



2010-05-04

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

**Slutrapport**

**Förstudie vid Enheten för LIBRIS-systemen och databaslicenser**

**Marja Haapalainen  
Christer Larsson  
Henrik Lindström  
Anders Söderbäck**

## **National Data Well – possibilities, challenges and demands. Final Report. Preliminary study by The Unit for LIBRIS and E-Licensing at The National Library of Sweden (KB)**

### **Summary**

The long-term aim for a data well should be to create a sustainable and integrated infrastructure where KB as well as the libraries can possess necessary control over the data, also for future today unforeseen demands. A multitude of synergetic effects can be reached if data wells are considered and developed in a wider perspective, i.e. as an integrated part of the libraries' infrastructure. Experiences from previous development projects points out that such a result best can be achieved through controlled in-house development in continuous dialogue with the libraries and the end-users.

The libraries' need for control and flexibility in order to meet also future challenges is the single most significant factor for a strategic decision regarding the creation of a national data well. This must also be weighed against the amount of resources required to fulfill other expected commitments, e.g. KB's extended commitment to provide a National Union Catalogue covering both research and public libraries and the management of e-resources resulting from the forthcoming law on e-deposit. The preliminary study indicates that a data well can be constructed to take a central part within many of these areas.

This report therefore **recommends** to focus on a scenario with a basically in-house developed data well providing control both over the collected metadata and the data exchange, possibilities to model services adapted to the actual demands, with as flexible and free access as possible. The details have to be sorted out in a later phase and the actual implementation must also be considered in relation to what personal and financial resources can be granted. A re-organization, also involving the libraries, will be required and it is estimated that some 5-10 extra persons will be needed to manage a data well that is limited to manage the licensed e-resources.

### **Sammanfattning**

Ambitionen med en databrunn bör vara att skapa en långsiktigt hållbar och integrerad infrastruktur med syfte att ge både Kungliga biblioteket och biblioteken den kontroll över biblioteksdata som behövs även för framtida, idag oförutsedda behov och krav. Genom att betrakta och utveckla databrunnar som en integrerad del av bibliotekens infrastruktur i vid mening uppstår flera positiva synergieffekter. Erfarenheterna av tidigare utvecklingsprojekt visar att detta säkrast åstadkommes med en kontrollerad egenutveckling i kontinuerlig dialog med biblioteken och användarna.

Bibliotekens behov av kontroll och flexibilitet för att möta även framtida utmaningar är den enskilt viktigaste faktorn för strategiskt vägval ifråga om en nationell databrunn. Behovet av en nationell databrunn måste också vägas mot de resurser som till exempel KB:s utökade uppdrag och hanteringen av kommande e-pliktleveranser kommer att ta i anspråk. Förstudien pekar på att en databrunn kan fylla en funktion inom flera av dessa områden.

**Rekommendationen** är därför, att arbeta vidare utifrån ett scenario med en i grunden egenutvecklad databrunn, med kontroll över både det insamlade datat och dess in- och utflöden, möjlighet att modellera tjänster efter aktuella behov och att skapa en så flexibel och fri åtkomst som möjligt. Detaljerna får kristalliseras ut i ett senare skede och även ställas i relation till de resurser (ekonomiska och personella) som kan tilldelas. En databrunn med fokus på det licensierade materialet beräknas i en driftfas kräva en organisatorisk förändring som dessutom involverar biblioteken, med en personell resursförstärkning på i storleksordningen 5-10 personer.

## Innehållsförteckning

1. Syfte med förstudien .....	4
1.1 Definition av databrunn .....	4
1.2 Frågeställningar .....	4
1.3 Metod .....	4
2. Bakgrund .....	4
2.1 LIBRIS och mega-aggregat .....	4
2.2 Samsök .....	5
2.3 Kungliga bibliotekets nationella uppdrag .....	5
3. Aktörer på marknaden .....	6
3.1 DEFF .....	6
3.2 NGC4LIB .....	6
3.3 Leverantörlösningar .....	7
3.4 Lösningar baserad på öppen källkod .....	7
4. Tekniska aspekter .....	7
4.1 Deduplicering .....	8
4.2 Bestånd och rättigheter .....	8
4.3 Export och import .....	8
4.4 Integrerad sökning .....	9
5. Metadatakvalitet .....	9
5.1 Metadataformat och miniminivå på metadata .....	9
5.2 Utökad metadatanivå .....	10
5.3 Analys av testinstallationen .....	10
6. Licensaspekter .....	10
6.1 Rättigheter att tillgängliggöra metadata .....	10
6.2 Tillgång till metadata .....	11
7. Diskussion .....	11
7.1 Nationell databrunn - en del i ett större sammanhang .....	11
7.2 LIBRIS samkatalog - en del av LIBRIS-systemen .....	12
7.3 Bibliotekens behov .....	12
7.4 Vägvalet .....	13
8. Scenarier .....	13
8.1 Scenario <i>Egenutveckling</i> .....	13
8.2 Scenario <i>Upphandling av färdig lösning</i> .....	14
8.3 Scenario <i>Ingen databrunn för licensierat material</i> .....	15
9. Rekommendation .....	16
Bilaga 1. ....	18

## 1. Syfte med förstudien

I förstudien undersöks vad det innebär att skapa en nationell databrunn, vilka tekniska aspekter som är viktiga att lyfta fram, och också vilka krav det ställer på det faktiska metadatat liksom rättigheterna till det. Studien är avgränsad till att undersöka huruvida en databrunn kan byggas upp i anslutning till LIBRIS-systemen. Vi har således inte gjort någon jämförelse mellan de produkter som finns på marknaden idag utan en mindre inventering av existerande system och databrunnsinitiativ.

Förstudien ska ge underlag till Expertgruppen för LIBRIS nationella system för vidare strategiska diskussioner om de olika möjligheter som finns att skapa en nationell databrunn.

### 1.1 Definition av databrunn

Betydelsen av begreppet ”databrunn” kan variera beroende på vad man lägger in i det. Vi har tagit till oss termen från danskans *databrønd*, (engelska *data well*). Målet i Danmark med en databrunn är att samla metadata bland annat för e-resurser från olika informationsleverantörer på ett ställe. Det behöver dock inte vara begränsat till e-resurser, tvärtom kan det inbegripa både fysiska och digitala resurser. 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<sup>1</sup> och export respektive import av data.

### 1.2 Frågeställningar

En viktig strategisk fråga är om vi själva ska utveckla en lösning för nationell databrunn från grunden eller om det finns färdiga lösningar att tillgå. Finns det möjligheter att kombinera viss egenutveckling med färdiga moduler eller deltjänster? Vad blir konsekvenserna om vi väljer det ena eller det andra? Dessutom har tillgången till metadata en central roll här, vilka är möjligheterna att få tag på det?

En annan viktig fråga är: Vad är målet med en databrunn för Kungliga biblioteket (hädanefter KB) och för biblioteken? Vilka behov uppfyller den och kan de behoven tillgodoses på annat sätt?

### 1.3 Metod

För att KB ska kunna ta ställning till frågan om en nationell databrunn undersöktes vilka förutsättningar det finns och vad det skulle innebära för KB att ta sig an denna uppgift. Förstudien omfattade:

- En marknadskartläggning av databrunnsleverantörer, såväl leverantörlösningar och lösningar baserade på öppen källkod som 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 tar också upp dessa frågor i samband med förhandlingar inför förnyelser av avtal med informationsleverantörer.

## 2. Bakgrund

Många bibliotek uttrycker 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. För att det ska fungera behöver metadata samlas på ett ställe. Detta medför att man måste få tillgång till metadatat vilket kräver förlagens medgivande. Metadatat måste sedan kunna hanteras, lagras och göras sökbar. Förlagen har inte varit så pigga på att lämna ifrån sig metadatat och biblioteken har inte haft infrastrukturen för det (Rochkind 2007). På senaste tiden har det dock börjat hända saker på marknaden, nya tjänster började erbjudas 2009 och flera frågor har kommit till KB om hur vi ställer oss till att skapa en nationell databrunn.

### 2.1 LIBRIS och mega-aggregat

LIBRIS samkatalog och söktjänst speglar bibliotekens bestånd när det gäller e-tidskrifter och e-böcker men artiklarna lyser med sin frånvaro. Det är önskvärt att även dessa blir sökbara på ett integrerat sätt. Frågan är om det ska ske via LIBRIS eller via bibliotekets lokala lösning. Många bibliotek är därför nu i färd med att undersöka de tjänster som de stora systemleverantörerna som till exempel Serials Solutions Proquest, Ebsco och Exlibris erbjuder. Leverantörerna kallar det ”data as a service” och biblioteken kan alltså prenumerera på

---

<sup>1</sup> Med integrerad sökning avses sökning mot ett centralt index där metadata från flera källor har samlats in. Jämför federerad sökning, där man söker parallellt mot de olika källorna eller databaserna via ett gemensamt gränssnitt (se även s. 5).

metadata via de så kallade ”mega aggregators” eller centrala index som systemleverantörerna underhåller och uppdaterar. Leverantörerna samlar in bibliografisk metadata för artiklar, e-tidskrifter och e-böcker från olika informationsleverantörer (dvs. 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, relevansrankning, klustring av sökresultat, sammanslagen träfflista etc. (Breeding, 2010)

Sådana mega-aggregat ger förstås stora möjligheter men innebär också att biblioteken knyts närmare de företag som byggt upp aggregaten. Särskilt om biblioteken, för att göra det traditionella biblioteksbeståndet sökbart tillsammans med mega-aggregatet, laddar upp även detta för indexering i leverantörens system. Man kan här observera två parallella trender som förefaller stå i motsättning till varandra. Den ena handlar om fri och öppen tillgång till bibliografiska data, där allt fler bibliotek menar att öppen tillgång till biblioteksdata både stärker bibliotekens position gentemot systemleverantörer och ger förutsättningar för nytänkande och innovation både i och utanför biblioteksorganisationer. Mega-aggregaten kan ses som exempel på en motsatt trend där data blir mindre tillgängligt än vad det var tidigare. Data samlas här i stora aggregat som inte är tillgängliga annat än via sökgränssnitt och leverantörskontrollerade API:er.<sup>2</sup> En konsekvens av detta skulle kunna bli att biblioteken skaffar sig en bättre position vad gäller tillgång till det traditionella biblioteksdatat samtidigt som man, vad gäller det elektroniska beståndet, knyts än hårdare till leverantörer och aggregatorer.

Vid sidan av de kommersiella systemleverantörernas tjänster finns även lösningar baserade på öppen källkod. 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. (Gillén, 2009)

Redan tidigare har det funnits databrunnsinitiativ både i Danmark och i 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.

## **2.2 Samsök**

Samsök är en nationell lösning för hantering av och åtkomst till e-resurser. Den lanserades 2006 och använder portalprogramvaran Metalib för samsökning i resurser och länkservern SFX för åtkomst till fulltext. Båda programmen är utvecklade av systemleverantören Exlibris. Det är KB 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 en 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. Det här är också en av anledningarna till att biblioteken söker efter något att ersätta det med. Hösten 2009 genomfördes en översyn av Samsök (Nelke 2009) och under våren 2010 startade en uppföljning av denna översyn. En arbetsgrupp utsedd av Expertgruppen för LIBRIS nationella system följer upp behovet av att ta ställning till hur Samsök ska utvecklas och integreras med övriga nationella tjänster för informationsförsörjningen, dvs. det längre strategiska perspektiv som enligt översynen saknas idag.

## **2.3 Kungliga bibliotekets nationella uppdrag**

KB 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.

---

<sup>2</sup> API: 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.

I detta sammanhang bör även nämnas den "nationella katalog" som lyfts fram som ett av de prioriterade områdena i KB:s uppdrag att fungera som nationell biblioteksmyndighet. Ett införlivande av folkbibliotekens poster i LIBRIS går inte att genomföra inom den tekniska lösning som idag används för LIBRIS samkatalog och söktjänst utan kräver en utökning av systemen som ligger i linje med en databrunnslösning.

### 3. Aktörer på marknaden

Vi har undersökt följande aktörer och deras produkter (produkt inom parentes): Exlibris (Primo Central), Serials Solutions Proquest (Summon), Ebsco (Discovery Service), Biblioteksdirektionen Lunds universitet (ELIN), Danmarks Biblioteks Center, DBC (Ting), Statsbiblioteket i Århus (Summa). Dessa aktörers tjänster undersöktes utifrån *hur* de fungerar som system för centralt insamlad metadata, avsikten var inte att göra någon jämförelse dem emellan. Samtliga aktörer, förutom Exlibris, har vi träffat för närmare genomgång av deras produkter. Med Exlibris har vi en löpande kontakt sedan tidigare då vi är kund till dem och därför har vi nöjt oss med e-postkommunikation som komplement till den dokumentation som finns om Primo Central. Dessutom användes som underlag även publicerade utvärderingar. (Mayfield, 2008; Stevenson, 2009)

Övriga aktörer på marknaden som vi valt att inte undersöka närmare är: OCLC (WorldCat Local), Encore, Endeca, TDOne, Danmarks Tekniska universitet/DTIC (DADS), Deep Web Technologies, Index Data. Eftersom vi inte inriktat oss på en jämförelse mellan befintliga system gjorde vi bedömningen att dessa aktörer, vid en avgränsning till databrunnar för e-resurser, särskilt artiklar, inte på något väsentligt sett skilde sig från de aktörer vi träffat och att en djupare undersökning av dessa därför inte skulle tillföra någon ytterligare kunskap.

#### 3.1 DEFF

En viktig aktör för oss i denna förstudie är Danmarks elektroniske fag- och forskningsbibliotek (DEFF) i Danmark. Under arbetet med denna förstudie har DEFF beskrivit processen kring sin upphandling av nationell databrunn, samt bidragit med information och underlag. 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. Hösten 2009 startade en nationell upphandling av databrunnsleverantör. Upphandlingen involverade fem leverantörer, varav tre kommersiella (Exlibris, Serials Solutions Proquest och TDOne) och två icke-kommersiella (Danmarks tekniska universitet och Statsbiblioteket i Århus i samarbete med Dansk bibliotek center). 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. (DEFF, 2008; Öhrström, 2008)<sup>3</sup>

#### 3.2 NGC4LIB

Vi skickade ut en förfrågan på diskussionslistan Next Generation Catalogs For Libraries (NGC4LIB) för att se om det finns initiativ till aggregering av licensierat metadata till nationella databrunnar ute i världen. Ett visade sig finnas i Australien, där nationalbiblioteket till sitt "discovery tool", Trove, planerar att även inkludera metadata för licensierat material. Andra liknande men inte nationella initiativ är ELIB Bremen, Staats- und Universitätsbibliothek Bremen och Ontario Scholars Portal. Den sistnämnda portalen drivs av Ontario Council of University Libraries som är ett konsortium för 21 universitetsbibliotek i Ontario. Den tjänsten är dock lite annorlunda då den inte erbjuder integrerad sökning i alla typer av material, utan sökingångarna är uppdelade på e-tidskrifter, e-böcker och artiklar. I övrigt är det ämnesportaler som är vanligt förekommande när det gäller

---

<sup>3</sup> Under rapportens skrivande har det kommit till vår vetskap att man har avbrutit upphandlingen under mars månad. Det som upphandlingen stupar på är det faktum att systemleverantörerna inte erbjuder exporter av metadata för lokal lagring, vilket är ett av villkoren i den nationella lösning som DEFF har specificerat. Det har uppstått en situation där det råder oklarheter om vilken intern IT-strategi man har hos de deltagande biblioteken och vilken slags lösning man egentligen föredrar. För att kunna föra en diskussion kring detta krävdes att upphandlingen avblåstes, i och med att två institutioner (DTU och Statsbiblioteket) både fungerar som avropare av upphandlingen och har lämnat ett anbud som svar på upphandlingen. Eventuellt blir det en ny upphandling senare där export ska vara en option istället för ett krav. Man antyder från DEFF att det inte är helt uteslutet att man kommer att välja en kommersiell leverantör som Exlibris, Serials Solutions Proquest eller Ebsco. Man ser det nämligen som en möjlighet att kombinera en leverantörlösning för just artiklar tillsammans med ett eget integrerat system för annat metadata. Dvs. man utnyttjar de API:er som leverantörerna erbjuder och bygger in sökfunktionalitet i det egna integrerade systemet.

aggregering av metadata, t.ex. Economist Online (<http://www.economistsonline.org>) och Io-port.net (<http://io-port.net>) liksom olika portaler för open access-material. (NGC4LIB, 2010; Blenke, 2009)

### 3.3 Leverantörlösningar

De leverantörer vi har pratat med erbjuder enbart helhetslösningar, det vill säga motsvarande det vi avser med databrunn integrerat med ett sökgränssnitt. För att möjliggöra sökning utan att använda deras gränssnitt erbjuder leverantörerna tillgång till sökfunktionalitet via API:er.

Det är i dagsläget inte möjligt att köpa endast metadata från en systemleverantör vilket var något av en utgångspunkt i förstudien när det gäller att få åtkomst till själva metadatat. Systemleverantörerna är villiga att samla in metadata (mot avgifter i några fall) men inte att exportera metadata för lokal lagring. Detta har flera orsaker. I många fall är systemleverantörerna förhindrade att exportera data pga. restriktioner i deras avtal med förlagen. Systemleverantörerna förefaller arbeta utifrån idén att sälja tillgång till sökning i aggregerat metadata och inte att sälja metadatat i sig. Systemleverantörernas linje är att biblioteken som deras kunder ska underlätta insamlandet av metadata för lagring centralt i leverantörernas mega-aggregat, men frågan är vad som händer med datat den dagen man vill byta leverantör? Ur ett nationellt perspektiv ser vi svårigheter med att förhålla sig till sådana villkor. Det står också emot de grundläggande principer som KB verkar för - öppet data.<sup>4</sup>

Ett annat grundläggande krav för en nationell databrunn är att metadatat som samlas där är publikt tillgängligt dvs. att metadatat måste kunna sökas av gemene man. Det är något som inte alla systemleverantörerna tillåter. Även detta faktum går emot KB:s principer. Se vidare avsnitt 6.1 *Rättigheter att tillgängliggöra metadata*.

Utifrån det insamlade materialet kan man konstatera att det inte är möjligt att skaffa en "färdig" lösning till en databrunn eftersom en leverantörlösning som det ser ut idag innebär att man måste investera i utvecklingsinsatser för att åstadkomma sökfunktionalitet mot den.

### 3.4 Lösningar baserad på öppen källkod

Två lösningar baserade på öppen källkod har undersökts, Ting och Summa. Ting är utvecklad för, och används för närvarande av, ett antal folkbibliotek i Danmark. DBC:s målsättning är dock att det ska kunna användas av alla typer av bibliotek. Summa används av Statsbiblioteket i Århus och ett antal folkbibliotek i Danmark. Statsbiblioteket har ingen översikt över användare utanför Danmark, men säger att både bibliotek i Sverige och i Norge har visat intresse.

Ting-mjukvaran publiceras under "GNU General Public License" (<http://gplv3.fsf.org/>). Ting är ett helhetskoncept i den meningen att DBC arbetar med att bygga upp en s.k. "open source community" kring Ting för att få biblioteken att samverka och medverka i utvecklingen. Dessutom har man en organisation som arbetar med innehållet dvs. med att fylla själva databrunnen med metadata från olika källor. DBC erbjuder också support och underhåll.

Summa är en systemlösning som bygger mycket på egenutveckling och som i första hand är utvecklat för Statsbiblioteket och dess behov. Det följer inte med något innehåll i form av metadata, utan det är en process man är hänvisad att bygga upp själv.

## 4. Tekniska aspekter

Vi har undersökt de tekniska frågorna kring uppbyggandet av en fullfjädrad nationell databrunn, med insamling, normalisering och uppdatering av data, hantering av bestånd och rättigheter. Våra slutsatser kring tekniska aspekter utgår till stor del från egna erfarenheter från LIBRIS och SwePub. Vi gjorde också en testinstallation

---

<sup>4</sup> En grundläggande princip för LIBRIS-samarbetet och för den överenskommelse som skrivits mellan KB och de deltagande biblioteken är att biblioteken utan förbehåll äger det data som hämtas från LIBRIS och integreras i egna system, vilket också innebär att LIBRIS-bibliotek är fria att lämna detta data vidare. LIBRIS arbetar också aktivt med att göra biblioteksdata fritt tillgängligt för att underlätta utbyte och samarbete mellan bibliotek, men också för att hitta användningsområden för biblioteksdata utanför den traditionella bibliotekssfären. Berlindeklarationen om Open Access (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>) behandlar vid sidan av vetenskapligt material även fri tillgång till rådata och metadata. Eftersom metadata är centralt för att ge åtkomst till OA-publikationer blir den fria åtkomsten till metadata lika viktig som den fria åtkomsten till publikationen i sig. Bibliotekssystem bygger på lagring och hantering av metadata och för utvecklingen av system är det, oavsett om systemen upphandlas eller egenutvecklas, viktigt att veta hur metadatat som systemet hanterar får användas.

där vi samlade in metadata från tre helt olika källor (Nature, HathiTrust och DOAJ) som samtliga gör sitt metadata fritt tillgängligt via OAI-PMH. Vi har också fått mycket information från DBC och Statsbiblioteket i Århus, Danmark.

En central del av systemet blir datahantering i form av normalisering/berikning/konvertering av metadata och deduplicering. I ett system där man samlar in i storleksordningen flera hundra miljoner poster kommer det att bli helt avgörande att detta till största del hanteras maskinellt.

Det kommer dock att krävas en hel del handpåläggning, inte minst vid introduktion av en ny datakälla. För att detta skall gå så smidigt som möjligt krävs att systemet även tillhandahåller bra administrationsgränssnitt. Här blir det en avvägning mellan hur mycket tid man lägger ned på utveckling av just administrationsgränssnitt jämfört med hur mycket "manuellt" arbete det blir med varje ny källa. Det blir också en avvägning om vilka metadataformat och leveranssätt man kan acceptera i förhållande till ambitionsnivå för täckningsgraden. Någon form av miniminivå både vad gäller metadatanivå och format måste dock definieras.

Enligt vad vi erfarit från bl.a. DBC, Biblioteksdirektionen i Lund och svar på diskussionslistan NGC4LIB kan man förvänta sig att data kommer i många olika former av varierande kvalitet. Även för källor som levererar enligt en viss standard, exempelvis OAI-PMH, finns en rad egenheter i implementation och metadataformat, vilket vi tydligt ser i testinstallationen. Frågan kvarstår alltså hur mycket som kan skötas maskinellt och hur mycket manuellt arbete varje källa kräver. (För vidare diskussion av detta, se avsnitt 5 *Metadatakvalitet*)

#### **4.1 Deduplicering**

Erfarenheter från arbetet med SwePub visar att ojämnheter i metadatat ställer krav på en deduplicering som inte bygger på exakt matchning. Detta problem och lösningar har diskuterats i en rad artiklar och rapporter (se exempelvis Hylton 1996 eller Lohrum et al. 1999). SwePub är en relativt kontrollerad och liten mängd data i ett gemensamt format, men redan här uppstår problem med ojämnheter i angivelser av titlar, författare och externa id:n såsom ISBN, DOI, PubMed-id, etc.

För att deduplicering skall fungera krävs att poster konverteras till någon form av gemensamt metadataformat. Detta underlättar även en smidig indexering, även om det är fullt möjligt att tänka sig att olika typer av material (dvs. olika metadataformat) indexeras på olika sätt. Vid sidan av konvertering till detta format är det även önskvärt att varje post sparas i sitt originalformat. Att spara alla originalposter underlättar även hanteringen av uppdateringar och borttagning av enskilda poster även om dessa ingår i ett dubblettkluster. Fördelarna med denna modell poängterades även av DBC.

En deduplicering måste vara generell i systemet och kräver en hel del insatser för att göras bra. Mycket av erfarenheterna från SwePub bör kunna återanvändas, men det skulle behövas större fokus på att komma bort från manuell finjustering av algoritmen till att med någon form av maskininlärning träna algoritmen.

#### **4.2 Bestånd och rättigheter**

Ytterligare en central del av systemet är hantering av bestånd och rättigheter. Det bör understrykas att datat måste kopplas till bestånd för att vara användbart. Denna berikning sker förslagsvis även den i själva lagringsdelen av databrunnen för att kunna ta ut de delar av data som hör till ett visst bibliotek. Eftersom beståndsuppgifter typiskt ändrar sig mer eller mindre regelbundet måste dessa kunna uppdateras smidigt på befintliga poster. Vad gäller e-resurser bör bestånd kunna hämtas från befintliga länkservrar och ERM-system, till exempel SFX. Länkservern bör också kunna användas för att ge tillgång till fulltexter.

I en modell där olika poster har olika rättigheter för vem som får söka/visa en post blir modellen med lagring av enskilda poster viktig för att kunna göra "rätt" post sökbar/visningsbar för "rätt" användare/bibliotek. Dock vore det rimligt att ställa som krav att åtminstone den grundläggande metadatan är åtkomlig för alla, men att visning av exempelvis abstracts kan vara belagd med vissa restriktioner.

#### **4.3 Export och import**

En nationell databrunn behöver kunna hantera export av data, inte minst till biblioteken. Denna typ av potentiellt omfattande export ställer krav på systemet, och analogt som vid sökning bör detta kopplas till bestånd/rättigheter. Även dumpar av och åtkomst till databrunnen i dess helhet (förutsatt att datat är fritt) vore önskvärt för att möjliggöra för andra aktörer att återanvända datat. (Se även avsnitt 6.1 *Rättigheter att tillgängliggöra metadata*).

Ur bibliotekens perspektiv kan behov finnas av att lägga till och administrera lokala avtal i databrunnen. Idealiskt vore om biblioteken själva hade möjlighet att administrera sina egna lokala resurser, vilket ställer krav på administrationsgränssnittet. (Se även avsnitt 8.1 *Scenario Egenutveckling*)

#### 4.4 Integrerad sökning

En av utgångspunkterna för denna undersökning var möjligheten att utifrån en databrunn kunna erbjuda s.k. integrerad sökning, där "allt" material kan sökas i ett och samma system med ett centralt index. Detta ställer stora krav på sökfunktionaliteten för att kunna erbjuda relevanta träffar som svarar mot användarens informationsbehov. Ett viktigt utvecklingsområde är att få metadata på "låg" och "hög" nivå att fungera ihop i sökningen, och även potentiellt i kombination med indexerade fulltexter.

En viktig del i detta arbete är att kunna utvärdera de söksystem vi bygger. För detta krävs någon form av testmängd, vilket är både svårt och tidskrävande att ta fram. DEFF-projektet "Værktøjer til integreret søgning" (Lykke, M. 2010), genomfört vid Danmarks biblioteksskola, har tagit fram en sådan testmängd, med bland annat exempel på sökfrågor med tillhörande träffar som manuellt bedömts som mest relevanta. Detta är visserligen gjort på ett avgränsat ämnesområde, men testmängden kommer säkert att vara mycket värdefull i utvecklingen av söksystem för en nationell databrunn.

## 5. Metadatakvalitet

En kritisk faktor i uppbyggnaden av en databrunn är kvaliteten på det metadata som kan samlas in till en databrunn. Det bibliografiska metadata behövs vara tillförlitligt och hålla en tillräckligt god och jämn kvalitet för att kunna bli generellt användbart vid såväl bearbetning (normalisering, deduplicering/klustring, skapande av relationer med mera) som stöd för integrerad sökning. Samma sak gäller administrativt metadata för åtkomst- och licensvillkor.

### 5.1 Metadataformat och miniminivå på metadata

Standardiserade och internationellt accepterade format (t.ex. MARC21/MARCXML, Dublin Core, MODS) är att föredra framför leverantörsutvecklade, eftersom det underlättar kontrollerad mappning till ett gemensamt metadataformat och därmed reducerar behovet av manuella finjusteringar. Vissa manuella insatser kommer sannolikt ändå att behövas när nya metadatakällor och format introduceras, inte minst därför att det förekommer variationer och ojämnheter i tolkningen av olika metadataelement även i standardiserade format (jfr även vad som sägs i avsnitt 4 *Tekniska aspekter*).

Det innebär också att det behöver finnas åtminstone ett minimum av definierade metadataelement. Metadata för e-böcker utgör det kanske minsta problemet, åtminstone från de stora leverantörer som erbjuder licensierade e-bokspaket (eBrary, Elsevier, Oxford Scholarship Online, Springer etc.). Här finns ofta MARC-poster av acceptabel till god kvalitet. Uppgiften blir då att i respektive avtal förhandla fram klausuler om leveranser av metadata, vilket diskuteras i avsnitt 6 *Licensaspekter*. Samma förhållande torde gälla licensierade e-tidskriftspaket, möjligen med undantag för att det kan förekomma variationer i behandlingen av tidskrifter som har bytt titel. Något annorlunda kan det förhålla sig med metadata för tidskriftsartiklar och andra typer av material.

Det insamlade metadata behövs också vara tillräckligt identifierbart för att möjliggöra ett integrerat och kontinuerligt uppdateringsflöde av bestånds- och rättighetsdata från annan källa, till exempel SFX (jfr avsnitt 4.2 *Bestånd och rättigheter*). Erfarenheterna av de överföringar av bibliotekens licensierade resurser från SFX till LIBRIS talar för att det går att skapa ett flöde som speglar aktuella förhållanden, förutsatt att källan till bestånds- och rättighetsdata kan kontrolleras. I ett integrerat scenario, där detta flöde kopplas ihop med det metadata som samlas in i en databrunn, skulle också den bibliografiska kvaliteten kunna förbättras avsevärt, även om man sannolikt fortfarande måste räkna med en kritisk undre gräns.

Exakt var denna kritiska gräns för acceptabel metadata går behöver analyseras och utvärderas i anslutning till uppbyggnaden av en databrunn. Rent generellt kan man dock förutse att det kommer att behövas tillräckligt säkra bibliografiska element för identifikation och bearbetning av respektive publikation, med typ av publikation, titel, författare och utgivningsår - eventuellt även språk och i tillämpliga fall ISBN/ISSN - som ett minimum. Användbarheten ökar också avsevärt om artikelmetadata dessutom inkluderar element om den tidskrift artikeln har publicerats i (värdpublikationen). Dessutom behövs ett antal metadataelement av tekniskt-administrativ karaktär, framför allt stabila identifierare eller länkar för åtkomst till fulltexten (ett alternativ kan

rent teoretiskt vara att själva fulltexten hämtas in och lagras tillsammans med metadatat) och eventuella åtkomst-/licensvillkor.

## 5.2 Utökad metadatanivå

Som ett riktmärke vid bedömningen av metadatakvalitet kan det dessutom ha ett värde att - utöver miniminivå - definiera en utökad metadatanivå. Den utökade nivån skulle då till exempel vara element av analytisk karaktär, såsom keywords/ämnesord, abstracts och innehållsförteckningar, men även olika typer av element som beskriver relationer till andra publikationer.

## 5.3 Analys av testinstallationen

En analys av testinstallationen (med metadata formaterat i Dublin Core, från Nature, HathiTrust och DOAJ) ger vid handen att flera av de miniminivåelement som diskuteras ovan finns med, men också att det förekommer ojämnheter i hur de har definierats. Det kanske mest slående är att det i detta begränsade källurval är svårt att se någon enhetlig hantering av information om värdpublikation i artikelmetadata (källor: DOAJ respektive Nature). Metadata om åtkomst- och licensvillkor är dessutom i allmänhet formulerade som en ren fritextanmärkning (t.ex. i källan HathiTrust). Det är naturligtvis för tidigt att dra några säkra slutsatser utifrån detta testurval, men frånvaron av en entydigt identifierbar värdpublikation kan rent hypotetiskt tänkas få negativa konsekvenser, åtminstone i vissa scenarier för hur man kan tänkas vilja använda artikelmetadata, till exempel i fråga om möjligheterna att koppla ihop med ett kontinuerligt uppdateringsflöde av bestånds- och rättighetsmetadata från annan källa, i fråga om möjligheterna att skapa säkra OpenURL:er och i fråga om att presentera artikelmetadata i en traditionell bibliotekskatalog. Å andra sidan förefaller såväl identifierare som fulltextlänkar vara stabila och tillförlitliga.

## 6. Licensaspekter

En viktig grundsten i en centralt koordinerad databrunn (oavsett om den är egenutvecklad eller upphandlad) är att göra de centralt upphandlade BIBSAM-avtalen tillgängliga. I dagsläget rör det sig om ett 30-tal avtal (se Bilaga 1). För att en databrunn ska vara relevant för ett bibliotek är det inte tillräckligt att de centrala avtalen ingår i databrunnen utan det behövs en möjlighet för enskilda bibliotek att få med även de lokalt förhandlade avtalen i den nationella databrunnen. Ett sätt att uppnå detta är att från centralt håll ta fram standardklausuler. Dessa används i de centrala förhandlingarna, men görs också tillgängliga för de bibliotek som vill använda dem i lokal förhandling.

### 6.1 Rättigheter att tillgängliggöra metadata

Licensaspekterna på databrunnen kan delas upp i två delar - hur data tillgängliggörs av leverantören samt vilka rättigheter som finns för att tillgängliggöra datat och att exportera det vidare. I båda fallen krävs antagligen någon form av standardisering för att hanteringen i databrunnen inte ska bli alltför komplex. DEFF utgår i sina förhandlingar från att metadata ska levereras. Utöver detta önskar man leverans av TOC, abstracts, fulltexter samt, i de fall detta förekommer, citeringsdata. För ett svenskt projekt bör finnas behov av att specificera en lägstanivå för metadata. Det finns också anledning att ställa krav kring på vilket sätt och hur ofta data kan levereras. Vissa grundkrav bör kunna specificeras kring leveransprotokoll, formatspecifikation osv., samtidigt som vi i de flesta fall kommer att tvingas anpassa oss till vad leverantörerna faktiskt erbjuder. Detta kommer åtminstone inledningsvis att kräva att personal med grundläggande system- och formatkompetens deltar i delar av förhandlingarna vid sidan av ordinarie BIBSAM-handläggare.

Mest intressant ur ett licensperspektiv är frågan om hur data kan tillgängliggöras. I DEFF:s förhandlingar anmodas leverantörerna ange vilka institutioner som har rättighet att lagra, få tillgång till och indexera data, samt vilka som kan få tillgång till det indexerade datat - i praktiken vilka som tillåts söka i en söktjänst men detta skulle också kunna innebära tillgång till API:er mot det indexerade datat. DEFF går i sina avtal inte längre än att kunna ge sökmöjligheter till auktoriserade användare vid DEFF-bibliotek. Ett grundkrav för en nationell lösning i LIBRIS bör dock vara att sökning kan tillåtas av gemene man utan inloggning eller annan autentisering, då KB som institution verkar för öppna system. Det är inte något som är självklart för de kommersiella systemleverantörerna och inte heller något de tillåter när det gäller artiklar som denna förstudie specifikt tittar på. För att en nationell lösning ska ge något mervärde jämfört med en upphandlad lösning bör även export av data (och på sikt även fulltexter) för lokal indexering hos de bibliotek som deltar tillåtas. En svensk databrunn hos LIBRIS men utan möjlighet att exportera data skulle visserligen ge LIBRIS mer flexibilitet kring utformningen av söktjänster och bättre kontroll över datat, men för biblioteken skiljer detta sig inte nämnvärt från en upphandlad lösning utan möjlighet att exportera data.

Då ett av målen med en nationell databrunn bör vara att kunna samarbeta med andra aktörer (t.ex. motsvarande initiativ i andra länder) är det också önskvärt att kunna dela och exportera åtminstone delar av datat vidare till dessa. DEFF berör i sina förhandlingar inte möjligheten att tillgängliggöra data utanför DEFF-biblioteken, men för ett svenskt databrunnsinitiativ bör målet vara så hög grad av öppenhet som möjligt. Vad gäller licenserade fulltexter är detta förstås inte möjligt, men möjligheten att exponera åtminstone delar av metadatat, exempelvis relationer mellan olika identifikatorer som kommit fram vid dedupliceringsarbetet, bör undersökas. För att uppnå en så bra global infrastruktur som möjligt bör ett av de grundläggande målen med databrunnen vara att verka för fri och öppen tillgång till metadata.

## 6.2 Tillgång till metadata

Utsikterna att få metadata från leverantörerna av de centrala avtalen kan, vad gäller det stora flertalet leverantörer, betraktas som goda. Varje avtal kommer däremot att kräva specifika arbetsinsatser för att hitta rätt kontaktpersoner och för att föra in relevanta skrivningar i avtalen. Vid enskilda möten i mars och under UKSG-konferensen den 12-14 april har diskussioner inletts med 14 organisationer: Association for Computing Machinery, Cambridge University Press, De Gruyter, Elsevier, Emerald, Institute of Physics, Oxford University Press, Portico, SAGE, Springer, Taylor & Francis, Thomson Reuters och Wiley. Samtliga dessa hade en positiv inställning till leverans både av metadata och av fulltexter: Man menade sig ha de tekniska system som behövs för leverans på plats. Två av förlagen (SAGE och Cambridge University Press) menade sig behöva ytterliggare underlag och diskussion kring fulltexter. Vad gäller Wiley behövs ytterliggare diskussion kring både metadata och fulltext, men det kan noteras att de skrivit avtal om leverans av metadata till DEFF. Utöver dessa 14 förlag finns metadata från Nature och Project MUSE tillgängligt redan idag. Det kan också noteras att DEFF skrivit avtal om tillgång till både metadata och fulltexter från Kluwer. Vad gäller övriga BIBSAM-avtal har DEFF ännu inte nått längre än vad vi har gjort.

Idag finns i BIBSAM-avtalet med Thomson Reuters inkluderat rätten att använda bibliografiska metadata från Web of Science i bibliotekens lokala publiceringssystem. Det bör, enligt information från Thomson Reuters, inte vara problematiskt att utöka denna rättighet till att inkludera lagring av bibliografiska data direkt i ett centralt system, dvs. en databrunn.

Vid ett möte med Springer den 16 mars hade Springer en mycket generös inställning till leverans av metadata och fulltexter för indexering. Detta bekräftades vid ett uppföljande möte under UKSG, där Springer även uttryckte stort intresse för ett närmare samarbete och erfarenhetsutbyte kring datahantering. En positiv respons från Springer kan användas som argument i förhandling med de andra leverantörerna. Vår bedömning är att det med största sannolikhet går att få tillgång till metadata rörande åtminstone de flesta BIBSAM-avtalen, och att möjligheten att få tillgång till metadata även vid lokala förhandlingar bör vara god.

## 7. Diskussion

Vi befinner oss i en situation som kan beskrivas som ett ”paradigmskifte” på marknaden för bibliotekssystem. Biblioteken behöver fatta beslut om hur man vill gå vidare med sina systemlösningar samtidigt som inga självklara alternativ finns att tillgå. I många fall avvaktar man ett beslut om färdriktning från KB. I andra fall implementerar man redan idag tillfälliga lösningar i väntan på en nationell sådan; Primo, Summa, Encore.

Vi vill framhäva att KB och biblioteken behöver kommunicera med varandra för att gemensamt hitta en framtida inriktning för hantering av stora mängder data, om det ska ske på ett nationellt eller lokalt plan. Tillsammans behöver vi komma fram till ett beslut om vägval, eftersom det lär bli svårare ju längre det dröjer, då övergången från dessa tillfälliga lösningar kan riskera att bli permanenta.

Förstudien är begränsad så till vida att den är inriktad på *hur* en nationell databrunn faktiskt kan genomföras, men en större fråga har väckts: Hur ska den framtida infrastrukturen för informationsförsörjningen fungera?

### 7.1 Nationell databrunn - en del i ett större sammanhang

Frågan om insamling av metadata i en nationell databrunn påverkar i högsta grad frågan om infrastruktur för informationsförsörjning i vid mening och får konsekvenser för hur denna ska utvecklas för att passa kommande behov. Det har blivit tydligt under denna förstudie att de olika komponenterna hänger ihop: KB:s nationella ansvar för informationsförsörjningen och e-resursstrategier, LIBRIS samkatalog och söktjänst som navet i detta tillsammans med andra nationella tjänster som SwePub och Sondera, den underliggande tekniska infrastrukturen, Samsök som ett initiativ att underlätta e-resurshanteringen för biblioteken, KB:s nya utökade uppdrag som bland

annat kommer att innebära en nationell katalog, KB:s insamling av e-pliktleveranser och ökande digitalisering. En nationell databrunn måste alltså ses i det här komplexa sammanhanget. Förstudien är begränsad till att undersöka möjligheterna att samla in metadata för e-resurser specifikt, men det står klart att det vore en kortsiktig lösning att utöka det redan idag insamlade metadatat med endast metadata för e-resurser. En strategi behöver formuleras som ger en långsiktig nytta och inte bara ser till dagens behov.

## **7.2 LIBRIS samkatalog - en del av LIBRIS-systemen**

Tidigt framgick att vi står inför ett viktigt strategiskt vägval som rör LIBRIS samkatalogs fortsatta utveckling och innehåll. Vi måste ställa oss frågan: Hur länge kan LIBRIS vara relevant för biblioteken? I den överenskommelse som finns mellan KB och LIBRIS-biblioteken anges att "innehållet i LIBRIS ska vara aktuellt och spegla LIBRIS-bibliotekens samlingar." Idag finns dock en stor brist i och med att LIBRIS inte speglar bibliotekens bestånd av elektroniskt material. Enligt SCB:s forskningsbiblioteksstatistik användes under 2008 (senast publicerade statistiken) 72% av det totala förvärvsanslaget till inköp av databaser, elektroniska tidskrifter och e-böcker. 2008 pekas ut som en viktig brytpunkt i användningen av elektroniska resurser, inte minst då 85% av bibliotekens periodikasamlingar enligt statistiken är digitala.

För denna diskussion kan det vara nyttigt att skilja mellan olika roller hos LIBRIS-systemen. LIBRIS fungerar, betraktad ur olika perspektiv, som samkatalog, som sök- och fjärrlånetjänst och som infrastruktur för import, export och maskinell bearbetning av metadata. Samarbetsfilosofin utgör ytterligare ett perspektiv. Traditionellt har dessa områden varit överlappande. Det är innehållet i samkatalogen, uppbyggt via samarbete, som gjorts sökbar i söktjänsten och som har bearbetats och exporterats till LIBRIS-bibliotekens lokala system. Samstämmigheten mellan dessa olika perspektiv har redan idag, i och med SwePub, börjat luckras upp och kanske behöver man i diskussionen om en databrunn skilja på dessa aspekter. Samkatalogen blir en del av en vidare infrastruktur som kan användas för att skapa olika typer av tjänster. Samarbetsaspekten bör kunna bevaras på samtliga dessa nivåer.

Utifrån detta blir den naturliga följdfrågan: Kan LIBRIS fortsätta att vara en relevant infrastruktur för forskningsbiblioteken? Ett svar på denna fråga kan vara genom att bygga upp en nationell databrunn. Då är frågan om vi ska utveckla en lösning från grunden eller om det finns färdiga lösningar som går att använda. Eftersom vi har en tradition av egenutveckling var det naturligt att undersöka frågan om nationell databrunn utifrån detta perspektiv och se vad det faktiskt innebär att göra det i praktiken, samt vilka konsekvenser detta får i jämförelse med upphandling. Vi har tittat på frågan utifrån tekniska aspekter liksom metadata- och licensaspekter och ser inga oöverstigliga hinder, men kan konstatera att det naturligtvis skulle bli resurskrävande. Däremot innebär systemleverantörernas produkter inte heller "färdiga" lösningar. Även en upphandling innebär att mycket utvecklingsarbete ändå tillkommer med anpassningar mot dessa produkter för att få dem att fungera med LIBRIS. Resultatet blir en slags hybrid där en kommersiell produkt i mångt och mycket kommer att skapa begränsningar i utvecklingen av LIBRIS.

## **7.3 Bibliotekens behov**

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. Under rapportens skrivande har det kommit till vår vetskap att man inom Sveriges universitets- och högskoleförbund, SUHF, har börjat diskutera en eventuell avsiktsförklaring när det gäller utvecklingen av en nationell databrunn, och också finansieringsfrågan. Man menar att biblioteken har samma behov, att den minsta gemensamma nämnaren för biblioteken är just metadata, alla behöver i princip samma metadata när det gäller e-resurser. Det finns vinster med att samla in metadata på en nationell nivå istället för att biblioteken gör det på lokal nivå.

I resultatet från projektet LIBRIS som lokal OPAC (L-OPAC) får vi antagligen många svar om vilka behov biblioteken har idag. Projektet bedrevs av fyra universitetsbibliotek (Göteborg, Lund, Uppsala och Linnéuniversitetet) i samarbete med KB och där undersöktes möjligheterna att använda LIBRIS som gränssnitt för den lokala OPAC:en. I förhållande till denna förstudie finns här flera frågor: Finns någon poäng med LIBRIS som lokal OPAC utan nationell databrunn? Vad är alternativet till databrunn? Kan KB erbjuda samma lösningar som andra leverantörer? Innebär en nationell databrunnslösning att KB konkurrerar med kommersiella leverantörer och är detta i så fall KB:s roll, eller ska man istället satsa på att komplettera de tjänster de kommersiella leverantörerna erbjuder? Ur ett mer kortsiktigt perspektiv kan en nationell databrunn betraktas som en konkurrent till kommersiella system. Då handlar frågan om en avvägning mellan kostnader och förväntad vinst.

Kortsiktigt är det sannolikt att en kommersiell leverantör kan erbjuda attraktiva lösningar på specifika problem.

Ur ett långsiktigt perspektiv däremot är det dock rimligt att KB:s uppdrag bör vara att upprätthålla en infrastruktur som är gynnsam för de svenska offentligfinansierade biblioteken och för allmänheten. En sådan infrastruktur bör möjliggöra konkurrens mellan olika leverantörer men också uppmuntra flexibilitet och nytänkande både hos bibliotek och hos andra aktörer. Upphandling av system är alltså inte ett självändamål, och bibliotekens infrastruktur kan inte definieras utifrån vad kommersiella leverantörer erbjuder utan måste utgå från vad biblioteken behöver.

#### **7.4 Vägvalet**

Vi bedömer det alltså som sannolikt att LIBRIS samkatalog kommer att påverkas, oavsett vilken väg man väljer att gå. Genom att integreras i en större mängd metadata för elektroniskt material får den traditionella samkatalogen en fortsatt betydelsefull - om än mindre central - position. Å andra sidan är det enligt vår bedömning sannolikt att samkatalogen, om den inte integreras med det elektroniska beståndet, isoleras och blir av marginell nytta för forskningsbiblioteken.

Vägvalet kommer naturligtvis också att påverka den färdriktning överlag man önskar att KB ska ha i framtiden dvs. om man ska fortsätta med egenutveckling eller i högre grad köpa färdiga tjänster. Detta vägval kommer också att påverka hur man väljer att arbeta vidare. Ska man gå i riktning mot att bli en upphandlingsorganisation eller ska man fortsätta vara en utvecklingsorganisation?

## **8. Scenarier**

Utifrån våra undersökningar ser vi två huvudsakliga scenarier: *Egenutveckling* och *Upphandling av färdig lösning*. Ett tredje scenario är *Ingen databrunn för licensierat material*.

### **8.1 Scenario Egenutveckling**

#### **Beskrivning**

LIBRIS samkatalog utvecklas till en integrerad tjänst där alla typer av metadata samlas centralt och kan återanvändas inte bara i en nationell söktjänst men även i lokala kontexter. Det blir möjligt att göra tvärsnitt i materialet och erbjuda portaler inom olika ämnen och inriktningar.

#### **Vad krävs?**

De metadata som idag finns samlade i SwePub, Uppsök och LIBRIS samkatalog utgör i praktiken en databrunn enligt vår definition. Men för att det här scenariot ska bli realistiskt och hållbart måste en helt ny teknisk infrastruktur byggas upp som ligger till grund för samtliga LIBRIS nationella system. Ett system som hanterar enorma flöden av metadata, både maskinella och manuella. Det behöver också utvecklas verktyg och administrativa gränssnitt för att hantera metadataflödet. Det dagliga arbetet med insamling, mottagande, uppdatering och indexering kommer att kräva en hög grad av arbetsinsatser som inte ryms inom dagens organisation. Det är arbetet med metadataflödet som kommer att utgöra den stora utmaningen i det här scenariot, utöver arbetet med licensförhandlingarna. I metadataflödet utgör arbetet med analys och normalisering/konvertering en stor del, då materialet kommer från många olika informationsleverantörer i lika många format och leveransmetoder. Vi kan förväntas oss mycket manuellt arbete med att "lägga upp" nya källor i inflödet, trots att mycket borde kunna skötas maskinellt.

#### **Resursbehov**

När det gäller arbetet med metadataflödet så kan detta lösas genom en utökning av personalstyrkan hos LIBRIS, men eftersom även bibliotekens lokala avtal med informationsleverantörerna ska inkluderas ser vi det som en nödvändighet att involvera biblioteken i insamlandet av metadata i form av ett upplägg som liknar dagens samkatalogisering i LIBRIS dvs. ett mer decentraliserat samarbete med biblioteken i form av t.ex. ett "community", vilket kräver administration och organisation. Det är svårt att uppskatta resursbehovet i löpande drift på KB, men som en jämförelse: På DBC arbetar idag ett 10-tal med utveckling och ett femtal med metadataflödet. Serials Solutions Proquest uppger att man har ett team på ca 15 personer som arbetar med utveckling och ca 15 personer med metadataflödet.

Omfattningen av licensarbetet skulle öka jämfört med det arbete som bedrivs idag på KB, när det gäller BIBSAM-avtalen. Dagens arbete behöver kompletteras med förhandlingar om tillägg till avtal som tar upp leverans av metadata samt rättigheterna till det. DEFF uppskattade för sina ca 200 avtal en halvtidstjänst som ägnar sig åt detta. På KB har vi 35 avtal i dagsläget och man befinner sig också i en omorganiseringsprocess då hela licensverksamheten centraliseras från att ha varit distribuerad på ett femtal universitetsbibliotek i landet. Antalet licenser kan komma att ändras.

## Nyttan

Det finns stora vinster med att centralt samla metadata inte bara motsvarande BIBSAM-avtalen, utan även för bibliotekens lokala avtal. Att exkludera dem skulle innebära en splittrad lösning där även biblioteken, förutom KB, måste utveckla sina infrastrukturer för att kunna hantera metadataflöden. Det blir inte heller en samlad lösning för bibliotekens e-resurser, vilket är det som är önskvärt.

## Sammanfattning av konsekvenser

- Egen projektorganisation som även involverar biblioteken.
- Utökning av personalresurser om 5-10 personer för den löpande driften.
- Investering i nya maskinvaror.
- Initialt arbete med att bygga upp en ny teknisk infrastruktur för att hantera databrunnar.
- Initialt arbete med att bygga upp interna verktyg och administrationsgränssnitt för administration av metadataflödet.
- Initialt arbete med att bygga upp externa verktyg och administrationsgränssnitt för biblioteken för administration av metadataflödet.
- Utveckling av nytt nationellt gränssnitt samt utveckling av API:er för att möjliggöra lokala anpassningar.
- Löpande arbete i form av insamling/mottagande av metadata, uppdatering och indexering.
- Löpande leverantörskontakter och administration kring licenser.
- Löpande administration av bibliotekens samlingar/källor, dvs. exporter/importer av data
- Hantering av bibliotekens bestånd, urval och metadata nivå.
- Administration/organisation kring ”community” för insamling av metadata.

## När kan Scenario *Egenutveckling* bli verkligt?

Först måste en helt ny teknisk infrastruktur byggas upp som berör LIBRIS nationella system som helhet och där tanken om databrunn är en del. Vi har också KB:s utökade uppdrag att ta hänsyn till som kommer att innebära utveckling av en ny nationell katalog dvs. en ny parameter i samlingen av nationella system. Dessutom behöver hänsyn tas till KB:s kommande hantering av e-pliktleveranser och den ökande mängden digitaliserat material. Det är mycket en fråga om prioriteringar och naturligtvis finansiering. Det är viktigt att veta vad biblioteken faktiskt kommer att satsa på när det gäller lösning för integrerad sökning, som det handlar om för deras del. En påverkande faktor är också huruvida biblioteken kan tänka sig att vara med och finansiera denna lösning.

## 8.2 Scenario *Upphandling av färdig lösning*

### Beskrivning

LIBRIS söktjänst vidareutvecklas och kompletteras med funktionalitet så att den fungerar mot en nationellt upphandlad leverantörlösning. Metadata lagras hos leverantören och insamling av nya källor beställs också där.

### Vad krävs

KB bygger upp ett system med en systemleverantörs API:er för att på så vis göra metadata sökbar i en nationell kontext.

När det gäller insamling av metadata, så begränsas det till artikeldata motsvarande BIBSAM-avtalen.

### Resursbehov

Förutom extra resurser för en nationell upphandling/nationellt ramavtal krävs initialt arbete med utveckling av sökfunktionalitet via API:er för ett nationellt sökgränssnitt. Omfattningen av nuvarande licensarbete skulle öka precis som i *Scenario Egenutveckling*.

### Nyttan

Det här är ett scenario som behöver undersökas vidare för att se om det verkligen motsvarar och uppfyller bibliotekens behov, och om det är hållbart i ett längre perspektiv. En viktig frågeställning handlar då om de tjänster leverantörerna erbjuder kring den samlade aggregeringen av data. Idag erbjuds sök- och länkservertjänster, men det är inte säkert att det är just dessa tjänster som behövs på längre sikt.<sup>5</sup> I en situation

<sup>5</sup> Se t.ex. den nedåtgående trend för biblioteken som sökingång som noteras i Ithaka-rapporten ”Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies” (Schonfeld, R. C. & Housewright, R. 2010). I denna undersökning

där relevant metadata finns aggregerat hos leverantören blir det svårare för bibliotek att undersöka, testa och samarbeta kring tjänster och gemensam infrastruktur. Här kan också noteras att en situation där metadata aggregeras av ett fåtal globala aktörer inte självklart gynnar biblioteken. En situation där bibliotek är beroende av systemleverantörer för dataaggregering innebär en situation där konkurrens mellan leverantörer i hög utsträckning sker på leverantörernas villkor. Vad gäller aggregeringen blir man beroende av systemleverantörernas kommunikation och avtal med förlagen, vilket innebär att biblioteken, och därmed den offentligfinansierade verksamheten, får en sämre position vad gäller att påverka marknaden i den riktning man önskar.

### **Kommentar**

Tjänsten Samsök, 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. Med nuvarande licens får medlemsbiblioteken möjligheter till integrerad sökning i metadata från fria och licensierade artiklar, e-tidskrifter och e-böcker. Den nuvarande nationella licensen för programvaran, som Kungliga biblioteket står för, innefattar inte integrering av bibliotekens lokala samlingar (bibliotekskataloger etc.). Den fortsatta utvecklingen av Samsök är något som diskuteras i en separat utredning som nämndes tidigare, men vi vill understryka att Samsök är en tjänst som ingår i den nationella infrastrukturen för informationsförsörjningen och därför bör utredas i skenet av detta och inte som en isolerad företeelse. Så som Samsök fungerar idag är det en tjänst för medlemsbibliotek och inte en fullständig nationell lösning.

Ur ett rent tekniskt perspektiv blir det svårt att se hur LIBRIS skall bygga en integrerad tjänst som delvis bygger på externa sökningar via API:er. Idén med databrunnen är att komma runt just den federerade sökningens tillkortakommanden. Ur teknisk synvinkel är det därför rimligare för biblioteken att integrera allt data hos en leverantör för att få en sammanhållen lösning.

### **Sammanfattning av konsekvenser**

- Nationell upphandling alternativt nationellt ramavtal.
- Tre-partsförhållande mellan: KB, systemleverantör och biblioteken.
- Biblioteken är medfinansiärer
- Metadata lagras hos leverantören. Vid insamling av metadata från nya utvalda källor till exempel regionala, måste det beställas hos leverantören.
- Löpande leverantörskontakter för förändrade krav och samordning från KB och från biblioteken.
- Biblioteken är hänvisade till färdigt användargränssnitt från systemleverantören, alternativt
- Utveckling av sökfunktionalitet via API:er för ett nationellt sökgränssnitt.

### **När kan Scenario *Upphandling av färdig lösning bli verkligt?***

Erfarenheter från DEFF visar att det förberedande arbetet med kravspecifikation är mycket omfattande, utöver själva upphandlingsprocessen. Därtill kommer efterarbete med utveckling av sökfunktionalitet om det ska anpassas och fungera i ett nationellt gränssnitt. Nu har också upphandlingen i Danmark avbrutits vilket borde vara en varningssignal om komplexiteten i det hela.

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

#### **Beskrivning**

Det finns en möjlighet att inte skapa en nationell databrunn för licensierade e-resurser. Istället skapas en nationell databrunn med inriktning på att samla metadata som kompletterar de behov som biblioteken har idag av artiklar. Det skulle kunna vara svenska resurser såsom Uppsök, SwePub, e-pliktleveranser, digitaliserat material, audiovisuella medier, BURK, Svensk Mediedatabas. Här ligger satsningen på bra kvalitet över data som vi har möjlighet att ha kontroll över och också i många fall unikt data. 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.

---

observeras hur forskare inom samtliga vetenskapliga discipliner i allt lägre utsträckning betraktar biblioteken som utgångspunkt för informationssökningar. Bibliotekens infrastrukturella roll för tillgängliggörandet av och åtkomst till vetenskapligt material värderas dock väldigt högt. Det är därmed inte självklart att det är utvecklandet av bättre söktjänster som är framtiden för forskningsbiblioteken.

## Kommentar

Eftersom fokus i denna förstudie var att undersöka vilka möjligheter det finns att skapa en nationell databrunn, för att på så vis hitta en lösning för åtkomst av e-resurser, särskilt artikeldata, har vi valt att inte gå djupare in på detta spår. Det skulle kunna vara en möjlig väg, men den löser inte problemet med att göra artiklar sökbara och ger inte heller en samlad lösning för biblioteken.

## 9. Rekommendation

Utifrån de erfarenheter som gjorts inom ramen för LIBRIS, Samsök och BIBSAM-konsortiet ser vi ett behov av att e-resurser hanteras så optimalt som möjligt. Bibliotekens behov av kontroll och flexibilitet är också den viktigaste faktorn här för att tanken om databrunn ska bli en hållbar lösning för framtiden och inte en nödlösning på kort sikt.

Erfarenheterna av tidigare utvecklingsprojekt visar att med en kontrollerad egenutveckling i kontinuerlig dialog med biblioteken och användarna är utsikterna goda för att skapa en långsiktigt hållbar och integrerad infrastruktur med fokus på att nå en situation där både LIBRIS och biblioteken har den kontroll över metadata som behövs även för framtida, idag oförutsedda problem. Genom att betrakta och utveckla databrunnar som en integrerad del av bibliotekens infrastruktur i vid mening uppstår flera positiva synergieffekter.

Behovet av en nationell databrunn måste också vägas mot de resurser som till exempel KB:s utökade uppdrag och hanteringen av kommande e-pliktleveranser kommer att ta i anspråk. En databrunn kan fylla en funktion inom flera av dessa områden, särskilt den ”nationella katalogen”.

Vi vill framhäva att det är viktigt med en kontinuerlig dialog med biblioteken och användarna för att vi ska få en hållbar lösning. Det är också nödvändigt att hitta bra samarbetsformer kring databrunnen, till exempel genom en utvidgning av nuvarande modell för katalogiseringssamarbete i LIBRIS. Biblioteken behöver delta i insamlandet av metadata. Det kan också vara av intresse att samarbeta över nationsgränser när det gäller utarbetande av standarder för metadataformat, hur det ska levereras, metadatakvaliteten med mera.

En viktig faktor är äganderätten till metadatat. Vi behöver kunna förfoga och hantera datat själva, för att det ska kunna utnyttjas effektivt både nationellt och lokalt. Ett grundläggande villkor för samarbetet kring LIBRIS samkatalog har varit att data ska vara öppet och att biblioteken ska äga det data som finns i de egna systemen. Det ger möjligheter att modellera tjänster efter aktuella behov och att skapa en så flexibel och fri åtkomst som möjligt. Vi ser **Scenario Egenutveckling** som den långsiktigt hållbara modellen för att åstadkomma detta.

## 10. Litteratur och källor

Blenkle, Martin; Ellis, Rachel; Haake, Elmar (2009) "Next-generation library catalogues: review of E-LIB Bremen". *Serials* 22(2) 178-181.

Breeding, Marchall (2010) "The state of the art in library discovery 2010". *Computers in libraries* 1(30) 31-34.

DEFF databrönde for artikler og e-bøger: Rapport med anbefalinger til Styregruppen fra de to afklaringsprojekter (2008). DEFF, Programområdet informationsforsyning.

Gillén, Camilla et al (2009) "Ny söktjänst för alla typer av resurser, Projektrapport Göteborgs universitetsbibliotek".

Hylton, J A (1996) "Identifying and merging related bibliographic records," M.S. thesis, MIT. Published as MIT Laboratory for Computer Science Technical Report 678.

Lohrum, S., Schneider, W., Willenborg, J. (1999), "De-duplication in KOBV" (<http://sc.hlrn.de/Publications/Reports/SC-99-05.pdf> 2010-04-07)

Lykke, M et al (2010) " Slutrapport: DEFF forskningsprojekt: Värktöjer til integreret sögning".

Mayfield, Ian; Humphreys, Linda (2008) "Next generation library catalogues: reviews of ELIN, WorldCat Local and Aquabrowser". *Serials* 21(3) 224-230.

Nelke, Margareta (2009) "Samsök under luppen - en översyn av organisation, användning och nytta samt några scenarier för framtida utveckling". I.C. at Once

NGC4LIB mailing list archives, February 2010:

<http://serials.infomotions.com/ngc4lib/archive/2010/201002/index.html>

Rochkind, Jonathan (2007) "(Meta) search like Google". *Library Journal*, 2(15).

<http://www.libraryjournal.com/index.asp?layout=articlePrint&articleID=CA6413442>

Schonfeld, Roger C.; Housewright, Ross (2010) Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies" <http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/faculty-survey-2009>

Stevenson, Karen (2009) "Next-generation library catalogues: Reviews of Encore, Primo, Summon and Summa". *Serials* 22(1) 68-82.

Öhrström, Bo; Mikkelsen, Lise (2008) "Danmarks nationale databrönde for elektroniske materialer: oplæg vedrørende organisation, styring m.m. DEFF.

### Undersökta produkter i denna förstudie

Ebsco Discovery Service: <http://www.ebscohost.com/discovery/>

ELIN: <http://elin.lub.lu.se/elinInfo>

Primo Central: <http://www.exlibrisgroup.com/category/PrimoCentral>

Summa: <http://wiki.statsbiblioteket.dk/summa/>

Summon: <http://www.serialssolutions.com/summon/>

Ting: <http://gnit.dk/>

### Databrunnsinitiativ

DADS: <http://www.dtic.dtu.dk/English/infosog/dads.aspx>

ELIB Bremen: [http://elib.suub.uni-bremen.de/frs\\_projekt\\_elib.html](http://elib.suub.uni-bremen.de/frs_projekt_elib.html)

Ontario Scholars Portal: <http://www.scholarsportal.info/>

Trove: <https://wiki.nla.gov.au/display/LABS/2.+Trove>

## Bilaga 1.

### Förteckning över BBSAM-avtal: Fulltext- och bibliografiska resurser

Publisher	Package/Database
American Chemical Society	ACS
American Institute of Physics	AIP
American Psychological Association:	PsycINFO, PsycARTICLES, PsycBOOKS, PsycCRITIQUES, PsycEXTRA.
Association for Computing Machinery	ACM Portal. ACM
BMJ Publishing Group	BMJ e-tidskrifter
Brill Academic Publishers	Brill Journals
Cambridge University Press	Cambridge Journals Online
Columbia University Press	Columbia International Affairs Online
Elsevier Engineering and Chemistry	Refex Engineering
Elsevier Engineering Information Inc.	Compendex
Elsevier Engineering Information Inc.	Inspec
Elsevier Science	Cell Press e-tidskrifter
Elsevier Science	ScienceDirect
Elsevier Science	SCOPUS
Emerald Group Publishing	Emerald Management Xtra Plus
Institute of Electrical and Electronics Engineers	IEEE Xplore. IEEE
Institute of Physics	IOP
John Wiley & Sons, Inc.	Wiley-Blackwell Journals
Johns Hopkins University Press	Project MUSE
Karger	Karger tidskrifter
Kluwer Law International Journals	Kluwer Law International Journals
Mediearkivet Svenska AB	Mediearkivet
Mary Ann Liebert	Mary Ann Liebert e-tidskrifter
Nature Publishing Group	Nature Journals & Academic Journals
Newsline Group AB	Affärsdata, Presstext
Oxford University Press	Oxford Online Databases
ProQuest	Cambridge Scientific Abstracts Databases
ProQuest	Dissertations & Theses database (PQDT)
Sage Publications	SAGE
Springer Verlag	SpringerProtocols
Springer Verlag	Landolt-Börnstein
Springer Verlag	Lecture Notes in Computer Science (LNCS)
Springer Verlag	SpringerLINK
Thieme	Thieme EbookLibrary
Thomson Reuters	Karnov
Thomson Reuters	Web of Knowledge: Web of Science, ISI Proceedings, Journal Citation Reports, Essential Science Indicators, Derwent Innovations Index