

Styringsdokument for prosjektet «Mapping mot norsk WebDewey» ved Universitetsbiblioteket i Oslo (UBO)

Styringsdokumentet gir retning for virksomheten i prosjektet på et overordnet plan. Det uttrykker målsetting og begrunnelse for prosjektet, samt prinsipper og rammefaktorer for hvordan prosjektarbeidet skal gjennomføres. Beslutninger som tas innen prosjektet skal være forankret i styringsdokumentet. Sentrale begreper er definert i del 4.

1 Prosjektmål: Hva vi skal gjøre i prosjektet «Mapping mot norsk WebDewey»

Hovedmål:

Mål for prosjektet er å mappe vokabularene Humord og Realfagstermer mot norsk WebDewey.

Delmål:

- 1) Prosjektet skal bidra til videreutvikling av metodikk for mapping mot Deweys desimalklassifikasjon.
- 2) Mappingarbeidet skal utføres som en datastøttet intellektuell prosess. Prosjektet skal resultere i utvikling av programvare for mapping av vokabularer, inkludert vedlikehold av mappings i forbindelse med endringer i de involverte vokabularene.

2 Formål: Begrunnelse for prosjektet

Formålet med prosjektet «Mapping mot norsk WebDewey» er å etablere interoperabilitet mellom vokabularer på tvers av språk og bibliografiske baser. Dette skal sikres i form av en felles verbal inngang til de involverte vokabularene via norsk WebDewey.

De mappede vokabularene skal legge grunnlaget for utbedring av sluttbrukertjenester i form av søk på tvers av samlinger og språk, samt muligheter for å orientere seg i emnebaserte metadata før søk. Bibliotekpersonalet skal tilbys støtte i indekseringsarbeidet i form av koblingspunkter mellom valgte indeksertermer og aktuell klassifikasjon, og omvendt.

Gjennomføring av prosjektet skal omfatte en kvalitetsheving av kildevokabularene i form av revisjoner og rettelser som foretas i forbindelse med mappingarbeidet. Prosjektet skal bidra til berikelse av norsk WebDewey i form av utvidelse av antall verbale søkeinnganger, på grunn av større granularitet i Humord og Realfagstermers faglige domener. Prosjektet vil også berike norsk WebDewey med bygde numre, i tillegg til kvalitetssikringsarbeid i form av tilbakemeldinger til WebDewey-redaksjonen på klasser der det er behov for revisjoner.



Behovet for kontrollerte vokabularer for norsk fagterminologi strekker seg utover bibliotekfeltet. Mappingene som etableres i prosjektet vil publiseres som åpne lenkede data, tilgjengelige for andre tjenester på Internett.

3 Bakgrunn

Ideen om mapping av norske vokabularer kom opp i forlengelsen av en diskusjonen om scenarier for anvendelse av norsk WebDewey, i forbindelse med et mappingseminar i regi av Nasjonalbiblioteket (NB) og Høgskolen i Oslo og Akershus (HiOA) i 2012. Erfaringen om at flere engelskspråklige vokabularer var mappet til den amerikanske utgaven av WebDewey, vekket et ønske om å mappe Humord og Realfagstermer til norsk WebDewey. Sluttbrukere ved UBO har flere kontrollerte vokabularer å forholde seg til i forbindelse med informasjonssøk, og vil dra nytte av en felles verbal inngang til informasjonsressursene.

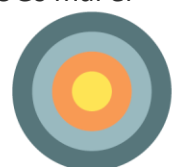
Universitetsbiblioteket gjennomførte i perioden mars 2014 til mars 2015 et pilotprosjekt om utvikling av metodikk for mapping av tesaurusen Humord mot norsk WebDewey. Dette utgjorde en separat aktivitet innenfor det større prosjektet «På vei mot en generell norsk tesaurus». Prosjektet «Mapping mot norsk WebDewey» bygger videre på pilotprosjektet, samt to tidligere prosjekter i regi av Realfagsbiblioteket ved UBO. Det første av disse foretok en sammenlikning av terminologi mellom Realfagstermer og NTNUs tesaurus TEKORD. Det andre prosjektet utforsket metoder for mapping mellom Realfagstermer til aktuelle deler av den foreløpige norske oversettelsen av Deweys desimalklassifisering.

Inneværende prosjekt startet i mars 2015 og omfatter tre vokabularer: *Humord* er en tesaurus som dekker humaniora og samfunnsvitenskap med tilgrensende fagområder. Den er strukturert med hierarkiske og assosiative relasjoner, og består av ca. 18 500 begreper, samt 8 500 synonymer til disse. Humord-tesaurusen er tuftet på et samarbeid mellom universitetene i Oslo, Bergen og Tromsø med HumSam-biblioteket som koordinator og sekretariat. Humord brukes også av Den internasjonale Ibsen-bibliografien og Senter for studier av Holocaust og livssynsminoriteter. *Realfagstermer* er et kontrollert, delvis prekoordinert emneordsvokabular som i hovedsak dekker naturvitenskap, matematikk og informatikk. Vokabularet inneholder ca. 15 000 emneord som kan brukes frittstående eller kombinert til strenger, samt ca. 2000 synonymer til disse. Vokabularet er flerspråklig (bokmål, nynorsk, delvis engelsk). Realfagstermer er utviklet og blir vedlikeholdt av Realfagsbiblioteket. UBB bruker også Realfagstermer, og er representert i redaksjonen. *Norsk WebDewey* er en fullstendig oversettelse av DDC 23 (Dewey Decimal Classification). Den inneholder ca. 44 000 klasser og utvides stadig med nye bygde numre. Norsk WebDewey er kun tilgjengelig på nett, og den blir løpende oppdatert.

4 Definisjoner

DDC: Forkortelse for Dewey Decimal Classification. Norsk forkortelse: DDK, dvs. Deweys desimalklassifisering.

EDUG: Forkortelse for interesseorganisasjonen «European DDC Users Group». EDUGs mål er



å fremme utviklingen av Deweys desimalklassifisering, i et samarbeid mellom europeiske brukerinstusjoner og OCLC. EDUGs hjemmeside ligger på <http://edug.pansoft.de/tiki-index.php>.

ISO: Forkortelse for International Organization for Standardization. I dette prosjektet bruker vi ISO 25964, spesielt del 2 som omhandler mapping, men også del 1 som omhandler tesauri.

Kildevokabular: Et vokabular som utgjør startpunktet når man søker et korresponderende begrep i et annet vokabular (dvs. målvokabularet). I inneværende prosjekt er Humord og Realfagstermer kildevokabularer.

Mapping: Mapping er en aktivitet hvor det etableres relasjoner mellom begreper i to ulike kontrollerte vokabularer. Man mapper fra et kildevokabular og finner korresponderende begreper i et målvokabular. Mapping brukes også til å benevne resultatet av en mappingaktivitet, dvs. relasjonen mellom et begrep i ett vokabular og en eller flere begreper i et annet vokabular.

Mappingtyper og mappingrelasjoner: ISO-standarden for mapping (ISO 25964-2) opererer med tre mappingtyper: Ekvivalens, hierarkisk og assosiativ. Disse mappingtypene gir opphav til fem mulige relasjonstyper (der tilsvarende SKOS-relasjonsbetegnelser er tilføyet i parentes): eksakt ekvivalens (skos:exactMatch, forkortes =EQ), tilnærmet ekvivalens (skos:closeMatch, ~EQ), overordnet mapping (skos:broadMatch, BM), underordnet mapping (skos:narrowMatch, NM) og relatert mapping (skos:relatedMatch, RM). Underordnet mapping anvendes aldri ved mapping til Deweys desimalklassifisering.

Målvokabular: Et vokabular der man søker etter korresponderende begreper til begreper fra utgangsvokabularet (dvs. kildevokabularet). I inneværende prosjekt består målvokabularet av norsk WebDewey.

SKOS: Forkortelse for Simple Knowledge Organization System, et utvekslingsformat beregnet på den semantiske veven (W3C) for å representere tesauri, klassifikasjonsskjemaer og andre typer kontrollerte vokabularer.

Vokabular/kontrollert vokabular: Et sett av termer, emneord eller koder som representerer begreper og relasjonene mellom dem, og som kan bli brukt i informasjonsgjenfinning.

5 Prinsipper

5.1 Grunnlagsdokumenter

Styringsdokumentet tar utgangspunkt i prosjektsøknader, tildelingsbrev og rapportering for pilotprosjektet og inneværende prosjekt, begge med NB som bevilgende myndighet. De nedfelte prinsippene er basert på EDUGs anbefalinger for mapping til/fra Deweys desimalklassifisering, som igjen er basert på ISO-standarden for mapping (ISO 25964-2). Styringsdokumentet og grunnlagsdokumentene for dette skal brukes som basis ved utarbeidelsen av prosjektinterne dokumenter, f.eks. mappingrutiner, eksempelsamling og beslutninger knyttet til rettinger i kildevokabularene.



5.2 Grunnleggende forutsetninger

I tråd med EDUGs anbefalinger brukes «nav-modellen» som overordnet struktur, slik modellen er beskrevet i ISO 25964-2 punkt 6.4. I denne modellen utgjør norsk WebDewey et sentralt koblingspunkt – et «nav i et hjul» – som øvrige vokabularer mappes mot. Vi skal bruke de fire mappingrelasjonene som er aktuelle ved mapping mot WebDewey, dvs. eksakt ekvivalens, tilnærmet ekvivalens, overordnet mapping og relatert mapping. Differensiering av mappingene er begrunnet i et sluttbrukerperspektiv.

EDUGs anbefalinger åpner for at man i et prosjekt kan beslutte å mappe visse typer assosiative relasjoner. Eksempel: Hvis man har vedtatt å mappe mellom yrkesutøvere og fagområde, vil kildebegrepet Leger bli mappet til klassenummeret for Medisin, i tillegg til nummeret for Leger. Inneværende prosjekt har ikke vedtatt mapping av slike forhåndsdefinerte typer av assosiative relasjoner. Dette prosjektet bruker relatert mapping for å etablere mappinger mellom assosiativt relaterte begreper i enkeltstående tilfeller, eller når vi er i tvil om relasjonstype men likevel ønsker å foreta en mapping. Prosjektet skal heller ikke etablere relatert mapping for assosiative relasjoner i Humord og Realfagstermer, da disse er et anliggende innen hvert av kildevokabularene.

Retningen på mappingene skal i mappingverktøyet kodes enveis fra hvert av kildevokabularene til WebDewey. Eksakt ekvivalens er pr definisjon toveis.

Vi skal unngå å etablere sammensatte ekvivalensmappings fordi det ikke lar seg uttrykke i eksportformatet SKOS. I situasjoner der et begrep i kildevokabularet har flere faglige plasseringer i WebDewey, skal vi etablere flere uavhengige mappinger for å unngå sammensatt ekvivalens. Tilfeller der kildevokabularet har flere uavhengige begreper for noe som er uttrykt som ett sammensatt begrep i målvokabularet, skal ignoreres i mappingarbeidet (dette berører i praksis generelle se-henvisninger i Humord).

Mappingarbeidet skal kvalitetssikres ved at hver mapping skal ses på av to personer; én utfører og én kontrollør. Ved tvil/faglig uenighet bringes en tredje person inn.

5.3 Omfang

Prosjektet skal mappe kildevokabularene universelt mot målvokabularet, dvs. mot alle hovedklasser i WebDewey. Dette er begrunnet i indekseringspraksis for kildevokabularene ved UBO, pluss med tanke på en eventuell kommende norsk generell tesaurus der Humord og Realfagstermer er tenkt å inngå. For hvert kildebegrep skal vi mappe til alle faglige plasseringer i WebDewey. Hierarkisk kontekst i Humord skal brukes til begrepsavklaring, men ikke til å avgrense hvilke faglige plasseringer vi skal mappe til i WebDewey. Indekseringspraksis i bibliografiske data brukes til supplerende begrepsavklaring. Vi skal mappe til både hovedtabellnumre, hjelpetabellnumre (unntatt H3) og til bygde numre. Vi skal ikke mappe til klassenumre for «tillatt løsning» eller til numre som har gått ut av bruk.

Prosjektet skal mappe alle begreper i kildevokabularene, inkludert formtermer, geografiske navn og strenger (begrenset til strenger av typen «innholdsbeskrivende emneord – innholdsbeskrivende underemneord», ca. 900 forekomster i Realfagstermer). Når et begrep



finnes både som hovedtabell- og hjelpetabellnummer, skal vi mappe til begge plasseringer. Det gjelder bl.a. lingvistiske begreper og begreper for det enkelte språk.

Vi skal mappe til bygde klassenumre når et kildebegrep ellers ikke ville latt seg uttrykke i målvokabularet pga. at Humord og Realfagstermer har større granularitet på sine spesialfelter i forhold til WebDewey. Nummerbygging i forbindelse med mapping er også i tråd med UBOs praksis om å indeksere så dypt som mulig. Klassenumre som bygges i forbindelse med mappingarbeidet skal registreres i nummerbyggingsmodulen i WebDewey. For geografiske navn skal prosjektet bygge numre for historie og geografi i hovedtabellen, i tillegg til å mappe til geografitabellen i hjelpetabell 2. For folkegrupper skal vi bygge numre i sosiologitabellen (under 305), i tillegg til å mappe til hjelpetabell 5 for etniske og nasjonale grupper.

Prosjektet skal etablere retningslinjer for hvordan vi skal forholde oss til atypiske trekk i Humord som tesaurus, dvs. håndtering av avvik fra tesaurusstandarden (ISO 25964-1). Strukturell opprydning i Humord (topptermer, hierarkier) er ikke en del av prosjektet, men legges til et eventuelt kommende prosjekt for etablering av en norsk generell tesaurus.

5.4 IKT-verktøy og tekniske presiseringer

Mappingverktøyet «ccmapper» skal utvikles av prosjektets tekniske personale i samarbeid med det tyske firmaet Pansoft. Programvaren skal håndtere vedlikehold av mappinger ved endringer i kilde- og målvokabularene. Vi forutsetter åpne kildekoder. Mappingene skal gjøres tilgjengelige i SKOS (uten lokale tilpasninger) og Marc. Det skilles mellom teknisk løsning i ccmapper kontra eksportformat.

De mappede vokabularene skal framkomme i WebDewey på en slik måte at foretrukne termer fra kildevokabularet representerer begrepene i kildevokabularet.

Mappingrelasjonstype skal spesifiseres sammen med mappingene som vises i WebDewey. Synonymer fra kildevokabularene skal være søkbare og synlige i WebDewey.

6 Ressurser

6.1 Økonomi

Prosjektet er i tildelingsbrev fra NB bevilget 2 086 000 kroner i 2015. For 2016 har prosjektet fått tildelt ca. 1,8 millioner kroner, inkludert budsjetttrammen i forbindelse med kontrakten med Pansoft. I tillegg kommer UBOs egenandel i prosjektet, som for en stor del består av avsetting av personalressurser til prosjektet.

6.2 Bemanning

Personalressurser til mappingarbeidet skal hentes inn i prosjektet ved avsetting og frikjøping av UBO-ansattes tid (bibliotekarere og fagreferenter), samt ved eksterne ansettelser fra aktuelle miljøer som besitter Dewey- og mappingkompetanse. Det er også aktuelt å utlyse prosjektstillinger som skal ivareta faste oppgaver for UBO-ansatte som frikjøpes til prosjektet.

Kjennskap til Deweys desimalklassifikasjon og tesauri vektlegges framfor kompetanse i kildevokabularenes faglige domene. Fagreferentene på UBO trekkes inn ved behov for



begrepsavklaringer. Prosjektet skal gjennomføre internopplæring i mapping, klassifikasjon (inkludert nummerbygging) og tesauri.

6.3 Tidsramme

Prosjektet er 2-årig regnet fra 1. mars 2015. Bevilgede midler kan distribueres over en lengre periode. Prosjektet skal arbeide etter en milepælplan som vedlegges styringsdokumentet.

6.4 Fremtidige ressursbehov

Etter at vokabularene er mappet, vil mappingene kreve kontinuerlig vedlikehold som følge av endringer i kilde- og målvokabularene. Prosjektet skal etablere retningslinjer for oppdatering av mappingene, innholdsmessig og med hensyn til teknisk løsning. Vedlikeholdet skal utføres av UBO, som loggfører endringer i kildevokabularene. NB melder endringer i WebDewey som får konsekvens for mappingene utført av UBO.

Prosjektet skal utarbeide en plan for hvordan mappingkompetansen i UBO-personalet skal vedlikeholdes etter at prosjektperioden er utløpt.

7 Ytterligere informasjon

7.1 Liste over prosjektmedlemmer m/kontaktinformasjon

Prosjektleder:

Unni Knutsen, Seksjonsleder ved HumSam, Seksjon for tilvekst og katalog
E-post: unni.knutsen@ub.uio.no, Mobil: +47-95777978

Prosjektdeltakere mapping:

Berit Sonja Hougaard, HumSam
Vibke Stockinger Lundetræ, HumSam
Mari Lundevall, Realfagsbiblioteket
Lembi Viola Kuldvere, Realfagsbiblioteket
Grete Seland, HumSam

Prosjektdeltakere IKT:

Dan Michael Heggø, Realfagsbiblioteket
Are Dag Gulbrandsen, USIT

Prosjektet samarbeider tett med Nasjonalbiblioteket v/Elise Conradi, redaktør for Norsk WebDewey.

7.2 Prosjektsidene

Prosjektets hjemmesider ligger på UiOs intranettsider: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/>.



8 Referanser

EDUG (2015, 9. oktober). *EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)*. Versjon 20151009. Hentet fra http://edug.pansoft.de/tiki-download_file.php?fileId=113

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

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

Mapping mot norsk WebDewey

Prosjektsøknad: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/prosjektsoknad.pdf>

Tildelingsbrev: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/tildelingsbrev-mapping.pdf>

Statusrapport og søknad for andre prosjektår: https://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/rapportering_nb_2015.pdf

På vei mot en generell norsk tesaurus? Delprosjekt *Metodikk for mapping av Humord mot WebDewey*

Prosjektsøknad:

<http://www.nb.no/Bibliotekutvikling/Utviklingsmidler/Prosjektoversikter/Paagaende-prosjekt/Paagaar-t.o.m.-2015>

Tildelingsbrev: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/avsluttet/tesaurus-mapping/delte-dokumenter/tildelingsbrev-tesaurus.pdf>

Sluttrapport: <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/avsluttet/tesaurus-mapping/delte-dokumenter/prosjektrapport-tesaurus-mapping.pdf>

