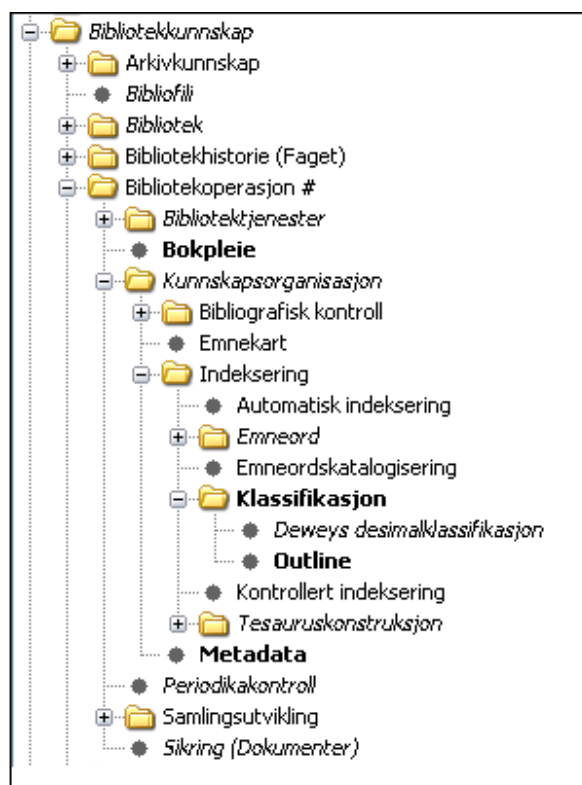


Bibliografisk og emnemessig beskrivelse av UBOs samlinger

Rapport fra en prosjektgruppe

UBO trenger emneord og klassifikasjon som sikrer god gjenfinningskvalitet og som muliggjør samsøk i trykte og elektroniske samlinger.

Systemene må også kunne utnyttes mot semantisk web. Løsningen er å gjøre Humord og Dewey til bibliotekets hovedsystemer, supplert med MeSH og NLM.



Endelig et system som gir meg en samlet oversikt over emnet.

Oslo, 15.februar 2010

Innhold

1 Innledning	1
1.1 Mandat og arbeid	1
1.2 Bibliografisk beskrivelse	2
1.3 Elektroniske dokumenter	3
2 Resultater fra kartleggingen	4
2.1 Emneord	5
2.2 Klassifikasjon	6
2.3 Emnebeskrivelse i elektroniske dokumenter	7
2.4 Oppsummering og konsekvenser	8
3 Hvorfor bruke tid på emnebeskrivelse	9
3.1 Emnebeskrivelsens funksjoner	10
3.2 Argumenter for fortsatt emnebeskrivelse ved UBO	11
3.3 Forutsetninger	12
4 Emnesystemer ved UBO – forslag og tiltak	13
4.1 Ansvar	14
4.2 Bytte av system	14
4.3 Klassifikasjon	14
4.4 Emneord	15
4.5 Prosjektforslag	17
5 Samarbeid og samvirke	19
5.1 Nasjonalt	19
5.1.1 Innen BIBSYS-området	19
5.1.2 Nasjonalbiblioteket	19
5.1.3 Biblioteksentralen A/L	20
5.1.4 SINTEF	20
5.1.5 Enhet for digital dokumentasjon	22
5.2 Internasjonalt	22
5.2.1 Sverige	22
5.2.2 Finland	22
5.2.3 Tyskland – Sveits	23
5.2.4 Overganger mellom klassifikasjonssystemer	23
5.2.5 Semantisk web	23
6 Oppsummering	24
Appendiks	28
A Forslag til Humord-funksjonalitet i BIBSYS Ask	28

B Arbeidsnotater	29
B.1 Norsk komité for klassifikasjon og indeksering — NKKI	29
B.2 A/L Biblioteksentralen	30
B.3 Svenska ämnesord (SAE) – notater fra møte i Kungliga biblioteket	31
B.4 FinnONTO — Finsk nasjonalt ontologiprojekt	33
C Andre relevante prosjekter	33
C.1 KoMoHe	33
C.2 Renardus	34
C.3 Høgskolen i Oslo/JBI	34
D Artikkelresymé	35
Referanser	38

1 Innledning

Prosjektgruppen ble oppnevnt av Bibliotekdirektøren i brev av 26. mars 2009. Gruppen har bestått av Kirsten Al-Araki, Marit Almo, Knut Hegna (leder)[7].

Fristen for fullføring ble satt til 31. desember samme år. I november ble fristen forlenget til 15. februar 2010.

Prosjektgruppen føler at den ikke har hatt tilstrekkelig tid til å studere faglitteraturen og internasjonal utvikling godt nok til å kunne gi en fullgod utredning av samtlige mandatets punkter.

Det samme gjelder gruppens forslag til tiltak og prosjekter.

Vi håper likevel at rapporten gir tilstrekkelig grunnlag for å bestemme en retning på det videre arbeidet.

1.1 Mandat og arbeid

I oppnevningensbrevet konkretiseres mandatet i følgende fire punkter:

1. Gruppen skal kartlegge emnesystemer som er i bruk ved UB

Kommentar: Kartleggingen ble gjennomført før sommeren 2009 og en foreløpig rapport ble da oversendt ledergruppen med tanke på at den kunne være nyttig for andre arbeidsgrupper på UBO. Kartleggingen følger som eget vedlegg til denne rapporten[2].

2. Gruppen skal innhente erfaringer med samordning av emnesystemer i andre land

Kommentar: Gruppen har besøkt Kungliga biblioteket i Stockholm for å studere arbeidet med *Svenska ämnesord*. Vi har også besøkt Semantic computing research group i Helsinki for å studere *FinnONTO*-prosjektet (se appendiks B.3 og B.4).

En i gruppen har deltatt på *European conference on digital libraries* og i tilknytning til det også på *NetWorked Knowledge Organization Systems Workshop* (september 2009).

To i gruppen deltok på HiO/JBIs *Kunnskapsorganisasjonsdagene* 28. og 29. januar 2010.

Gruppen deltok ikke på IFLAs satelitt-konferanse i Firenze i august 2009 som ble arrangert av IFLA Classification and Indexing Section, men har studert foredrag derfra som har relevans for prosjektet, se [5, 9, 17, 26].

I appendiks D gir rapporten et resymé av noen utredninger som vi oppfatter som interessante i vår sammenheng, se [8, 12, 19, 20, 25].

3. **Gruppen skal vurdere muligheten for å slå sammen/legge ned enkelte av systemene som er i bruk i dag**

Kommentar: Se gruppens forslag i avsnitt 4 og 5.

4. **Gruppen skal samarbeide med andre fagmiljøer om metadata**

Kommentar: Gruppen har hatt møter med representanter for Norsk komité for klassifikasjon og indeksering (NKKI), A/L Biblioteksentralen, HiO/JBI og SINTEFs faggruppe for språkteknologi.

Gruppen har også hatt kontakt/samtaler med professor Jan Tore Lønning i *Gruppen for Logikk og naturlig språk* ved Institutt for informatikk; Christian-Emil Ore som er forskningsleder ved *Norsk dokumentasjonssenter for språkteknologi* og faglig leder ved *Enhet for digital dokumentasjon* (UiO/ILN) og med Andrea Gasparini ved Digit som har spesiell interesse for emnekart.

Knut Hegna er blitt medlem av Samordningsgruppa for Språkrådets Dataterminologiprojekt.

Utover arbeidet med mandatet arrangerte prosjektgruppen – på bibliotekdirektørens oppfordring – et frokostmøte den 26. november 2009 om bibliografisk registrering med innledere fra Nasjonalbiblioteket og BIBSYS. Møtet samlet rundt 50 tilhørere. Etter frokostmøtet ble det arrangert et uformelt møte mellom NB, BIBSYS og representanter fra UBO.

I desember 2009 leverte gruppen et innspill til UBOs budsjett for 2010. Dette gjaldt forslag til finansiering av en konkret utvikling i BIBSYS for å gjøre Humord til en integrert del av BIBSYS Ask. Forslaget vil ha stor overføringsverdi til seinere system. En noe revidert versjon av innspillet er gitt i appendiks A.

I forbindelse med utarbeidelsen av budsjettinnspillet laget gruppen et navigeringsverktøy for Humord som et frittstående front-end-program for emnesøk i BIBSYS Ask (dels i Humord-feltet, dels i flere emnefelt og endelig i en kombinasjon av emnefelt og ord-i-tittel) og ulike søk mot BiblioFILS Samsøk-tjeneste. Programmet er vedlagt prosjektrapportens e-versjon.

1.2 Bibliografisk beskrivelse

Tittelen for prosjektgruppen nevner *bibliografisk beskrivelse*, men mandatet spesifiseres til å gjelde emnesystemer. Vi har valgt å holde oss til de fire emnepunktene. Her vil vi bare peke på noen problemområder når det gjelder øvrig bibliografisk beskrivelse. Disse bør det gripes fatt i på overordnet plan:

1. Samme materiale registreres i flere databaser, noe som medfører dobbeltarbeid (BIBSYS, X-port, DUO, Frida).
2. Det mangler standardisering av registreringsformater systemene i mellom (BIBSYS, X-port, DUO, Frida).

3. Utveksling av data systemene i mellom er ufullstendig og/eller manglende. Det er viktig å fastslå hvilke(t) system(er) som skal være primærkilde for ulike data, slik at oppdateringer der kan følges opp av de andre basene.
4. Store mengder katalogposter for elektroniske dokumenter får summarisk behandling ved import (se også avsnitt 1.3 nedenfor).
5. Det mangler samordning av registrering (katalogisering) av elektroniske og trykte dokumenter.
6. Det er viktig å følge nasjonale og internasjonale standarder for å kunne stå friere i valg av systemer og for å kunne drive effektiv utveksling av data (import og eksport). Dette berører katalogiseringspraksis (færre avvik) og datamodell (overgang til MARC 21) i BIBSYS.

I neste avsnitt behandles noen av problemene med samordning mellom trykte dokumenter i bibliotekets *eie* og pakker av elektroniske dokumenter man kjøper *tilgang* til.

Pakkevis import av emnebeskrivelse av e-dokumenter blir behandlet i avsnitt 2.3.

1.3 Elektroniske dokumenter

Elektroniske dokumenter omfatter både løpende ressurser og monografier. Dokumentene kan være reproduksjoner av trykte versjoner eller “født digitalt” uten tilsvarende versjon i papirbasert eller annen form (for eksempel CD-Rom).

Nettdokumenter blir anskaffet som:

- pakker (forlag, agenter/aggregatorer)
- enkeltkjøp (forlag, agenter/aggregatorer)
- digitaliseringsprosjekt (nasjonalbibliotek, Google)
- gratisressurser (open access, offentlige dokumenter)

De er tilgjengelige via leverandørenes servere og grensesnitt og leveres i ulike fysiske formater.

Etterspørselen etter elektroniske dokumenter øker. Både forskere og studenter ønsker rask tilgang til den informasjonen de trenger.

Det er viktig for UBO å eksponere sine tilbud. Det fremgår bl.a. av policydokumentet til prosjektgruppen som har undersøkt hvordan “UBOs samlinger skal reflektere UiOs forskning”¹.

I sluttrapporten under hovedpunktet om eksponering sier de bl.a.:

¹Se <http://www.ub.uio.no/prosjekter/samlinger/> (2010-02-08)

- De elektroniske tjenestene må markedsføres bedre, og de må også få mer fleksible grensesnitt.
- Vi må se på muligheter for å lage skreddersydde, automatiserte nyhetslister for enkelte forskere ut fra faglige prioriteringer og ønsker.

Begge punktene er avhengige av gode gjenfinningsmuligheter. Særlig når det gjelder elektroniske dokumenter, er disse usikre i dag. Uten emneord og klassifikasjon er det ikke mulig å tilby fullstendige, faglige oversikter over tidsskrifter og monografier.

Bibliotek verden over arbeider med å få til et samarbeid om deling av metadata og krav til bedre kvalitet fra leverandørene. Dette krever samordninger og systemer som snakker sammen/kan mappes mot hverandre. Problemet er særlig tydelig i forbindelse med den store mengden elektroniske dokumenter som nå blir levert uten kvalitetssikrede data:

- for klassifikasjon
- for emneord
- for autoritetsdata
- forskjellige alfabet/ulike transkripsjonsregler

De største problemene med emneord og klassifikasjon er ikke på den tekniske men på den menneskelige siden. Forskjellige (faglige) kulturer har forskjellige terminologier og begrepsstrukturer. UBO bør tilstrebe en så samordnet beskrivelse som mulig innen sitt nedslagsfelt. Dette vil lette kommunikasjonen til andre systemer.

Metadata må i størst mulig grad hentes fra kilder som holder høy standard og fra de mest brukte emnesystemene slik at samsøk og overganger kan muliggjøres. I all hovedsak vil det dreie seg om å kople seg på nasjonalbibliografiske data², selv om de ikke *alltid* er de beste.

En betingelse for at fellesdata vil bli akseptert av alle, er gode standarder og sentral autoritet. Helt lokale data vil kunne legges til av hvert enkelt bibliotek.

2 Resultater fra kartleggingen

En detaljert versjon av kartleggingen foreligger i et eget dokument[2] som heretter blir kalt *Kartleggingen*. I dette avsnittet vil vi oppsummere resultatet.

Kartleggingen viser at UBO i dag registrerer emnedata i tilsammen 12 ulike MARC-felt i BIBSYS. I tillegg kommer de obligatoriske MARC-feltene 610-630 som gjelder emneinnførsler for personer, korporasjoner og titler.

²Slik sammenkøpling kan gjøres ved at postene inkluderer identifikatorer fra nasjonalbibliografiene: LC-NR, BL-nr osv.

Emneord					
	UHS	UJUR	UMH	UMN	E-pub^a
LCSH (650)^b					Import
MeSH (660)			Alle fag		
Lokalt kontrollerte (687)		IMR ^c		INF ^d	
Frie nøkkelord (691)	SV-fag	UJUR ^e		Alle unntatt INF	
Humord (698)	HF-fag og teologi				

^aGjelder import. For titler behandlet manuelt brukes de samme systemene som for trykte dokumenter. Se ellers Kartleggingen, avsnitt 6.

^bLCSH – Library of Congress subject headings – er ikke søkbart i BIBSYS.

^cInstitutt for menneskerettigheter bruker en engelsk tesaurus.

^dInformatikkbiblioteket bruker sin egen kontrolliste.

^eUJUR bruker et kontrollert vokabular.

Tabell 1: *Bruken av emneordsfelt i avdelingene og for e-publikasjoner. MARC-felt-tag i parentes.*

2.1 Emneord

En oversikt over bruk av emneordsfelter fins i tabell 1. Her er noen kommentarer til enkelte deler av tabellen:

- UBO bruker i alt 5 ulike MARC-felt for emneordsregistrering (650, 660, 687, 691 og 698).
- Frie nøkkelord brukes delvis ved alle avdelinger bortsett fra UMH.
- Emneord fra kontrollerte vokabularer brukes ved alle avdelinger. Bare to av disse (MeSH³ og Humord) har egne MARC-felt med autoritetskontroll og autoritetslister i BIBSYS.
- Kontrollerte emneord for Informatikkbiblioteket (INF) og Institutt for menneskerettigheter (IMR) registreres i et felles MARC-felt for lokale, kontrollerte emneord (687). Det er ingen autoritetskontroll av det som legges inn.
- Emneord fra kontrollerte vokabularer registreres også som frie nøkkelord (UJUR).
- De kontrollerte emneordene blir dårlig utnyttet i BIBSYS Ask:

³Medical Subject Headings.

- Autoritetslistene er ikke tilgjengelige for brukerne i BIBSYS Ask.
- Strukturen i emneordsystemene utnyttes ikke ved søking.
- Synonymer har hittil ikke vært søkbare⁴.
- Kontrollerte emneord fra LCSH importeres til BIBSYS via poster i e-bokpakker, men er foreløpig ikke søkbare i BIBSYS.

Biblioteket nedlegger mye arbeid i emneordsindeksering, men manglende samordning vanskeliggjør gjenbruk av registreringsdata og skaper gjenfinningsproblemer. Sammen med mangelfull funksjonalitet i søkeprogrammene gjør dette at brukeren får relativt lite igjen for bibliotekets arbeidsinnsats.

2.2 Klassifikasjon

En oversikt over bruk av klassifikasjonsfelter fins i tabell 2 på neste side. Her er noen kommentarer til enkelte deler av tabellen:

- UBO bruker 7 klassifikasjonssystemer med egne MARC-felt i BIBSYS.
- I tillegg brukes flere lokale systemer. Disse registreres dels som lokal-klassifikasjon i 087, dels som oppstillingssignaturer.
- De internasjonale systemene Dewey og UDK brukes ved flere avdelinger, men praksis er ikke samordnet. Flere utgaver er i bruk både av Dewey og UDK⁵.
- Det finnes lokale registre både til klassifikasjonstabellene og til klassifikasjonspraksis. De er av varierende kvalitet, men bare ett – registeret til Dewey-praksis ved UHS – er overført til BIBSYS og oppdateres løpende.
- Ingen registre er tilgjengelige for brukerne i BIBSYS Ask.
- Strukturen i systemene utnyttes ikke ved søking.

Heller ikke her blir bibliotekets registreringsdata tilfredsstillende utnyttet: En flora av systemer fører til lite gjenbruk. Mange klassifikasjonskoder er fullstendig utilgjengelige for brukerne fordi det mangler en verbal søkeinnang (register). Der det finnes registre⁶, ligger systemene for disse utenfor Ask og er ikke integrert i søkesystemet. De lokale registrene er ikke samordnet terminologisk eller systemmessig.

⁴Seint på høsten 2009 ble synonymer fra Humord gjort søkbare, delvis som resultat av prosjektgruppens arbeid med budsjettinnspillet.

⁵Se avsnitt 7 i Kartleggingen.

⁶UHS har register til Dewey, UJUR har register til L-skjema, instituttbibliotek på UMN har diverse registre på WWW.

Klassifikasjon					
	UHS	UJUR	UMH	UMN	E-pub
NLM (060)			Alle fag		Import
CR (062)				INF	
UDK (080)	Arkeologi			Geologi, geografi, kjemi, muséene	
Dewey (082)	Alle fag unntatt arkeologi	Krimino- logi		Felles- biblioteket, biologi	Import
UBO (083)	Eldre samlinger. Oppstilling teologi	L-skjema (jus)	G-skjema (medisin) eldre samlinger	Biologi, geofysikk, geografi	
MSC (084)				Matematisk	
Lokal klassifikasjon (087)		IMR NIFSP ^a		Biologi, farmasi, fysikk/ astrofysikk ^b	

^aNordisk institutt for sjørett, avdeling for petroleumsrett

^bFysikk og astrofysikk bruker Inspec-klassifikasjon

Tabell 2: Bruken av klassifikasjonsfelt i avdelingene og for e-publikasjoner. MARC-felt-tag i parentes.

2.3 Emnebeskrivelse i elektroniske dokumenter

Importerte data for e-bøker inneholder ofte emnebeskrivelse fra store, internasjonale systemer: emneord fra LCSH og MeSH og klassifikasjon fra Dewey og NLM (National library of Medicine). Dette betyr ikke nødvendigvis at emnebeskrivelsen holder høyt fag-faglig og/eller bibliotekfaglig nivå.

Emnedata i SFX er av varierende kvalitet og importeres ikke til BIBSYS.

Problemer:

- Emneordene er på engelsk.
- LCSH er ikke søkbare i BIBSYS.
- Det er avvik fra navneformer (personer, organisasjoner).
- Det er avvik fra norske transkripsjonsstandarder, bla i forbindelse med navneformer.

	Antall emneordsfelt brukt	Antall klassifikasjonsfelt brukt	Totalt antall felt brukt
E-pub	2	2	4
UHS	2	3	5
UJUR	2	3	5
UMH	1	2	3
UMN	2	6	8 ^a
UBO^b	5	7	12

^aInklusive diverse lokale systemer.

^bUBO-tallene er *ikke* aritmetisk sum av hver kolonne

Tabell 3: Bruken av emnefelt i avdelingene og for e-publikasjoner.

- Lokal Dewey-praksis samsvarer ikke alltid med importerte Dewey-data.

Når det i BIBSYS finnes parallelle fysiske versjoner av samme tittel, får hver versjon sin egen objektpost. Siden det skal registreres lenker mellom dem, vil det i disse tilfellene finnes en indirekte emnetilgang til den elektroniske versjonen via den papirbaserte.

Blir papirposten fjernet, for eksempel ved kassering, kan også emnetilgangen til den elektroniske versjonen forsvinne. Enkel kopimulighet av emnedata mellom objektposter burde kunne løse dette.

Se også avsnitt 6 i Kartleggingen.

2.4 Oppsummering og konsekvenser

Kort oppsummert ser situasjonen ut som i tabell 3. En så kaotisk situasjon har mange ulemper:

1. Gjenfinningsproblemer

I en stor base som BIBSYS finner en alltid “noe” om et emne. Problemet blir dels å finne de “riktige” dokumentene, dels å få en samlet oversikt over hva biblioteket har. Når samme emne beskrives med ulike klassifikasjonskoder, ulike synonymer og på flere språk, blir emnesøk som å finne “nåla i høystakken”. En får problemer både med fullstendighet, presisjon og relevans.

Av samme årsaker blir det problemer med samsøk i trykte og elektroniske samlinger.

2. Dårlig utnyttelse av registrerte emnedata

Registrering har gjenfinning som mål. I dag registrerer biblioteket store mengder emnedata som ikke utnyttes ved søking: Synonymer som ikke er søkbare, klassifikasjon som mangler register, autoritetslister som ikke

er tilgjengelige for brukerne, ingen utnyttelse av strukturen i emnesystemer ved søking.

Dette er delvis systemproblemer, men med en mer samordnet emnepraksis vil biblioteket også stå sterkere i krav overfor BIBSYS og andre systemleverandører.

3. Liten gjenbruk

Et stort antall systemer og manglende samordning mellom systemene gjør det vanskelig å gjenbruke registrerte emnedata på en effektiv måte.

Kartleggingen avdekker tre behov:

- *Mengden av systemer må reduseres.*
- *Der samme system brukes av flere avdelinger må bruken samordnes og standardiseres.*
- *Det må lages overganger (mapping) mellom de systemene man bestemmer seg for å bruke.*

3 Hvorfor bruke tid på emnebeskrivelse

Emnebeskrivelse har en lang tradisjon i bibliotekene og er en viktig målsetting for bibliografisk registrering.

Charles A. Cutter formulerte i 1904 målsettingen for emnegjenfinning i bibliotek katalogen slik[10]:

- To enable a person to find a book of which the [...] subject is known.
- To show what the library has [...] on a given subject, or in a given kind of literature, and
- To assist in the choice of a book as to its character.

I 1904 var gjenfinningsspørsmålets vinkling det enkelte biblioteks samling og bare det. Man kunne i en slik situasjon tillate seg å bygge opp spesifikke emnesystemer for hver enkelt fysiske samling.

Vinklingen er fortsatt viktig, men økt tverrfaglighet, deling av katalogdata, tilgang til elektroniske boksamlinger og datakommunikasjon har reist spørsmålet på ny, men nå med et mye videre siktemål og nedslagsfelt.

Det er liten tvil om at den engangs arbeidsinnsatsen biblioteket gjør med bibliografisk og emnemessig beskrivelse ved anskaffelsen fører til "evig" gjentatt besparelse av tid og penger for den som søker informasjon og kunnskap i en tid med overflod.

3.1 Emnebeskrivelsens funksjoner

Klassifikasjon

- samler litteraturen i klasser
- kan brukes til systematisk, emnedelt hylleoppstilling og muliggjør “browsing” langs hylla og på skjerm
- Søking
 - En verbal emneinngang (“fanekort”, register) er nødvendig som grensesnitt mot brukerne.
 - Systemets struktur kan utnyttes til hierarkisk søking og navigering.

Kontrollerte emneord

- gir presis emnebeskrivelse av det enkelte dokument
- gir mulighet til å fange opp emner på tvers av klasser og disipliner
- er fleksibelt (emneord kan integreres i felles søkefelt, kfr. BIBSYS Ask)
- muliggjør søk på synonymer og relaterte emner (se-også-relasjoner)
- tesauri muliggjør navigering i emne- og begrepsstrukturer.

I dag stilles det spørsmål ved behovet for emnebeskrivelse i bibliotekene. Flere store bibliotek legger ned sine gamle emnesystemer og satser i stedet på:

- internett, “googlesøk”, fulltekstsøking i e-dokumenter
- brukerproduserte metadata
- kjøp av store databaser (e-bøker, e-tidsskrifter, artikkelbaser) med leverandørproduserte metadata
- datasystemer med innhøster (aggregator) og sentralisert søk (kfr Summa og Primo)

Samtidig ser vi at det drives mange prosjekter i en annen retning, dvs gjenbruk av emnesystemer og utvikling av overganger mellom emnesystemer.

Dette drives delvis fram av frustrasjon over treffmengden og ordningen i de kommersielle systemene (som Google), men også i et ønske om å bringe tilbake det bibliotekene delvis mistet ved overgangen til data: struktur og relasjoner mellom begreper.

Informasjonsflommen tvinger nå også fram bedre metoder for å lenke strukturerte ressurser på Web. Skal bibliotekenes ressurser spille noen rolle i utviklingen, må de både strukturere sine egne data bedre og stille dem til rådighet på nettet på en annen måte enn før.

I et foredrag på en IFLA-konferanse i fjor foreslo for eksempel Lois Mai Chan at bibliotekene for det første burde stille sine kontrollerte vokabularer til rådighet for systemer som driver tjenester med brukergenererte emnedata (folksonomier) for å bidra til samordning av terminologi og gi støtte for synonymkontroll. For det andre burde man se på muligheten for å lage overganger mellom de brukerproduserte emnedataene til bibliotekenes systemer[9].

Yi og Chan har gjennomført en studie for å se på muligheten for å mappe folksonomier til Library of Congress subject headings[29].

3.2 Argumenter for fortsatt emnebeskrivelse ved UBO

Krav til gjenfinningskvalitet (økt presisjon og økt fullstendighet)

- Titler er ofte lite pregnante, av og til direkte villedende (støy).
- Fritekstsøk gir problemer med presisjon og relevans.
- Leverandørproduserte metadata representerer ulike systemer. De er vanligvis ikke på norsk og dessuten av varierende kvalitet.
- Brukerproduserte metadata preges av tilfeldighet og gir ikke samme gjenfinningskvalitet som de kontrollerte vokabularene.

Språklige hensyn

- Norsk språkpolitikk har som hovedmål “å sikra det norske språkets posisjon som komplett og samfunnsberande språk i Noreg”. Rapporten *Mål og mening* (St.meld. nr.35 2007-2008)[11] peker bla på behovet for norske fagtermer i arbeidet med å opprettholde og videreutvikle norsk fagspråk. I samsvar med dette må norsk språk også være standard for litteratursøking i norske databaser.
- Samlingene ved UBO representerer et stort språklig mangfold både når det gjelder trykte og elektroniske samlinger. Norskspråklig emnebeskrivelse muliggjør emnesøk på tvers av språk i vår egen samling.
- Klassifikasjon er også språkuavhengig og bygger bro til andre samlinger.

Trykte samlinger er fortsatt viktige

- De trykte dokumentene vil fortsatt være viktige i overskuelig fremtid (varierer med fagområde).

- Ikke-vestlig litteratur foreligger primært i trykt form.
- Eldre samlinger er sentrale forskningsobjekter, kfr humaniora.

Den semantiske weben

Det er viktig at vi innretter oss slik at UBOs systemer kommer i inngrep med nasjonale og internasjonale systemer og at vi dermed legger til rette for å samordne både utover (at vi kan søke i andres data) og innover (at andre kan søke i våre).

Kontrollerte vokabularer (ontologier) spiller en sentral rolle i arbeidet med samordning/utnyttelse av bibliotek- og webressurser i større sammenhenger enn de samlingene biblioteket *eier* og det materialet biblioteket gir *tilgang* til⁷.

Målsettinger og tiltak for videre arbeid

Målsettingene med emnebeskrivelsene kan kort beskrives slik:

- sikring av gjenfinningskvalitet generelt
- bruk av norsk språk ved emnesøk
- effektiv gjenbruk og utnyttelse av emnedata
- samvirke med andre miljøer (f.eks arkiv og bibliotek)

3.3 Forutsetninger

Det trengs smidige, moderne verktøy for å kunne drive effektiv og kvalitativt god emnebeskrivelse med den dokumentmengden vi har, uavhengig av fysisk form. Slike verktøy har vi ikke i dag og forslag på dette feltet vil måtte inngå som kravspesifikasjoner til et nytt system. Siden det i all hovedsak ikke er noe poeng å komme med endringsforslag til BIBSYS på dette tidspunktet, vil en del av forslagene måtte formuleres som kravspesifikasjoner til nytt system:

- enkel gjenbruk av tidligere beskrivelse for nyere utgaver og andre dokumenter
- enkel gjenbruk av andres registreringer (fra andre kilder)
- enkel oppslags- og kontrollmulighet i egen og andres praksis
- enkelt å gjøre gjennomløpende endringer
- rike søke- og oppslagsmuligheter i klassifikasjonstabeller og registre

⁷Se f.eks FinnONTO-prosjektet i appendiks B.4.

- enkelt å forfølge henvisninger
- verktøy for å redigere begrepshierarkier
- Det må være mulig å knytte autoriteter til alle (relevante) felt i den bibliografiske og emnemessige beskrivelsen.
- Systemet må takle synonymer både i registrering og søking.
- Det må være mulig å åpne faglig baserte underavdelinger av et autoritetsregister.
- Register til klassifikasjon må integreres i sluttbrukergrensesnitt.
- Autoritetssystemet må tillate bruk av flere språk (f.eks. engelsk) parallelt med og som supplement (latin) til hovedspråket (norsk).

Verktøyene må inkludere funksjonalitet som i dag er normalt i de fleste systemer: stavekontroll, kontroll av syntaktiske mønstre, kontroll av numeriske verdier, kontroll mot ulike registre.

4 Emnesystemer ved UBO – forslag og tiltak

De tiltakene som trengs for å tilfredsstille behovene som Kartleggingen har avdekket, kan i all korthet beskrives slik:

- Antall systemer må reduseres.
- Bruken av samme system må standardiseres og samordnes over avdelingsgrensene.
- Det må lages overganger mellom systemer i bruk, både for trykte og elektroniske samlinger.
- Det må satses på etablerte systemer som:
 - kan mappes mot internasjonale systemer
 - kan brukes mot semantisk web
 - ivaretar norsk språkbruk

Disse generelle forslagene blir konkretisert i fortsettelsen.

4.1 Ansvar

Det må oppnevnes eller tilsettes en person som får som ansvar å

- lede arbeidet med emnebeskrivelse på UBO-plan, praktisk og redaksjonelt
- koordinere UBOs arbeid i forhold til nasjonale og internasjonale bibliotek- og språkmiljøer
- følge opp funksjonalitet i forhold til anvendt datasystem
- sørge for kontinuerlig opplæring av nytilsatte i bruk av systemene for internt og eksternt bruk
- sørge for kompetansevedlikehold av seg selv og lokalt ansvarlige

Personen må rapportere til Bibliotekdirektøren og må rekrutteres raskt for å sikre kompetanseoverføring.

4.2 Bytte av system

Der man velger å gå over til nytt system – enten det gjelder klassifikasjon eller emneord – stilles man overfor problemet med å knytte sammen søkingen i gammel og ny samling. Det er to måter å løse dette på:

- foreta en retrospektiv registrering – manuelt, automatisk eller ved en blanding av disse
- lage en overgang (mapping) mellom gammelt og nytt system i søkesystemet

Avdelingene må avgjøre hva som er mulig, nødvendig og tilstrekkelig.

Viktig: Behovet for konverteringsoppdrag hos BIBSYS må avklares før man går over til én felles bibliotekkode for UBO.

Der man velger å beholde sitt system, må man om nødvendig lage overganger til UBOs hovedsystemer, f.eks. mellom Dewey og UDK, mellom Dewey og Inspec osv.

4.3 Klassifikasjon

Forslag:

Dewey innføres som klassifikasjonssystem for alle fag hvor det ikke er tungtveiende, faglige grunner for å bruke et annet system. NLM bør fortsatt brukes for medisin.

UHSs Dewey-register (USVD) utbygges for flere fag og gjøres til et UBO-register.



Figur 1: Trenger vi en slik “enhet” ved UBO?

Tiltak Dewey:

- Organisering: Arbeidet med Dewey må organiseres på lokalt plan og samordnes for hele UBO.
- Kompetansebygging og -vedlikehold: Det må sørges for kontinuerlig opplæring i bruken av Dewey for å sikre enhetlig og konsistent praksis.
- Standardisering: Bruken av systemet må standardiseres med internasjonal praksis (gjelder for eksempel 800 litteratur).
- Terminologien i UHS Deweyregister må oppgraderes (standardisering, synonymhåndtering og samordning med Humord).
- Det bør inngås et terminologisk samarbeid med Nasjonalbiblioteket i forbindelse med oversettelsen av WebDewey til norsk (se 5.1.2).
- På sikt bygges UHS lokale register ut til et felles Dewey-register for hele UBO. Arbeidet bør organiseres som et prosjekt.
- Retningslinjer og hjelpemidler: Det må utarbeides felles retningslinjer. (UHSs Dewey grunnotat[13] revideres).
- **Ansvar:** UHS i samarbeid med andre avdelingsbibliotek ved UBO

4.4 Emneord

Forslag

Det satses på de etablerte systemene Humord og MeSH.

Begge vokabularer tilpasses formater som kan mappes mot importdata og samordnes med semantisk web.

Norskspråklige kontrollerte emneord innføres for hele UBO (real-fag og jus integreres i tesaurusen Humord som selvstendige del-hierarkier).

Tiltak MeSH

- Tesaurusen oversettes til norsk (arbeidet er i gang).
- Det pågår også arbeid med å etablere tesaurusen som en ontologi.

Tiltak Humord

Driftsansvaret for Humord overføres til UBO og dette medfører:

- Det inngås avtaler med BIBSYS og deltakerne i Humord-samarbeidet om utvikling og bruk av tesaurusen, inklusive økonomi.
- Nye juridiske avtaler om rettigheter og samarbeidsforhold må formuleres.
- Det oppnevnes en koordinator for arbeidet med Humord.
- **Ansvar:** UBOs ledelse i samarbeid med BIBSYS, UB Bergen og UB Tromsø.

Lokalt ved UBO medfører forslagene:

- Organisering: Arbeidet med emneord må organiseres lokalt og samordnes for hele UBO.
- Kompetansebygging: Det må gis nødvendig opplæring i emneordsindeksering.
- Retningslinjer: Det må snarest mulig utarbeides felles retningslinjer for arbeidet.
- Norskspråklige vokabularer etableres for realfag og jus:
 - Real-fag: Et omfattende prosjekt er i gang. Frie emneord (691) konverteres til lokale kontrollerte emneord (687). Det er utviklet et eget konverteringsprogram for arbeidet. Programmet har overføringsverdi og kan brukes i forbindelse med tilsvarende arbeid ved andre enheter.
 - Jus: Arbeidet må baseres på den eksisterende emneordslista ved UJUR og planlegges og gjennomføres som et prosjekt.
 - Begge vokabularer ombygges til tesaurusstruktur og integreres som selvstendige faghierarkier i tesaurusen Humord.

- Etablering av Humord som ontologi og mapping mot LCSH (se prosjektforslag nedenfor).
- **Ansvar:** Avdelingsbibliotekene.

Tesaurusen Humord må videreutvikles:

- Den eksisterende tesaurusen bygges ut for SV-fag. Arbeidet organiseres som et prosjekt.
- Nødvendig etterarbeid i forbindelse med flertallsprosjektet (2009) fullføres.
- Terminologibruken samordnes med Dewey-registeret der det er naturlig.
- Innplassering av andre fagområder (real-fag, jus) forberedes.
- **Ansvar:** UHS og Humordkoordinator i samarbeid med UMN og UJUR.

Oppgradering av hjelpemidler etc:

- All dokumentasjon i forbindelse med Humord overføres til UBOs tjenermaskin og vedlikeholdes der.
- Humord nettsted revideres og forenkles.
- Humord håndbok[27], indekseringsregler[1] og andre hjelpemidler revideres.
- De eksisterende juridiske avtalene om rettigheter[4] og samarbeidsforhold[3] erstattes av nye (se ovenfor).
- **Ansvar:** Humordkoordinator i samarbeid med BIBSYS og UBOs web-redaktør.

4.5 Prosjektforslag

Viktige mål for samordning av UBOs emnesystemer er standardisering, mapping mellom systemer og derved samsøk både i trykte og elektroniske samlinger. Et annet viktig mål er utnyttelse av vokabularer i arbeidet mot semantisk web.

Under forutsetning av at forslagene ovenfor følges opp, vil resultatet være en norsk tesaurus som omfatter de fleste av UBOs fagområder unntatt medisin. For medisin er det naturlig å satse på videre bruk av den etablerte, internasjonale tesaurusen MeSH. Arbeid med å oversette MeSH til norsk og med etablering av en ontologisk struktur er allerede i gang.

Det videre arbeidet med et felles vokabular for UBO bør realiseres i form av et prosjekt. Nedenfor skisseres en ramme for et slikt prosjekt.

Forslag:

Tesaurusen Humord omarbeides til en ontologi.

Det etableres en struktur for Humord-baserte indekstermer som kan mappes mot LCSH. Et alternativ til fulle LCSH termstrenger kan for eksempel være basert på FAST⁸. Det kan også tenkes en mapping slik det gjøres i Tyskland[17]. Dette må utredes.

Begrunnelse:

En ontologistruktur åpner muligheter for:

- å ta inn termer fra flere språk
- å utnytte vokabularet mot semantisk web

Indekstermer med LCSH-basert struktur:

- åpner for samordning av søk i trykte og elektroniske dokumenter ved mapping mellom Humord og LCSH fra importerte data
- gir mulighet for mapping også mot MeSH og andre prekoordinerte emneordsystemer
- gir større mulighet også for mapping mot Dewey (for eksempel via LCSH)

Standardisering av emneordsystemene i BIBSYS nye modul for autoritetsfiler gir mulighet for:

- å registrere termstrenger med struktur som kan registreres i MARC 21
- å ha autoritetskontroll fra tesaurusen mot de enkelte delfelt

Emneordsområdet er forsømt på nasjonalt plan i Norge. UBO er Norges største fagbibliotek og har nødvendig kompetanse på et bibliotekfaglige og fag-faglige plan.

Dette er et prosjekt som bør ha interesse også utenfor UBO i bibliotekmiljøet og i andre miljøer enn bibliotek. Det kan derfor være nødvendig å trekke inn andre institusjoner, enten i et direkte samarbeid eller som bidragsyttere i forbindelse med finansiering.

Naturlige samarbeidspartnere er Nasjonalbiblioteket og BIBSYS, men det bør også søkes IT- og systemkompetanse utenfor bibliotekmiljøet. Ut fra prosjektgruppens undersøkelser vil det være interessant å søke samarbeid med SINTEF og Enhet for digital dokumentasjon (UiO/ILN), men dette krever finansiering enten gjennom omprioriteringer internt eller ved søknad om midler fra eksterne kilder. Det vil derfor være naturlig å søke støtte fra andre universitets- og høgskolebibliotek, ABM og Forskningsrådet.

⁸FAST (Faceted Application of Subject Terminology) er et kontrollert vokabular basert på LCSH, men med en enklere struktur[28]

5 Samarbeid og samvirke

5.1 Nasjonalt

Dersom UBO tar tak i problemene slik det er foreslått i forrige kapittel, vil biblioteket kunne fungere som et sentrum også på nasjonalt plan ved å tilby sine systemer til fri bruk.

5.1.1 Innen BIBSYS-området

Emneord

Mange BIBSYS-bibliotek har utviklet kontrollerte emneordslister.

Riksantikvarens bibliotek har for eksempel utarbeidet en tesaurus (Ordnøkkel) som dekker deres faglige nedslagsfelt⁹. Det er et visst terminologisk sammenfall mellom denne tesaurusen og Humord, mens strukturene avviker.

UMB arbeider med AGROVOC – en tesaurus for jordbruk (FAO) som er i ferd med å konverteres til en ontologi.

Bergen offentlige bibliotek samarbeider med Nasjonalbiblioteket, Biblioteksentralen og andre musikkmiljøer i Norge om en ny tesaurus for musikk. Termene skal brukes både for beskrivelse av musikk (musikktrykk, lydfestinger) og som tradisjonelle emneord. Tesaurusen ferdigstilles i løpet av 2010.

UBO bør etterhvert innby andre bibliotek til terminologisk samarbeid og også forsøke å dra nytte av det arbeidet andre har gjort på fagfelter der UiO/UBO har mangler eller ikke er dekket. Samtidig må vi stille vårt eget arbeid til rådighet for andre.

Klassifikasjon

I BIBSYS ligger det emnedata av forskjellig karakter og system på mange millioner poster. Hver enkelt post kan ha blitt beskrevet med ulike emneordssystemer og ulike klassifikasjonssystemer.

Disse enestående dataene *kan* danne grunnlag for å lage overganger mellom systemer basert på statistisk analyse. Jo oftere to beskrivelser opptrer sammen, desto sterkere kan vi si at beskrivelsene (f.eks. DDC og UDK) betegner samme begrep.

Det er også mulig å strekke analysen enda lenger ved å lage overganger mellom overganger utfra ulike delmengder av poster.

Det bør settes i gang forskning på disse dataene for å se om de gir grunnlag for å lage automatiske koplinger mellom ulike klassifikasjonssystemer.

5.1.2 Nasjonalbiblioteket

Norsk komité for klassifikasjon og indeksering har vist interesse for UBOs emnearbeid og er interessert i å få til et samarbeid, ikke minst i forbindelse

⁹Se <http://ordnokkelen.ra.no/multites> (2010-02-08)

med oversettelsen av WebDewey til norsk (se avsnitt B.1).

Her kan UBOs faglige kunnskaper og kunnskaper om fagterminologien komme til nytte.

Det vil også kunne være mulig å utnytte vokabularet i UBOs Deweyregister til samme formål. Der poster i basen er indeksert både med Humord og Dewey vil vokabulartilfanget bli ytterligere beriket til støtte for oversettelsen.

Selve utarbeidingen av en norsk fullstendig versjon av WebDewey vil representere et knutepunkt både som bindeledd direkte gjennom klassifikasjonskoder og fra klassifikasjonskoder til andre emnesystemer som MeSH, LCSH.

5.1.3 Biblioteksentralen A/L

Vi har gjort små forsøk som viser at det allede nå er et visst sammenfall i terminologibruk mellom Humord og de emneordene som er knyttet til postene i folkebibliotekene og som de i stor grad har fått via Biblioteksentralens emneord til Dewey-klassifikasjon.

Naturlig nok gjelder dette i de øvre delene av Humord-tesaurusen og ikke på de mer fagspesifikke.

Det bør undersøkes med Biblioteksentralen om det er mulig med et terminologisk samarbeid.

Dersom Biblioteksentralen kan ha nytte av Humord, bør den stilles til rådighet for dem. Det samme gjelder dersom systemleverandører ønsker å bruke tesaurusen og dens struktur til å utbedre sine søkegrensesnitt og tilby nye tjenester.

I forbindelse med utarbeidelsen av tesaurusprogrammet for denne rapporten har f.eks Bibliotek-Systemer As (BiblioFil) vist interesse for å prøve ut dataene.

Se også notater fra møte med representanter for Biblioteksentralen, appendiks B, avsnitt B.2.

5.1.4 SINTEF

Prosjektgruppen har hatt diskusjoner med representanter for SINTEFs faggruppe for språkteknologi¹⁰ som har kompetanse og mulighet til å utføre oppdrag. Noen eksempler på mulige samarbeidsprosjekter:

- forprosjekt for å undersøke kontaktpunkter mellom de forskjellige klassifikasjons- og emnesystemene som finnes per i dag på UB (og nasjonalt), basert på beskrivelsene i BIBSYS-basen.

Dette kan i neste fase utnyttes til å lage overganger (mapping)

¹⁰Vi møtte Diana Santos, Luís Costa, Nuno Cardoso, Luís Miguel Cabral og Bjørn Skjellaug (forskningsjef).

- mellom gamle og nye systemer
- mellom de systemene vi bestemmer oss for å bruke
- mellom våre og eksterne systemer
- undersøke muligheten for å konvertere (halvautomatisk) f.eks Humord-terminologien til WebDewey ved hjelp av maskinoversettelsesteknikker (og noen tilleggsregler), og ved bruk av allerede ferdig-klassifiserte nye elektroniske bøker eller artikler
- en empirisk studie av variasjoner i sammensatte emneord eller søkeord
- kjøre diverse typer “case studies” innen forskjellige (fag)områder, for eksempel rundt mulighetene for
 - å bruke synonymer
 - å bruke hierarki
 - å bruke kontekstindikatorer/disambigueringsteknikker
 - å la brukerne velge – og velge bort – alternative beskrivelser/emner
 - å ha en annen type visualisering av og navigering i publikasjonsmengde
- rådgivning: SINTEF kan med utgangspunkt i eksisterende faglitteratur utarbeide en oversikt over problemene og forskjellige måter å takle dem på. Det vil si en “state of the art” slik at de avgjørelsene som UBO senere treffer, blir gjort i lys av kjent kunnskap.

Andre områder som SINTEF kan engasjeres i omfatter blant annet:

- utvikle et meta-system (metasøk), som bruker alle de andre systemene, og er det som sees av bruker, med teknikker av ontologi-integrasjon og lingvistiske teknikker
- språkvask
- identifisere mulige inkonsistenser eller gap inni ett og mellom flere emneordsystemer
- bruke synonymer og kanskje også oversettelser til andre språk for å finne flere kontaktpunkter og relatere søk
- studere brukerloggene for å skjønne hva folk er interessert i

Slike prosjekter og studier vil ha interesse for andre deler av bibliotekvesenet og burde kunne oppnå ekstern finansiering.

5.1.5 Enhet for digital dokumentasjon

Gruppen har snakket med Christian-Emil Ore ved *Enhet for digital dokumentasjon* ved Institutt for lingvistiske og nordiske studier. Ore deltar i arbeidet med kopling av ontologier, spesielt mellom bibliotek (FRBR¹¹) på den ene siden og museum (CIDOC CRM¹²).

Ore er åpen for samarbeid med UBO om liknende problemstillinger (ontologi-overganger), men også dette er avhengig av prosjektfinansiering.

5.2 Internasjonalt

UBO er ikke alene om å ønske løsning på problemet med kombinasjon av ulike emnebeskrivelsessystemer. Det foregår konkret arbeid på mange steder for å finne metoder for løsning eller for en helt konkret kopling mellom systemer enten det foregår manuelt, automatisk eller halvautomatisk.

Interessant er det også at det drives tverrspråklig samordningsarbeid.

5.2.1 Sverige

Kungliga biblioteket (KB) driver prosjektet *Svenska ämnesord* (SAE). Arbeidet startet i 1999 med utgangspunkt i registeret til klassifikasjonssystemet og en lokal emneordliste. Faglig har SAE tyngdepunktet i humaniora og samfunnsvitenskap og er lite utbygd innen realfagene.

SAE bygges gradvis ut med parallelle LCSH.

I forbindelse med nesten fullstendig oversetting av WebDewey til svensk koples nå også SAE til Dewey (3-årig prosjekt). KB har kommet med en generell anbefaling til svenske bibliotek om overgang til Dewey. Målsettingen er standardisering og gjenbruk.

Se nærmere i appendiks B, avsnitt B.3 på side 31.

5.2.2 Finland

I Finland har man på nasjonalt nivå satset på å utarbeide en tverrfaglig nasjonal ontologi med utgangspunkt i Nasjonalbibliotekets finsk-språklige tesaurus (YSA/Allärs¹³).

Prosjektet (FinnONTO¹⁴) ledes av *Semantic Computing Research Group* (SeCo) og finansieres av et konsortium med 38 ulike organisasjoner innen muséer, bibliotek, helseorganisasjoner, offentlig forvaltning, media og utdanning.

¹¹FRBR – Functional Requirements for Bibliographic Records. En datamodell for bibliografisk informasjon utviklet av IFLA.

¹²CIDOC CRM – (Conceptual Reference Model) er en ontologi for begreper og informasjon i tilknytning dokumentasjon av kulturarv og museumsobjekter.

¹³Se <http://vesa.lib.helsinki.fi> (2010-02-08)

¹⁴Se <http://www.seco.tkk.fi> (2010-01-18)

Siktemålet er å kunne tilby både en generell ontologi på finsk (og via parallellen *Allårs* på svensk) og et verktøy for å utnytte den i emnebeskrivelse. I dette inngår også å utvikle metadata-standarder innen ulike anvendelser.

Et viktig skritt i prosjektet var å transformere Nasjonalbibliotekets thesaurus (YSA) til en ontologi (YSO). Transformasjonen gikk først og fremst ut på å presisere relasjonene i thesaurusen til å skille mellom generiske og partitive hierarkiske relasjoner[15].

YSAs stilling som en nasjonal ressurs har gjort det mulig å bruke den som “semantisk terminologisk lim” mellom mange andre finske tesauri.

Slik er den finske kulturthesaurus integrert med YSA (blant annet ved hjelp av ontologiverktøyet *Protégé-2000* (utviklet ved Stanford University)[24].

FinnOnto-teknologi utnyttet i en del pilotanvendelser: *eCulture* – er den neste generasjonen av MuseumFinland; *eHealth*; *eGovernment*; *eLearning* og *eCommerce*.

Det endelige målet er å lage en infrastruktur for semantisk web i Finland.

5.2.3 Tyskland – Sveits

Gjennom MACS-prosjektet lages det lenker mellom kontrollerte vokabularer på engelsk, fransk og tysk¹⁵. Samarbeidspartnere er Deutsche Nationalbibliothek, British Library, Bibliothèque Nationale de France og Swiss national library. Arbeidsgrunlaget er det tyske *Schlagwortnormdatei* (SWD), det franske *Répertoire d'autorité-matière encyclopédique et alphabétique unifié* (RAMEAU) og det engelske *Library of Congress Subject Headings* (LCSH).

I den grad norsk WebDewey blir koplet til LCSH åpner dette muligheter for samsøk på mange språk med utgangspunkt i norsk fagterminologi. Det blir et spørsmål om å bygge de riktige systemene rundt disse dataene.

Deutsche Bibliothek har dessuten vedtatt å bruke MARC 21 som utvekslingsformat og har utarbeidet konkordansetabeller mellom det tidligere tyske formatet MAB2 og MARC 21¹⁶.

5.2.4 Overganger mellom klassifikasjonssystemer

I Tsjekkia leder Nasjonalbiblioteket et arbeid for å lage overgang mellom UDC og DDC. Dette arbeidet bør studeres nærmere med tanke på å utnytte tilsvarende overgang mellom avdelinger på UBO og mellom DDC og UDK på nasjonalt plan.

5.2.5 Semantisk web

Skaperen av World Wide Web, *Tim Berners-Lee*, lanserte i 2006 *Linked data* som en automatisk metode for, på semantisk grunnlag, å lenke sammen data[6]. *Linked data* er en praktisk tilretteleggelse av semantisk web

¹⁵Se prosjektets hjemmeside <https://macs.hoppie.nl/pub/topic/about-macs> (2010-01-04)

¹⁶Se <http://www.d-nb.de/standardisierung/formate/marc21.htm> (2010-01-04)

ved hjelp av URI-er som viser videre til andre URI-er på en konsistent måte. Slik kan man få tilbud om å navigere videre både manuelt og automatisk. Målet er å synliggjøre, dele og kople sammen data på semantisk web.

Libris lanserte i 2009 sin bibliotekatalog som “linked data”. Det betyr at bibliotekatalogen åpnes for andre for lenking, gjenbruk av data og brukermedvirkning, og at det lenkes fra bibliotekatalogen til andre ressurser på web.

UBiT 2010 har et prosjekt hvor målet er å gjøre autoritetsdata i BIBSYS bedre tilgjengelige for brukerne ved hjelp av semantiske teknologier. I praksis betyr dette at en i prosjektet konverterer autoritetsdata i BIBSYS til *linked data*-filosofien.

6 Oppsummering

Arbeidet med emnedata har lenge vært et nedprioritert område både i norske fagbibliotek og hos systemleverandørene. Emnedata spiller en viktig rolle i gjenfinningen og har et stort utviklingspotensiale, både innenfor biblioteksystemene og mot semantisk web, for eksempel i form av emnekart. Prosjektgruppen håper derfor gjennom sine innspill også å bidra til å heve statusen for arbeidet med kunnskapsorganisering.

UBO har som Norges største universitetsbibliotek kompetanse både bibliotekfaglig, fag-faglig og IT-faglig og dermed forutsetninger for å bygge opp et miljø som kan ivareta videre utvikling og samarbeide med andre miljøer både nasjonalt og internasjonalt.

Et stort grunnarbeid er allerede gjort. Det er i bibliotekets og brukernes interesse at dette arbeidet blir utnyttet og utviklet:

- Det arbeides med å bygge norske vokabularer (Humord, UMNs emneordsprosjekt, UJURs emneordsliste, UHSs Deweyregister, oversettelse av MeSH).
- Det arbeides med å samordne bruken av systemer (Dewey, NLM, MeSH, Humord).
- Det er utviklet programvare lokalt som kan utnyttes i videre arbeid både lokalt og nasjonalt
 - Et eksempel på tesaurusprogram er vedlagt denne rapporten og ble også levert med budsjettinnspillet i desember.
 - Et hjelpeprogram for å konvertere data fra felt for frie nøkkelord til felt for kontrollert vokabular.

Biblioteket bør satse på etablerte systemer og standardisere bruken av dem slik at de blir compatible med internasjonale systemer og standarder. Gjennom dette legges det til rette for:

- effektiv gjenbruk av emnedata inklusive importdata, både innen UBO og innen BIBSYS
- bedre utnyttelse av emnedata ved søking (synonymer, navigasjon i systemstrukturer)
- overganger mellom systemer (konvertering av registreringsdata og “mapping” mellom systemer)
- samsøk i samlinger, både for trykte og elektroniske dokumenter
- standardisering og effektivisering av arbeidet med systemkrav overfor BIBSYS
- utviklingsprosjekter på nasjonalt plan, som er på linje med arbeid som utføres i andre land i Europa

Forslagene ovenfor er godt forankret i UBOs årsplan for 2010:

- UB skal videreutvikle forskningsbiblioteket (1.1).
- UB skal gjøre bibliotekets tjenester mer synlige, brukervennlige og lettere tilgjengelige for forskere og studenter ved UiO (2.1).
- UB skal styrke arbeidet med bibliografisk og emnemessig beskrivelse av samlingene (2.4).
- UB skal ha en effektiv og profesjonell organisasjon... (3.3)
- Felles UB-identitet (3.4).

Under forutsetning av at forslagene i rapporten følges opp, vil resultatet være:

- en norsk thesaurus (Humord) som omfatter de fleste av UBOs fagområder
- en engelskspråklig, internasjonal thesaurus (MeSH), som er under oversettelse til norsk, for fagområdet medisin
- Dewey-klassifikasjon med et felles, norskspråklig register for de fleste fag
- NLM klassifikasjon for medisin (internasjonalt system)

Dermed er det lagt til rette for utvikling og samarbeid på mange plan og med mange aktører.

Forslagene *koster* i form av arbeidsinnsats, direkte bevilgninger og ikke minst politisk vilje. Avslutningsvis er det derfor viktig å peke på betydningen av å ha visjoner, både i det daglige arbeidet og i langsiktig planlegging. Det bidrar både til motivasjon, kreativitet og utvikling.

Prosjektoversikt

Arbeidsgruppen omtaler mulige prosjekter flere steder i denne rapporten. For oversiktens skyld er disse samlet i følgende tabeller:

Prosjektforslag som angår Dewey.

Forslag	Formål	Ansvar, samarbeid	Ref.	Komm., status
Oppgradering av UHS Dewey-reg.	Standardisering. Språkvask	UHS, SINTEF	4.3, 5.1.4	Oppstart 2010
Felles Deweyregister for UBO	Samordning, gjenbruk, standardisering	UBO, NB	4.3	2011
Samarbeid om overs.av Web-Dewey	Standardisering og gjenbruk av terminologi. Mapping mot LCSH.	UBO, NB, Biblioteksentralen	4.3, 5.1.2, B.1	2010-2012
Samordning av Dewey-registre ved UBO og UBB	Gjenbruk, standardisering, samsøk	UBO, UBB, NB, BIBSYS	B.1	

Prosjektforslag som angår Humord.

Forslag	Formål	Ansvar, samarbeid	Ref.	Komm., status
Utarbeide vokabular realfag	Innplassering i Humord	UMN, UHS	4.4	Prosjekt pågår. Ferdig: 2011
Utarbeide vokabular jus	Innplassering i Humord	UJUR, UHS	4.4	Oppstart 2010
Utarbeide vokabular SV-fag	Innplassering i Humord	UHS	4.4	2010
Tesaurusprogram	Utnyttelse av emnedata. Samsøk	UBO, BIBSYS, Bibliofil o.a.	5.1.3, A	2010
Humord til ontologi. Termer med LCSH-struktur.	Flerspråklighet. Semantisk web. Standardisering i BIBSYS. Mapping mot LCSH.	UBO, BIBSYS, NB o.a.	4.5	

Prosjektforslag som angår MeSH.

Forslag	Formål	Ansvar, samarbeid	Ref.	Komm., status
MeSH til norsk	Norsk søkespråk, Mapping mot Humord		4.4	Prosjekt pågår
MeSH til ontologi	Semantisk web. Flerspråklighet.		4.4	Prosjekt pågår

Konverterings- og utviklingsprosjekter

Forslag	Formål	Ansvar, samarbeid	Ref.	Komm., status
Frie emneord til Humord (SV-fag)	Samsøk. Mapping LCSH	UHS, BIBSYS	4.2, 4.4	Delprosjekt 2011
Frie emneord til lokale til Humord (real-fag)	Innplassering i Humord	UMN, BIBSYS	4.2, 4.4	Delprosjekt 2011
UBO-klass.til Dewey	Standardisering, samsøk	UHS, UMN	4.2	2010-2011
UDK til Dewey	Standardisering, samsøk	UMN, UHS, BIBSYS	4.2	2010-2011
Div. prosjekter: datastudier, mapping, grensesnitt og språkteknologi		SINTEF, UiO/ILN	5.1.4, 5.1.5	

Appendiks

A Forslag til Humord-funksjonalitet i BIBSYS Ask

1. Tesaurusverktøy

Dette er et verktøy for sluttbrukers navigasjon i tesaurusen. Det kan systemmessig legges *ved siden av* eller være *integrert i* Ask, men det må være mulig å se tesaurusverktøyet sammen med trefflister.

- (a) Visning: tesaurusvisning a la emnehierarkiet i emneportalen.
- (b) Forhåndsvisning av antall treff på hver term, evt aggregert på delhierarkier.
- (c) Søk i tesaurus skal kunne være automatisk trunkert på venstre og høyre side, men man skal også kunne søke eksakt (ved hjelp av egen syntaks).
- (d) Søk på synonymer i tesaurus skal gi direkte treff på foretrukne termer.
- (e) Treff på søk i tesaurus (foretrukne termer): det skal enkelt kunne vises hierarkisk plassering av et treff.
- (f) Definisjon av foretrukne termer skal enkelt kunne vises.
- (g) Videresøk i katalogen både på enkeltstående foretrukken term eller hele hierarkiet under termen (explode-funksjon) må være mulig.
- (h) Tilgang til se-også-henvisninger: det skal være funksjonalitet for samtidig søk i katalogen på foretrukken term og se-også-henvisninger.

2. Søk i BIBSYS Ask

- (a) Ved valg av søkefeltet "Humord" må man få se lenke til tesaurusverktøyet eller tesaurusen må synliggjøres på annen måte.
- (b) Frasesøk og enkeltordsøk må samordnes i samme humord-søkefelt, f.eks ved å implementere markering av fraser med anførselstegn.
- (c) Høyre og venstre-trunkering må være mulig.
- (d) Søk på synonymer skal gi treff uten omvei for bruker (dvs direkte videresøk på foretrukken term). Forklaring må gis bruker.
- (e) Systemet må gi lenker fra foretrukne termer inn i tesaurusverktøyet for å hjelpe brukeren med videre navigasjon.
- (f) Systemet må gi brukeren mulighet til å sende søket mot andre emnefelt i tillegg til HUME-feltet.

- (g) Ved store treffmengder (f.eks større enn 1000 treff) på en term som har underinndeling må man bli gjort oppmerksom på muligheten av å gå dypere i hierarkiet.

Funksjonaliteten i begge punktene må i det minste gjøres tilgjengelig for alle bibliotek som bruker Humord (og deres sluttbrukere).

B Arbeidsnotater

B.1 Norsk komité for klassifikasjon og indeksering — NKKI

Notater fra møte med Ingebjørg Rype (leder av NKKI).

NB har bestemt seg for å foreta en fullstendig oversettelse til norsk av WebDewey (DDC utgave 22). Dette arbeidet er ventet ferdigstilt i løpet av 2012.

Norsk WebDewey vil bli koplet til Biblioteksentralens emnestrenger til Dewey på samme vis som den engelske versjonen av WebDewey er koplet til Library of Congress Subject headings og MESH. Dette er koplinger dels på manuell (intellektuell) basis, dels på en blanding av manuell og statistisk grunnlag og dels på rent statistisk grunnlag. Det kan også være aktuelt å kople til andre norske emneordsystemer, enten manuelt eller statistisk.

(Dette gjør f.eks at man kan få treff på et klassennummer (530.0285) ved søk "bioinformatics" i WebDewey til tross for at tabellene ikke har denne termen. Derimot fins termen i LCSH og der er den koplet til Dewey-kode. Akkurat denne koplingen var ikke helt vellykket faglig sett, i og med at bioinformatikk anses som en ny disiplin og ikke bare som en anvendelse av datamaskiner innen biologi (hj.tab 0285).)

I oversettelsesarbeidet vil det bli samarbeidet med OCLC og med Kungliga biblioteket i Sverige. Det kan også være interessant å samarbeid med norske produsenter av emneregistre til Dewey (f.eks UBO og UBB). Det vil være interessant å samarbeide med andre produsenter av fagterminologi/emneord.

Grunnlaget for oversettelsesarbeidet vil være DDK5.

I Tyskland foregår det et langsiktig (10 års perspektiv) arbeid for å kople de tyske nasjonale emnesystemene til LCSH, til franske emnesystemer og til Dewey.

Det er foreløpig ikke avklart hvordan norsk WebDewey vil bli tilgjengeliggjort, men en mulig modell kan være at NB betaler en nasjonal lisens.

Det ble opplyst at det i Tsjekkia foregår et arbeid for å matche UDK med DDC.

Generelt er emnearbeidet som foregår på UBO interessant for NKKI og det ble uttrykt ønske om samarbeid.

B.2 A/L Biblioteksentralen

Møte med Kjartan Vevle og Kjersti Feiring Myrtrøen ved Biblioteksentralen.

BS leverer bøker med tilhørende katalogdata til folkebibliotekene, også de store som Deichmanske bibliotek, Stavanger bibliotek og til noen høyskoler.

Katalogdataene inkluderer autoritetsdata fra tre registre:

- persondata (register inneholder over 112 000 poster organisert som strenger for å skille mellom navn som ulike aktører (forfatter, illustratør, m.fl.) og emne. Marc-felt: 100, 600, 700 og andre personfelt
- emneord i streng koplet med Dewey som plasseres i MARC-felt 650 i postene. Registeret omfatter drøye 26 000 strenger. Synonymer legges også i postene. BS følger Hjortsæters[14] retningslinjer
- stedsnavn (bruker Utenriksdepartementets liste, med unntak av bynavn der Språkrådet er kilde)

Dataene utnyttes ulikt av bibliotekene. Det er avhengig av hvilke muligheter systemleverandørene har lagt opp til. Ingen systemleverandører har foreløpig lagt opp til å utnytte autoritetsregistrene til en blafunksjon som et pedagogisk hjelpemiddel for brukerne.

BS har inngått avtale med Nasjonalbiblioteket om levering data på norsk forlagsprodusert materiale til BIBSYS, men her følger klassifikasjon med, men ikke emneord.

Noen bibliotek har fått BSs autoritetsregistre overført for bruk lokalt i forbindelse med egenkatalogisering. Det er et problem at man i enkelte tilfeller ikke bruker autoritetsdataene konsekvent.

I et prøveprosjekt for Gjøvik bibliotek oppdateres lokal kopi av BSs autoritetsregistre fortløpende online.

Dewey brukes veldig forskjellig i bibliotekene. Et forslag til samordning og for å oppnå sømløst emnesøk på nasjonalt plan, er å opprette et eget "autoritetsfelt" for et nasjonalt Dewey i BIBSYS og andre systemer. Lokale klassekoder kan eventuelt legges i egne felt for de som har behov for det. Tanken ble nevnt av Kjersti, som også er medlem av NKKI.

BS har samarbeid med OCLC og en rekke andre aktører. De driver kommersielt, men ønsker at deres registre skal kunne brukes av andre i mye større utstrekning enn i dag, og de er åpne for samarbeid med UBO (evt i tilknytning til BIBSYS).

Vi fikk en demonstrasjon av hvordan autoritetsdataene brukes i registreringer. Systemet omfattet blant annet mulighet for å lage registreringsprofiler for ulike dokumentkategorier (som man selv kunne sette opp). Raske oppslag og kopiering fra autoritetsposter til katalogposter og posisjonsbestemt hjelp/dokumentasjon gjorde dette til et effektivt og arbeidsbesparende registreringsprogram. Programmet er laget som en frittstående applikasjon som

kommuniserer med aksjonsbasen og autoritetsregistre. Det er altså ikke en web-applikasjon slik det legges opp til av BIBSYS.

B.3 Svenska ämnesord (SAE) – notater fra møte i Kungliga biblioteket

- tradisjonelt emneordsystem bygget opp over tid
- LCSH-struktur (prekoordinerte termer)
 - SAE har en helt konsekvent syntaks, noe enklere enn LCSH
 - Aut.lista går mot enkelttermene (ikke termstrengen)
- Prosjektstart 1999
 - utgangspunkt: Registeret til klassifikasjonssystemet og en lokal emneordliste ved Kungl.Biblioteket.
 - opplæring: Lois Mai Chan, Univ.of Kentucky (LCSH-ekspert)
 - prosjektet totalt sett meget ressurskrevende (beskrevet av Pia Leth og Ingrid Berg på IFLA-konferanse[18])
- omfattende regelverk basert på internasjonale regler (RAMEAU, IFLA, LCSH med tilpasninger)
 - entall/flertall: Flertall er “standard” og praktiseres på en pragmatisk måte (man ser bort fra eventuelle meningsforskjeller entall/flertall.)
- SAE bygges gradvis ut med parallelle LCSH.
 - målsetting: Gjenbruk gjennom mapping og forenklet oversettelse av engelske termer i bibliotekene
 - LCSH nå søkbare i Libris (via aut.lista).
 - problemer med retting
 - VIAF¹⁷: termer rettes via unik id.
- omfang: Nå ca 40.000 termer, inkl. geografiske emneord
- tilvekst: Alle kan foreslå nye termer (ca 15 nye pr uke). Godkjennes av en redaksjonsgruppe.
- brukes av:
 - nesten alle fagbibliotekene i Sverige

¹⁷Virtual International Authority File, internasjonalt autoritetssystem for navn, se også [23]

- noen av de store folkebibliotekene slutter seg nå til. Ellers har folkebibliotekene sitt eget opplegg med kjøp av metadata inkl. emnebeskrivelse (kfr norske Bib.sentralen).
- Generell anerkjennelse i Sverige. Bygges og vedlikeholdes av Kungliga biblioteket. Oppnår gjennom dette autoritet.
- Det brukes mye ressurser på opplæring/kursvirksomhet. (Emneordsindeksing er ikke inkludert i den svenske bibliotekarutdannelsen.)
- Andre emneordssystemer:
 - MeSH: Det finnes en svensk versjon.
 - AGROVOC: Oversettes også til svensk.

Søkbarehet i Libris

- Termene (emneordene) er søkbare som enkeltord. Strukturen bør utnytted bedre.
- Ønske: Forbedret grensesnitt for emnesøk.
- Synonymer er søkbare.
- “Ämnesträdet” er basert på det gamle SAB-systemet og derfor nokså “grunt”. Trestrukturen ikke godt nok utnyttet.

Klassifikasjon

- Deweyprosjektet (3-årig): SAE kombineres nå med Dewey med en generell anbefaling til bibliotekene om felles overgang til Dewey
- Standardisering og gjenbruk er også her målsettingen
- WebDewey oversettes til svensk (nesten i sin fullstendighet).
- Hjelpemidler: Bla oversettelsesprogram
- Eksisterende terminologi i SAE/LCSH sammenholdes med tilsvarende Dewey-klass. (hvilke koder er brukt i forbindelse med visse termer)
- Konverteringstabell SAB/Dewey: <http://export.libris.kb.se/DS/>

Fremtidsaspekter

- Gode muligheter for mapping SAE/LCSH/DEWEY og gjennom dette utnytting av eksisterende emnebeskrivelse på katalogposter (både nasjonalt og internasjonalt)
- Store muligheter for gjenbruk og oversettelse av leverandørproduserte metadata, for eksempel på e-dokumenter.

- Sikrer bruk av svensk språk og terminologi

Nyttige hjelpemidler

- People, places & things : a list of popular Library of Congress subject headings with Deweynumbers. – ISBN 0-91068-69-5. – Dublin, Ohio: Forest Press, 2001.
- Cataloger's desktop [electronic resource] / Library of Congress (beta-lingsdatabase)

B.4 FinnONTO — Finsk nasjonalt ontologiprosjekt

Vi valgte å besøke prosjektledelsen representert ved *Eero Hyvänen* og *Katri Seppälä* fra Semantic Computing Research Group i Espoo.

Prosjektet er godt presentert i foredraget *Elements of a National Semantic Web Infrastructure—Case Study Finland on the Semantic Web*[15] og det skal ikke gjentas her.

Utgangspunktet var YSA (den finske nasjonale generelle tesaurusen som inneholder 27 000 termer). Denne er konvertert til en ontologi med en presisjon hva angår hierarkiske relasjoner (YSO). Opp mot denne ontologien er det knyttet en rekke spesialontologier.

Vi merket oss at det ennå ikke er bygget opp en organisasjon for å vedlikeholde det semantiske nettverket som disse ontologiene representerer. Tanken var at det skulle oppdateres fire ganger i året. Siden Nasjonalbiblioteket tar seg av oppdatering av YSA var håpet at de også ville ta seg av ontologivarianten YSO.

I forbindelse med prosjektet *HealthFinland* er det tanken å lage en overgang mellom YSO og MeSH.

Finske folkebibliotek bruker to klassifikasjonssystemer som begge har sitt utgangspunkt i UDC. Helsinki folkebibliotek bruker den ene varianten, resten bruker den andre. Det er tanken å lage en overgang mellom UDC og YSO.

Vi fikk en demonstrasjon av portalen *CultureSampo* – en sammenstilling av materiale fra ca. 20 finske museer, bibliotek, arkiver og liknende, samt Wikipedia og Panoramio. Den inneholder bla. Semantic Kalevala med lenker til bilder, radioprogram osv. Biografibasen inneholder 120 000 personer med relasjoner, temasøk på kart osv. Systemet inneholdt mye hjelp til videresøk og skjema for kommentarer og tilleggsinformasjon fra brukere.

C Andre relevante prosjekter

C.1 KoMoHe

KoMoHe¹⁸ er et tysk prosjekt som har som mål semantisk integrasjon av ulike informasjonssystemer for å forbedre emnesøk i heterogene databaser[21].

¹⁸KoMoHe = **K**ompetenzzentrum **M**odellbildung und **H**eterogenitätsbehandlung

I første fase tok prosjektet for seg 25 kontrollerte vokabularer innenfor 11 fag og på 3 språk for å lage overganger (terminologi-mapping) mellom dem. Overgangene ble først og fremst gjort på intellektuell basis av fagfolk.

Prosjektet definerte disse typer overganger (relasjoner) mellom vokabularer:

- *Ekvivalens*
dvs at termer fra ulike vokabularer dekker samme begrep
- *Hierarki*
dvs at en term i ett vokabular dekker et bredere eller smalere begrep enn en term fra et annet vokabular
- *Assosiasjon*
dvs en sideordnet relasjon (se-også)
- *Null*
dvs at en term fra ett vokabular ikke på noen av de øvrige måter kan relateres til et bestemt annet vokabular

Prosjektet genererte mer enn 513 000 relasjoner i 64 overganger. I snitt brukte man 4 minutter på hver relasjon.

I neste fase var hensikten å studere effekten av bruk av overgangene for å søke i heterogene kilder (15 databaser med 10 forskjellige vokabularer og rundt 2,5 millioner poster)[22].

C.2 Renardus

Renardus-prosjektet hadde lite med bibliotekataloger å gjøre, men er interessant som et forsøk på å samordne søk i forskjellige emneportaler (subject gateways) i Europa. Ti bibliotek deltok i startfasen med til sammen 13 portaler.

Prosjektet som var finansiert under EUs femte rammeprogram, ble av ulike årsaker ikke videreført, men metodikken for samordning er likevel interessant, ikke minst for å overvinne språkbarrierer.

Samordningen hadde et klassifikasjonssystem som basis (Dewey) som de ulike deltakersystemene “mappet” sine ressurser til. Klassifikasjonssystemet fungerte dermed som ei bru mellom portalene.

C.3 Høgskolen i Oslo/JBI

Prosjektgruppen har hatt et møte med Ragnar Nordlie for å høre om hans erfaringer/undersøkelser om brukeratferd i informasjonssøk.

Sammen med Nils Pharo gjennomføres et kontinuerlig prosjekt som retter seg mer mot brukeratferd i forhold til informasjonssøk mot tjenester som Wikipedia og ikke spesifikt bibliotekataloger.

Siden undersøkelsen er basert på Wikipedia, mente Nordlie at den kanskje var mindre relevant for oss som ser på gjenfinning i en bibliotekskatalog. Han mente derfor at han antagelig ikke kunne tilføre prosjektet vårt særlig mange nye tanker, men uttrykte interesse for det.

Nordlie nevnte North Carolina State University¹⁹ som bl.a. har en “browse subjects”-mulighet som kanskje kunne være av interesse²⁰.

D Artikkelresymé

Sammendrag av rapporter fra Research Information Network, London, OCLC og Library of Congress

Mange sluttbrukere foretrekker å starte søk i Google og lignende fremfor bibliotekenes nettsider og kataloger. Krav om endringer og bedre tilgjengelighet er derfor lansert fra bibliotekhold. Samtidig oppfordres bibliotekene til samarbeid om produksjon av bibliografiske data og standardiseringer, både for å forbedre verktøyene og for økonomisering.

- **Research Information Network, London**

Creating catalogues: bibliographic records in a networked world. June 2009[25]

Rapporten inneholder en undersøkelse av produksjonen av bibliografiske poster i engelske forskningsbibliotek, både for trykte og elektroniske dokumenter. Den følger produksjonen fra utgiver via mellommenn til sluttbruker. Konklusjonen er at et samarbeid mellom alle i kjeden, og en felles, åpen platform, er ønskelig. Men det betinger to viktige forutsetninger:

- en bedre forståelse av forretningsmodellene i forsyningskjeden: utgivere, aggregatorer, agenter, bibliografiske verktøy, nasjonalbibliotek, universitetsbibliotek og andre bibliotek, Google o.a.
- behovet for klart definerte standarder og kvalitet på postene i hvert ledd av kjeden, hvordan disse kravene kan møtes og av hvem. Uten en forståelse og erkjennelse av behovene til hver enkelt er det stor fare for at duplisering vil fortsette.

Copyright til bibliografiske data er diskutert. Det er nevnt som et problem at verken British Library, OCLC eller RLUK (Research Libraries UK) i dag tillater fri deling og gjenbruk av bibliografiske data. Rapporten nevner spesielt elektroniske publikasjoner som et felt der samarbeid er av stor viktighet. Den nevner også at utgivere bør gjøre metadata på artikkelnivå lettere tilgjengelig i standardformat.

¹⁹Se <http://www.lib.ncsu.edu/> (2010-02-08)

²⁰Denne fungerer som en slags emneportal som ender opp med fagsider a la UBO

Det er store utfordringer på agendaen i mange land, særlig i forhold til digitalisering av informasjonsressursene. Et prosjekt som nettopp er lansert av Research Information Network²¹ skal se på fremtidige scenarier for bibliotek i høyere utdanning og forskning og hvordan informasjonsbehovet kan dekkes. Prosjektet har tittel: “Academic libraries of the future” og skal gå over 18 mnd. Det er et samarbeid mellom British Library, Joint Information Systems Committee (JISC), Research Information Network (RIN), Research Libraries UK (RLUK) og Society of College, National and University Libraries (SCONUL).

- **OCLC**

Online catalogs: what users and librarians want. 2009[8]

OCLC inneholder mer enn 130 millioner katalogposter og samlingene til 70.000 bibliotek på 470 språk i WorldCat. Databasen VIAF (The Virtual International Authority File) er et produkt av autoritetsfilene for navn fra Library of Congress, Deutsche Nationalbibliothek og Bibliothèque Nationale de France. Etter hvert har flere bibliotek vist sin interesse, bla. Kungliga biblioteket.

Undersøkelsen ble gjennomført av OCLC i 2008. Utgangspunkt var at sluttbrukeres forventninger er sentrale, og at både grensesnitt og datakvalitet er viktig. Siden ansatte i bibliotek utgjør en stor del av brukerne, var de med som en egen gruppe.

Resultatene fra gruppene av sluttbrukere (studenter, forskere, andre) og bibliotekansatte ble sammenlignet. For sluttbrukere viser det seg at rask tilgang til dokumentene er like viktig som gjenfinning. Lenking til sammendrag/fulltekst og lignende kom også høyt på ønskelistene. Det samme gjaldt sammenhengen mellom søkerresultater og relevans. Nøkkelord toppet listen over søkevalg, men mulighetene for avansert søk var viktig i forbindelse med raffinering av søket, og navigering og håndtering av store treffmengder. Sluttbrukere brukte også avansert søk når de søkte etter kjent materiale. Mulighetene for mer presise søk ga bedre resultater.

End users want to be able to do a simple Google-like search and get results that exactly match what they expect to find. At the same time, they appreciate the ability to conduct advanced, guided searches so that they do not have to scroll through pages of results to find the items they need. They view faceted narrowing of results as an effective way to reduce what may be unmanageable results lists.

Undersøkelsen bemerker bl.a. at brukerne er vant til å søke i Google og nettbokhandlere som Amazon, og forventer at katalogen tolker sø-

²¹<http://www.rin.ac.uk/news/press/vision-academic-library-and-information-services-future> (2010-02-01)

kene på samme måte. For eksempel bør søketermer som kan ha flere betydninger likevel generere riktig dokument.

Ikke uventet fokuserte bibliotekansatte på datakvalitet, men også de plasserte innholdsfortegnelser og sammendrag høyt på listen over ønsker.

- **Library of Congress**

Sammendraget bygger på rapportene:

- *Study of the North American MARC records marketplace.* [12]
- *On the Record : report of the Library of Congress Working Group on the future of bibliographic control. January 9, 2008*[19]
- *Response to On the Record. June 2008* Offisiell kommentar til *On the Record* av Deanna B. Marcum)[20]

Rapporten *On the Record* er konsentrert til temaene:

- brukere og bruk av bibliografiske data
- strukturer og standarder for bibliografisk kontroll
- økonomi og organiseringen av bibliografisk kontroll

Arbeidsgruppen sier i et innledende sammendrag:

The Working Group envisions a future for bibliographic control that will be collaborative, decentralized, international in scope, and Web-based. The realization of this future will occur in cooperation with the private sector and with the active collaboration of library users. Data will be gathered from multiple sources; change will happen quickly; and bibliographic control will be dynamic, not static.

Det betyr et bredt samarbeid når det gjelder metadata og bibliografisk kontroll. Felles anstrengelser og samarbeid er nøkkelen til et godt resultat for bibliotekene i fremtiden. Rapporten kommenterer også Library of Congress sentrale rolle i bibliotekverdenen, som produsent av katalogdata og bibliografisk kontroll. Rapporten inneholder en rekke anbefalinger, bl.a. om internasjonalisering av autoritetsfiler, standarder osv.

Når det gjelder LCSH sier rapporten:

Subject analysis is a core function of cataloging, and Library of Congress Subject Headings have great value in providing controlled subject access to works.

Men rapporten sier også at LCSH ikke egner seg for noviser, men er best for den som allerede har kunnskaper om temaet de søker. Rapporten inneholder flere anbefalinger om forbedringer, bl.a. samarbeid med andre kontrollerte vokabularer og tesauruser.

Både denne rapporten og rapporten som har undersøkt produksjon og distribusjon av MARC-poster i USA og Canada konkluderer med at dobbeltarbeid bør unngås. Den viser også til at det er de store bibliotekene med mer komplekse samlinger som fremhever viktigheten av autoritetskontroll, og at det er en generell misnøye med data fra leverandører som ikke tilbyr LC-poster. LC-postene er helt klart rangert høyest, sammen med poster fra OCLC og Amicus (canadisk samkatalog).

Et annet spørsmål er om MARC-postene snart har utspilt sin rolle, eller bør suppleres/utvides. Men uavhengig av dette er argumentene for standardisering sterke, både når det gjelder kontrollerte vokabularer, klassifikasjon, emner osv. Dette vil også redusere kostnader.

Referanser

- [1] Marit Almo. *Humord : regler for indeksering til BIBSYS*. BIBSYS, 2004.
http://www.bibsys.no/files/pdf/handbok/humord_indekseringsregler_handbok.pdf (2010-01-05).
- [2] Marit Almo, Kirsten Al-Araki og Knut Hegna. *Kartlegging av emnesystemer ved Universitetsbiblioteket i Oslo*. UBO, 2010.
<http://www.ub.uio.no/prosjekter/emnesystemer/ubo-emnesystemer.pdf> (2010-01-12).
- [3] Avtale mellom BIBSYS og universitetsbibliotekene i Bergen, Oslo og Tromsø om bruk og drift av Humord.
<http://www.bibsys.no/files/out/humord/dokumenter/bibsysavtale.htm> (2010-02-08).
- [4] Avtale mellom universitetsbibliotekene i Bergen, Oslo og Tromsø om forvaltning av opphavsrett og databaserett (rettigheter) til tesaurusen Humord.
<http://www.bibsys.no/files/out/humord/dokumenter/rettighetsavtale.htm> (2010-01-05).
- [5] Marie Balikova. Focusing on user needs: new ways of subject access in Czechia. I *Looking at the Past and Preparing for the Future* [16].
<http://www.ifla2009satelliteflorence.it/meeting2/program/assets/Balikova.pdf> 2010-01-08).
- [6] Tim Berners-Lee. Linked data. Nettdokument, 2006.
<http://www.w3.org/DesignIssues/LinkedData.html> (2010-02-03).

- [7] Bibliotekdirektøren. Prosjektgruppe for bibliografisk og emnemessig beskrivelse av UBOs samlinger: oppnevningensbrev. <http://www.ub.uio.no/prosjekter/emnesystemer/136559-1.pdf> (2010-01-15), 26.mars 2009.
- [8] Karen Calhoun, Joanne Cantrell, Peggy Gallagher og Janet Hawk. Online catalogs: what users and librarians want. Rapport, OCLC, 2009. <http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf> (2010-02-01).
- [9] Lois Mai Chan. Social bookmarking and subject indexing. I *Looking at the Past and Preparing for the Future* [16]. <http://www.ifla2009satelliteflorence.it/meeting2/program/assets/Chan.pdf> (2010-02-03).
- [10] Charles A. Cutter. *Rules for a dictionary catalog : special report on public libraries*. Government printing office, 4 utgave, 1904.
- [11] Det kongelige kultur- og kyrkjedepartement. Mål og mening : ein heilskapleg norsk språkpolitikk, 2008. St.meld. nr. 35 (2007-2008) <http://www.regjeringen.no/pages/2090873/PDFS/STM200720080035000DDDPDFS.pdf> (2010-01-15).
- [12] Ruth Fischer og Rick Lugg. Study of the North American MARC Records Marketplace. Rapport, Library of Congress, 2009. http://www.loc.gov/bibliographic-future/news/MARC_Record_Marketplace_2009-10.pdf (2010-02-01).
- [13] Grunn-notat for Dewey-klassifikasjon og registerarbeid ved Bibliotek for humaniora og samfunnsvitenskap . <http://www.ub.uio.no/uhs/uhsintra/sek-gr/dewey/grunnotat-dewey.html> (2010-01-05).
- [14] Ellen Hjortsæter. *Emneordskatalogisering : innholdsanalyse, emnerepresentasjon og lagring*. ABM-media, 3. utgave, 2009.
- [15] Eero Hyvönen, Kim Viljanen, Eetu Mäkelä, Tomi Kauppinen, Tuukka Ruotsalo, Onni Valkeapää, Katri Seppälä, Osma Suominen, Olli Alm, Robin Lindroos, Teppo Kansala, Riikka Henriksson, Matias Frosterus, Jouni Tuominen, Reetta Sinkkilä og Jussi Kurki. Elements of a National SemanticWeb Infrastructure–Case Study Finland on the Semantic Web. I *First International semantic computing conference*. IEEE, September 2007. <http://www.seco.tkk.fi/publications/2007/hyvonen-et-al-elements-2007.pdf> (2010-01-18).
- [16] IFLA Classification and Indexing Section. *Proceedings*. IFLA, August 2009.

- <http://www.ifla2009satelliteflorence.it/meeting2/program/program.html>
(2010-01-08).
- [17] Yvonne Jahns. 20 years SWD – German subject authority data prepared for the future. I *Looking at the Past and Preparing for the Future* [16].
<http://www.ifla2009satelliteflorence.it/meeting2/program/assets/Jahns.pdf> (2010-01-08).
- [18] Pia Leth og Ingrid Berg. Subject Indexing in Sweden – the creating of a national system based on international standards in a country that often wanted to go its own way . I *World Library and Information Congress: 70th IFLA General Conference and Council*. IFLA, IFLA, 2004.
http://archive.ifla.org/IV/ifla70/papers/041e-Leth_Berg.pdf (2010-01-06).
- [19] Olivia M. A. Madison og José-Marie Griffiths. On the Record : Report of The Library of Congress Working Group on the Future of Bibliographic Control. Rapport, Library of Congress, 2008.
<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf> (2010-02-01).
- [20] Deanna B. Marcum. Response to On the Record. Rapport, Library of Congress, 2008.
http://www.loc.gov/bibliographic-future/news/LCWGRptResponse_DM_053008.pdf (2010-02-01).
- [21] Philipp Mayr og Vivien Petras. Building a terminology network for search: the KoMoHe project. I *International conference on Dublin Core and Metadata applications*, 2008.
<http://arxiv.org/pdf/0808.0518> (2009-12-11).
- [22] Philipp Mayr, Philipp Schaer og York Sure. Intra- and interdisciplinary cross-concordances for information retrieval. I *The 8th European networked knowledge organization systems (NKOS) workshop*, 2009.
<http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2009/abstracts.html#mayr>(2009-12-11 – abstract).
- [23] OCLC. VIAF (The Virtual International Authority File), 2009.
<http://www.oclc.org/research/activities/viaf/default.htm> (2010-01-06).
- [24] The Protégé Ontology Editor and Knowledge Acquisition System, 2009.
<http://protege.stanford.edu/> (2010-01-06).
- [25] Research Information Network. Creating catalogues: bibliographic records in a networked world. Rapport, Ken Chad Consulting and

- Sero, 2009.
<http://www.rin.ac.uk/our-work/using-and-accessing-information-resources/creating-catalogues-bibliographic-records-network> (2010-02-01).
- [26] Magdalena Svanberg. Dewey in Sweden : leaving SAB after 87 years. I *Looking at the Past and Preparing for the Future* [16].
<http://www.ifla2009satelliteflorence.it/meeting2/program/assets/Svanberg.pdf> (2010-02-08).
- [27] Norsk termbank i samarbeid med UBO, Fakultetsbiblioteket HF. *Humord : håndbok*. BIBSYS, 1996.
http://www.bibsys.no/files/pdf/handbok/humord_handbok.pdf (2010-01-05).
- [28] Diane Vizine-Goetz. FAST Headings as Tags for WorldCat. I *Looking at the Past and Preparing for the Future* [16].
<http://www.ifla2009satelliteflorence.it/meeting2/program/assets/VizineGoetz.pdf> (2010-01-15).
- [29] Kwan Yi og Lois Mai Chan. Linking folksonomy to Library of Congress subject headings: an exploratory study. *Journal of documentation*, 65(6):872–900, 2009.