

Revisjon av bokprosjektet, del 1

Innhold:

1. Innledning
2. Initiativet til revisjon av prosjektet
3. Oppsummering fra arbeidet i fase I
4. Status for Bokprosjektet august 2009

Vedlegg:

- Prosjektskisse fra BIBSYS
- ”Registrering av vitenskapelig publisering i monografier og artikkelsamlinger.” Brev fra Kunnskapsdepartementet til statlige universiteter og høyskoler 19.12.2006.
- Prosjektplan for revisjonen av bokprosjektet

1. Innledning

Våren 2005 ble det satt i gang et prosjekt for formidling og vedlikehold av data fra bibliografiske datakilder. Dette ble omtalt som "Bokprosjektet" og tok utgangspunkt i at BIBSYS-data skulle kunne brukes i systemene for forskningsdokumentasjon. Bokprosjektet er et initiativ fra Kunnskapsdepartementet, som ble vedtatt implementert fra 1. januar 2007, og som bygger videre på "Vekt på forskning" fra 2004. Bokprosjektet tar utgangspunkt i at data fra BiBIBSYS skal importeres inn i ITAR-tjenesten og deretter lastes inn i registreringssystemene FRIDA og Forskdok. Behovet for en revitalisering/revisjon av bokprosjektet bygger på erfaringene etter driftsperioden i 2008 hvor det har vært rapportert om manglende data fra ITAR-importene.

NSD har på initiativ fra referansegruppen¹ for ITAR nedsatt en arbeidsgruppe som har hatt som mål å jobbe tettere på bokprosjektet for å finne mer ut om hvorfor dataene er mangelfulle, og å kartlegge de kritiske prosessene. Målsetningen med revitaliseringen er å gjøre konkrete tekniske forbedringer av bokimportene, komme med forslag til tiltak som vil forbedre datakvaliteten samt vurdere bokprosjektets fremtid.

Denne rapporten gir en oppsummering fra arbeidet med revisjonen av bokprosjektet og tar for seg mer om initiativ og bakgrunnen for bokprosjektet, og resultatet av kartleggingsprosessen så langt. I siste del gir arbeidsgruppen sine anbefalinger om tiltak og det videre arbeidet.

2. Initiativet til revisjon av prosjektet

På møte i ITARs Referansegruppe i mars 2009 la BIBSYS frem en prosjektskisse til revisjon av Bokprosjektet der bakgrunnen for forslaget ble oppsummert slik:

- Ikke alle poster er registrert som vitenskapelig publisering i BIBSYS holder god nok bibliografisk kvalitet. En del poster er mangelfullt eller feil registrert.
- Det skjer en underrapportering fra BIBSYS til ITAR. Bare omlag 50 % av de postene som er registrert som vitenskapelige monografier blir eksportert til ITAR. Dette kan skyldes forhold nevnt i punkt 1 eller tekniske feil i eksportprogrammene.
- Ikke god nok utnyttelse av dataene i BIBSYS. Dette skyldes både mangler i eksportprogrammene og begrensninger i xml-skjemaet for overføringsformatet.
- Brukerne av Frida og Forskdok ønsker i størst mulig grad å kunne bruke ferdigregistrerte, importerte poster framfor egenregistrering.

Et kort tilbakeblikk på bokprosjektet

"Vekt på forskning" fra 2004 anbefaler at forskningsdokumentasjonssystemene bruker BIBSYS som bibliografisk datakilde til rapportering. Våren 2005 ble det satt i gang et prosjekt for formidling og vedlikehold av data fra bibliografiske datakilder. Dette "Bokprosjektet" ble beskrevet som den høyest prioriterte oppgaven for Felles tjeneste i 2005. Tre utfordringer ble identifisert i arbeidet med å tilveiebringe data fra BIBSYS til ITAR:

¹ Referansegruppen for ITAR består av representanter for BIBSYS, Forskdok, FRIDA og UHRs Publiseringsutvalg og NSD

- Definisjon og avgrensning av omfanget av bøker som skal analyseres.
- Tilrettelegging av nye felt i BIBSYS' bibliotekbase, og system for automatisk viderefremming av publikasjonsdata til ITAR.
- Etablering av rutiner for registrering på bibliotekene.

I februar 2006 slo UHRs Publiseringsutvalg fast at arbeidet med å bruke fagbibliotekenes registrering av monografier og artikkelsamlinger som kilde til data, ikke har ført fram, og at nye framgangsmåter må vurderes. Det ble skissert fire ulike alternative løsninger for registrering av bokpublisering:

1. Videre arbeid med modellen som er skissert i Vekt på forskning, som innebærer utvidet registrering ved første gangs registrering i fagbibliotekene.
2. Norart primærregistrerer bokutgivelser ut fra pliktavlevering
3. NSD videreutvikler metoden som ble anvendt i 2005 for å fange opp bokpubliseringer.
4. Bokpubliseringer egenrapporteres.

I mai 2006 ble det vedtatt i Publiseringsutvalget at 'Publiseringsutvalget går for en løsning der bokpubliseringer registreres av bibliotekene gjennom BIBSYS fra og med 2007. Det forutsettes at departementet pålegger bibliotekene denne oppgaven. Publiseringsutvalget anmoder BIBSYS om å gi prosjektet prioritet. Særlig gjelder dette forprosjektet høsten 2006.'

Kunnskapsdepartementet sendte så 19.12.2006 ut et brev til landets universiteter og høyskoler der institusjonene ble pålagt å følge opp de behovene og retningslinjene som er beskrevet i anbefalingen fra UHR (eget vedlegg). Helt konkret innebar pålegget at bibliotekene skulle registrere følgende tilleggsdata på de bibliografiske postene:

- a) Hvorvidt publikasjonen er vitenskapelig
- b) Antall forfattere
- c) Institusjonstilhørighet. Institusjonene identifiseres gjennom oppslag i DBHs institusjonsregister.
- d) Forfatternavn og -relasjon til institusjon.

I 2007 ble det laget registreringssystem for ITAR-attributt i BIBSYS' bibliotekbase. Til tross for at XML-schema for ITAR-uttak fra BIBSYS ble omarbeidet utover året, var det høsten 2007 fremdeles strukturfeil i uttakene. Spesielt var det problemer med forholdet mellom artikler og artikkelsamlinger, og med referanser til forlag og forlagskoder. I november 2007 hadde BIBSYS ferdigstilt det tekniske opplegget for import av bokdata. NSD påpekte at datakvaliteten fremdeles hadde svakheter, men dataene ble lastet inn i testversjonen av Frida, slik at superbrukerne kunne vurdere nytten av de importerte dataene.

Analyse av datakvalitet desember 2007

I desember 2007 undersøkte NSD datakvaliteten på bokdataimporten. Det ble funnet at:

- Mengden data var mindre enn forventet. Det ble konkludert med at registreringsrutinene ikke er tilstrekkelig gjennomførte på institusjonene.
- Det ser ut til at importen bare omfatter vitenskapelige publikasjoner, slik instruksjonen til institusjonene sier.
- Det mangler dataelementer på mange av postene. Det ble konkludert med at dette skyldes registreringspraksisen i BIBSYS.
- I en del tilfeller var det registrert feil årstall i forhold til hva som er oppgitt på forlagene. Kan skyldes andreutgaver av bøker i pliktavlevering.

Konklusjonen av analysen var at datakvaliteten ikke var tilfredsstillende pr. desember 2007. Det ble likevel anbefalt at datasettet ble tatt i bruk i Frida og Forskdok, fordi dette ville tvinge fram bedre registreringspraksis og uttaksrutiner i BIBSYS. Følgende konkrete tiltak ble foreslått:

- Nye spesifikasjoner for forskningsbibliotekarer.
- Gjenta pålegg om pliktavlevering av bøker og registrering i biblioteksystemene.
- Systematisk gjennomgang av bibliografiske referanser på NSD. Oppfølging av dataleveransene med rutiner for gjennomgang og påføring av manglende verdier på NSD.
- Opprettelse av et eget "Publiseringsutvalg" for bøker.

På ITAR-møte i juni 2008 kom man til enighet om at de bibliografiske dataene fra BIBSYS var kvalitetsmessig holdbare for inkorporering i Frida og Forskdok. I november samme år kom det så fram at dataomfanget i importen var bemerkelsesverdig lite. Følgende sannsynlige årsaker ble påpekt:

- UB på UiB og UiT har ikke startet systematisk registrering med sikte på import til ITAR. Opplegget er fremdeles heftet med uklarhet i forhold til definisjoner, ansvarsforhold og rollefordeling. Feil og mangelfull registrering forekommer.
- Prinsippet om pliktavlevering til bibliotekene er i liten grad fulgt opp.

Selv om en tok disse forholdene i betraktning, ble dataomfanget betraktet som mindre enn forventet. Problemet med mangelfulle og få data har preget bokprosjektet helt siden oppstarten i 2005. Viktigheten av at registreringspraksisen ved bibliotekene tilpasses ITAR-import er påpekt jevnlig siden 2005. Forventningen fra desember 2007 om mangelfulle data ville framtvinge bedre registreringspraksis i i Frida og Forskdok og bedre uttaksrutiner i BIBSYS, kan ikke sies å ha blitt innfridd.

Revisjonen av bokprosjektet

På ITAR-Referansegruppemøte i mars 2009 la BIBSYS fram et forslag til revisjon av bokprosjektet, der formålet var å se prosjektet i sin helhet og finne ut hvorfor dataene er få og mangelfulle. Etter møtet i ITAR Referansegruppe den 19.03.2009 utarbeidet NSD en prosjektplan for revisjonen av bokprosjektet. Planen innebar en inndeling av revisjonen i to faser. Første fase er en kartleggingsfase der målet var å finne ut konkret hvilke avvik som oppstår i overføringen av bibliografiske poster fra BIBSYS til ITAR. Resultatene av fase 1 skal fremlegges på et eget møte om bokprosjektet 9. september.

Det ble også foreslått at Frida og Forskdok skulle ha en brukerrepresentant hver i arbeidsgruppen. Hensikten var at forskningsdokumentasjonssystemene skulle få komme med sine innspill og erfaringer som brukere av dataene fra bokprosjektet. Fra mars til august 2009 ble fase 1 gjennomført. En testoverføring ble gjennomført, og graden av informasjonssvinn i

dataoverføringen ble kartlagt. I denne prosessen kom BIBSYS med forslag om å påføre ITAR-kode på de bibliografiske postene i BIBSYS.

NSD har også holdt møter med representanter fra Forskdok og Frida, der vi fikk viktige innspill om forbedringspunkter.

UHRs Biblioteksutvalg ble i april orientert om revisjonen av bokprosjektet. NSD vil følge opp bokprosjektet i samarbeid med UHRs Biblioteksutvalg i det videre arbeidet.

3. Oppsummering fra arbeidet i fase I

Kartlegging av svinn i dataeksport

Etter møtet i ITARs Referansegruppe den 19.03.2009 konsentrerte NSD seg om å koordinere en testoverføring av bibliografiske poster fra BIBSYS til ITAR. Hensikten var å kontrollere dataene i tre stadier i overføringen: i BIBSYS, i XML-filen som NSD mottar, og i ITAR.

Hovedresultatet for juni måned er at bare **26%** av det samlede datapotensialet fra BIBSYS kom velberget inn i ITAR. Det ble talt opp og funnet 513 vitenskapelige monografier i BIBSYS, og 224 av disse kom med i XML-filen som ble produsert samme dag for eksport til ITAR.

Resultatet viser at kun **44%** av de vitenskapelige monografiene og antologiartiklene i BIBSYS ble plukket ut til forsendelsen fra BIBSYS.

I all hovedsak skyldes dette mangler i både registreringspraksis og i datautplukk i BIBSYS-systemet. Blant de 513 postene i BIBSYS falt 18% av på veien fra XML-skjemaet og inn i ITAR. Konkret bør USIT her forbedre og dokumentere xml-skjemaet slik at det ikke slipper igjennom poster med feil eller ufullstendige data. Hovedresultatet for juni måned er at bare **26%** av det samlede datapotensialet fra BIBSYS kom velberget inn i ITAR.

Det betyr at det fins 513 monografier i BIBSYS Bibliotekbase som bibliotekarene har merket som vitenskapelige. I utgangspunktet burde da alle vært eksportert til ITAR. Men det er også slik at en del av disse må siles vekk pga. manglende ISBN eller annen feilregistrering. Poenget er at det virker urimelig at bare 224 blir igjen etter kvalitetssikring. Det er dette forholdet vi undersøker.

Forslag fra BIBSYS om påføring av ITAR-kode på bibliografiske poster

Som en del av arbeidet med å kartlegge de kritiske prosessene fra BIBSYS til ITAR henvendte BIBSYS seg til NSD med forslag om at BIBSYS legger på ITAR-kode på objektene i BIBSYS Bibliotekbase, slik at de mer forutsigbart blir med på eksporten via Bokprosjektet.

NSD ga tilslutning til denne løsningen, og påpekte også at det må følge med en kanaltypekode, siden ITAR-koden ikke er unik mellom periodika og forlag. NSD gikk ut fra at ITAR-koden kunne legges på automatisk, og at det dermed ganske raskt i fase 1 kunne gjennomføres en testoverføring der dette ble utprøvd. Hensikten var å se om det ble vesentlig mindre datasvinn med denne ordningen. BIBSYS orienterte om at dette ikke kunne gjøres umiddelbart, men at det trengtes tid og ressurser til arbeidet. BIBSYS satte likevel sitt eget utviklings- og vedlikeholdsteam

i gang med forbedring av eksportprogrammet til Bokprosjektet. Arbeidet er planlagt ferdig i august. Rapportene venter på resultatet av dette arbeidet

Tilbakemeldinger fra brukere

Formålet med kontakten med Forskdok-ansvarlige og Frida-superbrukere er å lære av de erfaringene som de har gjort seg som brukere av dataene fra bokprosjektet. Kontakten har bestått i e-postkontakt og møte med Forskdok-ansvarlige fra utvalgte institusjoner. Fra Frida-superbrukerne har NSD mottatt skriftlige rapporter innsamlet av USIT. I tillegg har disse gitt muntlige tilbakemeldinger til NSD på Frida-arbeidsgruppemøtet i Oslo den 15.06.2009.

Fra Forskdok-ansvarlige har NSD mottatt tilbakemeldinger fra UiS, UMB og NLA. NSD har også hatt et uformelt møte med Forskdok-ansvarlige ved NHH, NLA og HiB i NSDs lokaler den 27.06.2009.

Tilbakemeldingene fra Forskdok-ansvarlige var at den importen som forekom, kom for sent, og at de hadde rukket å egenregistrere posten først. Noen av de Forskdok-ansvarlige ved institusjonene driver også registreringsarbeidet for BIBSYS ved institusjonen. De sitter dermed i begge ender av prosessen. Det ble uttrykt at bibliotekarene i BIBSYS-registreringsprosessen nok ikke var tilstrekkelig bevisste på den senere bruken av posten i Forskdok-sammenheng. Det er dermed et forbedringspotensiale i å tilpasse registreringsarbeidet i BIBSYS slik at postene passer bedre til bruk i ITAR og i Forskdok. Forskdok-representantene påpekte behovet for ny informasjon og tettere oppfølging av pålegget til bibliotekene om registrering av tilleggsverdier på de bibliografiske postene i BIBSYS.

Det ble også fra de Forskdok-ansvarlige etterlyst import fra andre kilder enn BIBSYS, ISI og Norart.

Frida-superbrukere har gitt tilbakemeldinger om Bokprosjektet til USIT institusjonsvis. USIT oppsummerer disse tilbakemeldingene slik:

- Det finnes en del bøker uten tittel
- Institusjonsnummer er ikke alltid riktig, og ofte innblandet NSD-koder. En tilleggsfunksjon for institusjonsnavn ville gjøre det lettere å sjekke at det ble riktig. Nå kan det være at et nummer kan bli knyttet galt opp til en institusjon siden NSD-kodene og ITARs institusjonsnummer kan overlappe. Det finnes felt i utvekslingsformatet hvor institusjonsnavn kan legges.
- Det foreligger en del bøker uten ISBN som muligens skal være tidsskriftnummerserier, men hvor tidsskriftnavn ikke er oppgitt. Finnes felt i utvekslingsformat hvor dette kan legges.
- Man får ikke ISSN /tidsskrift på Tidsskriftserier. Dette har vært i utvekslingsformatet siden 15. september 2008 (i tillegg til tidsskriftreferanse som har vært der i flere år).

Hvis en følger forslaget fra BIBSYS om å påføre ITAR-kode på de bibliografiske postene i BIBSYS, kan de to siste problempunktene elimineres, og det første problempunktet trolig reduseres. Serien med ISSN vil da identifiseres gjennom ITAR-koden.

Det ble også på møte i Fridas arbeidsgruppemøte i Oslo den 15.06.2009 gitt muntlig tilbakemelding til NSD om Bokprosjektet. Her ble det påpekt at bibliotekarer som registrerer nye monografier i BIBSYS, ikke gjennomfører denne registreringen med hensyn til senere bruk av posten i forskningsdokumentasjonssystemene. Det var en klar tilbakemelding fra dette møtet at forbedringer i rutineene for førstegangsregistreringen av monografier i BIBSYS vil være det viktigste tiltaket for å få Bokprosjektet til å fungere.

Ut fra tilbakemeldingene fra brukerne samlet sett, er deres viktigste tilbakemelding at registreringsrutinene ved bibliotekene er for lite tilpasset ITAR-bruk, og at de trengs å følges opp. Brukerne henviste til brevet fra Kunnskapsdepartementet i 2006, og etterlyste oppfølging av dette. Ut over dette meldte brukerne om manglende data (spesielt ISBN, ISSN og tittel). Dette tyder på at forslaget om påføring av ITAR-kode ved første gangs registrering er hensiktsmessig.

4. Status for Bokprosjektet august 2009

I slutten av august 2009 har vi testet den nye versjonen av eksportprogrammet fra BIBSYS.

I forhold til de 513 monografier i BIBSYS Bibliotekbase som bibliotekarene har merket som vitenskapelige burde alle vært eksportert til ITAR. En del av disse må siles vekk på grunn av manglende ISBN eller annen feilregistrering. Det virker urimelig at bare 224 blir igjen etter kvalitetssikring. Det er dette forholdet BIBSYS nå undersøker, og resultatet av dette vil foreligge snart.

Hovedkonklusjonen fra revisjonen av bokprosjektet så langt kan oppsummeres slik:

- Kvaliteten på bokdata fra BIBSYS til ITAR var pr juni 2009 ikke tilfredsstillende.
- BIBSYS- nye eksportprogram har forbedret kvaliteten
- Tilbakemeldingen fra brukerne og forslaget fra BIBSYS tyder på at det er behov for et forbedret informasjonsopplegg og tettere oppfølging av førstegangsregistreringen av monografier og antologiartikler i BIBSYS.
- Det største datasvinnet skyldes mangelfull registreringen av bibliotekene i BIBSYS samt mangelfull kunnskap og rutiner for registrering av data som brukes i finansieringssystemet.

På bakgrunn av arbeidet i arbeidsgruppen foreslås det en satsing på følgende punkter:

1) *Nytt pålegg fra Kunnskapsdepartementet:*

Klare og nøyaktige rutiner for hvilken tilleggsinformasjon som skal registreres på de bibliografiske postene, og hvordan dette skal gjøres. Rutinene må pålegges bibliotekene gjennom et nytt brev fra KD.

2) *Opplæring av bibliotekarer:*

Gjennomgang og opplæring i ITAR-relatert registrering gjennom eget opplegg i regi av BIBSYS i samarbeid med NSD. Opplæringen bør omhandle både den konkrete registreringspraksisen og bakgrunnen for pålegget.

3) *Rapporteringsansvarlige ved bibliotekene:*

Oppnevning av en fast rapporteringsansvarlig ved hvert bibliotek for ITAR-relatert registrering ved bibliotekene, og sterkere forankring i ledelsesnivået ved bibliotekene.

4) *Tettere samarbeid om informasjonsoppbygget og nye registreringsbehov i UHRs biblioteksutvalg:*

UHRs biblioteksutvalg er orientert om revisjonen av bokprosjektet. Biblioteksutvalget bør også orienteres om det nye informasjonsoppbygget rundt førstegangsregistrering av monografier og antologiartikler.

Tiltak:

- NSD og BIBSYS utarbeider en oppfølgingsplan for bibliotekarer med fokus på opplæring og forståelse av bokregistreringer i finansieringssystemet
- BIBSYS utarbeider retningslinjer for interne rutiner
- Ressursbehovet må meldes inn fra de involverte slik at systemet kan implementeres.
- Opprette en utvalgfunksjon for bokpublisering.
- Nye retningslinjer og rutiner skal forankres på ledelsesnivå i institusjoner med bibliotek, ved at det utpekes en rapporteringsansvarlig ved hvert bibliotek. KD og NSD må underrettes om det.
- Et startseminar forut i regi av BIBSYS med støtte fra NSD. I denne sammenheng må retningslinjer og frister klargjøres.

Arbeidsgruppen ser nytten av et velfungerende system av bokimporter og går inn for at det satses videre på arbeidet med bokprosjektet. Utgangspunktet for det videre arbeidet i fase II er å legge til rette for de anbefalinger og tiltak som rapporten konkluderer med.

31.08.09 NSD/GE/BL/LPJ

Vedlegg

Prosjektskisse BIBSYS (vedlagt dokument)

Brev fra Kunnskapsdepartementet (vedlagt dokument)

Prosjektplan for Revisjon av bokprosjektet

Mye av behovene for tekniske utbedringer i tilknytningen til bokprosjektet er også relatert til driften av ITAR ved USIT. NSD sitter med ansvaret for kvaliteten i ITAR-tjenesten, og NSD ser på denne revisjonen også som en mulighet til å ta et nytt grep om rutine og funksjonene i ITAR. På denne måten ønsker vi å få klarhet i behovene for både nyutvikling og forbedringer i forhold til autoritetsregisteret for publiseringskanaler.

NSD ser for seg at det videre arbeidet nå deles inn i to faser, og at prosjektet følger en konkret oppfølgingsplan. Første fase bør være en kartlegging av hva de største problemene i bokprosjektet har vært, og hva som må gjøres for å utbedre dette. Dette kan gjøres ved å overvåke et eksempeldatasett på veien fra BIBSYS til ITAR, og på denne måten avdekke hvilke ledd som er svake og hvordan dette kan utbedres. Prosjektets neste fase blir å løse de problemene som er blitt påpekt i fase 1.

Fremdriftsplan for bokprosjektet:

Fase 1: Status for bokprosjektet, kartlegging av forbedringspunkter

April/mai:

- NSD, Bibsys og ITAR gjennomfører en testoverføring av bokdata fra Bibsys til ITAR.
- NSD sammenlikner det originale Bibsys-datasettet med ITAR-datasettet for å avklare graden av "informasjonssvinn" av data i overføringen.
- Møte mellom NSD, Bibsys, USIT og to representanter for Frida- og Forskdok-brukere. Innspill fra brukerrepresentantene om forbedringspunkter.
- NSD skriver rapport om status og forbedringspunkter for bokprosjektet.

Juni:

- Rapport med resultatene av fase 1 legges fram på møte i ITAR referansegruppe den 8. juni 2009. Fremgangsmåte for prosjektets fase 2 fastlegges i gruppen.

Fase 2: Utbedring av rutiner i bokprosjektet

- NSD fordeler oppgaver mellom seg selv, Bibsys og USIT for høsten 2009 i samsvar med den fremgangsmåte det er blitt enighet om i ITAR-referansegruppen. Det er på forhånd påpekt to grunner til at bokprosjektet ikke har fungert tilfredsstillende: dårlig registrering av bibliografiske poster i Bibsys og dårlige eksport- og importprogrammer og -rutiner i Bibsys og ITAR. Arbeidsoppgavene til NSD, Bibsys og USIT vil bli utformet innenfor disse problemfeltene, utfra en vurdering av resultatene en er kommet fram til i prosjektets fase 1.

Oktober:

- Bibsys og USIT leverer rapport om status for sine respektive arbeidsoppgaver i prosjektet til NSD.

November:

- NSD fremlegger rapport for Revisjon av bokprosjektet på møte i ITAR-referansegruppen den 5. november 2009. Rapporten skal også fremstilles for Styringsgruppen for DBH.

Forslag til medlemmer i arbeidsgruppen for bokprosjektet:

Gunnar Ellingsen (NSD) – Koordinatorfunksjon

Jan Erik Kofoed (Bibsys)

Grete Christina Lingjærde (USIT)

Brukerrepresentant (UiB)

Brukerrepresentant (HiO)