

# Mens vi venter på integrated search og databrønd

Af Eli Greve

**Det må være ideelt at samle flest mulige kilder i færrest mulige tjenester. Det må også være ideelt at lægge nyt ind i tjenester, brugerne kender og benytter sig af. For at opnå begge fordele, kan det være nødvendigt at indføre nye arbejdsgange. Og når det betyder bøv! for nogle få og er til glæde for de fleste, ja så må det være et rigtigt valg.**

Eli Greve  
Syddansk Universitetsbibliotek  
eligr@bib.sdu.dk



## Hvordan får brugerne den enklest mulige adgang til bibliotekets materialer?

For at få det størst mulige udbytte af satsningen på elektroniske ressourcer har Syddansk Universitetsbibliotek valgt at integrere ressourcerne i katalogen.

Tidsskriftabonnementerne kommer rimeligt nemt ind. Alle aktive full-text titler trækkes ved hjælp af SFX export tool i Aleph formatet ud af SFX. Et lokalt udviklet konverteringsprogram omdanner posterne til marc txt format og supplerer posterne med 001, 008 og 009 felter. Posterne lever nu op til de lokale krav til data og katalogen opdateres. Ikke opdaterede titler skal ud af bibliotekssystemet. De slettes af et sql-job. Bibliotekssystemet holdes ajour med data fra SFX ved opdatering med tidsskriftposter hver 14. dag. Arbejdsindsatsen på maks. 1 time pr. gang må betegnes som minimal.

Brugerne har således e-tidsskrifter søgbare sammen med de trykte tidsskrifter. Og jo, der er et særligt indeks til e-tidsskrifterne alene og de kan findes på hjemmesiden som Elektroniske tidsskrifter.

## Hvordan med e-bøger?

Ud fra betragtningen om at det skal være simpelt for brugerne at finde litteraturen, lægges e-bøger i katalogen. Brugerne fritages for overvejelser om materialets fremtrædelsesform og biblioteket stiller materialet til rådighed i et kendt værktøj. Brugere, som forventer at finde materialet som e-ressource udenfor katalogen, har via bibliotekets hjemmeside adgang via link til udbyders grænseflade.

Metadata til bøger er ikke umiddelbart så enkle at skaffe. Området er nyt, udbydere mange og løsningsmodellerne stort set lige så forskellige. Derfor bliver det også yderst interessant, hvordan samarbejdet om databrønden vil prioritere tidsskrifter/artikler overfor bøger.

## Er poster i marcformat fra udbyder en selvfølghed?

Billedet er meget blandet. Området er helt klart under udvikling og bærer præg af, at udbydere mangler viden om på hvilken måde, abonnenterne har eller kan have brug for at anvende data.

Nogle udbydere leverer marcposter i veldefinerede load-filer samt filer over titler, som ikke længere dækkes af licensen. Posterne overholder standarderne og kan importeres uden vanskeligheder. Det andet yderpunkt er pakker med en søgeflade som service. Metadata kan måske skaffes ved ihærdig indsats og med nogen forsinkelse.

## Mellem yderpunkterne er mange varianter

E-bogsbestanden på Syddansk Universitetsbibliotek udgør knap 90.000 pr. 31.12.2009. Og hvordan har vi fået det? Hvad har vi gjort, når udbydere ikke leverer poster i et eller andet marcformat.

Ebrary er absolut duksen i e-bogsligaen. Produktet er professionelt og gennemtænkt. Nye data filer lægges ud hver måned og dataene overholder standarder og formater. Også ACLS Humanities e-book leverer fine datafiler, men indtrykket skæmmes af organisatoriske navneændringer og omlægninger af URL-mønstre, som har nødvendiggjort import af de samme titler i flere omgange.

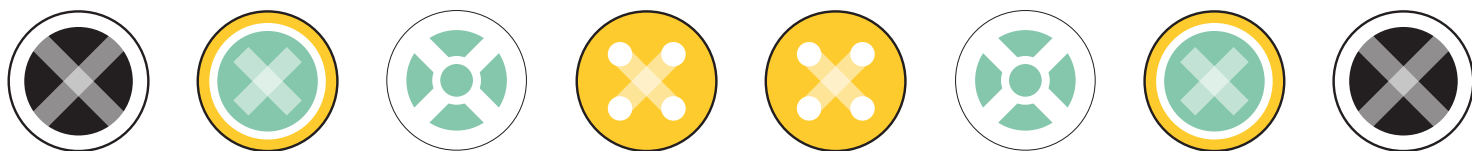
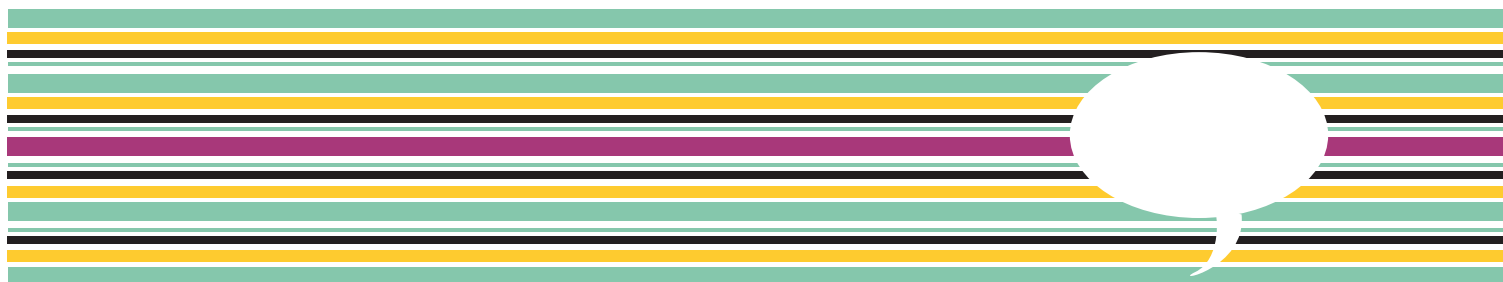
Udbydere med stærkt differentierede tilbud har komplekse strukturer for loadfiler. Det kan til tider være vanskeligt at afgrænse hvilke filer, der hører til det enkelte køb. Der kan være nogen usikkerhed, om udbydere vil publicere filer med de titler, som ved en fejl er lagt i marcfiler som fulltext. Det vil tiden vise. Særlige problemer opstår, når formater og standarder ikke overholdes.

Andre leverandørerne har valgt at løfte opgaven ved at stille websider til rådighed med mulighed for at downloade i Excel eller som kommasepareret fil. I heldigste tilfælde er dataene grupperet i periodevise loadfiler, eller dataene kan sorteres inden download, så kun titler siden sidste opdatering kommer med. Eksempelvis danner vi katalogposter på proceedings fra IEEE Explorer. IEEE har lagt titellister i Excel filer ud i grupperne 1998-2001, 2002-2004, 2005 to present. Filerne kan sorteres, dvs. det er muligt at isolere nye data, så kun tilvæksten skal behandles. Dataene omfatter konference titel, akronym, sted, dato, ISBN, ISSN og vigtigst en stabil URL. Der sker nu en berigelse af metadata med 001, 008, 009 felter suppleret af URL kommer ganske pæne marcposter ud af, når de har været gennem et lokalt konverteringsprogram.

## Resultatet er poster som sikrer genfindelighed

En ny måde at skaffe data på kom i brug ved ACM Association of Computing Machinery. ACM udgiver et lange række proceedings, samt i regi af en Special Interest Group. Proceedings fanges ikke af SFX, så der skal gøres noget andet, hvis brugerne skal kunne søge dem frem. De specifikke konferencetitler kan variere over tid, mens grupperne synes stabile. Marcposter var det ikke muligt at få. Vi fravalgte at investere tid på at foretage egentlig katalogiseringer. I stedet valgte vi at konstruere poster på basis af oversigten over ACM proceedings. Kort fortalt kopieres siden, gemmes som webside, kilden vises og kopieres til Note tab. Nu har vi basisoplysninger om de enkelte proceedingsrækker incl. URL, og kan danne marcposter. Dataene er

Brugerne har således e-tidsskrifter søgbare sammen med de trykte tidsskrifter. Og jo, der er et særligt indeks til e-tidsskrifterne alene og de kan findes på hjemmesiden som Elektroniske tidsskrifter.



sparsomme, men proceedingstitlerne er velkendte og citerede. Metoden til datafangst giver ikke mulighed for at vise konferencens år, men URL'erne går til proceedingtitlens oversigtsside og her folder det hele sig ud.

I værste fald oplyser leverandører intet, blot kan man af antallet af titler sammenholdt med eksisterende antal i katalogen og det faktum, at der er nye titler i listen, udlede, at der må være forsvundet noget. Hvilke titler, det drejer sig om, oplyses ikke. Det stiller særlige krav til arbejdsprocedurerne, når opdateringen af katalogen også skal sikre, at en række ukendte poster fjernes.

Arbejdet med e-bøger er kendetegnet ved ny pakke = ny løsning. Derfor udarbejdes og vedligeholdes et stamblad pr. pakke eller abonnement. Stambladene er detaljerede vejledninger om rettelse, hvilke konvertere skal benyttes, arbejdsgange og andre særlige forhold gældende for den enkelte pakke. Det er eksempelvis tilføjelse til URL'en, rettelse af materialetype,

som bestemmer visningen i katalogen, samt omdøbning af datafelter så konflikter med andre data undgås. For at holde styr på det hele registreres arbejdet med den enkelte pakke i en log.

Lokalt er kravet til bibliografisk niveau for e-bøger reduceret i forhold til andre materialer. Det er genfindeligheden og eksponeringen, der er det vigtigste.

I det hele taget er det en udfordring at få en ny pakke e-bøger. Er der marc poster? Hvor gode er de eller skal vi igen knække en kode og skrive ny konvertering? Set fra det enkelte bibliotek absolut ikke ensidigt, gentaget arbejde, men spredt på flere biblioteker begynder det at ligne. Det er her databrønden kommer ind, eller?

Metadata til bøger er ikke umiddelbart så enkle at skaffe. Området er nyt, udbyderne mange og løsningsmodellerne stort set lige så forskellige. Derfor bliver det også yderst interessant, hvordan samarbejdet om databrønden vil prioritere tidsskrifter/artikler overfor bøger.