samarbeid og idèutveksling om mulig framtidig bruk av norske emneord som linked data
- Gevinster i brukbarhet for bibliotekansatte og sluttbrukere som kan gjøres sømløs bruk av begge ressurser
- Gevinster knyttet til ressursbruk for vedlikehold og videreutvikling av systemene
- Erfaringer for emneordsarbeid som kan brukes inn i implementering av nytt biblioteksystem

Innsatsområde: Samarbeid

Skal det være en tesaurus før vi stenger?

Rapport fra en mulighetsutredning om norsk generell tesaurus

Oddrun Pauline Ohren, seniorrådgiver, Nasjonalbiblioteket.
Foto: Nasjonalbiblioteket

Kirsten Rydland, hovedbibliotekar, Nasjonalbiblioteket.
Foto: Mats Rydland

Nasjonalbiblioteket har et langsiktig mål om å kunne tilby bedre emnebaserte innganger til – og på tvers av –samlingene. En forutsetning for dette er gode og konsistente emnebeskrivelser i metadata, og den interne aktiviteten «Emner i NB» har i oppgave å komme med anbefalinger om god praksis på dette området. Emnebeskrivelser omfatter både klassifikasjon, generelle emneord, geografiske emneord, sjangre o.a. Når det gjelder emneord, konkluderer rapporten med at tesaurusen Humord vil danne et godt utgangspunkt. Humord brukes og utvikles av flere universitetsbibliotek, med Universitets-biblioteket i Oslo i spissen.

TEKST: **KIRSTEN RYDLAND** OG **ODDRUN PAULINE OHREN**

Digitalisering av dokumenter gir oss flere nye muligheter for gjenfinning via direkte søk i ressursene[1]. Det utelukker likevel ikke at det ligger en stor gevinst i bruk av kontrollerte emneord, ikke minst med tanke på presisjon.

1 Her: Fellesbetegnelse for dokumenter uavhengig av materialtype. En ressurs kan være en monografi, tidsskriftartikkel, avis, film, bilde, musikkalbum, musikkspor osv.

Utfordringa er i grunnen ikke å finne noe, men i å finne de riktige dokumentene i den store informasjonsflommen. Jon Bing sa det slik: «Å spørre om vi trenger bibliotekene nå som det finnes så mye informasjon, avdekker omtrent samme innsikt som å spørre om man trenger veikart nå som det er blitt så forferdelig mange veier»[2].

Nasjonalbiblioteket (NB) og Universitetsbiblioteket i Oslo (UBO) gikk i mars i år sammen om et arbeid for å utrede muligheten for en norsk, generell tesaurus. HumSam-biblioteket ved UBO bruker og drifter tesaurusen Humord. Tanken er at Humord kan danne basis for en ny, norsk tesaurus. I tillegg er det tenkt å integrere andre emnesystemer på ulike nivåer og i ulik grad. En eventuell ny tesaurus skal i rimelig grad dekke alle fagområder, uten å utvikle nye termer der det allerede finnes gode verktøy.

Forprosjektet ledes av NB, med deltakere fra NB og UBO. Det er lagt vekt på at deltakerne skal dekke nødvendige kunnskapsområder; primært bibliotekfag, informasjonsteknologi og språkteknologi. I tillegg fungerer NKKI[3] som rådgiver og diskusjonspartner undervegs.

Første møte var 5. mars i 2014, og prosjektet er planlagt sluttført i mars 2015.

HVA VIL VI MED FORPROSJEKTET?

Å etablere en ny generell tesaurus er en stor oppgave. Dette ettårige forprosjektet skal gi kunnskap om og oversikt over hva det innebærer å utvikle en norsk generell tesaurus basert på Humord. Det skal anbefale om tesaurusen bør utvikles, og hvordan det eventuelt kan gjøres. Arbeidet skal munne ut i en sluttrapport som vil være beslutningsgrunnlag når den endelige avgjørelsen skal tas.

Forprosjektplanen slår fast at arbeidet skal ende opp med konkrete resultater på fire områder:

1. En plan for hvordan første versjon av tesaurusen kan utvikles. Planen skal være konkret; den skal beskrive hva første versjon skal omfatte, inneholde en oversikt over hvilke aktiviteter som er nødvendige å utføre, og plassere dem i en tidsplan, angi organisering av arbeidet og deltakere, og gi et estimat over ressursbehov.

2 Morgenbladet 29.7.2005
3 http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/NKKI

2. Forslag til driftsmodell for en norsk generell tesaurus.

3. En pilot av en norsk generell tesaurus, dvs. «versjon 0.1». Med det skaffer vi oss erfaring med noen av de praktiske og tekniske problemstillingene som vil oppstå i et eventuelt hovedprosjekt.

4. Forslag til hvordan tesaurusen kan videreutvikles etter første versjon, og en prioritering av dette. Aktuelle områder for videre utvikling er:
   - mapping til andre vokabularer og emnesystemer
   - oversettelse til andre aktuelle språk i Norge (nynorsk, samiske språk, engelsk)
   - eventuelle utvidelser av fagområder

HVA FØRTE FRAM TIL FORPROSJEKTET?

I Nasjonalbiblioteket ble det i 2011 satt i gang et internt arbeid for å se på hvordan vi skal beskrive hva en ressurs handler om for å gi god gjenfinning for brukerne. Arbeidet fikk navnet *Emner i NB*.

Den interne gruppa som jobba med dette, kartla dagens situasjon i NB med hensyn på emnebeskrivelse[4]. Vi så også på hva som finnes av emnesystemer nasjonalt og internasjonalt.

Kartlegginga viste at arbeidet med emnesetting i NB er lite koordinert. Det brukes i liten grad standardiserte emnesystemer, og det finnes flere emneordslister utvikla for å passe til samlinga eller materialet de brukes for. Det er lite eller ingen samordning mellom listene. De aller fleste er flate lister uten bruk av henvisninger.

Også på UBO har en kartlagt praksis i bruk av emnebeskrivelse. Som på NB viser kartlegginga[5] i store trekk uensarta praksis og manglende samordning mellom systemene, i alle fall innen UBO som helhet. Innenfor enkelte fakultetsbibliotek har imidlertid emneordsarbeidet vært bedre koordinert. Humord er et resultat av dette.

4 Ohren, O. P., K. Rydland og I. Rype (2012). Emneinnganger i Nasjonalbiblioteket; Kartlegging av praksis for emnebeskrivelse. Oslo, Nasjonalbiblioteket. http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Tesaurus-forprosjekt/Emner-i-NB-Kartleggingsrapport-2012

5 Al-Araki, K., M. Almo og K. Hegna (2010). Bibliografisk og emnemessig beskrivelse av UBOs samlinger. Permanent lenke: http://urn.nb.no/URN:NBN:no-34030

HumSam-biblioteket ved UBO starta i 1994 arbeidet med et emnesystem på tesaurusform. Humord er fra begynnelsen av orientert mot humanistiske fag, men fra 2011 er også samfunnsfag innlemmet. I dag dekker den «humaniora og samfunnsvitenskap med tilgrensende fagområder»[6] . Flere institusjoner er deltakere i indekseringsprosjektet, bl.a. humaniora-avdelingene ved universitetsbibliotekene i Oslo, Bergen og Tromsø. Humord er integrert med BIBSYS, og alle BIBSYS-bibliotek kan delta i samarbeidet. Tesaurusen har i dag ca. 26 000 termer.

I tillegg til Humord finnes det også andre viktige emnesystemer både ved UBO og NB. Av disse kan nevnes Realfagstermene (UBO), Juridiske emneord (UBO), Norart (artikkelindeksering, NB) og emneliste til Samisk bibliografi (NB). Alle disse er flate emnelister.

I mai 2013 anbefalte[7] emnegruppa NB å gå inn for å utvikle en ny, norsk, generell tesaurus med utgangspunkt i Humord. Samme år søkte UBO om prosjektmidler til blant annet å videreutvikle Humord i retning av større faglig bredde. Både tidsmessig og innholdsmessig passet dette godt med en skisse til et forprosjekt for «Norske emneord» laget i forlengelsen av arbeidet med *Emner i NB*. Dette var bakgrunnen for at UBO ble tildelt utviklingsmidler til et slikt forprosjekt i samarbeid med NB.

Som oppsummering må vi kunne si at det ved begge institusjonene legges ned betydelig arbeid i emneord og indeksering. Men så lenge koordineringa av dette er såpass mangelfull som nå, er det vanskelig å utnytte både arbeidsressursene og resultatene (emnebeskrivelsen) optimalt. På den annen side er det de siste årene også tatt flere initiativer lokalt til bedre samordning. Derfor starter vi på ingen måte fra bunnen av, når vi nå utreder muligheten for en nasjonal ressurs.

HVA HAR VI GJORT?

I startfasen hadde vi behov for å presentere planene om et felles prosjekt mellom NB og UBO. Vi ønsket synspunkter fra bibliotekmiljøet på tanken om å utvikle en egen norsk generell tesaurus. Dessuten var det et poeng å vite at arbeidet hadde ei forankring ut over NB og UBO. I fellesskap arrangerte

6 http://www.bibsys.no/files/out/humord/om-tesaurusen.html

7 Ohren, O. P., K. Rydland og I. Rype (2013). Emneinnganger i Nasjonalbiblioteket; Anbefalinger om praksis for emnebeskrivelse Del 1: Materiale med verbalt innhold. Oslo, Nasjonalbiblioteket. http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Tesaurus-forprosjekt/Emner-i-NB-Anbefalingsrapport-2013

vi derfor et seminar for å presentere planene. Målgruppa var i første rekke fag- og forskningsbibliotekene. Også Biblioteksentralen var invitert som en sentral forvalter av et mye brukt emneordssystem, BIBBI, i tillegg var NKKI[8] invitert spesielt. Interessen viste seg å være stor: ca. 90 personer fra over 30 institusjoner deltok.

Tilbakemeldinger fra seminaret gjorde at vi gjennomførte en kvantitativ undersøkelse for å få mer kunnskap om bruken av emneord og emneordsystemer i norske bibliotek[9]. Seminaret ga oss også andre gode innspill som vi har tatt med videre. I tillegg har vi kartlagt bruk av generelle tesauri i en del andre land.

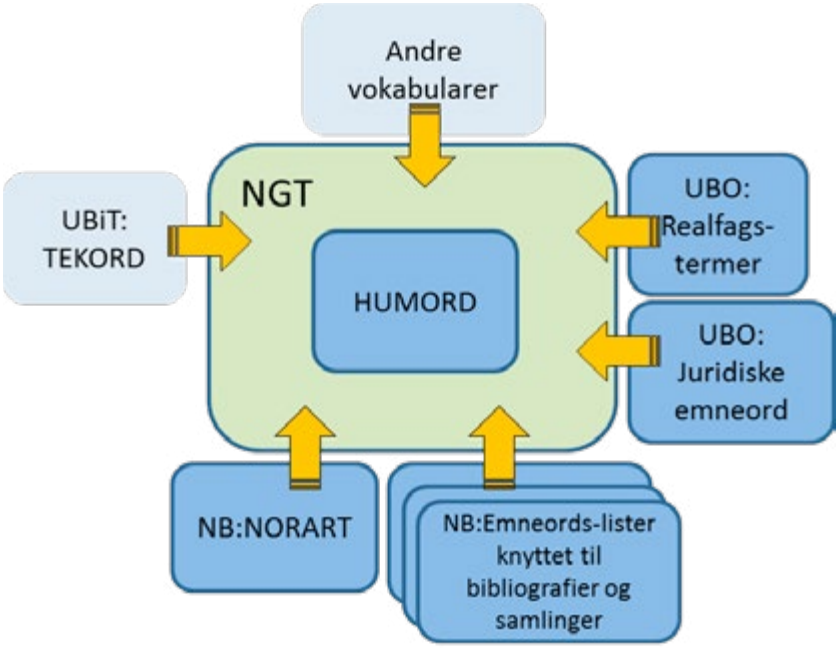
Selve utredningsarbeidet er organisert i arbeidsområder tilsvarende de fire områdene vi etter planen skal levere resultater på. Under hvert område er det igjen en rekke tema som skal avklares.

PLAN FOR UTVIKLING AV NORSK GENERELL TESAURUS, VERSJON 1.0

Første versjon av tesaurusen bør være ferdig i løpet av to –tre år, og vil i praksis bestå av emneordssystemene fra NB og UBO, med Humord som kjerne, se Figur 1.

8 Norsk komité for klassifikasjon og indeksering

9 http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Tesaurus-forprosjekt/Spoerreundersoekelse-om-bruk-av-emneord-og-emneordsystemer



Figur 1 «I begynnelsen var Humord» – Fra Humord til en helhetlig, flerfaglig tesaurus.

Omfang og avgrensing

Tesaurusen skal fungere i en verden der det finnes mange andre emnesystemer. Det er viktig å være bevisst på hvordan den skal forholde seg til disse. En kan tenke seg ulik grad av integrering, fra direkte innlemming av et vokabular i tesaurusen, til mapping (lenking) til begrep i andre vokabularer. Eksempelvis er det naturlig at alle emnevokabularer i bruk ved UBO og NB innlemmes i den nye tesaurusen, mens det etableres mappinger til andre større generelle vokabularer (f.eks. Dewey). På fagområder der det finnes etablerte vokabularer med en viss brukserfaring i Norge (f.eks. MeSH[10] og AGROVOC[11]) må det avgjøres om den nye tesaurusen skal utvikles for disse områdene, og eventuelt på hvilket detaljnivå. Det vil opplagt være viktig å finne gode tekniske løsninger slik at systemene kan fungere effektivt og sømløst sammen.

10 http://www.nlm.nih.gov/mesh/
http://www.kunnskapssenteret.no/prosjekter/medical-subject-headings-mesh-oversatt-til-norsk
11 http://aims.fao.org/agrovoc#.VG4c3_nF-vA

Tesaurusen skal dekke alle fagområder, men detaljeringsgraden innen de ulike områdene vil påvirkes av litteraturbelegget ved institusjonene som bidrar med nye begreper, i første omgang UBO og NB.

Tesaurusen skal inneholde generelle innholdsbeskrivende emneord. Geografiske emneord, formemneord og person-/korporasjonsnavn inngår ikke.

Tesaurusen er tenkt til postkoordinert bruk, det skal derfor ikke etableres emnestrenger.

Første versjon vil forekomme på bokmål, men den skal være forberedt for flerspråklighet fra første stund.

Vi ser for oss at tesaurusen skal foreligge åpent tilgjengelig på nett, både som nedlastbare datasett og for søk gjennom et sluttbrukergrensesnitt. Emnetermene skal være tilgjengelige som åpne lenkede data.

Valg av tesaurussystem og annen teknologi
Det blir viktig å finne et tesaurussystem som støtter distribuert arbeid, som håndterer lenkede data generelt, og gjerne har spesiell støtte for SKOS. Så langt er seks systemer identifisert og under evaluering. De fleste må bedømmes ut fra dokumentasjon og eventuell prøvelisens. For piloten brukes VocBench[12], et system med åpen kildekode som også brukes av AGROVOC og EUROVOC[13]. Foreløpig er det et åpent spørsmål om VocBench anbefales til et eventuelt hovedprosjekt.

Det er ønskelig å bruke språkteknologi i utviklingsarbeidet. Denne teknologien kan, gjennom søk og analyse av tekst, bidra særlig til å foreslå kandidater til nye emneord, både nye begreper og alternative termer til eksisterende begreper. Bruk av språkteknologi kan i tillegg fungere som støtte for å strukturere ustrukturerte (flate) vokabularer som skal inngå i tesaurusen.

Få av tesaurussystemene har innebygd språkteknologiske funksjoner. I miljøet rundt Språkbanken i NB foregår det imidlertid programvareutvikling som også kan være nyttig i tesaurussammenheng, se s. 64.

12 http://vocbench.uniroma2.it/
13 http://eurovoc.europa.eu/

DRIFTSMODELL

Tilbakemeldingene fra seminaret tidlig i prosessen, var at en anså det naturlig at NB tar ansvar for drift av en eventuell norsk tesaurus. Senere diskusjoner i forprosjektet viser imidlertid klart at selv om NB får hovedansvaret, må det være et nært og kontinuerlig samarbeid mellom NB og UBO (og eventuelle andre framtidige samarbeidspartnere) om dette.

Arbeidsgruppa er godt i gang med å meisle ut hvordan en generell norsk tesaurus kan driftes. Humord koordineres i dag fra HumSam-biblioteket ved UBO, og derfra har vi med oss erfaring som kan bidra i arbeidet. Detaljene ved forslaget til driftsmodell er ikke klare, men det er enighet om at arbeidsprosessen for behandling av begreper og termer må fungere raskt og ubyråkratisk, og vi må ha en teknisk drift som støtter opp om dette. Samtidig er det avgjørende at den faglige kvaliteten er på et høgt nivå. Både NB og UBO har mange ansatte med høg kompetanse på sine områder. Disse er en nødvendig ressurs i drift og videreutvikling av en eventuell norsk generell tesaurus.

ETABLERE PILOTVERSJON, VERSJON 0.1

Pilotversjonen av tesaurusen er en integrasjon mellom Humord, emneordene til Samisk bibliografi (SAM, 1458 begrep) og de samlede emneordene brukt i bibliografiene over forfatterne Bjørnson, Collett, Hamsun og Solstad (FBIB, 459 begrep). Siden SAM og FBIB er flate lister, består arbeidet i å plassere begrepene fra disse inn i Humords tesaurusstruktur. Så mye som mulig av arbeidet gjøres automatisk:

- Alle vokabularene er representert i SKOS-XL og lagt inn i VocBench som tre separate vokabularer. Foruten opplysningene fra grunndataene er det for alle begrep føyd til en proveniensnote som angir hvilket vokabular begrepet stammer fra.

- Emneord fra FBIB som betegner litterære karakterer er identifisert for seg og plassert på riktig sted i Humordhierarkiet.

- Leksikalsk likhet mellom foretrukne termer i FBIB/SAM og foretrukket term i Humord er identifisert som «eksakt match» og ansees som ferdig behandlet etter at en proveniens-note (f.eks.: «Kilde: Samisk bibliografi») er lagt inn som skos:editorialNote i det aktuelle Humord-begrepet.

- Leksikalsk likhet mellom alternative termer i FBIB/SAM og Humord, eller på tvers av foretrukket og alternativ term identifiseres som «nær match», er eksportert til et regneark og skal sees over og kodes manuelt for videre behandling.

- Termer i FBIB/SAM som ikke finnes i Humord (ca 750 termer) er eksportert til et regneark og må behandles/kodes manuelt for videre behandling. Dette består i – for hver term i FBIB/SAM - å identifisere relevant foreldre-begrep eller synonym i Humord, og kode dette på foreskrevet måte i regnearket.

I skrivende stund er den manuelle behandlingen nettopp startet. For å få et inntrykk av arbeidsmengden vil det bli gjort tidsregistrering fra dette arbeidet. Uten å foregripe resultatene av denne, er det klart at det trengs bedre systemstøtte enn det vi har i piloten gjennom VocBench, når større vokabularer skal integreres.

VEIKART FOR VIDERE UTVIKLING ETTER VERSJON 1.0

Som nevnt ser vi i dag for oss en første versjon ferdig i løpet av to – tre år fra arbeidet igangsettes. Ambisjonen er at denne dekker alle fagområder i rimelig grad (om enn på ulikt detaljeringsnivå). Dessuten at den fungerer som et brukbart verktøy for emnesetting og gjenfinning, i hvert fall i NB og UBO. Dernest er det viktig å ha tenkt gjennom hvor veien videre går, hvilke områder for videre utvikling som skal prioriteres. Her har forprosjektet kommet relativt kort, selv om noen elementer på veikartet peker seg ut: Lage mappinger til Dewey (og eventuelt andre vokabularer), oversettelse av hele eller deler av tesaurusen til andre språk brukt i Norge (nynorsk, samisk og engelsk) og utvidelse av fagområder i tråd med eventuelle nye samarbeidspartnere.

HVORFOR ER DETTE ARBEIDET VIKTIG?

Vi sier gjerne at vi lever i et informasjonssamfunn. Internett gir oss tilgang til en hel verden av informasjon i form av tekst, bilder lyd og levende bilder. Det er ikke lenger en tett knytting mellom innhold og publiseringskanal eller fysisk format, vi ser bildene våre på tv-en og fjernsyn på nettbrett eller pc. Internett er et daglig medium. For de som vet å bruke denne tilgangen er det ikke problematisk å finne noe om et emne. Men spørsmålet om hva vi finner, og om vi finner det vi trenger er ikke blitt mindre aktuelt, tvert i mot.

Det er også et språklig element her. Stortingsmeldinga Mål og Meining[14] slo fast at fagterminologi på norsk er viktig for å fremme presis faglig forståelse, og letter dessuten arbeidet med å popularisere spesialisert kunnskap. Samtidig erkjennes det at dette krever et særskilt nasjonalt terminologiarbeid. En tesaurus på norsk som brukes på universitetsnivå vil uten tvil være et godt bidrag i denne sammenhengen.

I bibliotekene har vi tradisjon for å bruke kontrollerte søkeord for å gjenfinne relevant materiale. Om vi klarer å kombinere bibliotekfaglige tradisjoner og tekniske muligheter på en god måte, bør vi kunne få til ei gjenfinning av interessante ressurser som er bedre og mer presis enn det vi gjør i dag. Sitatet fra Jon Bing om veikart og bibliotek er ikke blitt mindre relevant siden han skrev det i 2005. Forskning tyder dessuten på at kontrollerte og ukontrollerte (for eksempel fulltekstindeksering) emnesystemer utfyller hverandre[15] Vi ønsker derfor ikke å velge det ene eller det andre, enten kontrollerte emneord fra en tesaurus eller søk i fri tekst, vi vil ha både i pose og sekk. Vi må bare sørge for å legge til rette for å kombinere søkemulighetene på intelligente måter.

Artikkelen beskriver prosjektet «På vei mot en generell norsk tesaurus» i del to i faktaboksen på s. 45

14 Mål og meining; Ein heilskapleg norsk språkpolitikk. St.meld. nr. 35 (2007-2008).

15 Žumer, M., Ed.(2009). National bibliographies in the digital age: guidance and new directions. München, K.G. Saur, 2009.

Nasjonalt autoritets-register for person- og korporasjonsnavn

I Norge har vi manglet et nasjonalt autoritetsregister for person- og korporasjonsnavn. Behovet for et slikt register har flere ganger de siste 30 år vært etterlyst. Nå er det under oppbygging.

TEKST: **FRANK HAUGEN**

HISTORIEN

Så tidlig som i 1984 reiste daværende Norsk avdeling ved Universitetsbiblioteket i Oslo (UBO) spørsmålet om det burde bygges opp et nasjonalt register for person- og korporasjonsnavn. Spørsmålet ble aktualisert av at det nye regelverket for katalogisering (Katalogiseringsregler) ble introdusert i Norge og tatt i bruk for nasjonalbibliografien (Norsk bokfortegnelse). Både Nasjonalbibliografisk utvalg og Den norske katalogkomité, som ble opprettet året etter, ønsket å starte et forprosjekt for å se på hvordan et slikt register kunne bygges opp, men noe forprosjekt kom aldri i gang.

Frank Haugen, hovedbibliotekar, Nasjonalbiblioteket. Foto: Trine Adolfsen

Neste gang spørsmålet om et slikt register ble reist, var da Nasjonalbibliotekavdelinga i Rana ble opprettet i 1989. Kulturdepartementet påpekte behovet for et autoritetsregister, og saken ble utredet av Katalogkomiteen. I mars 1996 forelå utredningen «Nasjonalt autoritetsregister for person- og korporasjonsnavn».

Utredningen anbefalte at det ble tatt utgangspunkt i Biblioteksentralens autoritetsregister, og at Nasjonalbiblioteket og Biblioteksentralen samarbeidet om å tilpasse registeret til kravene for et nasjonalt autoritetsregister. Videre ble det anbefalt at registeret skulle driftes av Nasjonalbiblioteket som en integrert del av «det nasjonalbibliografiske systemet».

Det naturlige ville ha vært å basere registeret på nasjonalbibliografisk produksjon i Norsk bokfortegnelse. At det ikke ble konklusjonen, er først og fremst knyttet til at Norsk bokfortegnelse gjenspeiler nasjonalbibliografiens endringer i navneformpraksis gjennom tidene (fra 1971) og dermed ikke var homogent og konsistent nok. Biblioteksentralens register ble derimot ansett som mer egnet, med ca. 19 000 autoritetskontrollerte navneformer for personnavn.

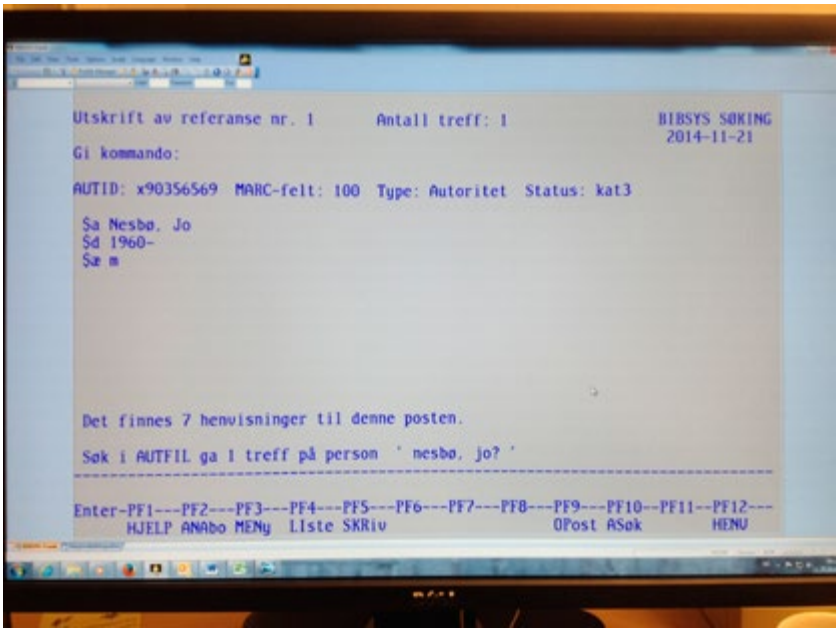
Heller ikke denne gangen skjedde det særlig mye i praksis. Man kan anta at årsaken til dette ligger i den sterke to-delingen som fantes i bibliotekvesenet på denne tiden. ABM-utvikling var ennå ikke etablert. Vi hadde Statens bibliotektilsyn på folkebiblioteksiden mens Riksbibliotektjenesten sørvet fagbibliotekene.

I forordet til utredningen er Katalogkomiteen kanskje inne på det samme (min utheving):

> «Den norske katalogkomité ønsker med denne utredningen å rette søkelyset på at norske bibliotek mangler et felles autoritetsregister for person- og korporasjonsnavn. Vårt mål har ikke vært å presentere et ferdig system - den oppgaven er for omfattende til å kunne gjøres innenfor rammen av komitéens virksomhet - men å trekke fram en del forutsetninger som bør knyttes til et slikt register og etablere et utgangspunkt for det videre arbeidet. *Slik vi ser det er de organisatoriske problemene større enn de bibliotekfaglige og tekniske problemene.»*

I 2002 startet Nasjonalbiblioteket en gjennomgang av bibliografisk produksjon og tjenesteutvikling, der også autoritetsregisteret ble tatt med. Katalogkomiteen blir bedt om å oppdatere utredningen fra 1996.

Komiteen anbefaler på nytt et samarbeid mellom Nasjonalbiblioteket og Biblioteksentralen for å få Biblioteksentralens register i samsvar med norske katalogiseringsregler, og å utrede en teknisk og organisatorisk driftsmodell.



Autoritetsfil

ENDREDE FORUTSETNINGER

Først i 2009 starter det praktiske arbeidet med å bygge opp et nasjonalt autoritetsregister for person- og korporasjonsnavn, men da med endrede forutsetninger.

Det skjer en god del mellom 2006 og 2009. Nasjonalbiblioteket flytter produksjon av spesial- og nasjonalbibliografier til biblioteksystemet BIBSYS, og i mars 2009 inngår Nasjonalbiblioteket en avtale med Biblioteksentralen om kjøp av metadata til sin nasjonalbibliografiske produksjon.

SKRITTVIS UTVIKLING

Grunnlaget er nå lagt for å produsere Nasjonalt autoritetsregister for person- og korporasjonsnavn i BIBSYS som en integrert del av nasjonalbibliografien, og fra januar 2010 begynner arbeidet med å befolke registeret. Bibliografiske poster fra Biblioteksentralen importeres daglig og Nasjonalbiblioteket produserer kvalitetssikrede data for resten av den pliktavleverte dokumentmengden. Omfanget av registeret knyttes til pliktavlevert materiale.

Det er effektivt å integrere produksjonen av det nasjonale autoritetsregisteret med nasjonalbibliografien. Nasjonalt autoritetsregister for person- og

korporasjonsnavn blir en delmengde av autoritetsregisteret i BIBSYS. Siden produksjonen foregår i et katalogsystem som deler data med andre bibliotek, er det nødvendig med noe teknisk tilrettelegging, blant annet funksjonalitet for «låsing» av autoritetspostene, slik at ikke andre BIBSYS-bibliotek skal kunne endre autoriteter merket som nasjonale.

Den norske katalogkomité kommer med råd og innspill underveis om hvilke felter registeret bør inneholde, avgrensning av omfanget, hva som bør registreres og behovet for endring av katalogiseringspraksis i BIBSYS.

OMFANG AV REGISTERET – KRAV TIL REGISTRERING

Siden registeret er basert på nasjonalbibliografisk produksjon, vil det også inneholde utenlandske personnavn og korporasjonsnavn. Norske forfattere registreres med leveår, og Nasjonalbiblioteket legger ned en god del ressurser for å finne fram til dette. Det samme gjøres for utenlandske forfattere, hvis informasjonen er lett tilgjengelig.

Når det gjelder korporasjoner, var en stor utfordring at man i BIBSYS biblioteksystem hadde sløyfet «Norge» som overordnet ledd for «offentlige organer innført som underordnet ledd», jf. Katalogiseringsregler 24.18. Denne praksisen endres, og Nasjonalbiblioteket tar på seg oppgaven med å oppdatere registeret. Alle nyregistrerte korporasjonsnavn av denne typen får «Norge» som overordnet ledd, men det gjenstår fortsatt mye arbeid med å oppdatere alle gamle data. For noen områder er dette gjort, bl.a. departementene.

For korporasjoner som har flere offisielle navneformer, der en av dem er samiskspråklig, gjøres det en vurdering av hvilken navneform som skal velges. For korporasjoner som har sitt virke rettet mot det samiske samfunn eller samiske interesser, er det naturlig å velge samisk navneform. Et typisk eksempel her er Sametinget. Den autoriserte innførselen i registeret er her: Norga. Sámediggi, ikke Norge. Sametinget.

«FLYING START»

Det var viktig at registeret raskt fikk et visst volum. Fra årsskiftet 2007/2008 begynte Nasjonalbiblioteket å registrere kjønn for norske forfattere i Norsk bokfortegnelse. Dette er data som brukes til statistikk. For å få en god start ble alle de «kjønnsmerkede» autoritetspostene låst og merket som «nasjonale». Antakelsen var at siden disse autoritetene var knyttet til registrering i Norsk

bokfortegnelse, ville navneformene være verifiserte. Dette medførte en «flying start» for registeret, og ved oppstart var det allerede generert rundt 20 000 nasjonale autoritetsposter for personnavn. Per. januar 2014 er dette tallet oppe i over 64 000 autoritetsposter for personnavn. Ved siden av tilveksten av nye navn gjennom registrering til nasjonalbibliografien, gjøres det en del oppgradering av eldre autoriteter i BIBSYS autoritetsregister til «nasjonal standard».

TILGANG TIL REGISTERET

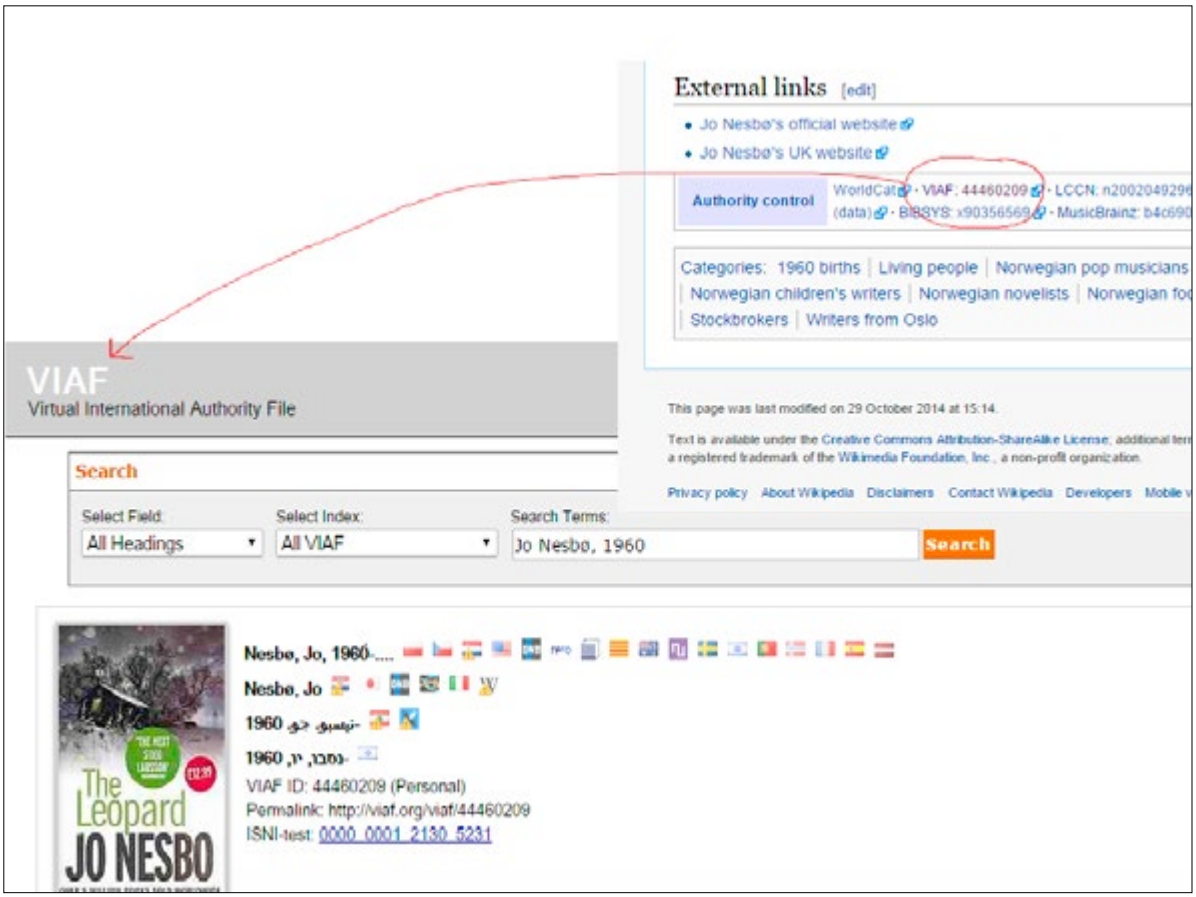
En ulempe med å produsere Nasjonalt autoritetsregister for person- og korporasjonsnavn i BIBSYS er at registeret er «innlåst» i et biblioteksystem. Registeret er ikke tilgjengelig som en separat tjeneste utenfor BIBSYS, noe som begrenser bruken av det for bibliotekmiljøet og andre miljøer som kunne vært interessert. Dette er også et poeng som Katalogkomiteen har trukket fram. Registeret må gjøres tilgjengelig som et frittstående, separat register.

VIRTUAL INTERNATIONAL AUTHORITY FILE (VIAF)

Fra januar 2013 sørger BIBSYS for eksport av nasjonale autoritetsdata til VIAF (Virtual International Authority File). VIAF er et samarbeid mellom først og fremst nasjonalbibliotek rundt om i verden der data fra de ulike landene kombineres i én autoritetstjeneste. Norske navneautoriteter er med dette blitt en del av et internasjonalt fellesskap der data gjøres tilgjengelig i flere formater, bl.a. MARC 21, men også som Linked Data i RDF-formatet. Det gjør at de kan utnyttes i nye sammenhenger, og i systemer som ikke holder seg med MARC-format. Wikipedia lenker blant annet til autoriteter i VIAF.

Det er også andre fordeler med deltakelsen i VIAF. En del autoritetsposter for personnavn blir automatisk tildelt International Standard Name Identifier (ISNI). Forutsetningen er at minst tre nasjonalbibliotek leverer autoritetsdata for samme navn til VIAF.

Men deltakelsen i VIAF gir heller ikke god nok tilgang til Nasjonalt autoritetsregister for person- og korporasjonsnavn utenfor BIBSYS. Til det er VIAF oppdatert alt for sjeldent (1 gang i måneden). For bruk i en nasjonal kontekst, bør registeret oppdateres minst hver natt.



Skjermbilde av Wikipedia og VIAF

NASJONALT AUTORITETSREGISTER FOR NAVN MED EGET GRENSESNITT

Det nasjonale autoritetsregisteret fortjener et eget grensesnitt for søking og verifisering av navneformer med oppdaterte data. Nasjonalbiblioteket har sett på ulike løsninger for dette. Det mest aktuelle nå er å lage en tilgang til registeret via et eget separat grensesnitt i Oria (Ex Libris' Primo). BIBSYS og Nasjonalbiblioteket jobber i skrivende stund med utviklingen av dette.

Det er også viktig å tilby data i maskinleselig form slik at de skal kunne tas i bruk i andre system. Autoritetsdata fra BIBSYS er tilgjengelig for innhøsting fra andre system i protokollene SRU og OAI-PMH. BIBSYS har også begynt å se på hvordan data fra BIBSYS biblioteksystem kan gjøres tilgjengelig som Linked Data (RDF). Hele BIBSYS' autoritetsregister ble gjort tilgjengelig

som Linked Data gjennom prosjektet *Rådata nå* – et samarbeid mellom NTNU Universitetsbiblioteket og BIBSYS, men disse dataene oppdateres ikke med nye poster.

VIDERE ARBEID

Nasjonalbiblioteket vurderer å starte tildeling av ISNI (International Standard Name Identifier). Hvis det blir aktuelt, må det innføres eget ID-felt i registeret for registrering av ISNI.

I forbindelse med innføring av nye katalogiseringsregler, RDA, vil det være hensiktsmessig å tilføye egne felter i personnavnregisteret for yrke og nasjonalitet. Dette er informasjon som tidvis registreres i notefelt i dag.

Både Nasjonalbiblioteket og Den norske katalogkomité har vært inne på tanken å la andre betrodde miljøer få delta i oppdateringen av registeret. Dette må det ses nærmere på, og ev. legges praktisk og avtalemessig til rette for. Nasjonalbiblioteket ønsker også å berike Nasjonalt autoritetsregister for person- og korporasjonsnavn med data fra Mavis, systemet som Nasjonalbiblioteket bruker til registrering av audiovisuelt materiale.

Det har også det seneste tiåret vært diskutert, både i bibliotekmiljøet i Norge og internt i Nasjonalbiblioteket, behovet for verksidentifikatorer for norske forfatteres verk. Problemet er at vi i Norge ikke har tradisjon for å registrere standardtitler for verk utgitt etter 1500. Det er stort sett bare for musikkmateriale vi har en utstrakt bruk av standardtitler. I tillegg brukes det selvsagt standardtittel for løpende ressurser, samlede verker, utvalg og Bibelen, men ikke for enkeltverker av en forfatter. Det betyr at det ikke umiddelbart finnes gode data i metadatapostene våre for å generere slike verksidentifikatorer i et nasjonalt autoritetsregister over forfatternes verker.

Veien videre her er fortsatt under vurdering i Nasjonalbiblioteket. Også her kan det være naturlig å knytte arbeidet opp mot registreringen til nasjonalbibliografien. Hovedproblemet ligger i å generere verksautoriteter retrospektivt, men noe data kan hentes fra andre kilder. I VIAF finnes det en god del slike verksinnførsler, basert på andre nasjonalbiblioteks data, også for norske forfattere. Deichmanske bibliotek har også gjennom flere prosjekt gjort mye godt arbeid som kan være utgangspunkt for et verksregister.

Digitalisering og klassifikasjon

Med Nasjonalbibliotekets digitale tekster kan Deweys desimalklassifikasjon benyttes til å utvikle vokabular for emnemodellering.

TEKST: **LARS G. JOHNSEN**

INTRODUKSJON

Hvordan kan Nasjonalbibliotekets samling av digitaliserte bøker benyttes i kunnskapsorganisering? Svaret er litt opp til hvilke oppgaver en søker å løse, alt fra klassifikasjon til konstruksjon av kontrollerte vokabular. Innenfor pilotprosjektet «Tesaurus forprosjekt»[1] skal vi se på det siste, og viser hvordan relevante innholdsord kan hentes ut fra digitale tekster ved å se på assosiasjoner mellom klassifikasjonskoder og ord. Metoden blir evaluert ved å vurdere resultatene i lys av emneord hentet fra BIBSYS.

Mens emneord gjerne er kontrollert for form og innhold vil ordene i en vilkårlig tekst ikke være mer kontrollert enn hva grammatikken, eller de språklige konvensjoner tillater. Innenfor kunnskapsorganisering tenker man gjerne på slike kontrollerte vokabular, eller emneord, og et underspørsmål blir om disse ordene kan høstes fra innholdsordene, eller vil de måtte konstrueres på annen måte?

1 http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Tesaurus-forprosjekt, Et samarbeidsprosjekt mellom Nasjonalbiblioteket og UBO

Emneord vil typisk være mer generelle enn innholdsord. I indekseringen av en bok kan ordene *kunst* og *skulpturer* benyttes som beskrivende emneord, mens boken selv handler om spesifikke figurer i Vigelandsparken, eller om Vigeland selv. Innholdsord vil typisk være lavere plassert i et semantisk hierarki enn det emneordene er, gjerne som semantiske rammer for flere aktører eller objekter; Vigeland er en aktør som lager skulpturer. Andre relasjoner mellom innhold og emne kan være i taksonomiske hierarkier som forholdet mellom *egg* og *næringsmiddel*, men også der emneordene er knyttet til begrepet *matlaging* som komponenter. Begrepet setter opp en ramme der objekter som *melk*, *mel* og *salt* inngår, mens begrepet *religion* setter opp en ramme for relasjoner mellom *gud*, *profeter* og *frelse*, samtidig som det står i et taksonomisk forhold til *kristendom* og *hinduisme*, for å nevne noe.

Lars G. Johnsen, forskningsbibliotekar, Nasjonalbiblioteket.
Foto: Nasjonalbiblioteket

En metode som skal foreslå typiske innholdsord fra en tekst, bør kunne relatere forslagene enten taksonomisk eller via de semantiske rammene for emneordene i den klassifiserte teksten.

De digitaliserte tekstene er gjort maskinlesbare via OCR-behandling[2], og er derfor tilgjengelig for ordtelling. Datagrunnlaget er Nasjonalbibliotekets samling av digitaliserte bøker som har fått en Deweyklassifikasjon (heretter DDK – Deweys desimalklassifikasjon).

Informasjonen som skal eskes ut er bøkenes implisitte informasjon om forholdet mellom innholdsord, klassifikasjonskode (Dewey) og emneord. Ordene i en klassifisert bok vil være indirekte koblet til bokens klassifikasjonskode; vi kan si at ordene arver klassifikasjonen fra boken de står i. Altså, et ord er koblet til en spesifikk klassifikasjonskode når boken det står i er slik koblet.

DATAINNSAMLING

De to hovedgruppene, emneord og innholdsord, er hentet ut fra BIBSYS og fra bokhyllamaterialet, henholdsvis. Titlene er fra en liste over digitalisert materiale frem til 2013. Fra disse velges bokmåls- og nynorsktekster, og kun de som er entydig klassifisert som bokmål eller nynorsk; flerspråklige verk er ikke tatt med. Totalt utgjør det ca. 240 000 titler, hvorav ca. 144 000

2 OCR (Optical Character Recognition) – gjenkjenning av bokstaver og ord fra bilder av skrift.

har et DDK-nummer. Fra siffergruppene i DDK beholdes første gruppe på tre siffer for videre analyse[3]. Totalt gir det 750 tresifferkombinasjoner av emneklassifikasjoner for enkeltverk.

INNHOLDSORD FRA DIGITALISERTE VERK

Verkene er telt opp på ord slik at hvert ord kobles til DDK-nummeret for verket. Ordene innenfor en og samme verk blir derfor, på en måte, dissosiert fra hverandre, og i stedet knyttet sammen med klassifikasjonen. Dermed kan ord fra forskjellige bøker relateres til hverandre ved at de faller inn under samme DDK-nummer. Samme ord kan kombineres med flere klassifikasjonskoder. Resultatet av prosessen er en database over kombinasjoner av DDK og ord der frekvensen angir hvor ofte DDK-nummeret forekommer med det aktuelle ordet. Her illustrert med ordene *egg* og *kano* for DDK 200 religion) og 759 (historie, geografisk behandling og biografier innen malerkunst), så *egg* forekommer 215 ganger i bøker klassifisert som religion, og *kano* forekommer 17 ganger i 759.

Tabell 1 Samforekomster mellom DDK-numre og innholdsord

Frekvens	DDK	Ord
215	200	egg
200	759	egg
17	759	kano
6	200	kvinner

For enkelhets skyld er det ikke tatt hensyn til flerordsuttrykk, som for eksempel den medisinske termen *cystisk fibrose*, eller fagtermer med ordet *syndrom* i seg. I en fullstendig analyse vil den type uttrykk ha en naturlig plass. Det er heller ikke foretatt noen automatisk grammatikalsk analyse av ordene, som inndeling i ordklasser eller lemmatisering[4]. I diskusjonen nedenfor gjøres det manuelt i hvert enkelt tilfelle.

STATISTIKK

Målet med den statistiske analysen er å finne meningsfulle innholdsord fra samlingen. For desimalgruppen DDK 799 (fiske, jakt, skyttersport), er det ca. 45 000 ordformer med frekvens over 20. Av disse skal vi hente ut et tyvetalls ord og sammenligne med frekvente emneord.

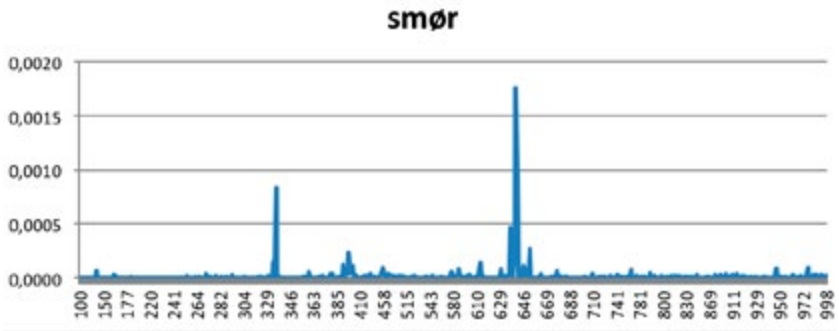
3 Desimalkombinasjonen 839.609 bidrar derfor kun med hovedgruppen 839 (norsk skjønnlitteratur)

4 Lemmatisering – erstatte et bøyd ord med dets stamme.

Statistikken gir et mål på variasjon i materialet. Ikke alle ord er likt fordelt på de forskjellige DDK-numre. De ordene som er typiske for en klassifikasjonskode antas å klumpe seg sammen under denne, og muligens også grupperes under forskjellige klassifikasjonskoder i varierende grad. For innholdsord er det viktig å ta med et slikt mål da ren frekvenstelling (relativ eller absolutt) ikke duger. Høyfrekvente ord som *og*, *en* og *på*[5] er jevnt fordelt over kategorier, og kan derfor neppe sies å være assosiert med en eller annen kategori: de tilhører alle grupper. De topp femti ordene i norsk er gjerne de samme på tvers av klassifikasjonskoder.

Det kan likevel være interessant å se på variasjonen i høyfrekvente ord, da den innbyrdes fordelingen mellom dem kan være bærer av informasjon. Enkelte grupper kan ha mer enn andre. Det er en forskjell i ratioen mellom preposisjonen *i* og konjunksjonen *og* på tvers av klassifikasjonskodene[6].

Ser vi på frekvensfordelingen (relativ frekvens) av ordet *smør* i figur 1, er det to klare topper, en for DDK 641 (mat og drikke) og en for DDK 336 (offentlige finanser). Visuelt er det lett å se hvilke klassifikasjoner *smør* tilhører.



Figur 1 Relativ frekvens for smør over DDK

Et kvantitativt mål for den type distribusjon får vi i såkalt PMI (Pointwise Mutual Information, ett av flere mål på samvariasjon)[7] for DDK og innholdsordet. I korthet beregnes PMI for frekvensdata ved å se på den predikerte samforekomsten med den faktiske, slik at ord som klumper seg

5 Også kjent som stoppord i søkemotorer.

6 I snitt er det 1.3 ganger så mange *og* som *i* med et standardavvik på 0.3. Flere frekvenslister innenfor forskjellige DDK har over dobbelt så mange *og* som *i*, noe som er over 2 ganger standardavviket.

7 Kenneth Ward Church og Patrick Hanks (1990). «Word association norms, mutual information, and lexicography". Computational Linguistics 16 (1): 22–29.

sammen under en kode vil gi høy score, mens ord som fordeler seg jevnt over kodespekteret vil gi lav score.

Logikken bak PMI prøve å fange inn egenskapene til visualiseringen i figur 1. Det er fire størrelser som inngår når vi tar for oss et ord og en kode i en samling:

1. Størrelsen på hele samlingen (T)
2. Antall forekomster O for det aktuelle ordet i hele samlingen
3. D, som er det totale antall ord i tekstene som er klassifisert med det aktuelle DDK-nummeret.
4. Antallet ganger ordet forekommer i tekster med koden, OD, som utgjør samforekomsten mellom kode og ord.

Den relative frekvensen for et gitt ord blir O/T, og den predikerte samforekomsten av ordet med DDK, blir da P = D*O/T, altså at ordet er jevnt fordelt over delgruppene av materialet. Det betyr at om et ord har en relativ frekvens på 0,005 % (*smør* ligger litt i underkant) i materialet, og koden forekommer 200 000 ganger, kan man forvente at ordet skulle forekomme 200000*0.00005 = 10 ganger under den klassifikasjonen. Målet for hvor nært ordet henger sammen med klassifikasjonskoden fås ved å dividere den faktiske forekomsten OD på den forventede, OD/P. PMI for frekvensdata defineres gjennom å ta logaritmen av den ratioen: PMI = log(DO/P). Den størrelsen kan benyttes til å sammenligne ord innenfor en gitt klassifikasjonskode, i tillegg til å finne beste klassifikasjon for et gitt ord.

Statistiske mål vil ikke uten videre plukke ut gode semantiske ord, ord som er relatert til emnet, da det kan være forskjellige grunner til at ord klumper seg rundt en spesifikk DDK. Det kan være idiosynkratiske skrivemåter for en forfatter, og et ord kan brukes sjelden. Et ord kan tilhøre en lite brukt skriftnorm som for eksempel former innenfor radikalt bokmål (e.g. flertallsformene *morfemene* vs. *morfema*). Selv om slike ord kan fungere som markører for klassifikasjonen trenger de ikke ha de semantiske egenskapene man er ute etter.

Innholdsordene bør ha en viss utbredelse innenfor koden. For å unngå at lavfrekvente ord dominerer, vektes PMI med et mål på spredningen av ordet. Vår kandidat er å vekte PMI med kvadratroten av frekvensen til samvariasjonen. Mens PMI måler hvor sterkt ordet er knyttet til klassifikasjonskoden vil kvadratroten av frekvensen gi et mål på spredningen, samtidig som høyfrekvente ord ikke får så stor vekt. For å hindre at spesielt særegne ord blander

seg inn er det kun ord med frekvens over 20 innenfor koden som beregnes.

Hvordan den vektede PMI fungerer, illustreres for DDK 641 i Tabell 2 og termen *smør* i Tabell 1 for de 5 med høyest score. Assosiasjonen sammenfaller med frekvensen i disse tilfellene bortsett fra *pepper,* som under DDK 641 som har en lavere frekvens enn *smør*, men likevel plassert foran sortert etter assosiasjon, se tabell 2:

Tabell 2 Smør

Frekvens	Assosiasjon	DDK	
73794	1753	641	Mat og drikke
13142	615	336	Offentlige finanser
51251	479	948	Skandinavia og Finland
1855	194	637	Framstilling av meieriprodukter og lignende
1078	183	642	Måltider og servering

Tabell 3 DDK 641

Frekvens	Assosiasjon	Ord
110528	2251	salt
100485	2160	dl
88474	1972	ss
64096	1910	pepper
73794	1753	smør

Hvor god denne sorteringen er får vi vite først når vi sammenligner med emneordene for klassifikasjonskoden. Mens DDK-nummeret i seg selv ikke bærer noen semantisk informasjon – det bidrar kun som en gruppeindikator for PMI – vil emneordene kunne sammenlignes med innholdsordene på et semantisk nivå.

EMNEORD FRA BIBSYS

Emneordene er hentet ut fra BIBSYS. For et (ukontrollert) utvalg på 120 000 titler over digitaliserte bøker er det registrert emneord for DDK fra MARC-

postene. På samme måte som for innholdsordene, blir bare de tre første sifrene i DDK-nummeret med. Likeså blir koblingen mellom verk og emneord, og emneordenes innbyrdes samforekomst, koblet fra hverandre. Emneordene relateres kun gjennom felles DDK. Forekomster av emneord og DDK er registrert og telt opp. Tellingen generer et datasett illustrert i Tabell 3, eksemplifisert for emneordene *likestilling* og *kvinner* (registrert både i kontrollert og ukontrollert vokabular):

Tabell 4 likestilling og kvinner

Frekvens	DDK	Tag	Ord
6	379	653	Likestilling
6	612	653	Kvinner
4	305	650	Likestilling
10	305	650	kvinner

Totalt er det ca. 620 000 forekomster av emneord fordelt på kontrollert vokabular (250 000) og fritt vokabular (370 000)[8].

Det er ikke foretatt noen analyse av emneordene utover opptelling. I evalueringen av innholdsord blir de vurdert sammen med de mest frekvente emneordene for en gitt DDK.

EKSPERIMENT OG EVALUERING

En gjennomgang av et sett emneord med et sett innholdsord, ett fra hver hundregruppe i DDK-serien viser at metoden for å velge ut innholdsord har noe for seg. Her skal vi vise et eksperiment med klassifikasjonskoden 799 (fiske, jakt, skyttersport). Settet med innholdsord som foreslås av det statistiske målet for en gitt klassifikasjonskode sammenlignes med de mest frekvente emneordene for koden. Sammenligningen skjer ved at innholdsordene vurderes som relevant i de tilfeller der de er semantisk relatert til emneordene.

Tabellen nedenfor illustrerer hvordan innholdsordene med høyest assosiasjonsscore for kode 799 er koblet til de mest frekvente emneordene for koden. Det er ikke skilt mellom kontrollert og ukontrollert vokabular, og det er valgt

8 I *MARC*-postene benyttes tag 650 for kontrollert vokabular, og 653 for ukontrollert

20 ord fra hver liste. De to listene er koblet sammen ved en skjønnsmessig vurdering av hvor godt, og på hvilken måte, emneordet henger sammen med innholdsordet.

I tabellen (Tabell 4) er både emne- og innholdsord gruppert sammen i celler. De er ikke gruppert i utgangspunktet, men der det synes å være en naturlig sammenheng deler de celle.

Venstre kolonne inneholder grupper av emneord, midtre innholdsord som er vurdert å være semantisk relatert, og høyre kolonne indikerer hvilke semantiske relasjoner det er tenkt på. Det er verdt å legge merke til at det av og til også er identitet mellom emneord og innholdsord, som gir tre typer relasjoner: ekvivalens, taksonomi og semantisk ramme.

Tabell 5 Emneord og innholdsord

Emneord fra BIBSYS	**Innholdsord fra Bokhylla.no**	**Semantisk relasjon**
Jakt, elgjakt, elg, (bjørnejakt)	*Jakt, jegeren, hund, rådyr, dyret, elg, elgen, viltet*	Ekvivalens (*jakt*), semantisk ramme (aktivitet *jakt* med aktører *jeger, rådyr, elg, vilt* redskap *hund*).
Fiske, ferskvannsfiske, sportsfiske	*Fiske, fisket, stanga, vannet, vann, snøret, fisker*	Ekvivalens (*fiske)*, semantisk ramme (aktivitet *fiske*, *fisker* redskaper *stang* og *snøre*)
Fiskefluer, fluebinding, Fluefiske	*Flua, flue, agn, agnet*	Taksonomi (*agn, flue, fiskeflue*) semantisk ramme (aktivitet *fluebinding*, produkt *flue*) (aktivitet *fluefiske,* redskap *flue, agn*)
Fisking, fritidsfiske, laksefiske	*Fisken, fisk, elva, laks, laksen*	Semantisk ramme (aktivitet *fisking, fritidsfiske* og *laksefiske*, aktør *fisk* og *laks*, sted *elv*)

Viltforvaltning, jegerprøven	*Fiskekort, jegere*	Semantisk ramme (aktivitet *jegerprøven* og *forvaltning,* aktører *jegere*) (*jeger* er agentiv for *jegerprøven* men passiv aktør for *forvaltning*)
Ørret	Ørret, fisken, ørreten, sjøørret	Taksonomi (*fisk, ørret, sjøørret*)
Alces, Norge, Friluftsliv, idrett/fritid		*Alces alces* er en artsbetegnelse for elg. *Friluftsliv* er over *fritidsfiske* i det taksonomiske hierarket.

Selv om listen er påvirket av fraværet av grammatikalsk analyse (jf. ørret, ørreten), illustrerer den forekomstene av semantiske relasjoner mellom innholdsord og emner.

Inntrykket tabellen gir, og de klassifikasjonskodene som er testet, sier at både taksonomiske relasjoner og semantiske rammer, i tillegg til ekvivalens er godt representert. Noe overraskende er det kanskje at den taksonomiske også ser ut til også å kunne gå andre veien, der emneordet er mer spesifikt enn innholdsordet (jf. *fiskefluer* som emneord vs. *flue* som innholdsord). Nå sier jo ikke materialet noe om hvilken betydning alle ordene har (*flue* kan jo beskrive en levende flue eller en kunstig), så rimelig sikre konklusjoner er det vanskelig å trekke.

OPPSUMMERING

Som eksperimentet viser, så ligger det mye god informasjon gjemt i det digitale materialet. Siden dataene kobler innholdsord til emneord i semantiske nettverk, vil metoden kunne være en støtte i konstruksjon av tesauruser, i tillegg til å kunne fungere som en støtte i aktuell klassifikasjon av nye bøker. Det siste vil være særlig aktuelt, gitt at nye bøker gjerne kommer i digital form. Samtidig er en betydelig del av det digitale materialet ennå ikke klassifisert. En kunne tenke seg en innsats der disse bøkene kunne gjennomgå en halvautomatisk tilordning av klassifikasjonskoder og forslag til emneord.

Kunnskapsorganisering av språkressurser

Språkressurser er en fellesbetegnelse på ulike typer språkdata som brukes i forskning innenfor et bredt spekter av områder, samt utvikling av språkteknologiske løsninger. Å utarbeide språkressurser er gjerne både arbeids- og kostnadskrevende, og fokuset på videre bruk utenfor det opprinnelige miljøet har derfor økt de siste årene. CLARINO er den norske delen av en felles europeisk innsats (CLARIN) som nettopp har som mål å støtte opp under slik gjenbruk, blant annet ved å legge til rette for gjenfinning basert på samordnede metadata.

TEKST: **ODDRUN PAULINE OHREN**

CLARIN OG CLARINO

CLARIN[1] er en såkalt ERIC[2] og har som oppgave å gjøre tilgjengelig digitale språkressurser og tilhørende verktøy for fagfolk og forskere innen humaniora og samfunnsvitenskap. Dette skal oppnås ved å etablere og drifte en felles, distribuert infrastruktur, som gir brukerne ett felles aksesspunkt til språkdataene på europeisk nivå. Infrastrukturen finansieres og bygges opp av medlemmene.

Oddrun Pauline Ohren, seniorrådgiver, Nasjonalbiblioteket.
Foto: Nasjonalbiblioteket

1 Common Language Research Infrastructure, se http://clarin.eu/

2 European Research Infrastructure Consortium, en type *internasjonal juridisk enhet* etablert av EU i 2009. Medlemmene i en ERIC er land/regjerninger i EU. Som ikke-medlem i EU har Norge status som observatør i CLARIN ERIC.

Hvert medlem må danne et nasjonalt konsortium, som i sin tur har ansvar for å bygge opp en CLARIN-senter-struktur i sitt land. CLARIN-sentrene skal inngå i den felles infrastrukturen og yte tjenester til fellesskapet, typisk ved å eksponere vel dokumenterte og tilrettelagte språkressurser i den felles infrastrukturen, eller ved å tilby spesielle infrastrukturtjenester som PID[3]-tjeneste, lagringsfasiliteter o.a.

Det nasjonale konsortiet i Norge kalles CLARINO[4] og er foreløpig et prosjekt støttet gjennom NFR-programmet INFRASTRUKTUR[5]. Deltakerne omfatter universitetene i Oslo, Bergen (koordinator), Trondheim og Tromsø, Handelshøgskolen i Bergen, Uni Research AS, UNINETT og Nasjonalbiblioteket. Nasjonalbibliotekets oppgaver er å

- tilby infrastrukturtjenester som
  - et nasjonalt metadataregister over de norske ressursene, med OAI/PMH-endepunkt slik at dataene kan høstes av andre. I dette ligger også å etablere hensiktsmessige metadataformater for beskrivelse av språkressursene, samt å støtte de andre partnerne i å lage gode metadata for sine ressurser.
  - langtidslagring for språkressursene i CLARINO
  - en PID-tjeneste

- inkludere Språkbankens ressurser og beskrive disse med hensiktsmessige metadata i henhold til de formater som anbefales i CLARINO

I denne artikkelen fokuserer vi på håndteringen av metadata og utarbeidelsen av metadataformatene som kan brukes til å beskrive språkressursene. Da må vi først si litt om hvilke typer data språkressurser egentlig er, dernest gjøre rede for teknologien som skal ligge til grunn for metadataene og metadataformatene utviklet i CLARINO.

SPRÅKRESSURSER – HVA ER DET EGENTLIG?

Språkdata/språkressurser utgjør et mangslungent landskap. Det kan omfatte digitale eller digitaliserte tekster, lyd- og videoopptak med transkripsjoner, ordbøker og konkordanser, historiske arkiver og datasett fra reaksjonstidsmålinger, dataverktøy for å analysere eller manipulere dataene regnes i seg selv som språkressurser.

3 Persistent Identifiers, se http://www.clarin.eu/content/persistent-identifiers
4 https://clarin.b.uib.no/
5 http://www.forskningsradet.no/prognett-infrastruktur/Forside/1224697900450

Dataene kan være rå data eller bearbeidet ved koding eller annotering.

Det verserer mange måter å kategorisere språkressurser på, men nedenstående er kategorier de fleste språkforskere har et forhold til:

- Korpus: Et korpus er en samling av brukt språk avgrenset/karakterisert på ulikt vis, for eksempel ved språk, medietype, språksituasjon, organisering, modalitet etc. Eksempler på slike er
  - Akustisk-fonetisk taledatabase for norsk – versjon 1.0 som inneholder innleste setninger på norsk, lest inn av personer med ulike dialekter. annotert på ord og fonemnivå (Språkbanken)
  - Nordisk dialektkorpus: talespråkskorpus med norske, svenske, danske, islandske og færøyske dialekter (Universitetet i Oslo)
  - Norsk dependenstrebank – Trebanker er tekstkorpus som er annotert syntaktisk og eventuelt morfologisk, og således danner syntakstrær. Norsk dependenstrebank består av to deler, en for nynorsk og en for bokmål. (Språkbanken)
  - Menota – Arkiv for nordiske middelaldertekster[6]: Inneholder digitaliserte middelaldertekster, kodet i henhold til TEI-standarden (Universitetet i Oslo)

- Leksikalske og konseptuelle ressurser: Dette er typisk samlinger av ord/begreper som er beskrevet på en strukturert måte, som ordbøker, leksika, terminologier, ontologier, tesauri, ordnett, etc. Eksempel:
  - Sametingets termsamling
  - Norsk ordnett
  - Frekvensordlister (1-grammer) basert på Norsk Aviskorpus til og med 2011.

- Verktøy og tjenester: Dette er applikasjoner, webservices etc utviklet for å analysere, bearbeide eller beskrive språkressurser på et eller annet vis, for eksempel stemmere, automatiske annotasjonsverktøy, søketjenester i korpus, rammeverk som organiserer andre verktøy i en arbeidsflyt, etc Eksempel:
  - Søkeverktøyet Glossa[7]

Mange språkressurser er komplekse data som er skapt i forskningsprosjekter og for spesielle formål, og i likhet med andre typer forskningsdata er de ofte

6 http://www.menota.org/tekstarkiv.xml
7 http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/glossa/

svært ressurskrevende å utvikle. Det har derfor lenge vært forsøkt å finne metoder og insentiver for å fremme gjenbruk av datasettene. Det er flere hindringer i veien:

- Å tilrettelegge sine data for gjenbruk krever betydelig arbeid ut over det man trenger til eget bruk, og med mindre det er satt av tid og penger til dette, blir det ofte vanskelig å få til.

- I akademia er det tradisjonelt bare publikasjoner som er meritterende. Man stiger ikke i gradene ved å legge til rette for andres forskning!

CLARIN/CLARINO er ett av mange initiativer som adresserer dette ved å beskrive ressursene med metadata og dermed gjøre dem tilgjengelige og ikke minst sitérbare.

METADATA FOR DATA

Som beskrevet ovenfor, utgjør språkressurser langt fra noen ensartet gruppe datasett. Det samme gjelder for så vidt *dokumenter*, men når slike skal beskrives kan vi støtte oss på en hundreårig bibliotektradisjon, og velprøvde – om enn omdiskuterte – metadataformater som etter hvert også har tatt inn over seg dokumenter på nye medier og innhold distribuert på nye måter.

Hvilke metadata som trengs for språkressurser er i utgangspunktet mindre avklart. Språkressurser er forskningsdata, og primærmålgruppen er forskere og utviklere, men også dette er en sammensatt gruppe. Utviklere av språkteknologiske løsninger (f.eks. talegjenkjenning, syntetisk tale, automatisk oversettelse), edisjonsfilologer, lingvister, datalingvister, historikere og samfunnsvitere kan ha svært ulike behov, både når det gjelder metadata og tilrettelegging av rådataene. Videre bør det tas høyde for en bredere sammensatt brukergruppe i framtida. Når ressursene og analyseverktøyene blir lettere tilgjengelige, kan det ligge til rette for bruk både i skolen og av den «interesserte allmennhet».

Likevel, metadata har samme funksjon for språkressurser som for andre informasjonsressurser – de skal beskrive ressursen slik at den enkelt kan gjenfinnes, og ikke minst må metadata inkludere alle opplysninger som er nødvendig for å avgjøre om dette er en brukbar ressurs i den foreliggende situasjon. Her er noen eksempler på behov en bruker kan ha:

- Finn språkressurser av alle typer, som er laget med tanke på meningsekstraksjon (opinion mining), men bare slike som tillater kommersiell bruk

- Finn opptak av andregenerasjons norskamerikanere utvandret fra Sørlandet som snakker norsk.
- Finn parallellkorpus på nynorsk og engelsk hvor annotering er utført manuelt (av mennesker)
- Finn opptak av tenåringsjenter som snakker et annet språk enn morsmålet sitt
- Finn alle termressurser hvor noen av de samiske språkene er representert

Det er flere metadataformater i bruk for å beskrive språkressurser, men alle med sin slagside. Etter vurdering av alle brukte/relevante formater ble det i CLARIN besluttet å gå en annen vei. I stedet for et fast metadataformat, ble det utviklet *rammeverk for utvikling og deling av komponentbaserte metadataformater*, kalt CMDI[8],[9](uttales «simdi»).

CMDI DELER METADATA INN I KOMPONENTER

Tre kjernebegrep er viktig for å forstå CMDI-rammeverket:

Profil: Det vi kan tenke på som et fullt metadataformat; en profil skal kunne brukes som skjema for å beskrive alle ønskede aspekter av de språkressurser det er designet for. Et eksempel er SpeechCorpusProfile[10] for beskrivelse av talekorpus. Profiler bygges opp av komponenter og/eller elementer.

Komponent: Et sett med opplysninger som til sammen beskriver et *aspekt* ved, *deler* av eller en *tilknyttet entitet* til språkressursen. Komponenter kan bestå av andre komponenter og elementer. Et eksempel er contakctPerson, som representerer ressursens kontaktperson, og inneholder felter (elementer) for for- og etternavn, epost, telefon, etc.

Element: Det vi kan tenke på som et metadatafelt, en atomisk egenskap. Elementer som ligger på profilnivå (direkte i profilen, ikke i en underliggende komponent), representerer typisk egenskaper ved ressursen som helhet, mens et element /felt i en komponent dypere i strukturen representerer en egenskap ved entiteten som omliggende komponent beskriver. Eksempelvis vil elementet etternavn i eksemplet over beskrive en egenskap ved kontaktperson, ikke ved språkressursen som sådan.

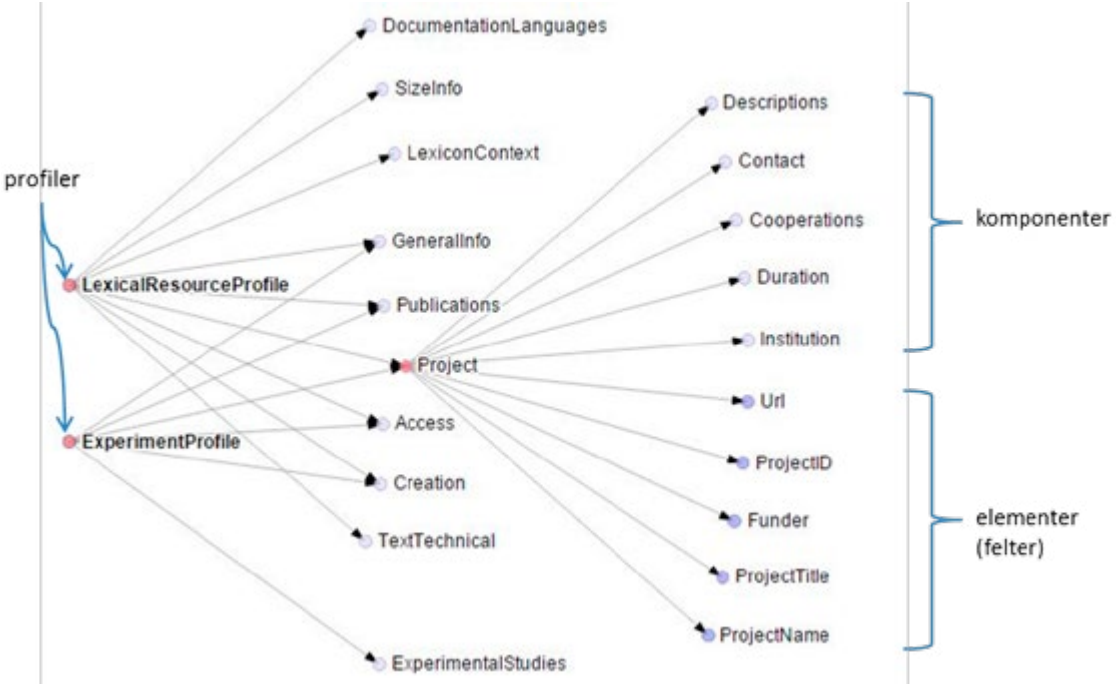
---

8 Component MetaData Infrastructure: http://www.clarin.eu/content/component-metadata

9 Broeder, D., O. Schonefeld, et al. (2011). A pragmatic approach to XML interoperability – the Component Metadata Infrastructure (CMDI). . doi:. Balisage: The Markup Conference. . Montréal, Canada, Balisage Series on Markup Technologies. **7**

10 http://catalog.clarin.eu/ds/ComponentRegistry?item=clarin.eu:cr1:p_1271859438166

Figur 1 viser et eksempel på to profiler, en for å beskrive leksikalske ressurser og en for å beskrive eksperimenter.



Figur 1 Profilene LexicaResourceProfile og ExperimentProfile, og nedbrytningen i komponenter på øverste nivå. Project-komponenten er ekspandert ett nivå lenger.

Figuren viser at hele seks komponenter er felles for de to profilene, - disse seks er altså *gjenbrukt* i flere profiler enn den de opprinnelig ble opprettet for. Komponenter og profiler forvaltes i CLARIN Component Registry (CCR)[11], hvor autoriserte brukere har et kladdområde til bruk under utarbeidelsen av komponent og profiler, før de publiseres og blir klare for gjenbruk. Det er også utviklet et grafisk navigasjonsgrensesnitt knyttet til CCR – nemlig SMC Browser[12], som viser hvordan de publiserte profilene, deres komponenter og elementer henger sammen. Figur 1 og 2 er laget i SMC Browser.

Med så stor fleksibilitet og frihet til å definere egne profiler, blir det svært mange felt i ulike komponenter som er tilnærmet synonyme. Samtidig er det viktig at alle metadataene innenfor CMDI-rammeverket er mest mulig

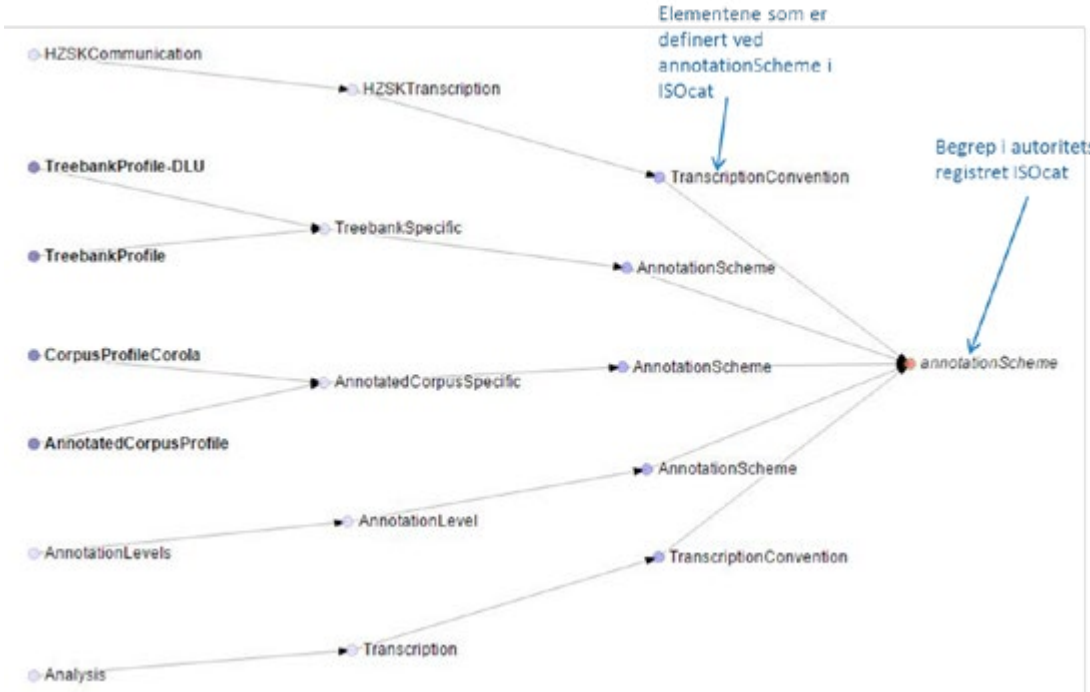
11 http://catalog.clarin.eu/ds/ComponentRegistry/#
12 Semantic Mapping Component Browser: http://clarin.oeaw.ac.at/exist/apps/smc-browser/index.html

semantisk interoperable, uavhengig av hvilke profiler og komponenter de er basert på. Problemet belyses best ved et eksempel: Det finnes flere komponenter som representerer informasjon om prosjekter, og de fleste av disse inneholder et felt for prosjekttittel. Men hvordan kan f.eks. en applikasjon – som skal håndtere metadata basert på ulike profiler –vite at feltene *projectName* i én komponent, *name* i en annen og *projectTitle* i en tredje står for det samme?

Måten dette er søkt løst på i CMDI er å *definere* elementene ved å knytte dem til et *autoritetsregister over begreper*. For tiden brukes ISOcat[13] til dette.

Hvert element skal altså knyttes til ett begrep ved å bruke PID-en til aktuelt begrep. Det tilknyttede begrepet indikerer hvordan elementet (og verdien av elementet i en konkret metadatapost) skal tolkes. Dersom det samme begrepet (samme PID) er knyttet til flere elementer i flere komponenter og profiler, betyr det at disse elementene og deres verdier skal tolkes på samme måte uavhengig av hva de kalles. I figuren nedenfor vises begrepet *annotationScheme*[14] fra ISOcat, som er knyttet til 6 elementer i flere ulike komponenter og profiler. Slik ser vi at elementene *AnnotationScheme* i komponenten *AnnotationLevel* og *TranscriptionConvention* i komponenten *Transcription* begge skal tolkes som *«Any systematically-documented descriptive or analytic notations applied to raw language data.»*, den ordrette definisjonen av begrepet *annotationScheme* i ISOcat.

13 Register over datakategorier brukt i språkvitenskapen: http://www.isocat.org/
14 http://www.isocat.org/datcat/DC-4085



Figur 2 Her vises hvilke elementer/felter som har betydning definert ved begrepet annotationScheme i ISOcat.

FLEKSIBILITETENS PRIS

CMDI-rammeverket er svært fleksibelt og rikt, ved hjelp av dette kan enhver leverandør av språkdata spesifisere en metadataprofil som passer med egne ressurser. Profilen kan enten spesifiseres fra bunnen av, eller helt eller delvis være basert på gjenbruk. Uansett skal knytningen til autoritetsregistret for begreper (ISOcat) sikre at metadataene blir interoperable med metadata skapt ved hjelp av andre profiler.

Dette er idealet, men i den virkelige verden er det mange utfordringer knyttet til denne modellen.

GJENBRUK VS. TILPASNING

CMDI-rammeverkets suksess er i stor grad avhengig av at deltakerne er innstilt på å gjenbruke komponenter, noe som igjen er avhengig av man tenker gjenbruk når man oppretter nye komponenter og profiler. Til nå har det sviktet på begge disse områdene, – i skrivende stund er det 162 publiserte profiler og 999 komponenter i CCR. Profilene er gjerne store og komplekse og verktøyene for innsikt og analyse er begrensede. Det krever derfor stor

innsats å vurdere eksisterende profiler for eget bruk. I tillegg er det flere som velger å ikke publisere sine profiler, for fritt å kunne endre på dem når som helst. Disse forblir skjult for det øvrige CLARIN-nettverket.

I praksis er det ofte slik at man finner en komponent som delvis passer, men at det er behov for endringer. I så fall må man kopiere komponenten og gjøre endringer på kopien. Det er jo også et slags gjenbruk, men fører til mange «nestenlike» komponenter.

ISOcat, som skal sørge for semantisk interoperabilitet mellom metadataene, går heller ikke fri for mangfoldssyken. Også der er det brukerne (ressurseierne) som definerer begrepene, og inspeksjon viser at det skorter til dels mye på evnen til å lage gode definisjoner. Ofte er definisjonene unødvendig spesifikke, i den grad at begrepene kun er brukbare i den kontekst de ble introdusert, noe som selvfølgelig reduserer interoperabiliteten.

CLARINOS TILNÆRMING

Virkelig gjenbruk av komponenter foregår i dag stort sett lokalt innenfor det enkelte miljø, for eksempel som vist i Figur 1, hvor begge profilene er utviklet av samme aktør. Dette er i stor grad også tilfellet i CLARINO. For å sikre at alle metadata fra CLARINO lar seg presentere i ett og samme grensesnitt, har vi sett det fornuftig å definere en felles kjerne i form av en egen komponent med generelle opplysninger. Denne er obligatorisk i alle profiler som skal brukes i CLARINO. I tillegg er det utviklet noen anbefalte profiler for de vanligste ressurstypene. Hver CLARINO-partner står likevel fritt til å bruke egne profiler, så lenge den obligatoriske komponenten er inkludert.

BIBLIOTEKENE KAN BIDRA

Utfordringene knyttet til manglende gjenbruk av metadatakomponenter er mye diskutert i CLARIN. Foreløpig er det tatt et organisatorisk grep, hvor hvert medlemsland utpeker noen som får et ekstra ansvar for begrepene som legges inn i ISOcat, samt å passe på at profilene og komponentene som utarbeides er rimelig fornuftige, og dessuten holde gjenbruks- og samarbeidsflagget høyt til enhver tid. I CLARINO er det Nasjonalbiblioteket som har dette ansvaret.

Et annet tiltak som er under gjennomføring er å flytte begrepsregistret over på en ny og enklere plattform basert på openSKOS[15], som i motsetning til ISOcat bare skal omfatte CLARINs begreper.

15 http://openskos.org/

På sentralt hold i CLARIN har det også vokst fram en erkjennelse av at bibliotekene har mye metadatakompetanse og i større grad enn nå bør involveres i arbeidet med språkressursene. I mange land utføres metadataarbeidet utelukkende av akademikere uten deltakelse fra institusjonens bibliotek. På CLARIN Annual Conference 2014 avsluttet direktøren med å skissere framtidige tiltak, og ett av disse var «Include libraries»!

Dette kan være en oppfordring til bibliotekene om aktivt å involvere seg i håndteringen av vitenskapelige data. Noen bibliotek i Norge har allerede sett en oppgave her, for eksempel Universitetsbiblioteket i Bergen, se s. 83. Forvaltning av forskningsdata er et relativt nytt område for bibliotekene, men det er ingen tvil om at bibliotekfaglig kompetanse, særlig om metadata, er av stor verdi på dette området.

Digitale vitenskapelige tekstarkiv – biblioteket i ny rolle

Universitetsbiblioteket i Bergen ønsker å være en mest mulig tjenlig infrastruktur for Universitetet i Bergen når det gjelder digitale informasjonsressurser. Flere forskningsmiljø, og biblioteket selv, merket vanskene med å opprettholde kompetanse for drift og utvikling av digitale tekstarkiv. Med dette som bakgrunn har vi utarbeidet og satt i permanent drift i biblioteket en ny teknisk støttefunksjon for digitale vitenskapelige tekstarkiv.

Karin Cecilia Rydving, seksjonssjef, Universitetsbiblioteket i Bergen

Rune Kyrkjebø, førstebibliotekar, Universtitetsbiblioteket i Bergen

TEKST: **KARIN CECILIA RYDVING OG RUNE KYRKJEBØ**

Humanistisk digital tekstbehandling er et felt med en lang historie i universitetsbyen Bergen. De siste tiårene har det vært aktive forsknings- og utviklingsmiljø i og rundt moderinstitusjonen vår. Arbeidet med XML[1]-koding av humanistiske tekster er særlig relevant for bibliotekfunksjonen vi her skal beskrive. På dette området har universitetet vårt spilt en viktig

1 http://www.w3.org/XML/

rolle. Det er bakgrunnen for at det i dag finnes viktige tekstressurser her som representerer et drifts- og støttebehov. Det kan nevnes at da Text Encoding Initiative (TEI) i år 2000 ble gjort til en internasjonal medlemsorganisasjon, var det med Universitetet i Bergen (UiB) som forslagsstiller sammen med University of Virginia. Disse universitetene ble de første konsortiepartnerne sammen med Brown University og Oxford University.[2] Det sentrale språk- og tekstteknologiske miljøet som er tett knyttet til Universitetet i Bergen, er den nåværende Computational Language Unit (CLU) ved Uni Computing (del av forskningsselskapet Uni Research).[3] Denne enhetens historie går tilbake i alle fall til 1973 da man startet NAVFs EDB-senter for humanistisk forskning. En betydelig aktør på feltet XML-kodede humanistiske tekster i Norge er ellers Eining for digital dokumentasjon (EDD) ved Universitetet i Oslo.[4] Dermed har vi grovt skissert noe av konteksten for vår søknad til Nasjonalbiblioteket i 2011 om et utviklings- og samarbeidsprosjekt kalt *Digitale Fulltekstarkiv* (DF-prosjektet). Nasjonalbiblioteket tildelte støtte til prosjektet. Prosjektet gikk over 2 år 2012–2013, se *Sluttrapport for prosjektet Digitale Fulltekstarkiv ved Universitetsbiblioteket i Bergen*[5] *for mer informasjon.*

BAKGRUNN: OPPARBEIDELSEN AV VITENSKAPELIGE XML-TEKSTRESSURSER

En bakgrunn for den nye biblioteksfunksjonen er også at Universitetsbiblioteket i Bergen (UB Bergen) på begynnelsen av 2000-tallet gjorde forsøk med framstilling av XML-kodede katalog- og dokumentvisninger for vårt eget spesialsamlingsmateriale. Støtte til et prosjekt i 2004 – 2006 for infrastruktur for diplommaterialet vårt fra årene mellom 1570 og 1660 kom fra Norges forskningsråd (NFR). Den tekniske siden av arbeidet hadde CLU sin forløper AKSIS/Unifob ansvaret for, og Historisk institutt ved vårt universitet deltok både i planlegging og utførelse. Det daværende Senter for middelalderstudier (CMS) ved UiB støttet og var med og utførte et annet prosjekt med mål om digitalisering og nykatalogisering av UB Bergen sine middelalderfragment med latinsk og norrøn tekst. Etter hvert ønsket UB Bergen selv å kunne være drifts- og vedlikeholdsmiljøet for disse ressursene.[6] [7]

2 http://www.tei-c.org/index.xml
3 http://www.computing.uni.no/units/clu
4 http://www.edd.uio.no
5 http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Prosjektoversikter/Avsluttede-prosjekt/Prosjekt-avsluttet-i-2013#Digitale
6 http://ub.uib.no/diplom/katalog/
7 http://ub.uib.no/fragment/list/

Behovet for å kunne ha en fungerende digital kurasjon av XML-tekstressurser ble etter hvert mer og mer følbart. For å ha et godt «etterliv» trenger ressursene en kontinuerlig oppmerksomhet både vitenskapelig og teknisk. Utviklingsprosjekter av tekstressurser har ofte, naturlig nok, fokusert sterkt på etableringen av selve ressursen, med vitenskapelig kvalitet, gjennomtenkt bruk av standarder, mål om god funksjonalitet med søk og visning, og ikke minst solid og kreativt programmeringsarbeid. Men selv om utviklingsfasen for en ressurs tar slutt, så tar ikke det tekniske behovet slutt. For å kunne vedlikeholde ressursen har det vist seg at standard IT-støtte ikke alltid er tilstrekkelig. Her ligger en pragmatisk hovedgrunn for vår nye bibliotekstjeneste: Dersom en vitenskapelig tekstressurs skal kunne vedlikeholdes trengs det en teknisk støttefunksjon. Dette merket vi når det gjaldt våre egne ressurser. Men viktigheten av funksjonen blir mye større når man tar i betraktning de andre tekstressursene som har akkurat det samme behovet. Signalene fra våre partnere i og rundt UiB var at man støttet ideen om denne funksjonen, og ønsket biblioteket velkommen til å gå inn i denne rollen.

FORSKNINGSBIBLIOTEKETS ARKIVANSVAR OG UNIVERSITETETS HÅNDTERING AV FORSKNINGSDATA

Bevaring og formidling av kulturhistoriske, faghistoriske, institusjonshistoriske arkiver og dokumenter er et ansvar som UB Bergen tar i sin strategiplan.[8] Ansvaret gjelder egne samlinger, men det fører også inn i aktivt samarbeid med fagmiljøene ved UiB med målet om å «styrke samarbeidet med fagmiljøene og imøtekomme deres behov for digitalisering, lagring og tilgjengeliggjøring av forsknings- og formidlingsarbeider». Den nye tekniske støttefunksjonen for digitale tekstarkiv er i samsvar med dette både en tjeneste internt i biblioteket og en tjeneste direkte overfor fagmiljøene utenfor– begge deler innenfor bibliotekets strategi.

Det har siden starten vært en motivasjon hos oss å komme videre i retning av nye verktøy for den digitale håndteringen av eget spesialsamlingsmateriale – manuskript, bøker og fotografi. Slikt arbeid er utviklingspreget, og utgjør i seg selv en god grunn til å ha teknisk kompetanse på huset. Den fysiske og digitale siden av arbeidet med arkivbevaring hører tett sammen, og for et større forskningsbibliotek blir ikke disse oppgavene mindre viktige i årene som kommer.

8 ttp://www.ub.uib.no/felles/dok/strategi/Strategisk-plan-UB-2010-2015.pdf

I et bredere bilde er det dessuten viktig at et forskningsbibliotek engasjerer seg i *datahåndtering* sammen med moderinstitusjonen. Sommeren 2012 publiserte den europeiske forskningsbibliotekorganisasjonen LIBER en rapport kalt «Ten recommendations for libraries to get started with research data management».[9] UB Bergen sin støttefunksjon til drift av XML-tekstarkiver sammenfaller med flere av punktene i denne autoritative anbefalingen. Dette handler om tekstressurser som samtidig er forskningsdata, og der intensjonen typisk er å beholde ressursene *dynamiske* også etter at de er etablert og prosjektperiodene er gått ut. Dette kan bare gjøres ved hjelp av en teknisk støttekompetanse som kan arbeide med XML, de relevante kodeskjemaene, og de nødvendige dataverktøyene.

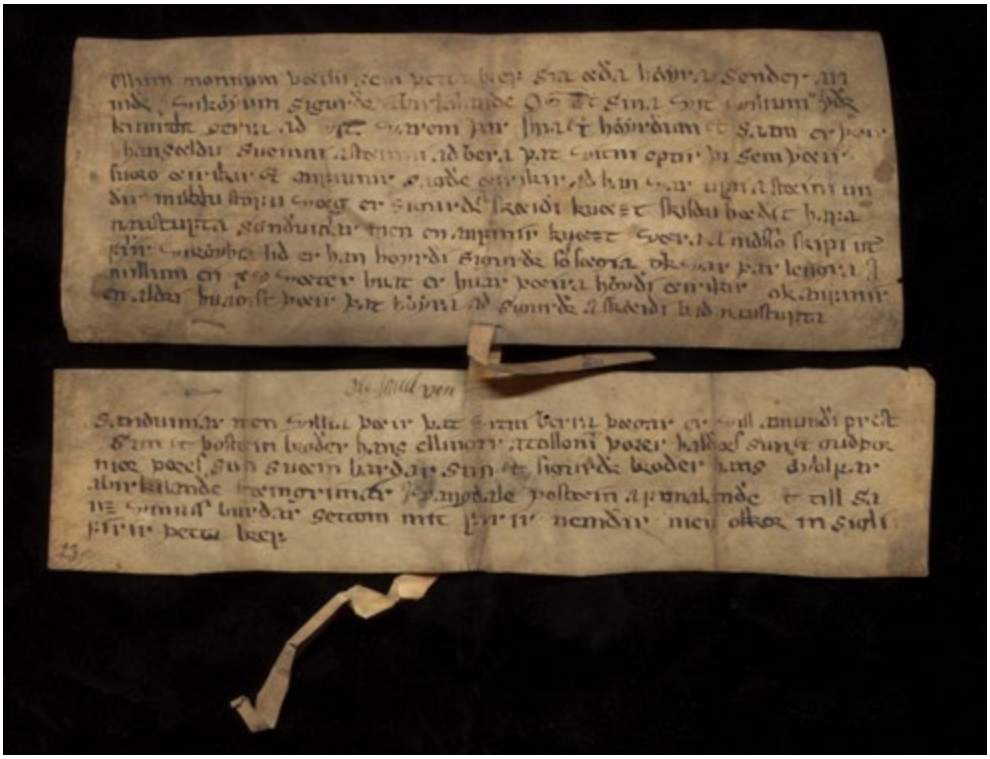
TJENESTEMODELL – TEKNISK OG ØKONOMISK

Modellen for vår tekniske støttefunksjon skiller mellom tre tjenestenivåer. Disse nivåene henger igjen sammen med finansiering, ut fra tanken om at utviklingsprosjekter i utgangspunktet trenger egne midler, mens daglig driftsstøtte er grunnbevilgningsfinansiert, i likhet med andre tjenester ved et universitetsbibliotek eller en IT-avdeling.

Tjenestenivå	Oppgave	Finansiering	Ansvar
Grunnleggende tjenester	Dag-til-dag vedlikehold, teknisk støtte. Enkel modifisering, justering av visningsformater, tillegg av nytt materiale	Grunnbevilgning fra moderinstitusjonen	Biblioteket ansvarlig
Mindre utviklingsprosjekt	Oppgradering, konvertering, sammenslåinger, videreutvikling	Egenbetaling fra fakultet eller institutt	Prosjektpartner ansvarlig
Større utviklingsprosjekt	Nyutvikling av ressurser og/eller verktøy, forskningsprosjekter	Ekstern finansiering	Prosjektpartner ansvarlig

Vi så ganske tidlig i prosjektet vårt at den grunnleggende støttefunksjonen begynte å fungere. CLU overførte til oss fragment- og diplomdataene som vi så håndterte selv, rett nok uten at vi klarte å videreføre all funksjonaliteten som var utviklet i de opprinnelige prosjektene. Prosjektpartneren vår Witt-

9 http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%20 2012%20v7%20final.pdf



Diplom (ca. 1320), Vikøy, (Kvam, Hordaland). Kilde: Spesialsamlingene, UB Bergen.

gensteinarkivet ved Universitetet i Bergen (WAB)[10] var fra starten av en aktiv bruker av støttetjenesten, og vi så at dette fungerte godt. Dette samarbeidet ble et godt eksempel på felles engasjement i forskningsdataene. UB Bergen tjente (og tjener) faglig på samarbeidet med WAB, særlig på grunn av felles deltakelse i prosjektet Digitised Manuscripts to Europeana.[11] Fem tusen sider av Wittgensteins *Nachlass* er nylig kommet med i Europeana som et resultat av dette prosjektet og av WAB sitt langvarige arbeid i Bergen med de grunnleggende XML-tekstdataene.[12] Den tekniske støttefunksjonen vår overfor WAB fortsetter. Ytterligere en arbeidspakke som har resultert i varige forhold, er samarbeidet med prosjektet Ludvig Holbergs Skrifter.[13] Her yter UB Bergen en teknisk redaktørfunksjon som bl.a. omfatter grunnleggende XML- og XSLT-tjenester.

10 wab.uib.no
11 dm2e.eu
12 http://wab.uib.no/wab_nachlass.page/
13 http://holbergsskrifter.no/

DATAMODELLERING OG DATABEHANDLING

I prosjektet som Nasjonalbiblioteket støttet var det også en arbeidspakke som handlet om å skissere oppgavefordeling og arbeidsmåte rundt datares-sursene. Vi tok utgangspunkt i en arbeidsfordeling der fagmiljøene er de som ivaretar vitenskapelig innhold og oppbygging av ressursen, og tar ansvar for videreutvikling. Grunnleggende IT-tjenester som serverdrift og backup er et naturlig ansvar for den sentrale IT-avdelingen. Universitetsbibliotekets rolle er da å sørge for langsiktig lagring og drift av de digitale tekstressursene, med egen kompetanse til å gjøre dette samtidig som det ytes teknisk støtte overfor brukerne og de faglig ansvarlige for ressursene.

For oss bekreftet prosjektperioden at denne oppgavefordelingen er rasjonell. Vi så at når en ressurs er ferdig utviklet og dokumentert, så er det ganske mye som kan gjøres innenfor det grunnleggende driftsnivået for å lette det videre arbeidet med ressursen. Samtidig ser vi at det er viktig å skille ut utviklingsoppgavene og være realistisk om at disse krever ekstra ressurser. Vi ser det slik at biblioteket med den kompetansen vi nå har bygget opp, kan være en partner og støttespiller i utviklingen av nye prosjekter, enten infrastrukturtiltak eller enkeltstående prosjekt.

ARBEIDSPAKKER

Etablere samarbeidsflate med WAB og støtte arkivet i drift, vedlikehold og digital formidling og visning

Utforme ny og dynamisk tjeneste med ulike typer materiale i et søkbart format for kunnskapsportalen GRIND

Overta det tekniske og faglige vedlikeholdet av diplom- og fragmentdatabasene

Utarbeide et bilde av hva som trengtes av tjenester og infrastruktur med tanke på framtidige drifts- og lagringsalternativer for ressursen Ludvig Holbergs Skrifter

Utrede framtidig løsning for drift og utvikling av digitale fulltekstarkiv

Se på institusjonelle arkiv som en del av framtidig infrastruktur

Det har utkrystallisert seg tre aktiviteter som er svært nyttige for vår nye bibliotekfunksjon: Datamodellering, datakonvertering, og aktiv bruk av åpne løsninger. Datamodellering er en viktig komponent siden det fort blir spørsmål om nettopp dette – for eksempel når eldre datasett skal korrespondere med nyere semantiske og strukturelle skjema. Eller når et datasett skal integreres med et annet. En aktuell problemstilling er å legge semantisk lenking (RDF / Linked Open Data) på både nyere og eldre dataressurser. Her forutsettes det å kunne arbeide både deskriptivt og normativt med datamodellene. Denne kompetansen ligger innenfor vårt konsept av den nye bibliotekfunksjonen. Datakonvertering er den neste komponenten. Det at vi på biblioteket selv kan omstrukturere datasett, for eksempel ved bruk av XSLT,[14] utløser mye framgang i arbeidet. Den nye bibliotekfunksjonen ville fort ha stoppet opp hvis vi ikke kunne gjøre vår egen databehandling, på metadata så vel som innholdsdata.

SAMARBEID MED FAGMILJØENE

Den nye tekniske støttefunksjonen har også gitt oss et tettere samarbeid med flere av de humanistiske fagmiljøene ved UiB. Dialogen har alltid vært til stede, men gjennom samarbeidet i prosjektet, og videre i drift, har biblioteket blitt en mer likestilt partner. Vi inviteres inn som prosjektdeltaker, samtalepartner i workshops og bidrar i publisering. Vi ønsker at denne utviklingen skal fortsette, og at vi skal få utvikle bibliotekets rolle som infrastruktur for universitetet i snittet mellom informasjonsbevaring, bruk og forskning. Arbeidet vårt med humanistiske tekstlige data er et eksempel på en naturlig rolle for forskningsbiblioteket, med en nærhet til dataene og en serviceoppgave overfor brukerne.

METODE OG VERKTØY I DF-PROSJEKTET

Vi bestemte tidlig i prosjektet at det som vi utviklet skulle videreføres som tjenester. En grunnleggende tanke var at data skulle åpnes, deles og gjøres tilgjengelige. Vi valgte også tidlig å mest mulig basere vårt eget utviklingsarbeid på Open Source-verktøy. Det viste seg at utviklingen av Open Source-muligheter raskt gikk fremover, og i løpet av prosjektet dukket det opp stadig bedre verktøy som vi kunne ta i bruk i forbindelse med den arbeidspakken vår som handlet om framtidige tekniske løsninger.

For å sikre at prosjektet skulle kunne fases over i drift, konstaterte vi at ny teknisk kompetanse måtte bygges og introduseres i biblioteket. Tidlig i

14 http://www.w3.org/Style/XSL/

prosjektet brukte vi en del relevant konsulentbistand for å etablere oss. Så ansatte vi XML-kompetent personale som viste seg å utfylle godt den sterke kompetansehevingen som prosjektet utløste hos andre medarbeidere. Vi opplevde at bibliotekets lange erfaring med metadata på denne måten ble løftet frem. Vi tror at det på feltet semantisk arbeid med tekstressurser og lenking kan ligge spennende muligheter for å bruke bibliotekenes tradisjonelle kompetanse på katalogisering og informasjonsorganisering.

I DF-prosjektet har vi arbeidet i tett dialog med våre brukere, og gjennomført hyppige leveranser. Nå i den tidlige driftsfasen for biblioteksfunksjonen fortsetter vi tjenesten overfor WAB, og vi er i et kontraktsforhold og yter lagrings- og driftstjeneste for prosjektet Ludvig Holbergs Skrifter (LHS). Når det gjelder LHS-dataene har vi Det Kongelige Bibliotek i København som samarbeidspartner.

VIRTUELLE FORSKNINGSMILJØ

Den tekniske støttefunksjonen vår er et eksempel på hvordan biblioteket kan bidra til å øke de digitale forskningsmulighetene. Dette er helt i tråd med Liber sine *Ten recommendations* som anbefaler at man bygger opp bibliotekkompetanse for å arbeide med forskningsdata, og at man gjør dette i partnerskap med forskere, forskergrupper, dataarkiver og datasentre. Ser vi på det humanistiske arbeidet med digitale tekster (*e-humanities*), kan det hevdes at dette er et felt der den tekniske kompetansen ikke er en adskilt støttefunksjon. Det er heller et gjensidig samarbeid med de vitenskapelige skaperne og brukerne av ressursene.[15] Dataarbeidet på driftssiden kan være med og åpne muligheter for nye forskningsspørsmål. Forskning og avansert programmering på sin side utvikler ideer som kanskje etter en tid blir nye tekniske muligheter på driftsnivået.

LINKED DATA

Prosjektet *Digitale Fulltekstarkiv* og prosjektet Linked Data – et nettverk![16] har også lagt grunnen for at vi nå utvikler en ny løsning – Marcus –for våre egne digitale spesialsamlinger. De gamle systemene var vanskelige og kostbare å vedlikeholde. De manglet også god funksjonalitet for å registrere og formidle våre samlinger.

15 http://balisage.net/Proceedings/vol13/html/Anderson01/BalisageVol13-Anderson01.html

16 http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Prosjektoversikter/Avsluttede-prosjekt/Prosjekt-avsluttet-i-2013#Linked%20Data - et nettverk!

Vi arbeider for å konvertere alt som er registrert av metadata i ulike system og format til Linked Data. Dette vil gjøre det mulig for oss ikke bare å knytte sammen alle våre egne samlinger og datasett, men også å dele våre data og bruke det som allerede eksisterer av Linked Data. Det gjenstår fortsatt arbeid med en teknisk infrastruktur, men vi har god progresjon. Løsningen og samarbeidsmulighetene, så vel mellom mennesker som mellom maskiner, som dette arbeidet gir oss, krever sin egen presentasjon. Vi håper å kunne fortelle mer om dette i en egen artikkel ved et seinere tidspunkt.

Vi slutter oss til LIBER sine anbefalinger og vil gjerne oppfordre til at man prioriterer arbeidet med aktuelle datasett i bibliotekene eller moderinstitusjonene – gjerne innen rammen av Linked Data. Alle er ikke forskningsbibliotek med fulltekstressurser som vi beskriver her, men de fleste har en eller annen form for datakilder som man ønsker å dele. Det er vårt overordnende budskap: Sats på arbeidet med egne data. Søk samarbeid.

FAKTA

Prosjektets tittel: Digitale fulltekstarkiv
Prosjektansvarlig: Universitetsbiblioteket i Bergen
Ansvarlig kontaktperson: Karin Cecilia Rydving
Prosjektstøtte fra Nasjonalbiblioteket: 2012: kr 930 000
2013: kr 780 000

Mål for prosjektet: Prosjektet har som mål å være med å bidra til at Universitetsbiblioteket i Bergen bygger opp sin rolle som felles infrastruktur for vitenskapelig oppbygde og forskningsrelevante elektroniske fulltekstarkiv ved Universitetet i Bergen.
Samarbeidspartnere: Wittgensteinarkivet og prosjektet Ludvig Holbergs Skrifter, begge ved Det humanistiske fakultetet, Universitetet i Bergen og Uni Computing
Innsatsområde: Samarbeid

«Linked data–et nettverk!»

«Linked data – et nettverk!» var et samarbeidsprosjekt mellom NTNU Universitetsbiblioteket (NTNU UB), Universitetsbiblioteket i Bergen (UB Bergen) og Bergen offentlige bibliotek som ble gjennomført i 2012 og våren 2013.

TEKST: **LENE BERTHEUSSEN**

I denne perioden var interessen for Linked data og tilhørende teknologier stadig økende, og ideen til prosjektet oppstod på en felles workshop i forbindelse med et tilsvarende prosjekt som var satt i gang ved NTNU UB.

Sammen så vi at vi hadde mye materiale som omhandlet samme hendelser/personer spredt på de ulike bibliotekene, samtidig som alle jobbet med å formidle dette til publikum på nye måter.

Alle som deltok ønsket å teste mulighetene som Linked data gir til å virkelig kunne vise sine unike ressurser til publikum i en større og mye mer givende sammenheng. Vi så også at vi hadde mye å lære av hverandre.

Vi tok tak i ressurser fra Sangerfestene i Bergen og Trondheim (1863 og 1883). Dette materialet inkluderte brev mellom kjente personer, manuskripter, bilder og kart samt noter tilknyttet Sangerfestene. Materialet ble digitalisert og ved hjelp av Linked data-teknologier fremstilt slik at det lettere kunne benyttes av publikum.

NETTVERKSBYGGING MED LINKED DATA I FOKUS

Lene Bertheussen, universitetsbibliotekar, NTNU, Universitetsbiblioteket
Foto: Nils Kristian Eikeland NTNU

Målet med prosjektet ble primært å skape et nettverk i fagbibliotekmiljøet med Linked data i fokus. Vi kunne øke vår kunnskap ved å lære av hverandre, dele kunnskap og erfaringer samt få nye erfaringer og sammen opparbeide oss den kunnskapen vi manglet.

Dette kunne vi så benytte for å vise Sangerfestene som en enhetlig hendelse til publikum på alle tre institusjonene.

Et viktig punkt var å skape en felles forståelse av prosesser, begrep og metoder vi benyttet oss av. Vi ønsket også å få en kompetanseheving som ville gjøre oss i stand til- og kanskje gi oss muligheten til å omstille oss til ny teknologi, nye verktøy og arbeidsmetoder.

Første treff med hele prosjektgruppa bestod i en workshop med en ekstern aktør som fortalte grunnleggende om Linked data og noen måter å arbeide med dette på. Det førte til at alle startet med samme utgangspunkt, og at vi i den grad det manglet, fikk et felles grunnleggende vokabular og forståelse for det vi skulle arbeide med. Deretter hadde vi flere workshoper med ulike aspekter ved temaet etter som vi så hva vi trengte.

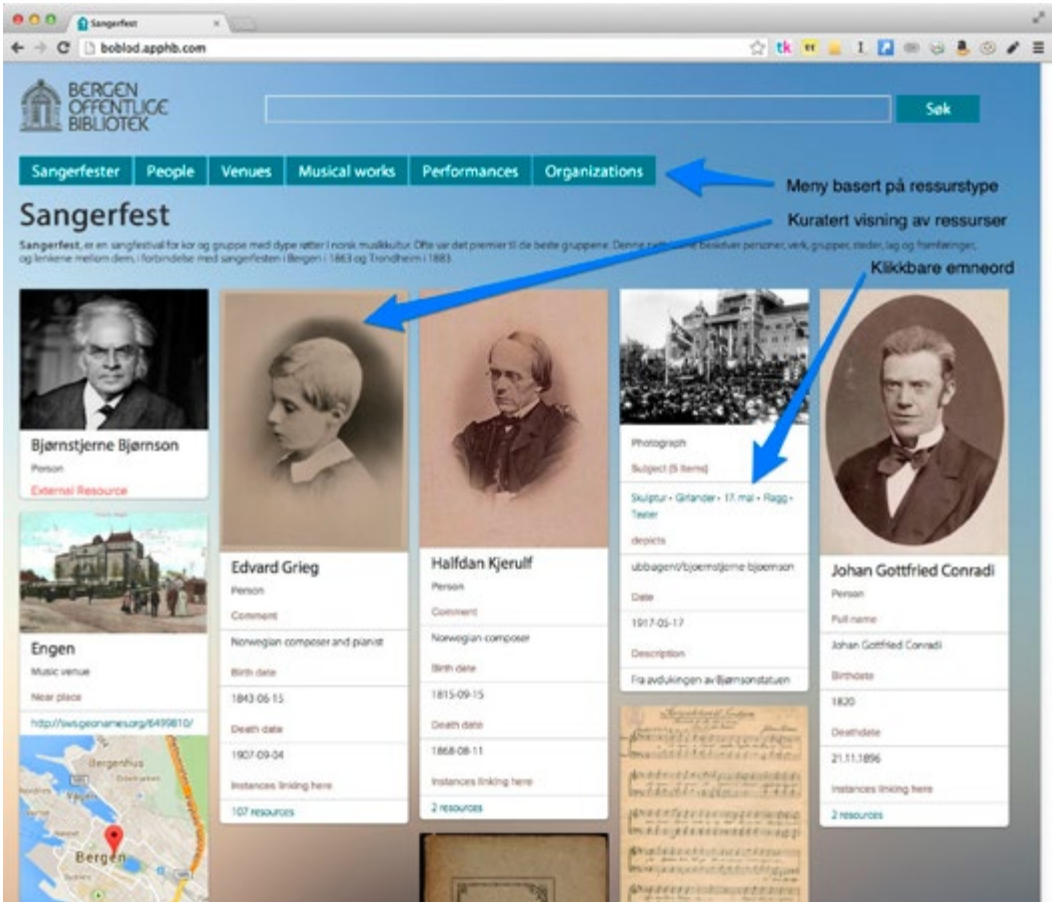
RESULTATER

Det mest synlige resultatet var en demo hvor vi kunne vise arbeidet vårt. Det gav oss også konkrete erfaringer med utfordringer knyttet til bruk av data fra ulike kilder. Demoen undersøkte muligheter og begrensinger med tanke på brukeropplevelser og publisering av data i spesialsamlinger.

Verktøyet inneholdt et enkelt CMS[1] som håndterer linked data, slik at bibliotekarer og arkivarer lett kunne sette opp nettsider for samlingene sine og velge ut hvilket innhold som skulle framheves.

Hovedansvaret for demoen ble lagt til Bergen offentlige bibliotek fordi de ved det tidspunktet, jobbet med en visualiseringsdemo i forbindelse med prosjektet «Bergens musikkhistorie». Til tross for samarbeid valgte alle de tre institusjonene på sikt ulike verktøy og ulike eksterne samarbeidspartnere.

1 Content management system: Et system for å redigere, lagre og publisere informasjon, i vårt tilfelle på web.



Skjermbilde fra demo: Sangerfest i kontekst.

Disse valgene tok vi med bakgrunn i erfaringer fra prosjektet samt muligheter for ressurser og forankringer ved de ulike institusjonene.

Alle de involverte har lært mye av dette samarbeidet. Nettverket har blitt større fordi vi også har knyttet kontakter med interesserte miljø og prosjekt på hver vår kant.

Medlemmene i gruppa ble godt kjent med hverandre og har kunnet spille på hverandre både i prosjektet og ellers der det var nødvendig eller lurt. Vi har kommunisert og delt vårt videre arbeid med hverandre og beholdt kontakten om ikke annet for å ha noen å dele frustrasjoner med.

Skulle mulighetene by seg, samarbeider vi gjerne igjen ved en senere anledning.

TING TAR TID!

Vi har erfart at det er vanskelig å forutse hvordan ressurser best bør utnyttes. Vi har endret posteringer av økonomiske ressurser underveis og brukte mer på eksterne konsulenter enn det vi først så for oss. Vi så også at det med fordel kunne ha vært satt av mer tid til egeninnsats. Det var mer tidkrevende å tilpasse seg ny teknologi og tankemåte enn først antatt. Prosjektet ble derfor forlenget noen måneder.

Hver institusjon har varierende muligheter for valg av og overgang til ny teknologi. Vi har ulike tilganger på ressurser og støtte fra våre enheter. Det er viktig å ha en tydelig forankring innad i institusjonen før et slikt prosjekt settes i gang.

Ting tar tid! Vær realistisk til hvor lang tid arbeidet tar. Det er ikke nødvendigvis et nederlag hvis arbeidet må endres eller tidsrammen forlenges for å nå ønskede mål. Selv om man kanskje må forkaste ideer, er tiden man bruker på forbedring av egne data alltid verdt innsatsen.

Med endringer i våre rammer opplevde vi Nasjonalbiblioteket som forstående og imøtekommende til våre ønsker underveis i prosjektet.

Prosjektet har bidratt til å øke fokus og bevisstheten på mulighetene Linked data teknologier og den endring i tankegang dette fører med seg. Det trengs derimot fortsatt en del innsats før metodikker og erfaringer fra denne typen prosjekt blir en del av de daglige arbeidsrutinene.

Vi ser at vi til stadighet støter på de samme problemene. Dette var blant annet problemstillinger knyttet til arbeidsprosesser, generelle driftsspørsmål og innhenting av kompetanse versus opplæring av intern kompetanse. Samarbeidet har vært givende, ikke minst fordi det gir ideer til også andre utviklings- og driftsoppgaver vi kan samarbeide om.

Institusjonenes daglige drift ble ikke påvirket i vesentlig grad, men det satte fart i endringen av fastgrodde tankemønstre og påvirket arbeidet for flere enkeltpersoner som var involvert. Det bidro også til at vi tenker annerledes på måten vi behandler data på fra grunn av. Det ville for eksempel være interessant å se forholdet mellom hvordan vi jobber med data og hvordan brukerne utforsker de samme dataene.

For hver av de involverte partene er det satt i gang endringsprosesser som vi ser har gjort seg gjeldende i større eller mindre grad. Dette innebærer alt fra interne arbeidsoppgaver og rutiner til integrasjon mot ulike systemer og leverandører, teknologivalg og strategi. Ved UB Bergen har kompetansen og erfaringene fra prosjektet vært viktige i utviklingen av nye tjenester gjennom prosjektet Digitale Fulltekstarkiv.

For mer informasjon om prosjektet henvises det til prosjektrapporten[2]. Vi oppfordrer alle å jobbe med å forbedre egne data og dele autoritetsdata som åpen linked data. Det vil forenkle arbeidet i kommende Linked data-prosjekt.

FAKTA

Prosjektets tittel: Linked data – et nettverk!
Prosjektansvarlig: NTNU Universitetsbiblioteket
Prosjektleder: Stein Johansen
Varighet: Ettårig 2012
Prosjektstøtte fra Nasjonalbiblioteket: kr 500 000

Mål for prosjektet: Å etablere og utvikle et nettverk innenfor «Linked Open Data» med utgangspunkt i aktive miljøer ved Berge Offentlige bibliotek, NTNU Universitetsbiblioteket og Universitetsbiblioteket i Bergen. Et viktig mål er å initiere kompetanseutvikling og omstilling knyttet til ny teknologi og nye og mer brukertilpassede arbeidsmåter i bibliotekene.
Innsatsområde: Samarbeid

2 http://tinyurl.com/obc8shm

Deichmanske bibliotek sier «Morn'a MARC!»

Deichmanske bibliotek har nylig gjort noen spennende valg for framtidig kurs. Det ene er å gå bort fra det biblioteksystemet som biblioteket har basert drift og tjenester på i flere tiår, og i stedet å gå for en ny løsning bygd rundt fri programvare-alternativet Koha. Det andre er å gå bort fra MARC som registreringsformat for katalogdata, og i stedet satse på linked data i RDF-format.

TEKST: **ASGEIR REKKAVIK**

Asgeir Rekkavik, førstekonsulent, Deichmanske bibliotek. Foto: Gyda Kjekshus

På Deichman har vi i flere år eksperimentert med utvikling av digitale sluttbrukertjenester som prøver å utnytte katalogdata til formidlingsformål. En av erfaringene vi har gjort, er at eksisterende katalogiseringspraksis, -regler og -formater ofte kan være til hinder for en slik bruk av bibliotekets data. Dette er betenkelig, for når stadig flere brukere i stadig større grad møter bibliotekene og bibliotektilbudet gjennom digitale grensesnitt, må bibliotekene ha muligheten til å tilby tjenester digitalt. Det er påfallende hvordan produksjon av metadata, som er et av biblioteksektorens ekspertisefelt og noe bibliotekene bruker mye ressurser på, kan resultere i data som er lite egnet for publikumsrettede digitale tjenester. Årsaken ligger i tankesett og standarder som er formet for en annen tidsalder og ikke tilpasset nye behov, og i treghet i utforming og implementering av nye løsninger.

Katalogisering handler tradisjonelt om å gjøre dokumenter i bibliotek-samlingen gjenfinnbare ved hjelp av oppslagsord. For noen tiår siden var gjenfinningsteknologien basert på fysiske kortkataloger, der kortene var organisert alfabetisk etter utvalgte oppslagsord. De øvrige opplysningene på kortene tjente ikke som oppslagsmuligheter, men bidro til at brukeren kunne identifisere riktig dokument. Med digitale bibliotekkataloger har vi fått flere muligheter til å bruke mer av dataene til søk og avgrensing, men tankegangen er i bunn og grunn den samme som før: å representere dokumentene i samlingen med oppslagsord. Både katalogiseringsregler og dataformater er gjennomsyret av et fokus på tekst, navn og termer. Det er interessant å se på terminologien som brukes i kunnskapsorganisasjonsfaget: Vi registrerer ikke emner, men *emneord*. Vi registrerer ikke forfattere, men *personnavn* som *hovedordningsord*. Vi registrerer ikke oversettere, illustratører og redaktører, men biinnførsler på *personnavn* med *funksjonsbetegnelser*. Det er ikke vanskelig å forstå hvorfor det har blitt slik. Siden formålet med metadata i bibliotek tradisjonelt har handlet om oppslag, søk og gjenfinning, er det først og fremst oppslagsord og søketermer vi har hatt behov for. Men når vi etter hvert ønsker å tilby brukerne noe mer enn rene søketjenester, og når metadata skal utgjøre et grunnlag for tjenester som sikter mot formidling, utforskning og oppdagelse, så kommer slike tekstorienterte data til kort.

Katalogiseringen i bibliotekene må bytte ut det fokuset på tekst, navn og ord som vi har i dag, med et fokus på de tingene som ordene beskriver. Vi må ha data som identifiserer og representerer personer, steder, verk, konsepter, objekter og hendelser. Dataene må si noe om hva disse tingene er, hvordan de relaterer seg til hverandre og hvor det finnes mer informasjon om dem. I stedet for lukkede baser av søketermer, trenger vi data som åpner opp til og henger sammen med det informasjonsuniverset som finnes utenfor bibliotekkatalogene, slik at vi får tilgang til og kan dra nytte av den informasjonen som finnes tilgjengelig andre steder.

Løsningen i Deichmanske biblioteks prosjekter har vært å konvertere MARC-poster til linked data i RDF-format. Dette er en teknologi som uttrykker og lagrer informasjon som et nettverk av data, hvor tingene som bekskrives representeres med unike identifikatorer mens relasjonen mellom dem uttrykkes med meningsbærende lenker. Selv når utgangspunktet for dataene er MARC-poster, gir konvertering til RDF med påfølgende viderebehandling og berikelse av de konverterte dataene, betydelig merverdi. For eksempel har det gjort det mulig å introdusere instanser som representerer verk i

Deichmanske bibliotek. Foto: Norsk design- og arkitektursenter

katalogen, slik at vi nå kan gruppere titler som er utgaver av samme verk. Dette er avgjørende for gode formidlingstjenester, men til tross for at problemstillingen er godt kjent og grundig behandlet i bibliotekverdenen over flere tiår, gir det anvendte katalogiseringsformatet ingen muligheter til å uttrykke slike sammenhenger mellom poster. Med linked data har vi fått en katalog med økt fleksibilitet og uttrykksfullhet og helt nye muligheter til å anvende egne og andres data.

Erfaringene med RDF og linked data på Deichmanske bibliotek har vært overbevisende. Med bytte av biblioteksystem ser vi derfor ikke lenger noen grunn til å holde fast ved MARC som registreringsformat.

FAKTA

Prosjektets tittel: marc2rdf – konvertering av MARC-data
Prosjektansvarlig: Oslo kommune kulturetaten Deichmanske bibliotek
Ansvarlig kontaktperson: Benjamin Rokseth
Varighet: Ettårig 2013
Prosjektstøtte fra Nasjonalbiblioteket: kr 310 000

Mål for prosjektet: Utvikle et gjenbrukbart rammeverk for konvertering av bibliografiske MARC-data fra bibliotekkataloger til RDF
Innsatsområder: - Samarbeid og partnerskap – Ny formidling

Linked data som verktøy for attraktive brukertjenester i musikkbibliotek

Bergen offentlige bibliotek fikk i 2012 prosjektmidler fra Nasjonalbiblioteket til å prøve ut Linked Open Data i formidlingsarbeidet av materialet fra Bergen Filharmoniske Orkester (BFO), som fyller 250 år i 2015. Vårt ønske er å finne gode metoder for å knytte mye av dette materialet sammen slik at både folk og forskere enkelt og brukervennlig finner materialet de trenger, på sitt nivå.

Siren Steen, avdelingsleder, Bergen offentlige bibliotek.
Privat foto

TEKST: **SIREN STEEN**

METADATA OG LINKED OPEN DATA

Målet med prosjektet er å gjøre bibliotekets metadata for spesialsamlinger fritt tilgjengelig som et datasett basert på Linked Data og formidle på beste måte innholdet som beskrives. Prosjektet skal utforske muligheten for å presentere eget og andres materiale på nye måter som viser sammenhengen mellom det digitale objektet og andre dokumenter i egen og andres samlinger. Prosjektet skal utvikle rike metadata ved å knytte sammen nettkilder med katalogdata, og tilgjengeliggjøre disse i åpne format som andre kan bruke og bygge videre på.

Det har vært jobbet mye med prototyping av søkegrensesnitt samt utvikling av presentasjons- og formidlingsgrensesnitt i samarbeid med eksterne spesialister på brukergrensesnitt og mobile flater. Å finne et fullgodt grensesnitt har vist seg å være en stor utfordring, og arbeidet med dette pågår ennå.

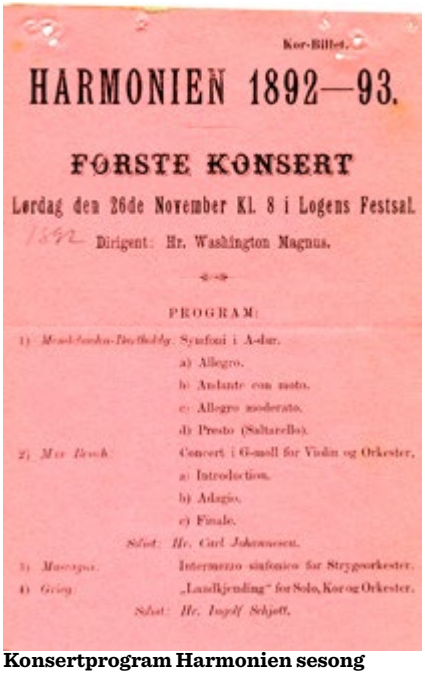
Mange praktiske aspekt ved Linked Open Data er ikke løst, eksempelvis vil bruk av ulike vokabular by på problemer når en skal samkjøre data i en webapplikasjon.

ARBEIDSMETODER

Prosjektet har internt vært et samarbeid mellom personale fra bibliotekets digitale avdeling og musikkavdelingen. Arbeidet med datamodellen har krevd sammensatt kompetanse med kunnskap om modellering og webteknologi på den ene siden og kunnskap om musikk og musikkregistrering på den andre. Det har vært gjort mye registrering for å prøve ut datamodeller. Dataene er konvertert med Google Refine[1]). Vi har opplevd at den største utfordringen har vært fravær av en god registreringsklient, det har derfor gått mye tid til strukturering og modellering av metadata.

Fra Musikfesten i Bergen i 1898. Christian Cappelen, Catharinus Elling, Ole Olsen, Gerhard Schjelderup, Iver Holter, Agathe Backer Grøndahl, Edvard Grieg, Christian Sinding, Johan Svendsen og Johan Halvorsen. Foto: Troldhaugen

1 http://code.google.com/p/google-refine/



Kor-Billet.

HARMONIEN 1892—93.

FØRSTE KONSERT

Lørdag den 26de November Kl. 8 i Logens Festsal.

Dirigent: Hr. Washington Magnus.

PROGRAM:

1) Mendelssohn-Bartholdy: Symfoni i A-dur.
a) Allegro.
b) Andante con moto.
c) Allegro moderato.
d) Presto (Saltarello).

2) Max Bruch: Concert i G-moll for Violin og Orkester.
a) Introduction.
b) Adagio.
c) Finale.
Solist: Hr. Carl Johannesen.

3) Mascagni: Intermezzo sinfonico for Strygeorkester.

4) Grieg: „Landkjending" for Solo, Kor og Orkester.
Solist: Hr. Ingolf Schjøtt.

Konsertprogram Harmonien sesong 1892-93, 26.november.

I prosjektet har det blitt utviklet en demo for å få konkrete erfaringer med utfordringer knyttet til bruk av data fra ulike kilder. Løsningen er utviklet i samarbeid med Fludo, et lite gründerselskap som spesialiserer seg på visualisering på mobile flater. Deler av applikasjonen er utviklet i samarbeid med prosjektet «Linked data - et nettverk!» som Bergen offentlige bibliotek deltok i i 2012/13, sammen med Universitetsbiblioteket i Bergen og NTNU Universitetsbiblioteket.

INFRASTRUKTUR, MODELLERING OG KONVERTERING AV DATA

Fokus på åpne bibliotekdata de siste årene har i hovedsak vært rettet mot infrastruktur, modellering og konvertering av data. Dette har vært nødvendig for å tilgjengeliggjøre dataene, men lite oppmerksomhet har vært rettet mot hvordan man kan lage attraktive tjenester og grensesnitt for brukerne, basert på dataene. Vi har utviklet en demo for å undersøke muligheter og begrensinger med tanke på brukeropplevelse og publisering av data i spesialsamlinger. I demoen er det mulig å dukke ned i ressursene i et «single-page» grensesnitt beriket med bilder, kart og tidslinjer, som er tilpasset for å kunne vises på ulike flater som mobil, nettbrett og skjerm. Løsningen inneholder også et verktøy for å sammenstille ulike data og et søk med mulighet for å filtrere resultatene.

Erfaringene fra arbeidet med grensesnittet bekrefter en del av antagelsene om fleksibiliteten som RDF og trippel-strukturen gir, og åpner opp for en ny måte å tenke på når det gjelder katalogisering, lenking og publisering av informasjon. Erfaringene har også gitt oss innsikt i de ulike utfordringene knyttet til å bygge tjenester på åpne data, spesielt på hvordan data fra ulike kilder kan utforskes i ett og samme grensesnitt.

NOEN FUNKSJONER

Demoen inneholder blant annet disse funksjonene:

- Søk, med filtrering basert på type og tag
- Visning av eksterne data fra andre kilder
- Trippelbasert CMS for oppretting og visning av samlinger

- Lenke til å editere ressurser
- Opphavsinformasjon på ressursene
- «Drilldown» i datasettet og eksterne datasett
- Mulighet for fremheving av spesielt interessant innhold
- Foto, tidslinje og kartvisning

Demoen er planlagt klar for visning våren 2015.

De fleste konsertprogrammene til BFO fra 1860-tallet er skannet, og for at disse skal være tilgjengelig mens vi jobber med registrering og publisering av tilhørende metadata, har vi publisert dem på Issuu[2].

Det er foreløpig bare årgangene fra 1939-1951 som er dataregistrert iht prosjektets mål.

Slik vi ser det knytter det store paradigmeskiftet i bibliotekverden seg til endring i, og spredningen av, metadata. I Norge trenger vi mange eksperimenter rundt de ulike fagfeltene for å finne optimale løsninger på dette og vi støter i ulike sammenhenger på behovet for mer kunnskap om Linked Data, og håpet er at vi ved enden av prosjektet våren 2015 kan gi et bidrag i feltet.

FAKTA

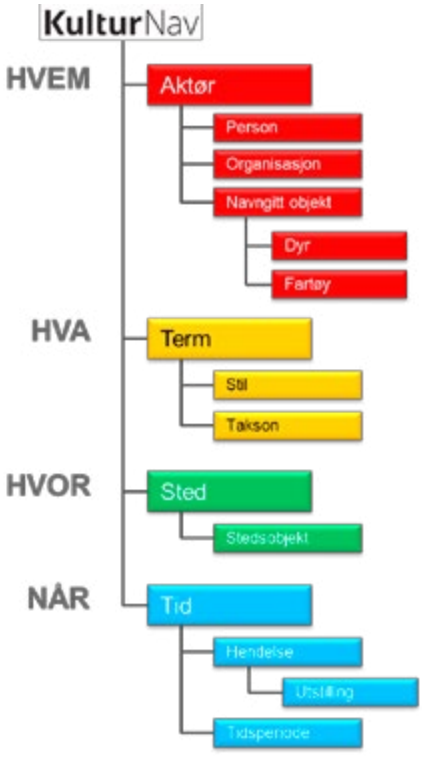
Prosjektets tittel: Bergens musikkhistorie
Prosjektansvarlig: Bergen offentlige bibliotek
Ansvarlig kontaktperson: Siren Steen
Prosjektstøtte fra Nasjonalbiblioteket: 2012: kr 375 000
2013: kr 200 000
2014: kr 175 000

Mål for prosjektet: Målet med prosjektet er å gjøre bibliotekets metadata for spesialsamlinger fritt tilgjengelig som et datasett basert på Linked Data og formidle innholdet som beskrives på en best mulig måte. Prosjektet skal utforske muligheten for å presentere eget og andres materiale på nye måter som viser sammenhengen mellom det digitale objektet og andre dokumenter i egen og andres samlinger. Prosjektet skal utvikle rike metadata ved å knytte sammen nettkilder med katalogdata og tilgjengeliggjøre disse i åpne format som andre kan bruke og bygge videre på.
Innsatsområde: Ny formidling

2 http://issuu.com/bergenbibliotek

KulturNav – et nettsted for autoriteter og felles terminologi om kulturarv

TEKST: **ULF BODIN OG TOVE WEFALD PEDERSEN**



I bibliotekssektoren har det i lang tid vært brukt kontrollerte og etablerte emneords- og autoritetslister i katalogiseringen. På mange måter arbeider museene i Norden også slik i sine samlingsforvaltningssystem, men museene har ikke hatt samme grad av koordinering når det gjelder å bruke felles ressurser. Med unntak av standarden Feltkatalogen (2002) [1] som handler om retningslinjer for å sikre at metadata om samlingene katalogiseres på en standardisert måte, har hvert museum bygget opp og forvaltet metadata. Museene har sine egne termlister og kontrollerte lister over personer, organisasjoner, steder osv.

Måten museene og andre kulturarvsinstitusjoner arbeider med metadataene på vil endre seg når de nå får tilgang til en ny løsning – Kultur-

1 Feltkatalog for kunst- og kulturhistoriske museer, 2002. Kulturrådet/NMU.

NAV[2] - som støtter arbeidet med felles lenkede åpne data (LOD). Løsningen lanseres i januar 2015, men har vært i skarp betadrift siden september 2013. KulturNav er et nettsted og en teknisk plattform (*software as a service*) for å skape, forvalte og distribuere felles flerspråklig, åpen terminologi og autoriteter med spesielt fokus på kulturarvdata i Norden. Utviklingen er finansiert av Norsk Kulturråd og ble påbegynt i 2011. Løsningen utvikles og driftes av KulturIT ANS, et selskap som eies av Norsk Folkemuseum og Lillehammer museum.

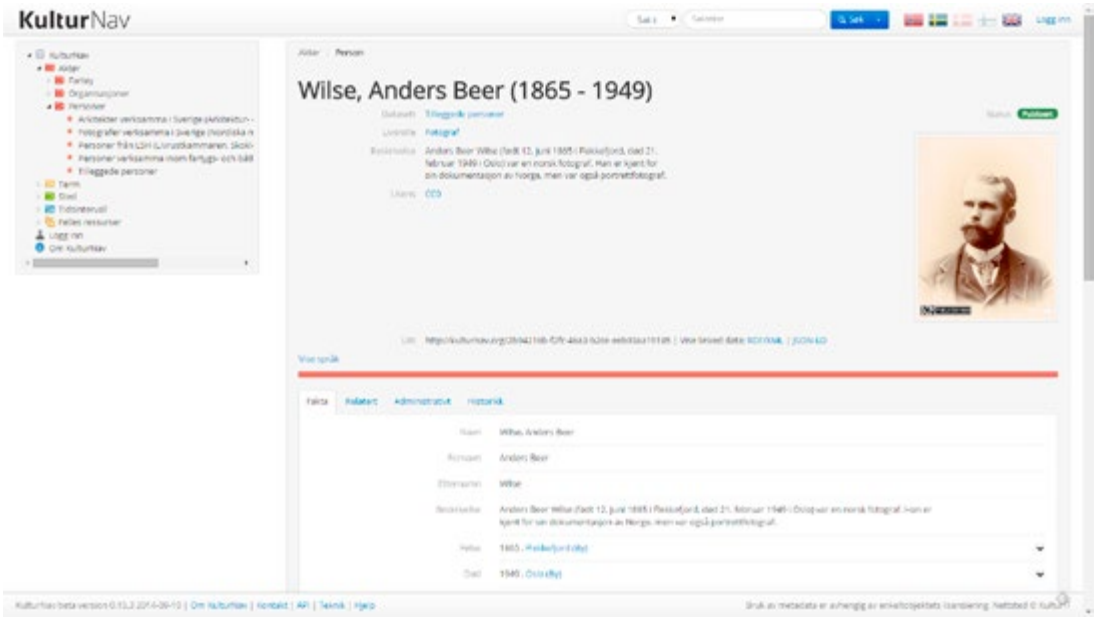
Ulf Bodin, konsulent, Kultur IT. Foto: Kultur IT

Tove Wefald Pedersen, avdelingsleder, Kultur IT. Privat foto

Løsningen håndterer kulturkonsepter som f.eks. benevnelser, materialer, teknikker, emneord og klassifikasjoner, samt autoritetsposter for personer, organisasjoner, fartøy, hendelser, tidsperioder,administrative inndelinger og stedsnavn. KulturNav gir hvert konsept/autoritet en persistent URI (*Universal Resource Identifier)*. Terminologi og autoriteter samlet på ett og samme sted kan brukes direkte ved katalogisering for å øke kvaliteten på informasjonen om samlingene. Dette gjør dataene enklere å gjenfinne og bruke både internt i organisasjonen og for publikum. Løsningen har direktekobling til samlingsforvaltningssystemet Primus som brukes av nærmere 150 museer i Norge, og koblinger mellom objekt og konsept/autoritet, f.eks mellom fotografier i museets samling og fotograf. Autoritetsposten for fotografen vil også være tilgjengelig i DigitaltMuseum[3].

Nettstedet er utviklet for å støtte arbeidet med forvaltningen av data på tvers av organisasjoner og fagområder. Det finnes også verktøy som gjør det mulig å knytte til seg brukere med fagekspertise utenfor sektoren for å komplettere og forbedre innhold i metadatene. Brukerne kan foreslå nye poster og komme med forslag til endringer og tillegg. Informasjonen er ordnet i datasett. Det enkelte datasettet har administrative funksjoner om forvaltning og tilhørighet, med en ansvarlig organisasjon som følger opp og således gir datasettet autoritet.

2 http://kulturnav.org
3 http://digitaltmuseum.no



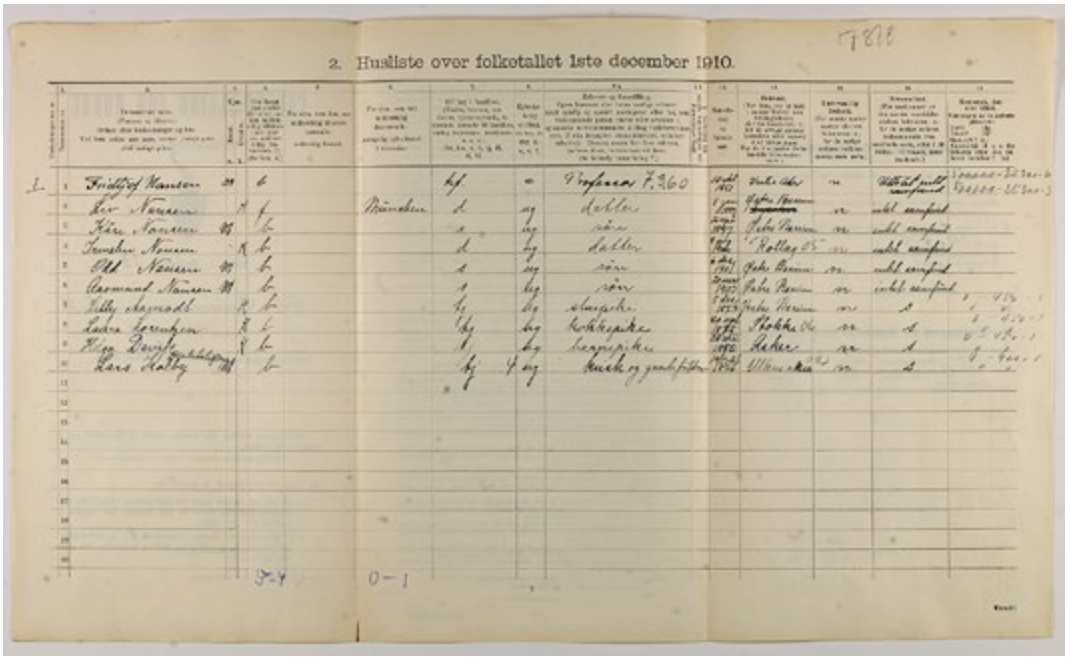
Autoritetsposten til Anders Beer Wilse i KulturNav, http:// http://kulturnav.org/2b94216b-f2fc-46a3-b2ce-eeb93aa19185.

Informasjonen i KulturNav er fritt tilgjengelig gjennom et åpent API og som lenkede åpne data (LOD) i *RDF/XML* og *JSON-LD* format. Plattformen er åpen for alle og bruk er gratis. Løsningen har også en annen viktig funksjon tilpasset museer som kanskje ikke har kommet så langt i arbeidet med egne vokabular og autoritetslister. Det er mulig å importere ferdige ressurser, som f.eks. datasett med personautoriteter eller emneord som forvaltes av andre bibiliotek, arkiv eller museer, og lagre disse i en «permanent cache» som «plassholdere» i KulturNav. Museene kan bruke disse postene direkte i katalogisering og på denne måten knytte museumsobjektene til nasjonalt og internasjonalt etablerte konsept/autoriteter.

Museenes arbeid med å skape og etablere felles lister i KulturNav er allerede i gang, om enn i en tidlig fase. Løsningen inneholder f.eks. Feltkatalogen 2002 (se ovenfor; over 60 datasett). Nasjonalmuseet for kunst, arkitektur og design og Norsk Folkemuseum leder et prosjekt hvor målet er å oppdatere deler av denne standarden. Det finnes to ulike ressurser med oversikt over fotografer, den ene ved Preus Museum i Norge og den andre ved Nordiska museet i Sverige. Begge datasettene vil bli tilgjengelige i KulturNav. Totalt blir dette 5000–6000 autoritetsposter knyttet til fotografer fra 1800- og

1900-tallet som gjøres tilgjengelig som lenkede data. Andre prosjekter det jobbes med nå er f.eks. datasett med emneord for fotografi og autoritetsdata for maritim kulturarv – fartøy, fartøyskonstruktører og verft.

KulturNav gjør det mulig for museer og andre kulturarvsinstitusjoner å arbeide mer effektivt gjennom felles metadata. Løsningen innebærer også at metadataene kobles opp mot lenkede åpne data som nå vokser hurtig fram på nettet. For mange museer betyr dette at de ved katalogisering vil bruke lenkede data uten at de behøver å tenke på den komplekse teknologien bak. For publikum betyr det bedre kvalitet på dataene, og at det blir enklere å søke i, finne og bruke dem. Dermed åpnes mulighetene for en samlet tilgang til kulturarvsinformasjonen som finnes i arkiv, bibliotek og museer.


2. Husliste over folketallet 1ste december 1910.

Folketelling 1910 publisert som lenkede, åpne data

Denne høsten har Riksarkivet sluttført et pilotarbeid som har sett på hvordan kulturdata kan tilgjengeliggjøres som lenkede, åpne data. Målet er å legge til rette for viderebruk og nye tjenester.

TEKST: **HÅVARD LUNDBERG OG GUNNAR URTEGAARD**

Utgangspunktet har vært folketellingen fra 1910 som ble gjort digitalt tilgjengelig og søkbar i 2010 etter at hundreårsklausulen for slike data hadde utgått. Folketellingen er en historisk skattekiste som er blitt brukt til forskning og slektsgransking.

Den inneholder detaljert informasjon om en hjemmehørende befolkning på 2 391 782 personer. Her kan man finne alt fra fødselsdato og fødested til yrke og eventuelle sykdommer en person hadde. I tillegg inneholder folketellingen informasjon om mer enn 330 000 bosteder fordelt på by og land. Her kan

man finne informasjon som antall etasjer og leiligheter, hvor mange som bodde der eller hva leieprisen lå på.

Håvard Lundberg, student, Universitetet i Oslo.
Foto: Cathrine Bjerknes

Gunnar Urtegaard, avdelingsdirektør, Riksarkivet.
Foto: Riksarkivet

LENKEDE, ÅPNE DATA?

All denne informasjonen for de daværende 20 fylkene og 659 kommunene er nå konvertert til og publisert som lenkede, åpne data på Riksarkivets servere.[1] Lenkede data er et sett med prinsipper som legger grunnlaget for et *Semantisk Nett* eller *the Web of Data*. Prinsipper som persistente og unike identifikatorer for ressurser og lenking til andre relevante data ble foreslått av Tim Berners-Lee allerede i 2006. Målet er å bygge et nett av data som er maskinlesbart, sammenkoblet, åpent og tilgjengelig for alle.

Pilotarbeidet som er gjennomført av Riksarkivet har hatt som mål å publisere kulturdata med denne teknologien. Allerede ved digitaliseringen av folketellingen ble det lagt ned mye arbeid i struktur og permanent identifikasjon av de ulike enhetene. Dette forenklet arbeidet med konvertering til Resource Description Framework (RDF) som er de facto standard for representasjon av kunnskap som lenkede data. I tillegg var over 80 prosent av alle bosteder i folketellingen kartfestet, noe som gjorde det enkelt å koble folketellingen til andre ressurser.

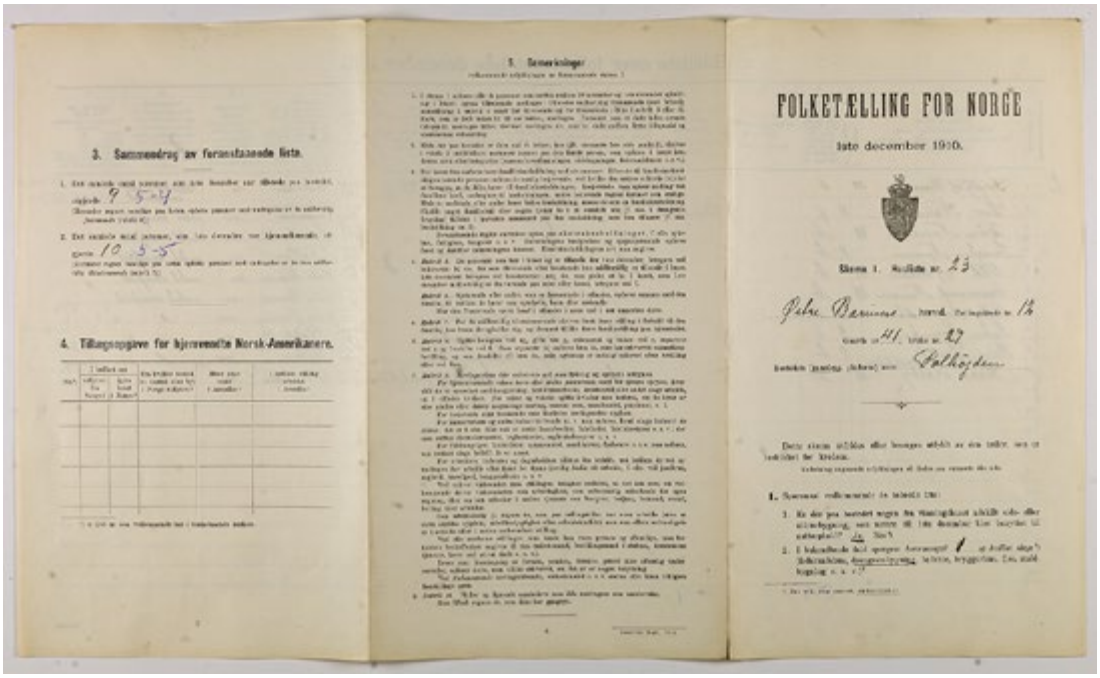
5 STJERNES DATASETT

Publiseringen av folketellingen har fulgt prinsippene for 5 stjernes lenkede, åpne data som ble formulert av Tim Berners-Lee i 2010:

- Tilgjengelig på internett (med en åpen lisens)
- Være i et maskin-lesbart format
- Være i et ikke-proprietært format
- Bruke åpne standarder for å identifisere ressurser
- Lenke til andre data for å gi kontekst

Dette betyr blant annet at hver person, bygning og leilighet har egne, unike IDer i form av persistente URIer. Dette danner grunnlaget for at folketel-

1 http://lod.riksarkivet.no

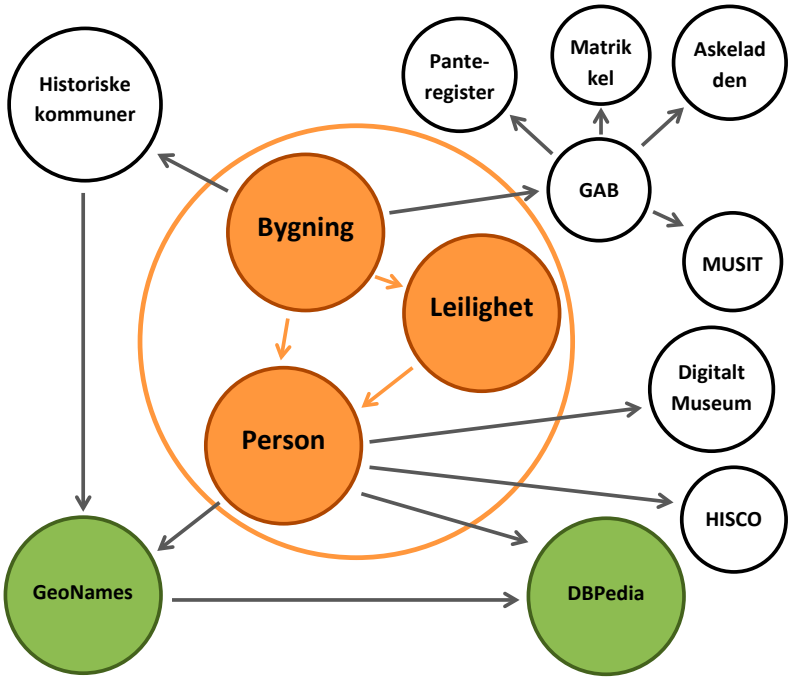


lingen av 1910 kan fungere som et autoritetsregister for historiske personer. Fordi folketellingen er historisk og inneholder informasjon som er domenespesifikk, var det ikke mulig å finne eksisterende ontologier som beskrev datasettet fullt ut. Derfor er det bygget en egen ontologi i OWL[2] som bygger på anerkjente standarder. I tillegg er det lagt ned mye arbeid i å avdekke og validere koblinger til andre relevante datasett.

TVERRSEKTORIELT SAMARBEID

Resultatet er en digital sandkasse som inneholder demodata fra 8 ulike datasett.[3] Folketellingen fra 1910 er brukt som utgangspunkt, Derfra er det opprettet koblinger til tilsvarende ressurser, eller til ressurser som gir mer kontekst (se figur 1). GAB-registeret[4] til Kartverket er et annet sentralt datasett, som gjør det mulig å koble ulike ressurser basert på kommune-, gårds- og bruksnummer. Det er utarbeidet ulike prototyper som viser hvordan datasettene kan brukes. De ulike datasettene og den semantiske representasjonen av dem er også dokumentert.

2 Web Ontology Language, se http://www.w3.org/TR/owl2-primer/
3 http://lod.riksarkivet.no/apps/
4 Grunneiendom, adresse og bygningsregisteret, i dag avløst av Matrikkelen



Figur 1 Oransje: folketellingen, grønn: publiserte datasett i LOD-cloud, hvit: demodatasett fra øvrige organisasjoner.

Prosjektet er gjennomført som et sandkasseprosjekt for å lære å se muligheter. Datasettene er publisert i samarbeid mellom Riksarkivet, Riksantikvaren, Kulturrådet, Kartverket og Kultur- og naturreiseprosjektet[5].

Avinet[6], som er teknisk partner i det pågående EU-prosjektet LoCloud[7], som koordineres av Riksarkivet, har spilt en viktig rolle i arbeidet med kartfesting og etablering av database for RDF-dataene (en såkalt «triple store»). LoCloud samler mer enn 30 europeiske land og har som mål å publisere kulturdata i Europeana[8], som er en felles, europeisk plattform og portal til europeisk kulturarvsmateriale. LoCloud prosjektet startet i 2013 og varer i tre år. Her spiller semantisk teknologi en viktig rolle.

5 Se http://lod.riksarkivet.no og http://locloudlod.avinet.no/sparql
6 http://www.avinet.no/
7 http://www.locloud.eu/About
8 http://europeana.eu/portal/

Nye katalogiseringsregler: Resource Description and Access (RDA)

De gamle katalogiseringsreglene er modne for utskifting. RDA skal både ta vare på den eksisterende bibliotekkatalogen og åpne for tilpasninger til dagens og framtidens digitale virkelighet.

TEKST: **NINA BERVE**

RDA (Resource Description and Access) er det nye katalogiseringsregelverket som skal erstatte de anglo-amerikanske katalogiseringsreglene (AACR2). RDA utvikles av Joint Steering Committee for the Development of RDA (JSC) som består av representanter for ulike bibliotek og bibliotekorganisasjoner i flere land. Etter ni års utviklingsarbeid gikk Library of Congress over til RDA 31. mars 2013. Også flere av de store, ledende bibliotek rundt om i verden har innført, eller er i ferd med å innføre denne nye standarden for katalogisering.

HVA ER RDA?

AACR2 oppdateres ikke lenger, og RDA er på mange måter en videreutvikling av AACR2. Det er lagt betydelig vekt på at metadata som er basert på AACR2 skal kunne fungere sammen med metadata som er registrert i henhold til RDA. RDA er tiltenkt å ha større internasjonal utbredelse og bli benyttet i flere miljøer enn bibliotek, og er således løsrevet fra visnings- og registreringsformater.

AACR2 er ordnet etter materialtyper og er lite fleksibel med hensyn til nye typer. For å gjøre regelverket mer fleksibelt og bedre tilpasset en digital virkelighet er det ordnet på en ny måte: Det er strukturert etter begrepsapparatet i FRBR (Functional Requirements for Bibliographic Records, Funksjonkrav til bibliografiske poster). Sentrale begreper her er «verk / uttrykk / manifestasjon / eksemplar» og brukerbehovene «finne / identifisere / velge / anskaffe». Fagterminologien er også endret i tråd med FRBR og moderne katalogiseringsprinsipper.

Nina Berve, hovedbibliotekar, Nasjonalbiblioteket.
Foto: Nasjonalbiblioteket

RDA legger større vekt enn AACR2 på å skape søkeinnganger i katalogen og tydeliggjøre relasjoner mellom dokumentene og hvilke roller ulike personer og korporasjoner har til dokumentene. Slik vil det også legges bedre til rette for at data lettere kan omformes til åpne, lenkede data.

En bibliografisk post registrert etter RDA vil ikke se så forskjellig ut fra én registrert etter AACR2, og på detaljnivået er de fleste reglene beholdt. Men det er noen endringer, og her kan blant annet nevnes:

- Kjerneelementer (Core elements) erstatter de tre beskrivelsesnivåene. Bare disse kjerneopplysningene er obligatoriske, og det er ellers til en viss grad valgfrihet og alternativer.

- Flere kilder er tillatt som hovedkilde for opplysningene, og man skal i større grad gjengi opplysningene slik de forekommer, uten forkortelser («take what you see»).

- Generell materialbetegnelse er erstattet med tre nye og mer omfattende elementer: Content type, Media type og Carrier type.

- Den såkalte «rule of three»-regelen er fjernet, og man kan registrere så mange biinnførsler som det er behov for.

Som det vil fremgå av noen av disse endringene, ligger det mer valgfrihet i de nye reglene. Hensikten er å gjøre dem brukbare for flere miljøer, og å legge til rette for at gjenbruk av data skal være enklere ved at variasjoner i registreringen bør kunne godtas. Men i dette siste ligger også en fare for at postene kan bli så ulike at det kan bli vanskelig å sjekke dem mot hverandre.

NYTT FORMAT

Når metadata registrert etter nytt og gammelt regelverk skal kunne fungere sammen, og når endringene ikke er så store, ser ikke dette ut som en revolusjon. Men det ligger et potensial i RDA til en bedre utnyttelse av metadataene, og dette vil kunne realiseres ved at MARC-formatet byttes ut med et nytt format som er bedre tilpasset nyere nettstandarder. At MARC-formatet er utgått på dato, har det lenge vært enighet om. Et nytt format var også en betingelse for at Library of Congress skulle ta i bruk RDA. Arbeidet med et slikt format er i gang, ledet av Library of Congress. Foreløpig har formatet fått navnet BIBFRAME. Det er ikke kjent når formatet kan bli ferdigstilt, og inntil videre oppdateres MARC 21 med felt det er behov for i RDA.

RDA TOOLKIT

RDA foreligger ikke som bok, men som nettverktøy: RDA Toolkit. I denne verktøykassa får man navigerbar RDA og andre hjelpemidler, og muligheter til å lage egne hjelpemidler. RDA oversatt til andre språk er også en del av pakken, og hittil er tysk og fransk oversettelse gjort tilgjengelig. Tilgangen til RDA er ikke åpen og gratis, og betalingen foregår gjennom lisenser. Lisensmodellen tar utgangspunkt i antall brukere, men konsortiemodeller er også mulig.

UTBREDELSE OG SAMARBEID

RDA er ment å skulle være en internasjonal katalogiseringsstandard, og fra 2015 er visjonen sågar «RDA: the global standard enabling discovery of content». Flere land enn de engelskspråklige har vært trukket med i utviklingen, gjennom direkte deltakelse eller gjennom høringer, og etter hvert er det blitt mindre amerikansk vinkling. De fleste europeiske landene med interesse for saken har gått sammen om en lokal brukergruppe: EURIG (European RDA User Group). Gjennom denne har vi større muligheter for å få gjennom endringer. Norge deltar i denne gruppen.

RDA I NORGE

'Katalogiseringsregler' er regelverket vi bruker i Norge i dag. Siden dette er en norsk oversettelse av AACR2 og sistnevnte ikke oppdateres, var det naturlig for oss å følge med på utarbeidelsen av RDA. Den norske katalogkomité (DNK) har fulgt utviklingen, og har sendt inn høringssvar til flere av kapitlene. Da RDA ble lansert og flere land tok det i bruk, ba Nasjonalbiblioteket komiteen om å vurdere om regelverket skulle tas i bruk i Norge. Komiteen anbefalte at RDA tas i bruk, at hele RDA oversettes, at det utarbeides norske retningslinjer for bruk og at behovet for norske tilpasninger vurderes. Vinteren 2014

vedtok Nasjonalbiblioteket dette. Argumenter som komiteen har lagt vekt på er blant annet:

- AACR2 oppdateres ikke lenger
- RDA gir bedre muligheter for å beskrive ressurser i ulike fysiske formater
- Mer vekt på søkeinnganger og relasjoner
- Bygger på FRBR og moderne katalogiseringsprinsipper
- Utbredelsen av et felles regelverk er gunstig for utveksling av metadata internasjonalt

Nasjonalbiblioteket har ansvaret for innføringen av RDA i Norge, i nært samarbeid med DNK og andre aktuelle aktører.

OVERSETTELSE OG TILPASNINGER

Nasjonalbiblioteket følger DNKs anbefaling om å oversette RDA til norsk, slik det ble gjort med AACR2. Selve teksten vil bli oversatt av en profesjonell oversetter, og er planlagt utført i 2015. Som et forarbeid til oversettelsen er DNK nå i gang med å oversette den bibliotekfaglige terminologien. Det gjelder glossaret, content/media/carrier-termer, relasjonstermer og en del tekst i eksempler. Slik vil oversettelsesarbeidet bli enklere for en oversetter. Den norske oversettelsen vil bli en del av, og tilgjengelig i, RDA Toolkit.

Som nevnt har RDA stor fleksibilitet og en del alternative regler. DNK er av den mening at vi i Norge bør ha mest mulig ensartet praksis, og vil derfor utarbeide retningslinjer for RDA i Norge. Slike retningslinjer vil gi en felles tilnærming når det gjelder katalogiseringsnivå, men vil samtidig gi stor fleksibilitet ved at de som har behov for det kan innlemme flere elementer i beskrivelsen. Komiteen vil også gjennomgå hva som tidligere har eksistert av norske avvik, men det er ønskelig å holde avvik og alternative regler på et minimumsnivå.

Gjennom samarbeidet i EURIG vil det også være mulig å bli enige om en europeisk praksis. Ønsker om endringer i regelverket får større gjennomslagskraft når man har EURIG i ryggen.

OPPLÆRING OG ANDRE FORBEREDELSER

I god tid før RDA er klar til å tas i bruk i Norge, skal det lages en plan for opplæring i det nye regelverket. Her skal Høgskolen i Oslo og Akershus (HiOA) involveres. Tidspunktet for opplæringen avhenger av den øvrige tidsrammen og progresjon for arbeidet med oversettelse.

Siden det nye regelverket er bygget opp rundt FRBR, vil kursing i FRBR også være en viktig del av opplæringen.

Det er en rekke andre aktører som må involveres eller holdes løpende orientert om arbeidet mot en innføring av RDA i Norge. Blant annet må Bibliotek-systemleverandørene få mulighet til å tilpasse sine interne MARC-formater i henhold til RDA.

Det er allerede innført en del nye felter og delfelter i MARC 21 som en konsekvens av RDA. En forutsetning for innføring av RDA i Norge er derfor at disse feltene tas inn i NORMARC og andre brukte MARC-formater i Norge.

AKTUELLE LENKER

Resource Description and Access (RDA)
http://www.rda-jsc.org/rda.html
RDA Toolkit
http://www.rdatoolkit.org/
DNKs side om RDA
http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/RDA-Resource-Description-and-Access-i-Norge
Funksjonskrav til bibliografiske poster (FRBR oversatt til norsk)
http://www.nb.no/Bibliotekutvikling/Kunnskapsorganisering/Den-norske-katalogkomite/Anbefalinger
Bibliographic Framework Initiative (BIBFRAME)
http://www.loc.gov/bibframe/
European RDA Interest Group (EURIG)
http://www.slainte.org.uk/eurig/index.htm

Bruk av logo s.112: «Logo used by permission of the Co-Publishers for RDA (American Library Association, Canadian Library Association, and CILIP: Chartered Institute of Library and Information Professionals)

Bibliografi + Nasjonalbiblioteket = sant

Bibliografisk arbeid tilhører kjernen av det faglige arbeidet som gjøres i bibliotekene. I Nasjonalbiblioteket legges kreftene først og fremst ned i Nasjonalbibliografien, med alle dens deler. Men i tillegg produseres også en rekke mindre spesialbibliografier.

TEKST: **HEGE STENSRUD HØSØIEN**

> «Singer på Notodden. I ferd med å leve sitt liv, og ennå med forventningene i blodet. For Singer var et bibliotek lag på lag av avleiret materiale, støvete bøker. Biblioteket var en labyrint, og katalogiseringssystemet var en måte å forholde seg til labyrinten på. Å kunne beherske det var en stor nytelse for ham.»
> (*T. Singer*, 1999, 92)

Hege Stensrud Høsøien, seksjonsleder, Nasjonalbiblioteket.
Foto: Nasjonalbiblioteket

Nasjonalbiblioteket har lang tradisjon for å lage slike spesialbibliografier. Tidligere ble bibliografiene utgitt i bokform. Mange bibliotekarer og litteraturinteresserte forholder seg den dag i dag til det enorme bio-bibliografiske arbeidet lagt ned i *Norsk Forfatter-Lexikon*, utgitt i flere bind av J.B. Halvorsen. Senere fulgte en periode der bibliografier ble laget i egne databaseløsninger, ofte i kombinasjon med trykte produkter. Et godt eksempel er Littforsk, bibliografi over norsk litteraturforskning, som ble oppdatert som database og samtidig trykket i *Norsk litterær årbok* frem til 2012. I dag produseres alle bibliografier

inne i felleskatalogen BIBSYS og vises frem i et eget grensesnitt. Med dette har Nasjonalbiblioteket fått en mer bærekraftig og effektiv modell.

Dagens integrerte løsning har mange fordeler. Alle Nasjonalbibliotekets bibliografiske metadata ligger i BIBSYS. Dette betyr at de som arbeider med spesialbibliografi, kan trekke veksler på katalogiseringsarbeid av høy kvalitet som er gjort i Nasjonalbibliografi og autoritetsregister, samtidig som data generert av spesialbibliografisk arbeid føres tilbake i Nasjonalbibliotekets oversikter og dermed blir tilgjengelig for bredere brukergrupper.

BIBSYS er katalog for Nasjonalbibliotekets boksamling. I arbeidet med spesialbibliografi hentes alle relevante trykk opp fra magasin og gås igjennom. Ofte er det mange tusen bøker som hentes opp fra kjelleren av flittige bokhentere. Nasjonalbibliotekets målsetting er å ha en *komplett* samling norske trykk. Og med komplett menes her komplett i bokhistorisk forstand. Dette betyr at Nasjonalbiblioteket skal ha ikke bare alle titler, men alle utgaver av alle titler, og alle bibliografiske varianter av utgavene.

Da vi laget Solstad-bibliografi i 2011, ble vi for eksempel minnet om noe vi egentlig visste fra før: Flere av førsteutgavene fra perioden tidlig i Solstads forfatterskap utkom samtidig i innbundet og heftet variant. I nettopp denne perioden var det samtidig vanlig praksis i norske bibliotek at heftet litteratur ble bundet inn i stive permer av bokbindere, slik at bøkene skulle vare lenger. De originale hefteomslagene ble ikke alltid tatt vare på, heller ikke i Nasjonalbiblioteket. De litt stive hefteomslagene kom nesten litt i veien mellom permen og bokblokken, og gjorde boken vanskelig å åpne og lese. I dag er holdningen en helt annen. Omslag og alt annet bokutstyr og paratekst som omgir hovedteksten, ses på som viktig og bevaringsverdig. Det arbeides kontinuerlig for å tette huller i Nasjonalbibliotekets samling, og spesialbibliografisk arbeid er en metode for å opparbeide den samlingskompetansen og oversikten som skal til for å kunne identifisere lakuner.

I forbindelse med det bibliografiske arbeidet blir også all litteratur som tilhører bibliografien digitalisert, hvis den ikke allerede er det, noe den i økende grad er. Litteraturen som har falt i det fri, eller som dekkes av

Bokhylla-avtalenen avtalelisensordning mellom Nasjonalbiblioteket og opphavsrettsorganisasjonen KOPINOR som gjør all norsk litteratur frem til år 2001 gratis tilgjengelig for norske ip-adresser – gjøres tilgjengelig for brukeren. Dermed blir alle bibliografiens titler digitalt tilgjengelige for forskere på Nasjonalbiblioteket, og størstedelen av dem også utenfor huset på forskerens eller brukerens egen datamaskin.

Nasjonalbiblioteket har som en viktig del av sitt mandat å være infrastruktur for norsk forskning. Til denne infrastrukturen hører også spesialbibliografisk arbeid. Samarbeidet med forskere har tatt ulike former, og har i sin enkleste form hatt karakter av en dialog rundt utvalgskriterier og emneord, og tips om titler.

Et eksempel på et mer ambisiøst samarbeid var forskningsprosjektet 'Mangfoldige offentligheter. Tidsskrifter, ytringsfrihet og produktiv sensur i Danmark- Norge 1720- 1814' ved Universitetet i Oslo. Prosjektet gjennomgikk en rekke danske og norske 1700-tallstidsskrift og så på aspekter som: Hvem skriver, hva skriver de om, litterær sjanger etc. Opplysningene forskerne var interessert i å kartlegge, var bibliografiske i sin natur. I samarbeid med Nasjonalbiblioteket ble et utvalg av de norske tidsskrifttitlene digitalisert, og det ble samtidig lagt til rette for en infrastruktur der forskerne kunne registrere deler av sine forskningsdata i et format som gjorde at opplysningene på en effektiv måte kunne integreres i en spesialbibliografi. I praksis betydde dette at opplysninger ble registrert i en form som lå tettest mulig opp til MARC-formatet. Dataene ble registrert i BIBSYS og forskerne bidro på denne måten med bibliografiske opplysninger både til en spesialbibliografi, til Nasjonalbibliografien, til Nasjonalbibliotekets katalog, felleskatalogen BIBSYS og til det nasjonale autoritetsregisteret. Bidragene til det nasjonale autoritetsregisteret blir i sin tur integrert med det internasjonale autoritetsregisteret VIAF (Virtual International Authority File), og BIBSYS' bibliografiske data og autoritetsregister publisert som linked open data. Slik blir forskningsdata tilgjengelige ikke bare for forskerne selv, og for det publikum som er interessert i nettopp deres forskning, men for et bredere publikum, som kanskje vil nyttiggjøre seg opplysningene på en måte vi i dag ikke kan forutse.

1814-bibliografien er et annet eksempel på arbeid der spesialbibliografi og forskning går hånd i hånd. Bibliografien samler dokumenter skrevet i perioden 1812-1814, og dokumenter som handler om denne perioden.

Bibliografien er en integrert del av forskningsprosjektet 'Propagandakrigen om Norge 1812-1814', som er basert ved Nasjonalbiblioteket, og som inngår i Forskningsrådets Grunnlovssatsning. I prosjektet er hundrevis av skrifter i årene rundt 1814 plassert i en politisk og bibliografisk kontekst, gjennom det som her kalles en politisk-bibliografisk metode. Dokumentene er fulgt på tvers av landegrenser, språk og sjanger, via gjenutgivelser, oversettelser og motskrifter. Gjennom forskningen er hittil ukjent og vanskelig tilgjengelig materiale brakt inn i bibliografien og katalogen, og om mulig anskaffet antikvarisk eller digitalt. På den måten har prosjektet også bidratt til digital tetting av huller ved samarbeid og innlån. Materialet er i stor grad gjort digitalt tilgjengelig. Forskningsbasert kunnskap om dokumentene, mange av dem utgitt anonymt og med sparsomme og noen ganger villedende bibliografiske opplysninger, er underveis tilført bibliografien og katalogen. Materialet i bibliografien er i neste omgang utnyttet i bøker, artikler og ulike former for forskningsformidling.

Da Nasjonalbiblioteket laget Hamsun-bibliografi i forbindelse med Hamsun-jubileet i 2008, fikk vi deler av datagrunnlaget fra Universitetsbiblioteket i Tromsø. Disse ble gjennomgått og integrert i Nasjonalbibliotekets egen løsning, men det var først med Prøysen-bibliografien, i anledning årets Prøysen-jubileum, at vi har kunnet realisere bibliografimodellens potensial for samarbeid. Prøysen-bibliografien er laget i tett samarbeid med Høgskolebiblioteket på Hamar, som i sin tur samarbeider tett med en Prøysen-forskergruppe ved Høgskolen i Hedmark. Bibliotekarer fra høgskolebiblioteket og Nasjonalbiblioteket har registrert bibliografiske data til bibliografien samtidig, og har slik virkelig kunnet trekke veksler på felleskatalogaspektet ved BIBSYS.

De siste årene, og særlig etter at Nasjonalbiblioteket fikk ansvar for forfatterjubileer som del av sitt mandat, er det produsert en rekke forfatterbibliografier. I tiden som kommer blir en av utfordringene å tilpasse produksjon av spesialbibliografi til biblioteksystemet Alma som BIBSYS-bibliotek skal ta i bruk fra 01.01.2016. Når vi føler at vi behersker dette, skal vi fortsette med å lage spesialbibliografier. Det er ikke lagt konkrete planer, men vi har flere ideer som vi leker med.

En av dem er å lage en stor nynorsk bibliografi. Kanskje sammen med Nynorsk kultursentrum og Ivar Aasen-tunet - og andre nynorskbibliotek? Tenk å kunne få oversikt over den nynorske litteraturen, digitalisere det som ikke allerede er digitalisert og gjøre den tilgjengelig og formidlingsvennlig. Kanskje det er slik vi skal bidra til feiringen av Vinje-jubileet i 2018?

En annen idé er å videreutvikle bibliografimodellen slik at den i større grad kan håndtere eksemplarspesifikke opplysninger. Det er på høy tid at vi skaffer oss mer kunnskap om nasjonal utbredelse av og tilstand for eldre bøker. Den norske nasjonallitteraturen er ung; de første bøkene ble trykket i Christiania fra 1643 og utover. Noen år senere kom et nytt trykkeri til i Halden. Nasjonalbiblioteket ville gjerne hatt med seg norske bibliotek på en registrering av alle bøker trykket i Norge på 1600-tallet.

I tillegg til spesialbibliografiene laget vi i fjor en filmografi over Norsk spillefilm (nb.no/filmografi). Utviklingsarbeidet som lå til grunn for dette gjør at vi fremover kan lage multimediale bibliografier, bibliografier som i tillegg til trykt materiale kan ha radioprogram, filmklipp, musikkfiler og fotografier. Hvis det fortsatt skal hete bibliografier, da?

Tidligere var det slik at hovedhensikten med en spesialbibliografi var å sikre en samlet fremstilling av for eksempel et bestemt forfatterskaps titler og utgaver, og alt som var skrevet om forfatterskapet. Med dagens digitale og bibliografiske løsninger, og med det store arbeidet som er lagt ned i digitalisering av nasjonallitteraturen, ser vi at behovet og bruken i noen grad endrer seg fra å gjenfinne enkelttitler til å behandle hele forfatterskap som korpus i forbindelse med kvantitative språk- og litteraturstudier. Denne endringen bør få betydning for hvordan vi tenker om spesialbibliografisk produksjon i årene fremover.

Ein nasjonalbibliografi for Sápmi

Samisk bibliografi i Noreg omfattar publikasjonar på samiske språk, dokument som handlar om samiske emne eller som er relevante i ein samisk samanheng. Frå 1993 har utvalet vore avgrensa til pliktavleverte dokument. I Sverige, Finland og Russland blir det utført tilsvarande bibliografisk arbeid for samisk litteratur utgitt i dei respektive landa.

TEKST: **GUNNAR HAGEN**

KVA ER SAMISK BIBLIOGRAFI?

I Noreg er det tre offisielle samiske språk: Nordsamisk, lulesamisk og sørsamisk. Det er store skilnader mellom desse tre språka, og grensene mellom språkområda følgjer heller ikkje landegrensene. Nordsamisk er det klart største språket og blir nytta i Noreg, Sverige og Finland, mens lulesamisk og sørsamisk finst både i Noreg og Sverige. Vidare er det andre samiske språk i Sverige, Finland og Russland.

Nokre døme på samiskrelevante emne er språk (forsking, opplæring), næringar (reindrift, fiske), rettsspørsmål, kunst (musikk, duodji, skjønnlitteratur), religion og misjonshistorie, kulturhistorie osb. Vi legg elles vekt på å registrere dokument med biografisk og lokalhistorisk innhald.

Samisk bibliografi tar opp mange publikasjonstypar: Bøker, periodika, artiklar i tidsskrift og bøker, småtrykk, lydbøker, notetrykk (frå 2010), musikk (frå 2011) og eit utval digitale dokument (frå 2014). Eit særtrekk ved bibliografien er det at i overkant av 70 % av innførslane er artiklar i tidsskrift og bøker. Vanlege monografiar utgjer drygt 25% av bibliografien. Denne fordelinga har halde seg ganske stabil.

Samisk bibliografi inneheld hausten 2014 vel 22 000 innførslar. Den årlege tilveksten har i seinare tid auka og ligg no på 1 400 - 1 500 postar.

Gunnar Hagen, hovudbibliotekar, Nasjonalbiblioteket.
Privat foto

KORT HISTORIKK

Historia om Samisk bibliografi begynte i Trondheim. Der hadde den sudettyske bibliotekaren Justin Siegert frå 1973 på hobbybasis tatt til å samle inn bibliografisk informasjon om samar og samiske forhold. Det vart frå 1979 løyvd offentlege midlar til arbeidet. Siegert vart pensjonert i 1982, og prosjektet vart ført vidare av Anders Løøv og Inga-Britt Blindh.

Det vart publisert to trykte bibliografiar frå arbeidet i Trondheim. Den første kom i 1984, dekte perioden 1966 - 1980 og hadde 2 117 bibliografiske innførslar. Neste bibliografi kom i 1989 og tidsrommet var utvida til 1945 - 1987. Innhaldet hadde auka til 4 127 postar.

Dessverre tok løyvingane slutt, og Samisk bibliografi fekk ein lakune for perioden 1988 -1992. Frå 1993 vart bibliografien fast plassert ved den nye Nasjonalbibliotekavdelinga i Rana (NBR). Her vart arbeidet leia av Eli Karin Svingeseth. Ho hadde tidlegare arbeidd ved Samisk spesialbibliotek i Karasjok. Den nye pliktavleveringslova frå 1989 gjorde det mykje lettare å få tilgang til dokument til bibliografien.

NBR publiserte ein trykt bibliografi for 1993 - 1994. Seinare kom bibliografien på CD-ROM, og etter kvart på internett som ein av Nasjonalbiblioteket sine Trip-basar. Ulempa var at basen ikkje vart kontinuerleg oppdatert.

Samisk bibliografi gjekk over frå registrering i Mikromarc til registrering i BIBSYS i 2010, og i 2011 vart postane frå Mikromarc importerte til BIBSYS. Importen omfatta også dei 4 127 postane frå Trondheim.

KVIFOR SAMISK BIBLIOGRAFI ER VIKTIG

Alle dei samiske språka er små, også samanlikna med norsk. Det er derfor viktig at dokument på samisk blir systematisk registrerte, og at denne informasjonen blir gjort allment tilgjengeleg. Dette er ei primær oppgåve for Samisk bibliografi.

Like eins er det viktig at vi uavhengig av språk dokumenterer samisk kulturell og litterær verksemd, samisk næringsliv, organisasjonsliv og politikk. Det er mange tema som er perifere for storsamfunnet, men sentrale i ein samisk samanheng. Viktige samiske merkeår er 1751 (lappekodisillen, dvs. grensa mot Sverige vart fastsett), 1917 (det første samelandsmøtet) og 1989 (Sametinget vart opna). Dette må bibliografien ha eit vake auge for.

Noreg har ratifisert ILO-konvensjon nr. 169 om urfolk og stammefolk i sjølvstendige statar. Konvensjonen peikar m.a. på det kulturelle mangfaldet urfolk representerer, og Samisk bibliografi er med og dokumenterer dette mangfaldet.

KVA ER SPESIELT MED SAMISK BIBLIOGRAFI?

Når vi arbeider med Samisk bibliografi, er det naturleg å sjå verda gjennom samiske briller. Med det meiner eg at bibliografien må spegle av det som er viktig for samar. I tillegg til solid samisk språkkunnskap, må ein òg ha så god kunnskap om samiske emne at klassenummer og emneord blir korrekte.

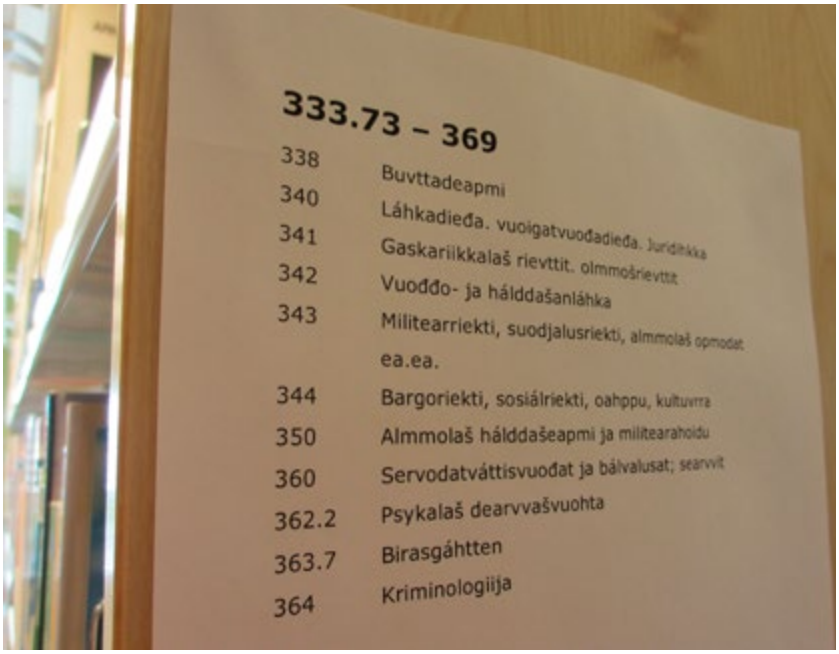
Utval

Utval til bibliografien er ei heilt grunnleggjande oppgåve. Vi går kvar veke gjennom alle nye pliktavleverte monografiar og hefte i monografiseriar. Dokument til bibliografien blir plukka ut. Her må vi ta stilling til om vi skal ta inn heile boka, eller berre enkeltartiklar eller kapittel. Vi vurderer også om det samiske aspektet er tydeleg nok.

Når det gjeld periodika, har vi ei liste på rundt 300 titlar der vi kvar veke går gjennom nye hefte. Lista inneheld for det første titlar på samisk, men det meste er periodika og årbøker som i varierande grad inneheld samiskrelevante artiklar.

Vi har ikkje noka absolutt nedre grense for omfang av ein artikkel, men vi tar ikkje notisaktig stoff. Derimot er vi opne for relativt korte artiklar når det gjeld personhistorie og lokalhistorie. Geografisk er lokalhistorie interessant frå Femundsområdet og nordover.

Sametingsbygningen, Karasjok. Foto: Siri K. Gaski



Frå biblioteket ved Samisk høgskole, Kautokeino. Foto: Siri K. Gaski

Klassifisering etter Dewey

Alle dokument registrerte frå 1993 og seinare, er klassifiserte etter Dewey. Unntaket er skjønnlitteratur der vi tok i bruk Dewey først frå 2013.

Samisk bibliografi følgjer i all hovudsak dei allmenne retningslinene for Dewey, men med nokre avvik. For det første bruker vi to klassenummer noko oftare enn reglane seier for å betre gjenfinninga av dokument med fleire emne. Vidare deler vi i mindre grad inn klassenummer etter folkegrupper. Grunnen er at dokument skal vere samiskrelevante. Samisk bibliografi kan i enkelte høve leggje større vekt på samiske aspekt i framstillinga enn *Norsk bokfortegnelse* (Norbok) og *Norske og nordiske tidsskriftartikler* (Norart) gjer.

Vi vurderer løpande om dei avvika vi har, er viktige nok til å halde dei oppe.

Klassifisering etter Løøv

I tillegg til Dewey har vi eit eige klassifikasjonssystem, Løøv, som er særleg tilpassa samiske forhold. Det har fått namn etter Anders Løøv som var sentral i arbeidet i Trondheim. Løøv-systemet består av 21 klassar som er delt inn vidare i varierande grad. Her er 15b til dømes nummeret for skolehistorie, mens 06b1a står for namnegransking, inkludert stadnamn.

Alle dokument i Samisk bibliografi er klassifiserte etter Løøv. Systemet blir også brukt i dei andre samiske bibliografiane og i andre samiske bibliotek.

Emneord

Alle faglitterære innførslar får i prinsippet emneord. Unntaket er ein del dokument der ein person, ein korporasjon eller eit geografisk område er emne for framstillinga. Skjønnlitterære dokument får ikkje emneord.

Samisk bibliografi har ei eiga emneordliste som har utgangspunkt i ei firespråkleg liste som vart utarbeidd ved Universitetsbiblioteket i Trondheim. Denne emneordlista har vi utvida etter behov, og vi har prekoordinert emneorda i emnestrengar. Nye emneord vurderer vi mot emneorda hos Biblioteksentralen og Humord.

Fordi bibliografien har ein del emne med mange aspekt og stort dokumentunderlag, kan emnestrengar vere nyttige for å halde orden og unngå å lage fleire emneord enn nødvendig. Slike område er reindrift og språk. I litteraturen om reindrift er det svært mange termar om ulike sider av næringa, og av omsyn til brukarane er det viktig at vi bruker ein detaljert terminologi. Vi har sjølvsagt vide termar som *Reinbeite*, *Reindrift* og *Reindriftsnæring* som er delt opp i underavdelingar, t.d. med *Forvaltning* og *Økonomi*, eller i form, t.d. *Forskrifter*. Dessutan har vi heilt spesifikke emneord som *Reingjeting*, *Reinpåkjørsler*, *Gjeterhytter* og *Reinlykke*.

Vi prøver å unngå samansette emneord av typen *Samisk litteratur*. Emneordet *Litteratur* blir da brukt. Derimot har vi oppretta emneordet *Norsk litteratur*. Men i blant endrar litteraturen seg. *Religion* har vi brukt for den førkristne trua. Det gjekk lenge greitt, heilt til vi fekk samiskspråklege dokument om både religion generelt og om førkristen tru. Vi kan ikkje ha to tydingar av eitt emneord. Dermed måtte vi ta i bruk *Samisk religion* for førkristen tru og la *Religion* stå for religion generelt.

Eit anna emneord vi helst unngår, er *Samer*. Det kunne vi sikkert ha brukt for dei fleste dokumenta i bibliografien. For dokument som handlar om både samar og kvenar, må vi også bruke *Samer* dersom vi bruker *Kvener*. Elles er det sjeldan aktuelt.

Geografiske emneord

Samisk bibliografi har også ei eiga geografisk emneordliste. Dei aller fleste stadene og områda ligg i Noreg, men område i Sverige og Finland er også godt representert. Vi legg stadnamna i eit hierarki med forma Fylke – Kommune – Stad, til dømes slik: *651 $a Troms ;; Lavangen ;; Spansdalen*. Når det gjeld stad innan ein kommune, er vi svært spesifikke og går ned til gard, husmannsplass eller buplass i utmarka. Dermed er bibliografien med og dokumenterer samisk historie, aktivitet og busetjing.

Vi må i ein del tilfelle tilpasse hierarkiet noko. Område som omfattar meir enn ein kommune, registrerer vi som *Nordland ;; Ofoten* dersom det er ein region, eller som *Nord-Trøndelag ;; Namsen* for andre naturformasjonar. Strekker eit område seg over to eller fleire fylke, går vi rett på området, t.d. *Trollheimen*. Like eins er det når ein naturformasjon ligg i to land: *Tanavassdraget*.

Vi lagar òg geografiske emneord for dei ulike samiske distrikta: *Sørsamiske, Lulesamiske, Markesamiske, Sjøsamiske og Østsamiske distrikter* fordi det kan vere nyttig for brukarane av bibliografien. Emneord for konkret stad kjem i tillegg.

Tidlegare brukte vi hovudsakleg atlas for å fastsetje skrivemåten for stadnamn. Vi har dei siste par åra tatt i bruk databasen Sentralt stedsnavnregister (SSR) som blir drive av Kartverket. Fordelen er at denne basen inneheld fleire stadnamn enn andre oppslagsverk, og at nye vedtak blir tatt inn løpande. Men det ligg òg ein del forelda vedtak der.

For geografiske emneord har vi eit par utfordringar. For det første gjeld det korleis vi registrerer det geografiske hierarkiet. Vår praksis er ikkje i samsvar med MARC21 som i felt 651 har eit omvendt hierarki, men med eitt overordna ledd. I dømet *Spansdalen* må vi velje om kommunen eller fylket skal vera dette leddet. Men MARC21 har òg eit felt 662 for å registrere hierarkiske geografiske emneord, og der er det delfelt for ikkje-juridiske naturformasjonar, t.d. elver, sjøar og fjell.

Den andre utfordringa dreier seg om likestilte, fleirspråklege stadnamn. Mange stader og lokalitetar har offisielt godkjende stadnamn på to eller tre språk: samisk - norsk, samisk - kvensk, norsk - kvensk, samisk - norsk - kvensk. Samisk står her for nordsamisk, lulesamisk eller sørsamisk alt etter kvar i landet staden ligg. Offentlege instansar som utarbeider geografiske

register, er pålagde å bruke alle godkjende former av stadnamnet. Det er lite heldig at Samisk bibliografi har geografiske emneord berre på norsk.

Vi arbeider med å finne løysingar på desse to utfordringane.

Andre emneinnførslar
Samisk bibliografi nyttar i ein viss grad sjanger-/formemneord, marc-felt 655. Vi meiner at ein slik verbal søkjeinngang er nyttig for brukarane. Vi har eiga emneordliste her.

Vi legg òg inn årstal som emne i marc-felt 648. Her bruker vi både enkeltståande årstal og årstal i spenn. Dette er ein ressurs vi ikkje har fått utnytta i søk.

Katalogisering
Da Samisk bibliografi gjekk over til å registrere postar i BIBSYS i 2010, tilpassa vi praksisen vår til BIBSYS sitt regelverk. Vi har dermed ingen særleg store avvik. Ein lett synleg skilnad er at vi relativt ofte legg på note om kva dokumentet handlar om. Det gjeld særleg for analyttar, og formålet er å gi brukarane nyttig informasjon slik at dei får dei dokumenta dei har bruk for.

OFTE STILTE SPØRSMÅL

Vi får ikkje så mange spørsmål om bibliografien, men det er gjerne to spørsmål som dukkar opp. Det første spørsmålet er: *Er ikkje alle dokumenta på samisk?*

Dette er det ikkje så mykje å seie om. Det er ikkje berre norskspråklege dokument i Norbok heller. Men så kjem det andre spørsmålet: *Kor mange dokument (eller kor stor del) er på samisk?*

Spørsmålet er rimeleg, og det skulle ikkje vera så vanskeleg, men så kan vi reflektere litt over kva som er meint med at eit dokument er på samisk. Skal heile dokumentet vera på samisk? Fleirspråklege dokument er ganske vanleg i bibliografien. Ofte er eit dokument parallellteksta på nordsamisk og norsk. Det hender at same teksten står på to eller tre samiske språk. Korleis skal vi rekne i slike tilfelle?

Den einaste måten å finne eit svar på, er å søkje på språk, og da er det språkkodane som blir talde. Vi får i alle fall som svar at 26 % av dei bibliografiske postane er på samisk. 22 % av postane er på nordsamisk, 2 % på kvar av sørsamisk og lulesamisk.

KVA MED FRAMTIDA?

Eit viktig tiltak er å gi ein felles søkjeinngang til dei samiske bibliografiane i Noreg, Sverige, Finland og Russland. Alle landa er samde om å gjennomføre dette.

Stadig meir av forskinga på samiske forhold blir publisert i utlandet. Det gjeld både forskarar frå Noreg og frå andre land. Samisk bibliografi bør ta opp eit rimeleg utval av desse dokumenta.

Og ikkje minst – det er viktig at Samisk bibliografi gjer seg kjend overfor brukarar og samarbeidspartar.

Kunnskapsorganisasjon som forskningsfelt

Kunnskapsorganisasjon som fagområde defineres og diskuteres i en egen artikkel i inneværende nummer av Bibliotheca Nova. I denne artikkelen rettes søkelyset mot kunnskapsorganisatorisk forskning (og utviklingsarbeid)[1]. Eksempler er primært hentet fra eget miljø ved Institutt for arkiv, bibliotek- og informasjonsstudier (ABI) på Høgskolen i Oslo og Akershus. For å sette denne forskningen inn i et større bilde presenteres innledningsvis noen betraktninger over interesser og bevegelser i feltet på et internasjonalt nivå.

TEKST: **KIM TALLERÅS**

HVA ER KUNNSKAPSORGANISATORISK FORSKNING?

Tuomaala, Järvelin og Vakkari (2014) har nylig publisert en artikkel som analyserer og kategoriserer bibliotek- og informasjonsvitenskapelig (B&I) forskning. Analysen tar utgangspunkt i forskningsartikler fra 2005, publisert i et utvalg sentrale tidsskrift. Artiklene sammenlignes videre med resultater fra

Kim Talleråas, stipendiat, Institutt for arkiv, bibliotek- og informasjonsfag, Høgskolen i Oslo og Akershus.
Foto: Joacim Anthony Hughes

1 En diskusjon av begrepene forskning og utvikling, eller *FoU* som det gjerne heter, knyttet til utviklingen innenfor kunnskapsorganisasjon ville vært interessant, men må av plasshensyn føres et annet sted. Her vises det til eksempler på arbeid innenfor et bredt FoU-spekter.

tilsvarende studier utført i 1965 og 1985. Denne historiske analysen gir ikke noe oppdatert tidsbilde for 2014, men viser hvor dagens B&I-forskning har sine røtter. Av 718 artikler (fra 2005) er nærmere 85 prosent plassert i fire hovedkategorier.

Den største av disse kategoriene (30,1 prosent av alle artiklene) er *Information Storage and Retrieval* og samler forskning innenfor klassifikasjon og indeksering, men også ulike tilnærminger til gjenfinningssystemer. 24 prosent av artiklene er plassert i kategorien *Scientific and professional communication.* Disse omhandler i stor grad bibliometrisk og webometrisk forskning som typisk undersøker henholdsvis siteringsstrukturer i vitenskapelige publikasjoner og nettverk av lenker i en større web-kontekst. Kategorien *Library and information services activities* (17 prosent) samler forskning om blant annet digitalisering, digitale dokumentsamlinger og administrasjonen av slike, mens forskningen innenfor *Information seeking* (12,3 prosent) er mer brukerorientert og fokuserer på søkeprosesser og søkepraksiser.

FLERE TRADISJONER

I slike undersøkelser er det selvsagt potensielle overlapp mellom kategoriene og flere av artiklene kunne nok vært plassert flere steder. Samlet speiler allikevel de nevnte kategoriene et sett av kunnskapsorganisatoriske tradisjoner som er listet opp i en europeisk undersøkelse av interesseområdene innenfor B&I-utdanninger (Broughton, Hansson, Hjørland, & López-Huertas, 2005). Disse tradisjonene er diskutert videre i Hjørlands oversiktsartikkel «What is knowledge organization?» (2008). Her finner vi igjen kjerneområder som klassifikasjon og indeksering, gjenfinningssystemer (*information retrieval*) og bibliometri, men også mer brukerorienterte (for eksempel *information seeking*) og humanistiske tilnærminger. Hjørland poengterer at disse tradisjonene i stor grad eksisterer uavhengig av hverandre og at de delvis vil ha ulike svar på kjernespørsmålet han stiller i artikkeltittelen. Dette både kan og bør problematiseres, men i denne sammenhengen gir tradisjonene og forskningen som springer ut av dem et tilstrekkelig grunnlag for å fastslå at rammene for kunnskapsorganisatorisk forskning er brede (uavhengig av interne sammenhenger og eventuelle parallelle tilnærminger). Her inkluderes ikke bare tradisjonelle verktøy og teknikker, men også anvendelser av dem i moderne gjenfinningssystemer samt hvordan de fungerer for sluttbrukere.

FRA SYSTEM TIL BRUKER

Over tid hevder Tuomaala, Järvelin og Vakkari å se nettopp en «sosial» utvikling i forskningen, fra systemer til brukere. De ser også en utvikling fra å studere bibliotekinstitusjonen og dens spesifikke innretninger til et mer generalisert blikk. Prebor (2010) bekrefter, gjennom en analyse av master- og doktorgradsavhandlinger fra begynnelsen av 2000-tallet, tendensen mot en økt interesse for «social aspects of information». Hun finner i tillegg at to tredjedeler av avhandlingene innenfor «library science» og/eller «information science» (tagger fra ProQuest Digital Dissertation and Theses database) ikke stammer fra tradisjonelle B&I-institusjoner og at denne forskningen blant annet har et mer generelt og teknologisk fokus.

Utviklingstrendene reflekterer nok til en viss grad at de kunnskapsorganisatoriske forskningsobjektene har blitt digitale, eller at de i alle fall inngår i en digital kontekst. Med det har de delvis vokst seg ut av de tradisjonelle (bibliotek)institusjonene og blitt overtatt av mer systemorienterte fagområder som informatikk og ingeniørstudier[2], hvor majoriteten av den systemorienterte gjenfinningsforskningen i dag foregår.

«ISCHOOLS»

Teknologiutviklingen har dermed også ført til reorienteringer av fagfeltet som i konkurranse med andre fagområder tar hevd på visse «digitale» perspektiver og (tverrfaglige) problemstillinger. En slik reorientering kommer for eksempel til uttrykk gjennom nettverket av såkalte *iSchools*. Nettverket oppstod blant ledende B&I-institusjoner i USA, men teller i dag akkrediterte medlemmer over hele verden, inkludert de største B&I-utdanningene i Skandinavia. I 2014 ble også ABI medlem av nettverket. I iSchool-charteret[3] heter det at: «The iSchools take it as given that expertise in all forms of information is required for progress in science, business, education, and culture. This expertise must include understanding of the uses and users of information, the nature of information itself, as well as information technologies and their applications». Hovedinteressen her ligger riktignok ikke i fordypninger i de nevnte områdene hver for seg, men i relasjonene dem imellom: «The iSchools are interested in the relationship between information, technology, and people».

2 Se Warner (2001) for en mye diskutert (krise)fortolkning av denne utviklingen: «The spread of modern information technologies [...] has also weakened the exclusivity and LIS`s claim to its established domains.»

3 http://ischools.org/about/charter/

TVERRFAGLIGE AMBISJONER

Holmberg, Tsuo og Sugimoto (2013) har studert interessene blant forskere ved iSchools, og resultatene viser at en slik posisjonering har et visst fotfeste i praksis. Av de mest brukte emneordene for å beskrive egen forskning finner vi *human-computer interaction* på topp. Lista beskriver også «sosio-teknologiske» interesser som *social media* og *information seeking*. Samtidig uttrykker den mer systemorienterte interesser som *information retrieval*, *information systems*, *data mining* og *digital libraries*. Dette bryter med den sosiale tendensen antydet ovenfor og kanskje også med ambisjonen om tverrfaglighet som er uttrykt i iSchool-charteret. Det er nok flere forklaringer på dette, både knyttet til sammensetningen av studier innenfor de enkelte iSchool-institusjonene og til selve reorienteringen mot det digitale som påvirker både fokus og forskerrekrutteringen. En oversikt utarbeidet av Wiggins & Sawyer (2012) fra 2009 viser i alle fall at 30 % av (fulltidsansatte) forskere ved iSchool-institusjonene har bakgrunn fra *computing*, og dessuten at øvrige ansatte kommer fra en rekke ulike fagområder. Forfatterne hevder videre at iSchools-nettverket dermed må betraktes som et *fler*faglig fenomen, men også at den tverrfaglige ambisjonen til en viss grad er i ferd med å etablere seg (i litt ulike retninger) på institusjonsnivå.

FORSKNING OG PRAKSISFELT

Samtidig med, og kanskje på grunn av reorienteringene og utviklingen skissert ovenfor, har det oppstått diskusjoner om hva som skal vektlegges innenfor B&I. Blant annet hevdes det (se for eksempel Gorman (2004)) at klassiske ekspertområder undermineres til fordel for generelle og mer teknologiorienterte fokus, og videre at dette bidrar til å skape en avstand mellom forskningsfeltet og bibliotekene som praksisfelt. Mot et slikt argument kan man kanskje trekke frem de mange konferansene (for eksempel Dublin Core, LODLAM og EMTACL) og tidsskriftene (for eksempel Code4Lib Journal, IFLA Journal, Library Hi Tech og D-Lib Magazine), som ikke dekkes av Tuomaala, Järvelin og Vakkaris utvalg, men som nettopp inkluderer kunnskapsorganisatorisk forsknings- og utviklingsarbeid med nære relasjoner til anvendt praksis. Disse arenaene viser også at det foregår mye forskning og ikke minst vesentlig utviklingsarbeid i praksisinstitusjonene og i regi av interesseorganisasjoner som IFLA.

For å bringe mer kunnskap inn i en videre diskusjon om forskningens retning og fokus, kunne det være spennende å utføre en bibliometrisk undersøkelse av overlappen mellom de som bidrar i de ovennevnte praksisnære kanalene,

Høgskolen i Oslo og Akershus. Foto: John Hughes/HiOA

de som skriver i tradisjonelle og velrennomerte forskningstidsskrifter som JASIST[4] og Journal of Documentation, og de som bidrar på iSchool-institusjonenes årlige konferanse, de såkalte iConference.

Oppsummert kan det hevdes at den kunnskapsorganisatoriske forskningen i 2005 utgjorde en vesentlig del av den totale forskningen innenfor B&I, forutsatt utvalget av tidsskrifter hos Tuomaala, Järveling og Vakkari og Hjørlands mangefasetterte tradisjoner som faglig avgrensning. Forskningen kan nok fortsatt defineres bredt på denne måten og utgjør fortsatt en vesentlig del av forskningsaktivitetene ved B&I-institusjoner, men den er også i en utvikling som peker i retning av teknologiens sosiale og brukerorienterte aspekter. Om dette fører til en samling om visse av (de brukerorienterte) tradisjonene, eller at forskningen knytter an til nyere tradisjoner (*digital humantities* anyone?), eller fører til enda flere parallelle perspektiver vil nye kartlegginger måtte vise.

4 Journal of the Association for Information Science and Technology

FORSKNING VED ABI

I det følgende gjøres det kort rede for pågående forskning ved ABI. Redegjørelsene knytter an til det nokså ferske doktorgradsprogrammet ved instituttet og til aktiviteten ved én av de etablerte forskningsgruppene.

DOKTORGRADSFORSKNING

Ved doktorgradsprogrammet kan flere av studentene skilte med problemstillinger relatert til kunnskapsorganisasjon. Tre av disse inngår i det såkalte TORCH-prosjektet (*Transforming organization and retrieval in cultural heritage*) som bygger på et samarbeid med NRK om analyser og (automatiserte) bearbeidelser av metadata.

Innenfor prosjektet utforsker Anne-Stine Husevåg nytteverdien av å ekstrahere egennavn fra naturlig språk i semi-strukturerte NRK-metadata. Hun studerer både bruken av og relasjonene mellom egennavn med tanke på rangering for indeksering og gjenfinning.

David Massey arbeider med bruk og evaluering av automatiske metoder for informasjonsuttrekk (*information extraction*). Han jobber mot det samme semi-strukturerte NRK-materialet som Husevåg, med fokus på algoritmer for navnegjenkjenning, kategorisering og disambiguering av entiteter som personer, steder og kulturelle produkter. Undersøkelser av hvorvidt eksterne datakilder som Wikipedia, språkdata og bibliotekkataloger kan brukes til å forbedre disse algoritmene står sentralt i arbeidet.

Undertegnede jobber også i siste instans med data fra NRK og resultatene av Massey og Husevågs (strukturerings)arbeider, men tar utgangspunkt i de mange nye standardene som for tiden utvikles for bibliografiske data. Her evalueres i hovedsak interoperabilitetseffekter ved å bruke såkalte semantiske teknologier og prinsipper for lenkede data. Et av evaluerings-casene baserer seg på et korpus av bibliografiske poster (i NORMARC-format) som transformeres til et utvalg standarder og videre lenkes til NRK-data. Et annet case undersøker gjenfinning på tvers av museums- og bibliotekdata, mens et tredje case etter planen skal omhandle hvordan de nye standardene blir mottatt i ulike utviklermiljøer.

TORCH-prosjektene baserer seg delvis på eksperimentelle metoder og vil benytte en felles infrastruktur for evaluering. I den anledning utvikles det et annoteringsverktøy som skal brukes til å utarbeide fasiter for mappinger,

transformasjoner, automatiske uttrekk og lenking. Dette utviklingsarbeidet involverer studenter på både bachelor- og masternivå som bistår med uttesting. Den metodiske infrastrukturen og det innledende arbeidet med annoteringsverktøyet er dokumentert i Tallerås, Massey, Husevåg, Preminger & Pharo (2014).

Av de øvrige studentene ved doktorgradsprogrammet jobber Maliheh Farrokhina med brukerinnganger til museums- og bibliotekdata basert på standardene CIDOC CRM[5] og FRBRoo (en bibliografisk utvidelse av CIDOC CRM). Arbeidet forsøker å identifisere kjerneelementer i standardene gjennom intervjuer av ulike typer brukere.

Leiv Bjellands prosjekt, «Archival classification as storytelling», tar utgangspunkt i at arkiv, i henhold til arkivteorien, er aggregerte helheter mer enn de er samlinger av enkeltdokumenter. Siden relasjonene mellom dokumentene i et arkiv er vesentlig for å kunne vurdere og fortolke det enkelte dokumentet, blir organisasjons- og gjenfinningsverktøy som klassifikasjon mer enn bare metadata; de kan forstås som en integrert del av arkivenes data. Prosjektet undersøker arkivarers forståelse for og bruk av metadatasystemer i arkivdanningen ut fra dette perspektivet.

Marit Kristine Ådland er ansatt ved ABI, men holder på med et doktorgradsprosjekt ved vår søsterorganisasjon, *Det informationsvidenskabelige akademi* (IVA), ved Universitetet i København. Hennes prosjekt undersøker tagging med utgangspunkt i en konkret nettside med kreftinformasjon. Foreløpige resultater viser at ulike brukergrupper har til dels motstridende syn på hva tagger skal være; kommunikasjon eller indeksering?

Grete Seland, som tidligere var ansatt ved ABI, har nylig disputert med avhandlingen «User revealment revisited : Knowledge formation in the prefocus stage of information-based work tasks» (2014) ved Aalborg universitet. Avhandlingen undersøker søkestrategier hos studenter i en tidlig og utforskende fase av informasjonsinnhenting som foretas i forbindelse med studierelaterte arbeidsoppgaver. Oppgaven bygger på teori om informasjonssøkeatferd, kognitiv lingvistisk teori og kognitiv psykologisk teori.

5 http://www.cidoc-crm.org/

Høgskolen i Oslo og Akershus. Foto: John Hughes/HiOA

METAINFO

Ved ABI er forskningen organisert i tre grupper hvorav *Metadatabaserte informasjonssystemer* (METAINFO)[6] samler den kunnskapsorganisatoriske aktiviteten. Dette avsnittet viser eksempler på noen av prosjektene og forskerne som inngår i denne gruppa.

KLASSIFIKASJON OG INDEKSERING

Samarbeidet med NRK har i tillegg til metadatastudiene nevnt ovenfor medført forskning på emneindekseringen av radioprogrammer. Tidligere ble denne utført av ansatte i NRKs metadataavdeling som benyttet et kontrollert vokabular. I 2012 endret organisasjonen politikk og ansatte involvert i programproduksjonen utfører nå indekseringen. Det kontrollerte vokabularet har blitt erstattet med forslag til termer og i prinsippet kan indekstermene velges relativt fritt. Analyse av den nye indekseringspraksisen viser blant annet at spesifisiteten og uttømmenheten er svært ujevn. En masteroppgave (Søbak, 2013) har blitt skrevet om dette og to artikler er under bedømmelse.

6 Se http://bit.ly/10R5SR6 for en mer detaljert beskrivelse av forskningsgruppa. De to andre gruppene samler forskning innenfor henholdsvis informasjon og samfunn (INFOSAM) og litteratur- og kulturformidling (LITKULT).

E-BØKER

HiOA utgir per i dag sju open access-tidsskrifter, med PDF som prioritert publiseringsformat (som i vitenskapelige tidsskrifter generelt). I et samarbeidsprosjekt mellom ABI, Institutt for informasjonsteknologi og Læringssenteret ble arbeidsflyten i tidsskriftet *Professions & Professionalism*[7] skreddersydd for e-bok-produksjon. Dette innebærer at alle artiklene kodes i XML-applikasjonen Journal Article Tag Suite (JATS)[8]. Deretter transformeres og konverteres de til tre ulike formater: HTML (for web), EPUB (for lese- og nettbrett) samt MOBI (for Amazon Kindle). Erfaringene fra prosjektet er publisert i Eikebrokk, Dahl, & Kessel (2014) og kildekoden for tilhørende transformasjons- og konverteringsverktøy er tilgjengelig for nedlasting under GNU General Public License v3.0[9].

De sentrale skandinaviske B&I-institusjonene har vist seg å dele en interesse for forsknings- og utviklingsprosjekter knyttet til e-bøker (se Balling et al. (2014) for en oversikt). Dette har blant annet ført til at ABI, IVA (København) og Högskolan i Borås samarbeider om et skandinavisk mastermene om e-bøker med oppstart våren 2016.

INFORMATION SEEKING

Professor Katriina Byström jobber med bruk av informasjonssystemer på ulike typer arbeidsplasser, med fokus på hvordan disse systemene oppleves i det daglige arbeidet. Foreløpige resultater antyder blant annet at systemene først og fremst brukes til de enkleste arbeidsoppgavene. Til mer komplekse og utfordrende oppgaver foretrekkes kollegaer som informasjonskilder. Andre resultater viser at læring anses som en positiv del av komplekse arbeidsoppgaver som tilhører egen profesjon, mens i enklere profesjonsrelaterte oppgaver, og dessuten i alle typer administrative oppgaver kobles ikke informasjonssøkingen til læring, men heller til frustrasjon. I videre forskning vil Byström undersøke hva dette innebærer for begreper som relevans og brukbarhet i tilknytning til systemevaluering (se Byström (2014) for mer om denne forskningen).

ARKIV

METAINFO inkluderer også forskning innenfor arkiv. Et eksempel på dette er Thomas Sødrings prosjekter basert på fri programvare. Her er formålet å

7 *Professions & Professionalism*: https://journals.hioa.no/index.php/pp
8 Journal Article Tag Suite (JATS): http://jats.nlm.nih.gov/
9 Jats2epub: https://github.com/eirikhanssen/jats2epub

gi innsikt i metoder for å måle datakvaliteten i kommunale arkiv og utvikle verktøy for øke dokumentfangsten fra slike. Et nytt prosjekt (også basert på fri programvare) som undersøker arkivets rolle i tjenesteorienterte arkitekturer er under planlegging.

GJENFINNING

Når det gjelder gjenfinning har ABI-forskere særlig arbeidet innenfor rammene av forskningsnettverket INEX (Initiative for the Evaluation of XML retrieval). I mange år var forskere ved ABI delansvarlige for det såkalte interaktive sporet (Nordlie & Pharo, 2012), der det blant annet ble utført eksperimenter for å studere hvorvidt sluttbrukere foretrakk å benytte hele eller deler av dokumenter for å løse sine gjenfinningsoppgaver. Data fra disse eksperimentene har videre blitt brukt for å prøve ut alternative målemetoder for brukeres innsats i en søkeprosess (Nordlie & Pharo, 2013; Pharo & Nordlie, 2013). Det er også gjort en del arbeid innenfor det såkalte bok-sporet (Preminger, Nordlie, & Pharo, 2010; Preminger & Nordlie, 2011), der det for tiden pågår eksperimenter med bokanbefalinger i sosiale kontekster for å utvikle bedre anbefalingssystemer. Her tas det utgangspunkt i diskusjonsfora hos LibraryThing.

ET BREDT OG VITALT FORSKNINGSOMRÅDE

Den pågående forskningsaktiviteten ved ABI må kunne sies å bekrefte de internasjonale utviklingstrekkene skissert ovenfor. Prosjektene representerer tilnærminger knyttet til ulike kunnskapsorganisatoriske tradisjoner, samtidig som de deler en grunnleggende brukerorientering (selv om enkelte av prosjektene er mer «tekniske» enn andre). Prosjektene er samlet sett rettet mot en rekke institusjoner, brukergrupper og system.

Å gjennomføre en fullstendig kartlegging av hva som skjer av forskning innenfor en slik bredde ville ha vært svært omfattende. Derfor fokuserer artikkelen på hva som skjer av konkret forskning ved én sentral institusjon i Norge. Samtidig er det mye som tyder på at fagområdet trives godt utenfor rammene av en klassisk B&I-institusjon som ABI. Praksisfeltet som forskningsarena er allerede nevnt, men det skjer også spennende forskning relatert til kunnskapsorganisasjon innenfor andre forskningsinstitusjoner.

Den tverrfaglige forskningsgruppa «Semantiske og sosiale informasjonssystemer»[10] ved Universitetet i Bergen inkluderer forskningstilnærminger som minner om flere av prosjektene ved ABI. Når det gjelder forskning på metadata finner vi et ekspertmiljø ved Institutt for datateknikk og informasjonsvitenskap ved NTNU i Trondheim. Her har blant annet førsteamanuensis Trond Aalberg publisert en rekke artikler som undersøker utfordringer ved metadatatransformasjoner (se for eksempel Aalberg og Zumer (2013) for et godt eksempel innenfor det bibliografiske området). Prosjektet «Archives in motion»[11], et samarbeid mellom Nasjonalbiblioteket og to institutt (IFIKK og IMK) ved Universitetet i Oslo, er et annet eksempel på forskning som vil kunne bidra med interessante kunnskapsorganisatoriske perspektiver. Det samme gjelder ressursene og miljøet rundt Språkbanken[12] ved Nasjonalbiblioteket.

Her kunne flere eksempler listes opp, og en mer systematisk kartlegging ville være interessant. Det ville også en mer inngående analyse av metodiske tilnærminger og tverrfaglige forankringer ha vært, og selvsagt kvalifiserte gjetninger om veien videre. I denne sammenhengen får det allikevel holde med noen konkrete eksempler som viser at kunnskapsorganisasjon i 2014 er et vitalt forskningsområde med et bredt nedslagsfelt.

LITTERATUR

Aalberg, T., & Žumer, M. (2013). The value of MARC data, or, challenges of frbrisation. *Journal of Documentation, 69*(6), 851–872. doi:10.1108/JD-05-2012-0053

Balling, G., Dahl, T.A., Mangen, A., Nilsson, S. K., Lund, H., & Höglund, L. (2014). E-bogen: Skandinaviske perspektiver på forskning og uddannelse. *Nordisk Tidsskrift for Informationsvidenskab Og Kulturformidling, 1*(3), 5–19.

Broughton, V., Hansson, J., Hjørland, B., & López-Huertas, M. J. (2005). Knowledge organisation: Report of working group 7. I Kajberg, L. & L.ørring, L. (Red.), *European Curriculum Reflections on Education in Library and Information Science*. København: Royal School of Library and Information Science.

Byström, K. (2014). Searching as a learning activity in real life workplaces. *Searching as Learning Workshop IIiX 2014*.

Eikebrokk, T., Dahl, T. A., & Kessel, S. (2014). EPUB as publication format in open access journals: Tools and workflow. *Code4Lib Journal*, (24).

Gorman, M. (2004). Whither library education? *New Library World, 105*(9-10).

Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization, 2/3*(35), 86–101.

10 http://www.uib.no/fg/ssis

11 http://www.hf.uio.no/ifikk/english/research/projects/archive-in-motion/

12 http://www.nb.no/Tilbud/Forske/Spraakbanken

Holmberg, K., Tsuo, A., & Sugimoto, C. R. (2013). The conceptual landscape of iSchools: examining current research interests of faculty members. *Information Research, 18*(3).

Nordlie, R., & Pharo, N. (2012). Seven years of INEX interactive retrieval experiments - lessons and challenges. *Lecture Notes in Computer Science*, (7488), 13–23.

Nordlie, R., & Pharo, N. (2013). Search transition as a measure of effort in Information Retrieval Interaction. I *Beyond the Cloud : Rethinking Information Boundaries.: ASIST 2013*.

Pharo, N., & Nordlie, R. (2013). Using «Search Transitions» to study searchers' investment of effort: Experiences with client and server side logging. I *Multidisciplinary Information Retrieval 6th Information Retrieval Facility Conference, IRFC 2013* (s. 33–44).

Prebor, G. (2010). Analysis of the interdisciplinary nature of library and information science. *Journal of Librarianship and Information Science, 42*(4), 256–267. doi:10.1177/0961000610380820

Preminger, M., & Nordlie, R. (2011). OUC's participation in the 2010 INEX Book Track. *Lecture Notes in Computer Science*, (6932), 164–170.

Preminger, M., Nordlie, R., & Pharo, N. (2010). OUC's participation in the 2009 INEX Book Track. *Lecture Notes in Computer Science*, (6203), 190–199.

Seland, G. (2014). *User revealment revisited: Knowledge formation in the prefocus stage of information-based work tasks*. Aalborg universitet, Aalborg.

Søbak, V. D. B. (2013). *Desentralisert indekseringspraksis : en studie av det semikontrollerte vokabularet i NRK*. Høgskolen i Oslo og Akershus, Oslo.

Tallerås, K., Massey, D., Husevåg, A.-S. R., Preminger, M., & Pharo, N. (2014). Evaluating (linked) metadata transformations across cultural heritage domains. In *Metadata and semantics research conference* (s. 250–261).

Tuomaala, O., Järvelin, K., & Vakkari, P. (2014). The evolution of library and information science 1965–2005: Content analysis of journal articles. *Journal of the American Society for Information Science and Technology, 65*(7), 1446–1462.

Warner, J. (2001). W(h)ither information science?/! *Library Quarterly, 71*(2), 243–255.

Wiggins, A., & Sawyer, S. (2012). Intellectual diversity and the faculty composition of iSchools. *Journal of the American Society for Information Science and Technology, 63*(1), 8–21. doi:10.1002/asi.21619

Prosjekter med støtte fra Nasjonalbiblioteket

Tema: Kunnskapsorganisering

2012: Universitetsbiblioteket i Bergen:
PhDportal.uib ..... kr 316 000

2013: Høgskolen i Hedmark:
Prøysen-bibliografi ..... kr 200 000

2014: Universitetsbiblioteket i Oslo:
Mapping mot webDewey ..... kr 2 086 000

2014: Universitetsbiblioteket i Bergen:
Digitale fulltekstarkiv integrasjon søk ..... kr 110 000

2014: Storbybibliotekene v/Sølvberget bibliotek og kulturhus:
En bok—en registrering ..... kr 130 000

2014: Biblioteket, Høgskolen Betanien:
Informasjonssøking i sykepleiers praksis ..... kr 176 000

2014: Sykehuset Innlandet:
Bruk av MeSH i fagprosedyrer ..... kr 180 000

2014: Universitetsbiblioteket i Bergen:
BIBSYS Bibliotekdatabase i semantisk web ..... kr 845 000

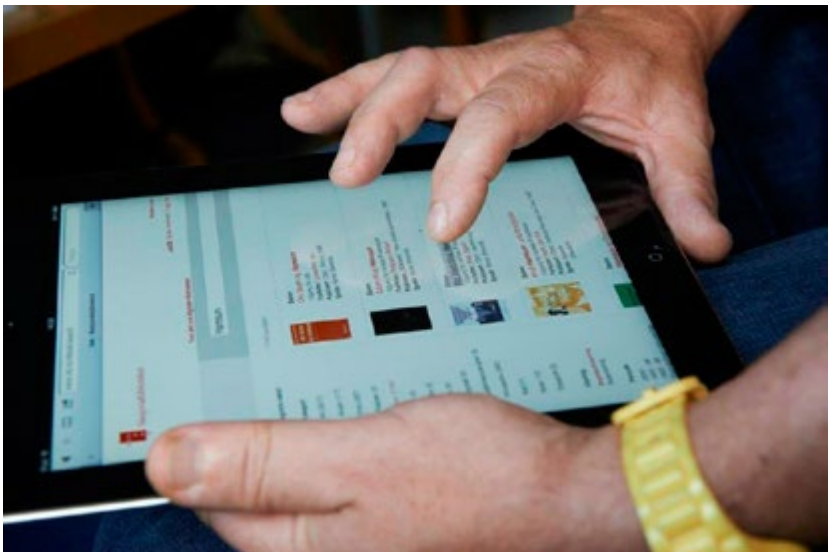


Foto: Ketil Born

Foto: Nasjonalbiblioteket

"A library's function is to give the public in the quickest and cheapest way: information, inspiration, and recreation. If a better way than the book can be found, we should use it."

MELVIL DEWEY