

## **Nationell databrunn - möjligheter och behov**

### **Underhandsrapport**

#### **Syfte med förstudien**

Förstudien ska ge underlag till Expertgruppen för LIBRIS nationella system för vidare strategiska diskussioner om de olika möjligheterna som finns för att skapa en nationell databrunn. Förstudien resulterar också i en förståelse för vad det innebär att skapa en nationell databrunn.

#### **Definition av databrunn**

Termen databrunn har vi tagit till oss från danskans *databrønd*, (engelska *data well*). Målet med en databrunn är att samla metadata bland annat från e-resurser från olika informationsleverantörer på ett ställe.

Med databrunn avser vi i den här rapporten: Centralt lagrat metadata för (främst licensierade) e-artiklar, e-tidskrifter och e-böcker som står till förfogande för LIBRIS-biblioteken. Databrunnen understödjer integrerad sökning och export respektive import av data.

#### **Frågeställningar**

En viktig strategisk fråga är om vi själva ska utveckla en lösning för nationell databrunn från grunden eller om vi ska köpa en färdig lösning eller ingen databrunn alls! Finns det möjligheter att kombinera viss egenutveckling med inköp av färdiga moduler eller deltjänster? Vad blir konsekvenserna om vi väljer det ena eller det andra?

Innan man beslutar om färdriktningen när det gäller e-resurser är det viktigt att undersöka vilka behov biblioteken faktiskt har i denna fråga och på vilket sätt en nationell databrunn, oavsett modell, tillfredsställer och uppfyller dem. I resultatet från projektet LIBRIS som lokal OPAC (L-OPAC) får vi antagligen många svar.

Andra frågor är:

- Ska KB/LIBRIS driva en nationell databrunn eller är det någon annans uppgift? Ska och kan KB konkurrera med andra leverantörer på det här området?
- Vad är målet med databrunn för KB? Vilka behov uppfyller den och kan de behoven tillgodoses på annat sätt?
- Vad ska en nationell databrunn innehålla? Fria resurser, lokala resurser, avtalsbundna resurser? Vem ska bestämma över det?
- Ska vi inleda samarbete över nationsgränser?
- KB:s samlingar? Hur ska man förhålla sej till e-pliktleveranser och KB:s digitaliserings samordning?
- KB:s utökade uppdrag, vad kommer det att innebära?

## Bakgrund

Flera frågor har kommit till KB/LIBRIS under år 2009 om hur vi ställer oss till en nationell databrunn. Frågorna bottnar dels i de nya tjänster som började erbjudas på marknaden under 2009, dels i bibliotekens behov av att göra alla sina (fysiska såväl som elektroniska och digitala) resurser och samlingar sökbara på ett ställe dvs via ett enda gränssnitt och med en gemensam sökingång.

Stora systemleverantörer som till exempel Serials Solution och Exlibris erbjuder vad de kallar centrala index som de underhåller och uppdaterar, Summon respektive Primo Central. De samlar in bibliografisk metadata för artiklar från e-tidskrifter och e-böcker från olika informationsleverantörer (förlag och aggregatorer) och indexerar datat i index som ger möjligheter till integrerade sökningar tillsammans med bibliotekens lokala resurser som också kan samlas in och indexeras. I och med att datat är samlat på ett ställe kan man erbjuda en samlad sökingång för alla typer av resurser, snabba sökningar, relevans ranking, klustring av sökresultat, sammanslagen träfflista etc.

Danmarks elektroniske fag- och forskningsbibliotek (Deff) startade redan 2007 en utredning om en nationell databrunn för elektroniska medier i och med att man såg en ökning av användandet av elektroniska resurser och att tillhandahållandet av dessa till användarna var bristfälliga. Man saknade en samlad sökingång för användarna. Under

2008-2009 pågick i Danmark ett projekt som undersökte hur en databrunn ska styras, organiseras, finansieras, databrunns innehåll med mera. Sedan hösten 2009 pågår en nationell upphandling av databrunnsleverantör. Syftet med en nationell databrunn i Danmark är att det ska fungera som ett datalager som ställs till förfogande för alla medlemsbibliotek och att det ska vara en grund för kostnadseffektiv insamling, bearbetning och export av gemensamt centralt data. Databrunnen ska understödja integrerad sökning, det ska finnas funktioner för export och import av data samt olika grader av databehandling.

Redan tidigare har det funnits databrunnsinitiativ både i Danmark och Sverige. Danmarks tekniska universitet, Danmarks Tekniske Informationscenter (DTIC) har under många år samlat metadata i databrunnen Digital Article Database Service, DADS, som lanserades redan 1998. Lunds universitet, Biblioteksdirektionen, har en motsvarande databrunn i Electronic Library Information Navigator, ELIN. ELIN utvecklades 2001 och används idag av ca 30 institutioner, de flesta svenska men även en del utländska.

Ett av Sveriges större universitetsbibliotek, Göteborgs universitet, genomförde under år 2009 en förstudie om en ny söktjänst för alla typer av resurser. Man undersökte olika söktjänster som tillåter sökning även på artikelnivå och som också integrerar bibliotekskatalogen. Efter förstudiens avslutande har man beslutat att gå vidare med en lösning baserad på öppen källkod, Summa, som utvecklas på Statsbiblioteket i Århus, Danmark.

Samsök är en nationell lösning för hantering av och åtkomst till e-resurser. Den lanserades 2006 och använder portalprogramvaran Metalib och länkservern SFX utvecklad av systemleverantören Exlibris. Det är KB/LIBRIS som underhåller och uppdaterar Samsök på licens för idag 36 institutioner. Samsök erbjuder federerad sökning dvs sökning mot flera källor i ett gemensamt gränssnitt. I motsats till ovanstående tjänster där man kan söka direkt i ett index med metadata från flera källor samlat på ett ställe, så skickas sökfrågor ut från Metalib till flera källor parallellt och hämtas tillbaka för presentation i sammanslagen lista. Federerad sökning är av förklarliga skäl betydligt långsammare än integrerad sökning och det är en viktig förklaring till varför tjänsten Samsök inte används i så hög utsträckning.

KB/LIBRIS utvecklade under 2009 en ny söktjänst för forskningspublikationer, SwePub, som är tänkt att integreras i LIBRIS, dvs göras sökbart via LIBRIS gränssnitt. Man kan välja olika metoder att integrera datat på, till exempel genom att skapa en databrunn som ligger bakom LIBRIS istället för att slå ihop datat från SwePub med datat i samkatalogen. Man kan också välja att göra en mer flexibel lösning som tillåter insamlande av metadata från flera olika källor, som sedan kan göras sökbart via LIBRIS. SwePub kan så att säga vara embryot till en nationell databrunnslösning. Detta är en viktig aspekt inte minst med tanke på att Kungliga biblioteket förbereder inför e-pliktleveranser och också har en växande digital samling som ska exponeras. Metadata för detta material kan samlas i en databrunn för att på så vis göras sökbart via LIBRIS.

## Metod

För att KB/LIBRIS ska kunna ta ställning till frågan om en nationell databrunn behöver det undersökas vad det finns för förutsättningar och vad det skulle innebära för KB/LIBRIS att ta sig an denna uppgift. Förstudien omfattar:

- En marknadskartläggning av databrunnsleverantörer, såväl leverantörlösningar som open source lösningar och hybrider av dessa.
- En enklare teknisk testinstallation utifrån det system som byggts upp för SwePub. Testerna syftar till att ge en uppfattning om möjliga tekniska komplikationer vid insamling, uppdatering och indexering av metadata.
- En undersökning av förekommande begränsningar i tillgången till - och rättigheterna att använda - det metadata som samlas in eller erhålls. Licensverksamheten vid KB/LIBRIS tar också upp dessa frågor i samband med förhandlingar inför förnyelser av avtal med informationsleverantörer.

## Några preliminära scenarier

En bit in i kartläggningen av databrunnsleverantörer arbetar vi med ett antal *preliminära* scenarier. Vi bedömer det som sannolikt att LIBRIS samkatalog kommer att påverkas, oavsett vilken väg man väljer att gå.

### **Scenario *Egenutveckling***

Det här är ett scenario där vi utvecklar en databrunn från grunden på KB/LIBRIS. Vi har högre grad av kontroll över metadatat eftersom datat lagras centralt på KB/LIBRIS. Vi har kontroll över vilka källor som ska ösas in i brunnen. Det är möjligt att bygga olika index för sökning. Det metadata som idag finns samlat i SwePub, Uppsök och LIBRIS samkatalog utgör i praktiken en databrunn enligt vår definition. I det här scenariot utökas detta redan idag insamlade metadata med ytterligare material (utifrån behov och ambitionsgrad) samtidigt som tekniska och administrativa rutiner för detta utvecklas.

Ambitionsnivån styr över hur resurskrävande detta scenario blir. Ska det finnas möjligheter att samla in lokala avtalsresurser/samlingar in i den nationella databrunnen kräver det förstås mer resurser än om man begränsar sig till till exempel artikeldata motsvarande de centrala avtal som licensverksamheten förhandlar om. Frågan är vad biblioteken är nöjda med? Fria resurser medför inga leverantörskontakter, men någon form av policy behöver upprättas för vad som ska samlas in. Även den tekniska infrastrukturen behöver anpassas för hantering av större mängder data.

Scenario *Egenutveckling* kan också kombineras med att KB/LIBRIS köper viss metadata för central lagring och på så vis slipper en del arbete med metadataflödet. Detta skulle naturligtvis medföra en upphandling av databrunnsleverantör med allt vad det innebär. Se vidare följande scenario.

### **Scenario *Upphandling av färdig lösning***

I detta scenario har vi en färdig databrunnslösning från en systemleverantör. Vi har ingen kontroll över metadatat eftersom datat lagras hos leverantören. Åtkomst till extern databrunn finns i bästa fall via API:er (Application Programming Interface dvs gränssnitt avsedda för maskin till maskin kommunikation, som möjliggör integrering och tillgängliggörande av till exempel en söktjänst i en helt annan kontext). Det kvarstår dock att undersöka i vilken utsträckning det går att köpa tillgång till själva datat då de flesta leverantörer erbjuder helhetslösningar dvs databrunn med integrerat sökgränssnitt som till exempel Serials Solution, Exlibris och Ebsco. Leverantörerna erbjuder oftast enbart tillgång till sökfunktionalitet via API:er och det behöver utredas vidare om det är en möjlig väg att gå.

I det här scenariot uppskattar vi mindre arbete med datahantering, men i gengäld behöver arbete läggas på en administrativ organisation för kontakter och krav ställande gentemot leverantörer. Går man in på det här spåret påverkar det förstås den färdriktning man önskar att KB/LIBRIS ska ha i framtiden.

Scenario *Upphandling av en färdig lösning* är en möjlig vidareutveckling av tjänsten Samsök, som kommer att gå över till att använda Primo (som alltså ska ersätta Metalib som avvecklas av systemleverantören) och i och med det få tillgång till Primo Central, Exlibris centrala index. Den nuvarande nationella licensen för programvaran, som Kungliga biblioteket står för, skulle då behöva utökas för att även innefatta integrering av bibliotekens lokala samlingar (bibliotekskataloger etc). Med nuvarande licens får biblioteken möjligheter till integrerad sökning i metadata från fria och licensierade e-artiklar, e-tidskrifter och e-böcker.

### **Scenario *Hybrid***

I scenario *Hybrid* har vi en modell där vi vidareutvecklar nuvarande samkatalog och kompletterar med ”databrunnsfunktionalitet” dvs vi har ingen nationell databrunn utan samkatalogen fungerar mot bibliotekens lokala lösningar. En annan hybrid modell som nämnts tidigare är förstås en kombination av scenarierna *Egenutveckling* och *Upphandling av färdig lösning*.

Samkatalogen LIBRIS vidareutvecklas genom att ”databrunnsfunktionalitet” byggs in via API:er. Det skulle innebära att samtidigt med sökning i LIBRIS så görs även sökning parallellt i olika index med artikeldata. Resultaten kan presenteras i en sammanslagen lista eller flera parallella listor enligt den modell man använder på Högskolan i Jönköping, där den lokala bibliotekskatalogen Julia skulle motsvaras av LIBRIS samkatalog.

Ett annat exempel är det sätt man gjort på Florida Center for Library Automation, FCLA, för universitetsbiblioteken i Florida, med tjänsten Mango. Där har alla deltagande biblioteken en egen ingång till bibliotekens samlade resurser, samtidigt som det finns en gemensam ingång i en ”Combined Catalog of the State University Libraries of Florida. Man använder Metalib:s API:er för att integrera bibliotekens e-resurser (artikeldata) i stället för att samla metadata i en databrunn.

Det här är lösningar som behöver undersökas vidare för att se om de verkligen motsvarar och uppfyller bibliotekens behov, och om de är hållbara i ett längre perspektiv.

### **Scenario *Ingen databrunn för licensierat material***

I följande scenario skapas ingen nationell databrunn för licensierade e-resurser av KB/LIBRIS. Istället skapas en nationell databrunn med inriktning på att samla metadata för svenska resurser såsom Uppsök, SwePub, e-pliktleveranser, digitaliserat material, audiovisuella medier, BURK. Här ligger satsningen på bra kvalitet över data som vi har möjlighet att ha kontroll över. Biblioteken erbjuds att exportera det datat till lokala lösningar som till exempel Primo/Ebsco/Summa/Summon. Biblioteken ansvarar således själva för metadata för e-artiklar.

En annan möjlig väg är att KB/LIBRIS undersöker möjligheterna att förhandla ramavtal för biblioteken på de leverantörlösningar som finns för central index.

- Hur ser behovet ut av att nationellt samla metadata för svenska resurser och biblioteken helt och hållet använder andra söktjänster?
- Vilka bibliotek har ett behov av ramavtal för centrala index, och är deras behov tillräckligt överensstämmande för att ett ramavtal ska kunna förhandlas?