

Fra søg til find – Primo i Danmark

Af Jørgen Madsen

På Det Kongelige Bibliotek har vi igennem de sidste 2-3 år været os stadig mere bevidste om behovet for et opdateret søgesystem. Dette behov har været foranlediget af to sideløbende udviklingstendenser.

Jørgen Madsen
Det Kongelige Bibliotek
jma@kb.dk



Vores bibliotekskatalog sakker i stigende grad agterud i forhold til den generelle udvikling af søgeapplikationer, der møder brugerne på nettet, og det bliver til stadighed tydeligere, at brugerne forventer at kunne søge ”det hele” ét sted og få tingene præsenteret samlet. Denne sidste del er desuden accelereret af, at de fysiske samlinger relativt set får mindre betydning, mens materiale tilgængeligt online – det være sig artikler, web-sider, billeder eller digitaliseret materiale – fylder mere og mere i brugernes bevidsthed og dækker en stadig større del af deres informationsbehov.

Vores nuværende katalog bærer præg af at være en afledning af bibliotekssystemet, som den er bygget direkte oven på. Alene navnet afslører, hvad det reelt er, vi tilbyder brugerne: Katalogen. I bund og grund har vi gjort vores katalogkort elektronisk tilgængelige og søgbare. Selv vores fremvisningsformater bærer fortsat præg heraf. Dette ikke være sagt som nogen kritik, eftersom udviklingen har været naturlig. Det har i sin tid været revolutionerende, at brugerne pludselig kunne sidde derhjemme og orientere sig i bibliotekets samlinger og oven i købet kunne bestille materiale. Vi må blot erkende, at den direkte kobling imellem bibliotekssystemet og de søgemuligheder, der pga. denne kobling møder brugerne, ikke længere er hensigtsmæssig og ikke fungerer tilfredsstillende. Der er umiddelbart fire kerneproblemer:

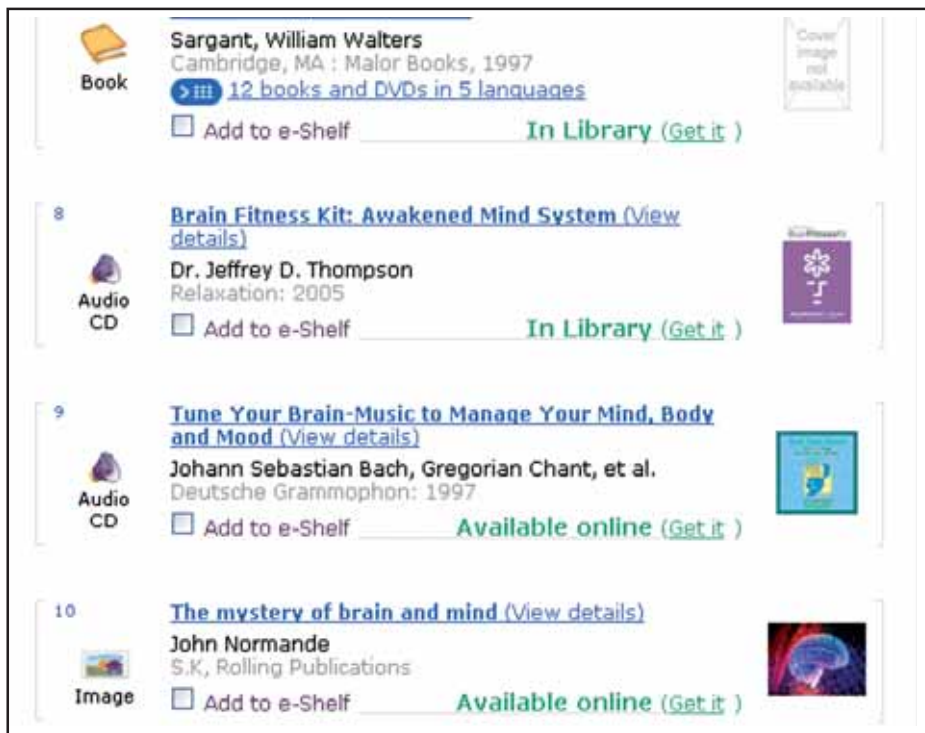
- Søgning foregår relativt langsomt – især når der er tale om søgninger, der finder mange poster.
- Katalogen fungerer bedst til verifikation. Brugere finder typisk det ønskede andetsteds (f.eks. via Google eller Amazon) og går så efterfølgende til bibliotekskatalogen for at se, om biblioteket har det pågældende værk, så de kan bestille det.
- Og afledt af ovenstående: Katalogen er ikke et sted, der indbyder til at gå på opdagelse. Der er ingen ”begejstringsskabende” effekt, der giver mulighed for at foretage de associative spring, der fungerer så godt på nettet. Hvem kender ikke følelsen af at gå i Amazon og skulle finde en bog om amerikanske præsidenter for tyve minutter senere at sidde og læse om franske rokokkjoler? Og vejen derhen har været helt naturlig. Den slags er vores nuværende søgesystemer ikke gearret til.
- Vi kan ikke integrere forskellige datakilder og -typer i samme søgesystem. På Det Kongelige Bibliotek har vi alene otte ALEPH-databaser, der skal søges hver for sig. Derudover har vi mindst fjorten forskellige øvrige applikationer (f.eks. Portrætrejstranten, Arkiv for Dansk Litteratur m.m.), der ligger som adskilte systemer med hver deres søgegrænseflade. Det betyder for det første, at brugerne ikke kan søge alt på én gang, og det betyder for det andet, at vi med ret stor sandsynlighed har store dele af vores samlinger, der er voldsomt undereksporerede. Dette er et problem såvel for os som institution, men især for brugerne. Vi skal med andre ord skabe et system, der giver brugerne det, de end ikke vidste, at vi havde – og som giver dem det, de slet ikke vidste, de havde brug for!

Kort om Primo

Vores situation på Det Kongelige Bibliotek var derfor den, at vi havde en ret god fornemmelse af problemstillingen. Vi havde identificeret vanskelighederne samt den funktionalitet, vi samlet set kunne ønske os. Der var blot ikke nogen løsning i sigte. Denne situation ændrede sig imidlertid for cirka halvandet år siden, hvor vi for første gang så en demo af Primo. Primo er under udvikling af firmaet Ex Libris, der også står bag vores bibliotekssystem, ALEPH. Det stod umiddelbart klart for os, at der her var tale om et nyt og revolutionerende koncept, der potentielt set kunne løse de udfordringer, vi stod overfor.

Det Kongelige Bibliotek tog derfor initiativ til, at der sammen med Danmarks Tekniske Videncenter, Aalborg Universitetsbibliotek og Det Administrative Bibliotek blev dannet et konsortium, der i første halvdel af 2006, via Fujitsu, satte sig i kontakt med Ex Libris med henblik på et muligt samarbejde. Samarbejdet kom endeligt på plads i løbet af efteråret 2006 i form af en aftale om at indgå et udviklingssamarbejde, der vil løbe i cirka to år til og med version 2 af Primo. Udviklingssamarbejdet er desuden støttet af DEFF.

Primo er et såkaldt integrated search-system. Filosofien i Primo går i sin enkelhed ud på at tage data fra forskellige steder og lægge dem ét samlet sted i et fælles format. Et meget vigtigt aspekt herved er, at man adskiller back office (bibliotekssystemet) fuldkommen fra frontend (søgedelen). Ved at gøre dette sikrer man sig en langt større grad af fleksibilitet, da ændringer foretaget det ene sted ikke umiddelbart har nogen indvirkning på det andet. Data høstes fra bibliotekssystemet (og andre kilder), og der arbejdes



Figur 1: Eksempel på en resultatlistevissning fra Primo. Bemærk, at forskellige materialer optræder sammen. Derudover er det direkte fra det korte format muligt at se tilgængelighed, at bestille samt at tilføje til e-hylde. Bogforsider vises, hvor det er muligt. Endelig grupperes materiale (FRBR), som det kan ses i den første post.

på den måde på en "kopi" i søgesystemet. Dette medfører, at der kan laves et fælles søgeindeks, så det hele søges og præsenteres på én gang i et fælles søgeresultat, hvor f.eks. bøger, artikler og billeder kan blandes i resultatlisten. Problemet med inkommensurable data er løst (se figur 1).

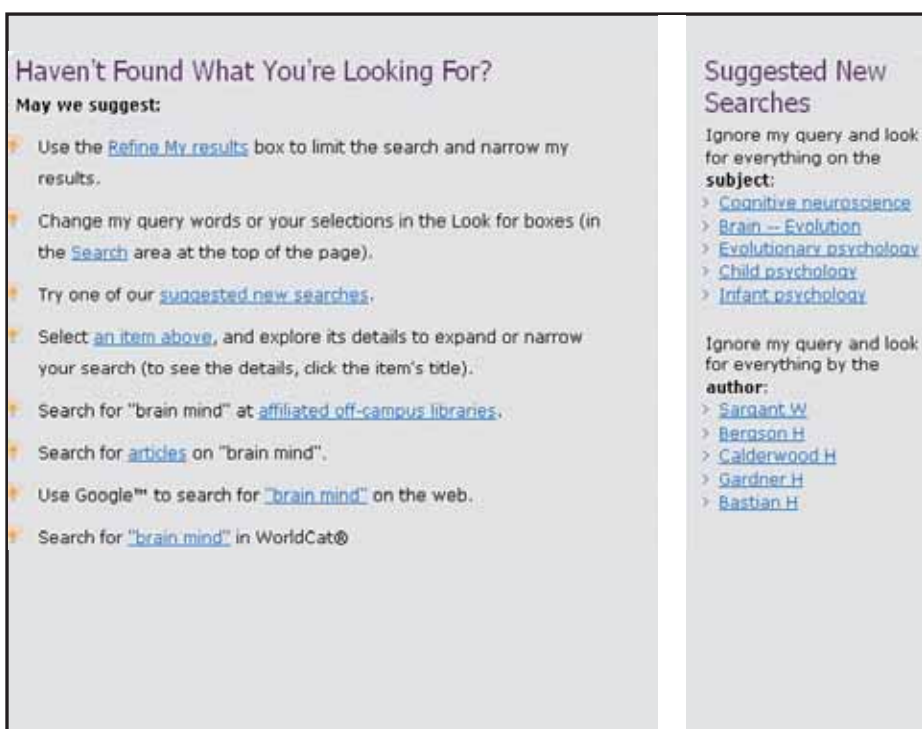
Samlingen af forskelligt materiale fra forskellige kilder i samme resultatliste er naturligvis i sig selv et kvantespring, men det løser ikke alle de skitserede problemer. Primo tager fat i de fleste af dem (se figur 2):

- **Hastighed:** Søgning i Primo vil være lynhurtig. Hastigheden vil være stort set lig Google, og søgeresultatet præsenteres efter brøkdele af et sekund. Dette vil også gælde for store

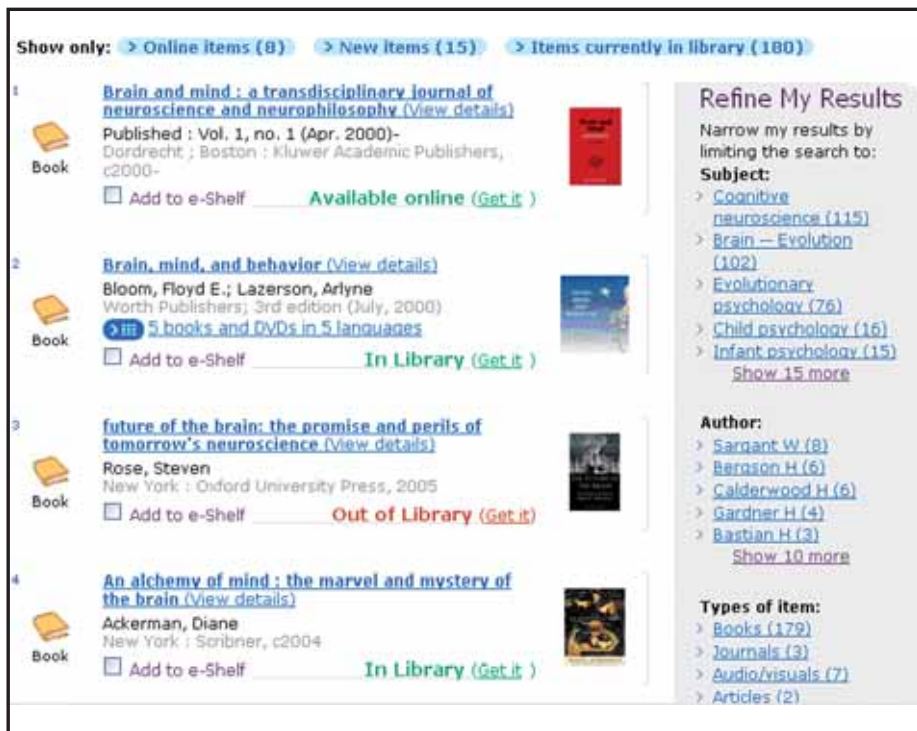
søgesæt. Det skyldes Primos arkitektur, hvor indekserne er bygget op på en sådan måde, at store søgninger vil foregå (stort set) lige så hurtigt som mindre søgninger.

- **Oplevelsesorienteret:** Et af de centrale aspekter i Primo er, at det skal være et sted, hvor brugeren har lyst til at gå på opdagelse. Det skal være et sted, hvor det er naturligt at foretage de førnævnte associative spring. Dette gøres ved blandt andet at integrere data mining (*Andre, der har søgt efter, har også søgt...*), forslag til nye søgninger, korrektion af f.eks. stavfejl (*Mente du...*) osv. Anden kontekstafhængig hjælp kunne tage sig ud som vist i figur 2.

- **Facetter:** Enhver søgning i Primo nedbrydes automatisk i logiske delmængder, der præsenteres ude til højre for resultatlisten. Søgessættet grupperes logisk efter f.eks. emne, forfatter, materialetype, sprog osv. Facetterne kan sættes op efter ønsker og behov. I praksis tager det sig således ud som vist på figur 3.
- **"Social computing"** – brugeren som katalogisator? Et sidste meget interessant aspekt ved Primo er muligheden for direkte brugerinddragelse i katalogen, den såkaldte social computing, som for alvor er



Figur 2: Kontekstafhængige søgeforslag i Primo, der hurtigt henleder brugerens opmærksomhed på andre muligheder ifm. søgningen.



Figur 3: Primos facetter kan ses til højre i skærbilledet under 'Refine My Results'. Et klik på en af disse facetter giver brugeren en delmængde af søgeresultatet. Facetterne kan sættes op efter ønske og kan evt. også gå på delsamling m.m. Bemærk også øverst i skærbilledet, hvor brugeren har mulighed for at afgrænse sin søgning til 'Online items', 'New items' eller 'Items currently in library'.

ved at vinde indpas på nettet. I Primo sker dette i form af reviews (anmeldelser) og tags (brugertildeling af emneord):

- Reviews: Fra det fulde format i en post er der mulighed for at give brugerne adgang til at skrive anmeldelser – og for at de kan se andres anmeldelser af det pågældende materiale. Det enkelte bibliotek vælger selv, om det er en funktion, man ønsker at stille til rådighed.
- Tags: Tags er blevet en stor succes på nettet, især for billeders vedkommende (jf. f.eks. Flickr). I Primo vil der være mulighed for at lade brugerne tildele deres egne emneord til posterne. Disse tags kan ydermere gøres søgbare, hvis det ønskes, og brugerne kan orientere sig

i de brugertildelte emneord i systemet som inspiration til videre søgninger (se figur 4).

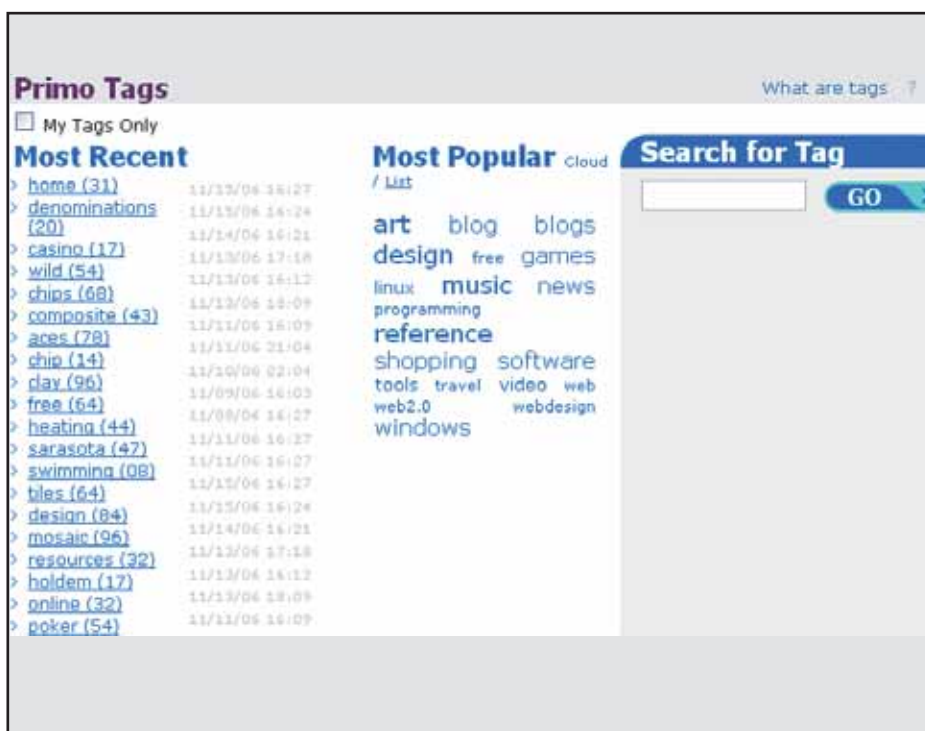
Det danske Primo-konsortium samarbejder konkret med Ex Libris om følgende:

- Håndtering af sign on-problematikker, der især er relevante i en konsortieløsning.
- Brugergrænsefladen i Primo ifm. integration af Ex Libris' metasøgningsværktøj MetaLib, der er fuldt integreret med Primo allerede i version 1.
- Skalérbarhed for meget store datamængder. Det er slutmålet, at Primo skal kunne indeholde metadata på alle licensbelagte artikler, som konsortieparterne har adgang til, og det vil give et indhold i Primo på langt over 100 mio. dokumenter.

- Primo og web services. Primo er baseret på en meget åben arkitektur og overholder alle internationale standarder. Det er derfor let at integrere dele af Primos funktionalitet i andre applikationer (f.eks. en Primo-søgeboks), og det er samtidig enkelt at implementere web services direkte i Primo, så man enkelt og med effekt kan bruge andre systemer som leverandør af tjenester til Primos grænseflade.

Kernen i Primo: PNX og pipes

Helt centralt i Primo er fællesformatet, den såkaldte PNX (Primo Normalized XML). Dette format er på mange måder, sammen med Primos pipes (se nedenfor), kernen i hele systemet og forudsætningen for arkitektur, funktionalitet mv. PNX'en er en helt ny måde at tænke formater



Figur 4: Brugertags i Primo. Brugerne kan se deres egne tags, og de kan søge i andres.

FAKTABOKS

Af Steen Pedersen
Fujitsu
steen.pedersen@dk.fujitsu.com

Efter opkøbet af Endeavor fra Elsevier er Ex Libris i dag en af verdens 2 største leverandører af løsninger til bibliotekerne globalt.

Ex Libris er i Danmark (og Sverige) repræsenteret af Fujitsu, der igennem en lang årrække har leveret IT baserede løsninger til bibliotekerne (oprindeligt under navnet Regnecentralen med løsningen RCLib).

I dag består produktporteføljen af ALEPH 500, MetaLib, SFX, Verde, Digi-Tool, Materialevalg, og en række andre løsninger.

Udviklingen af Primo er baseret på et udviklingssamarbejde med i alt 4 partnere. Udover det danske udviklingskonsortium består partnerne af:

- Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen, Tyskland (Publishing platform)
- University of Minnesota, USA (Consortia og Usability)
- Vanderbilt University, USA (Unicorn og MetaLib).

Som et led i samarbejdet mellem bibliotekerne og Ex Libris, er der etableret en Primo Charter Customer Group. Baggrunden er dels den store interesse og bevågenhed om Primo, dels behovet for

at imødekomme bibliotekernes ønske om at kunne påvirke i hvilken retning udviklingen af Primo skal gå.

Aftalen, der er indgået om Primo, indebærer som nævnt i artiklen favorable indtrædelsesvilkår for biblioteker under DEFF, herunder mulighed for yderligere reduktion i prisen såfremt det enkelte bibliotek indgår aftale i løbet af 2007.

For yderligere informationer om Primo, herunder oplysninger om pris, kontakt venligst Fujitsu på: info@dk.fujitsu.com Eller tlf. 4489 4489

Se også www.exlibris.co.il

på. I stedet for at trække på de samme format-data i forskellige sammenhænge går Ex Libris med PNX en helt ny vej, hvor nødvendige data i formatet sektionsopdeles og dubleres, således at hver sektions data bruges til et og kun et formål. Dette har helt afgørende betydning for f.eks. svartider og fleksibilitet. PNX'en sikrer, at al funktionalitet er umiddelbart tilgængelig, og at systemet er lynende hurtigt, fordi der ikke skal ske efterbehandling af data, når en søgning først er sendt af sted, og resultatet foreligger.

Dette aspekt har været af stor betydning for Primo-konsortiets ønske om at entrene med Ex Libris, da vores vurdering fra begyndelsen har været, at man med PNX har foretaget et afgørende og rigtigt strategisk valg. PNX'en er derfor forskellig fra den filosofi, man finder i nuværende standard- samt i traditionelle, selvudviklede systemer. Samtidig kan det med en vis sandsynlighed formodes, at PNX vil kunne gå hen og blive en slags de facto industri-standard pga. Ex Libris' internationale gennemslagskraft.

Pipes

Primo er desuden født med en sofistikeret metode til at høste, normalisere, berige og standardisere data fra en lang række datakilder og -formater. Dette kaldes i Primo 'pipes'. Disse pipes kan på forskellig vis via en lang række standardprotokoller (ftp, OAI, Z39.50, XML-gateways m.m.) loade data ind i Primo. Data hentes altså ind i Primo i helt

automatiseret form og transformeres og beriges i selve processen, hvilket reducerer alle arbejdstunge processer dramatisk. Dette vil have en afgørende indflydelse på driftsomkostninger, da der alene på denne konto i det lange løb vil kunne spares hele årsværk i forhold til selv at skulle udvikle og (ikke mindst) vedligeholde en sådan metode. Ydermere har vi garanti for løbende vedligeholdelse og videreudvikling, uden at konsortiet skal lægge arbejdstimer heri. Konsortiet vil desuden have mulighed for direkte at påvirke udviklingen af nye pipes.

Endelig er administrationen af Primo på forhånd en velintegreret del af systemet. Administrationen finder sted i form af et web-interface, hvor langt størstedelen af arbejdsopgaverne – fra grundlæggende opsætning over høstning af data til udseendet af søgegrænsefladen og statistikudtræk – kan håndteres via en række trin-for-trin wizards, der guider systemadministratoren igennem forløbet. Administrationsdelen er designet med henblik på, at de fleste arbejdsopgaver skal kunne varetages af (system)bibliotekarere.

Primo i Danmark: åben for alle DEFF-biblioteker

Primo er som system født med ægte konsortiehåndtering for øje. Det betyder, at den samme Primo-installation kan indeholde utallige Primo'er, der hver især trækker på præcis den mængde data, der er relevant for det konkrete bibliotek. Hvert bibliotek vil have sine egne

søgeindekser og sit eget "look and feel" mht. grafik, søge- og præsentationsmuligheder osv. Samtidig vil det være muligt at trække på andres data i Primo. Man kunne forestille sig, at f.eks. beslægtede kunstbiblioteker kunne have en interesse i at synliggøre hinandens samlinger, eller at institutbiblioteker ved Københavns Universitet ville være interesserede i, at brugerne kunne søge i samtlige institutbibliotekers samlinger på én gang. Med Primo vil dét være muligt, og søgningen vil være reel, idet posterne rent fysisk ligger i samme "base". Derfor er Primo heller ikke afhængig af det bagvedliggende biblioteks-system, men kan integreres op mod de fleste standardssystemer.

Det Kongelige Bibliotek er databasevært for cirka 100 eksterne biblioteker. Det har derfor været os særdeles magtpåliggende i vores forhandlinger med Ex Libris og Fujitsu, at det skulle være realistisk for små og mellemstore biblioteker at koble sig på Primo. Efter lange og svære forhandlinger nåede vi frem til en aftale. Aftalen indebærer, at små og mellemstore biblioteker under DEFF vil kunne tilslutte sig Primo under favorable vilkår.

Vejen frem imod et kørende driftssystem er den, at Primo i marts 2007 installeres på Det Kongelige Bibliotek. Denne installation danner udgangspunkt for en meget grundig testning af systemet og tilpasning til konsortieparternes data og øvrige behov. Det forventes, at Primo åbnes for publikum medio 2007.