

# Flere spørgsmål end svar på fremtidens digitale bibliotek

Konferencen "Well11" belyste, hvordan ideen om databrønne kan realiseres. Konferencen afslørede, at der ikke er nogen fælles opfattelse af, hvordan fremtidens digitale bibliotek ser ud, ligesom der er langt til at etablere et samarbejde eller en forretningsmodel bibliotekerne imellem. Vi ser nærmere på udfordringerne her. Slides, oplæg og præsentationer er tilgængelige på [www.well11.dk](http://www.well11.dk).

Af Jens Hofman Hansen [jhh@statsbiblioteket.dk](mailto:jhh@statsbiblioteket.dk), Statsbiblioteket

## **U**dfordring: Giv brugeren ét, relevansrangordnet søgeresultat – hurtigt

Eksposeringen af indholdet i databrøndene er meget afhængig af den søgeteknologi, der anvendes. Forstår og mestrer bibliotekerne således søgeteknologien, kan de påvirke eller realisere visionen om, at brugere skal opleve bibliotekernes data på en relevant måde via et Google-inspireret søgeresultat (ét søgeresultat, hurtigt svar, simpel afgrænsning).

Der opereres nu med tre variationer af søgeteknik, der har hver deres fordele og ulemper. "Federated search" er kendt som den oprindelige søgeteknik, hvor data kan ligge i mange forskellige databrønne, men hvor søgeresultatet også typisk opbygges asynkront og ikke er ægte relevansrangordnet.

Svaghederne ved federated search har længe været søgt løst ved at arbejde med "integrated search". Her lægges alle data på én og samme server og alting indekseres samlet. Det giver den åbenlyse fordel, at et dokument relevans let kan vurderes i forhold til den samlede datamængde. I praksis betyder det at kolossale datamængder kan gennemsøges og præsenteres relevansrangordnet på under ét sekund. Denne teknik er nu at betragte som mainstream.

Men der er et problem. At samle alle data ét sted strider mod de forretningsmodeller, der er forbundet med handel med metadata i øjeblikket. Metadata er en handelsvare, som ikke er frit tilgængelig, og det er den helt store ulempe ved rendyrket integrated search, hvilket også var afsættet for "Well 11".

Et nyt begreb, "virtual integrated search"<sup>1</sup>, blev således introduceret på konferencen. Begrebet dækker netop over søgning, der foregår i indekser, der ligger på flere forskellige servere (f.eks. bibliotekets egen og en ekstern dataudbyders), men opfører sig over for brugeren,

som om alle metadata var placeret i et fælles indeks. Det er i praksis muligt at opnå et ægte relevansrangordnet søgeresultat under disse forhold, hvis nogle forudsætninger er opfyldt. I dette tilfælde baserer begge søgeindekser sig på Lucene/SOLR-teknologien. Det er herved muligt at udveksle de såkaldte term-statistikker (tf-idf), der er et udtryk for, hvor relevant et ord er i et dokument og i den samlede tekstkorpus. Virtual integrated search er et nyt fundament for at vi kan give et værdifuldt søgeresultat til brugerne, men giver naturligvis ikke pr. automatik et for brugeren relevant resultat. For præcist hvordan ønsker man at "booste" den ene datakilde frem for den anden? Hvordan skal dråben af bøger optræde i havet af artikler?

Svagheden ved "virtual integrated search" er for så vidt det samme som styrken, nemlig at data kan ligge på flere servere. Løsningen er kun så hurtig som den langsomste server. Men hvor forretningsmodellerne forhindrer os i at indekser store og vigtige metadata-samlinger på vores egne servere, viser det sig altså, at der nu reelt alligevel er løsningsmuligheder.

Når data ligger på flere servere, kan vi ikke umiddelbart uniformere data på tværs, og det kommer f.eks. til udtryk i problemer med afgrænsningsmuligheder (f.eks. afgrænsning på emne eller genre) der kan være vidt forskellig fra datakilde til datakilde. Det er et problem, der skal løses ved yderligere tekniske landvindinger eller ved aftaler med de eksterne dataleverandører.

## **Udfordring: Forøg værdien af data ved at koble data sammen**

Alle oplægsholdere på "Well 11" berørte på forskellige måder, hvorfor det at koble data sammen har betydning for arbejdet med databrønne og brugeroplevelsen af data i det hele taget.

Richard Wallis fra den engelske biblioteksvirksomhed

Talis er varm fortalere for “linked data”, som er en teknisk simpel måde at tilgængeliggøre og relatere data på en maskinlæsbar form (adgangen foregår typisk via http-protokollen og sammenkoblingen via URI'er)<sup>2</sup>. Han fremstiller næsten “linked data” som en modsætning til søgning: “Libraries are trapped in the search paradigm”. Imidlertid er linked data ikke et alternativ eller en modsætning til søgning. Wallis' vægtning af begrebet illustrerer blot hvor kompleks og omfattende et område, bibliotekerne opererer i, når de skal servicere brugerne i deres arbejde med at tilegne sig og opbygge viden.

Linked data er langt hen ad vejen det pragmatiske bud på, hvordan vi realiserer det “semaniske web”, hvor data er koblet sammen på kryds og tværs efter betydning. Linked data er også en trend i omverdenen, som atter engang minder bibliotekerne om, at vi langt fra er de eneste, der arbejder med at strukturere data. Ifølge Richard Wallis må vi se på, hvordan omverdenen strukturerer i stedet for at anvende vores egne metoder (MARC-formatet ses igen og igen som overkomplekst, unyttigt, en joke i nutiden). Bibliotekernes data skal selvfølgelig spille sammen internt – men også med omverdenen.

Af samme grund dødsdømmer tilhængere af linked data i øvrigt også forretningsmodeller, der baserer sig på at fastholde f.eks. artikelmetadata på udbydernes servere. Spørgsmålet er så blot, hvornår alle relevante metadata kommer i frit flow – om nogensinde. Den tekniske løsning “virtual integrated search” er netop et forsøg på at lade data, der ligger på en ekstern server, komme i spil her og nu i stedet for at vente på den totale frigivelse. “Open” er et relativt begreb.

Richard Wallis pegede på, at bibliotekerne kun i meget ringe grad selv frigiver deres data, så andre kan bruge dem. Imidlertid kræver frigivelse af dataressourcer, og alle biblioteker vil gerne se et hurtigt eller sikkert afkast af den investering, vi foretager. Frigivelse af data kan potentielt set give et kæmpemæssigt afkast, men investeringen er både usikker og formentlig langsigtet. Frigivelsen rører også dybt ved vores forestilling om det digitale bibliotek – er det hele befolkningen der skal skabe det i mange forskellige udgaver, eller bibliotekerne, der skal skabe én samlende portal? Sidstnævnte er nok nemmest for de fleste biblioteksfolk at forestille sig, men ikke nødvendigvis den mest fornuftige i vor tid.

### Udfordring: Integrér brugerdata

Indtil nu har bibliotekernes databrønde ikke rigtigt indarbejdet brugerdata. Det står i skærende kontrast til udviklingen indenfor fremfindning og sammenkobling af




Foto: Jakob Boserup

data i den verden brugerne ellers befinder sig i – også de mest seriøse forskere og studerende. Claus Dahl (fra [classy.dk](http://classy.dk)) udtrykte det sådan, at “hvis I [bibliotekerne] tænker på jeres databrønde uden at indarbejde brugerdata, så arbejder I jer frem mod noget, som ligger 15 år tilbage”. Brugerdata er bibliotekernes mulighed for at foretage det næste tigerspring inden for relevansrangordnede søgeresultater. Kender vi brugerens søge- og adfærdshistorik, kan søgeresultatet tilpasses brugeren og andre brugere. Alle er efterhånden enige om vigtigheden af brugerdata, men selv de store udenlandske biblioteksvirksomheder (SerialsSolutions, OCLC, Ex Libris) kan foreløbig kun fortælle om lovende projekter, der stadig er i støbeskeen. Ja, der er ideer og mindre initiativer i biblioteksregi, men der er lang vej endnu.

### Udfordring: Arbejd sammen, hav en fælles vision og mission, eksekvér

Det blev ikke udtrykt så direkte på selve konferencen, men i den efterfølgende evaluering, besvarede deltagerne spørgsmålet “Efter at du nu har deltaget i konferencen, hvad mener du så, er bibliotekernes største udfordring i forhold til virtuelle databrønde?” Næsten samstemmende svarede deltagerne, at den allerstørste udfordring for bibliotekerne er at skabe en platform for samarbejde – for med det store kompleks af udfordringer, der er forbundet med at levere relevans til brugerne via databrønde, søgning og sammenkobling af data, står det klart, at et enkelt bibliotek ikke kan klare opgaven.

Deltagerne havde heller ikke en oplevelse af, at der er en fælles vision og mission for fremtidens digitale bibliotek. Spørgsmålet er, om vi skal vente til at finde modellen på en evt. “Well 12”-konference, eller om vi kan agere hurtigere end det. Bibliotekerne kritiseres nemlig gang på gang (også af undertegnede) for, at vi planlægger og planlægger – og når vi når til at eksekvere, har verden flyttet sig, og så vi igen må til at planlægge. 

### NOTER

<sup>1</sup> <http://sbdevel.wordpress.com/2011/02/04/virtual-integrated-search/>

<sup>2</sup> [http://en.wikipedia.org/wiki/Linked\\_Data](http://en.wikipedia.org/wiki/Linked_Data)