

Udredning om national opsamling af brugerskabte data

Der har på forskellige workshops været enighed om, at brugerskabte data bør opsamles og kunne tilgås via en national service. En fælles service for biblioteker, netbiblioteker og bibliotek.dk er betragtet som afgørende for momentum og for opnåelse af en kritisk masse. Endvidere var ønsket, at data skal kunne opsamles centralt og bruges lokalt – og omvendt.

Efter en drøftelse i Koordinationsgruppen for netbiblioteker har Styrelsen for Bibliotek og Medier besluttet at udbyde et forprojekt med henblik på at gennemarbejde et beslutningsgrundlag, der samtidig kan danne baggrund for et muligt udbud af et implementeringsprojekt af sådan national fælles løsning. Forprojektet gennemføres af et konsortium med en beløbsramme på 50.000 kr.

Resultatet af forprojektet foreligger nu i form af en afrapportering fra en workshop, hvor en række væsentlige spørgsmål er blevet drøftet blandt en kreds af medarbejdere fra folke-, uddannelses- og forskningsbiblioteker samt netbiblioteker mv. (se bilag).

Forprojektet er gennemført som en virtuel workshop af et konsortium bestående af Axiell, DBC (projektleder), Gentofte Bibliotekerne, Herning Bibliotekerne og Århus Kommunes Biblioteker.

Den virtuelle workshop er gennemført i 2 tempi; først en debat på <http://bsd.n.wordpress.com/>, opdelt i 17 forskellige temaer, dernæst en spørgeskemaundersøgelse blandt workshopdeltagerne. Bloggen er stadig åben, medens resultatet af spørgeskemaundersøgelsen vedlægges som bilag.

Det var Styrelsen for Bibliotek og Mediers oprindelige forventning, at afrapporteringen skulle indeholde en beskrivelse for en national opsamling af brugerskabte data. Som led i den endelige aftale mellem Styrelsen og konsortiet ønskede Styrelsen for Bibliotek og Medier endvidere overslag over økonomien i udvikling og drift af en løsning samt forslag til mulige forretningsmodeller både for udvikling og efterfølgende drift.

Konsortiet er mht. økonomien nået til enighed om, at det ville være konkurrenceforvridende såfremt konsortiet i forbindelse med den endelige beskrivelse afgiver overslag over økonomien. Desuden indeholder den endelige beskrivelse en række uafklarede spørgsmål, hvor svarene er med til at definere servicen, og som Styrelsen derfor bør forholde sig til forud for afholdelse af en udbudsforretning. Spørgsmål om forretningsmodeller er behandlet på workshoppen.

På konsortiets vegne, den 26. september 2008

Per Mogens Petersen
Projektleder

Den kursiverede tekst på de følgende sider afspejler konsortiets konklusion på baggrund af debat på bloggen og resultatet af spørgeskemaundersøgelsen

Indledning

Der etableres en national biblioteksservice for brugerskabte data, i det følgende omtalt som den nationale service eller blot servicen.

Den nationale service er en database, som kan modtage, opbevare og udlevere brugerskabte data. De brugerskabte data er først og fremmest data, som er indtastet af brugerne, dvs. brugerne af bibliotekernes sites, herunder bibliotekskataloger, netbiblioteker, fagportaler og bibliotek.dk. Dataene i databasen opbevares som poster under de enkelte materials id-numre.

Data fra servicen udleveres som poster til sites / portaler via en webservice eller som dataleverancer.

På workshoppen blev brugerskabte data i tilknytning til spørgsmål og svar på spørgetjenesterne (Biblioteksvagten og Spørg Olivia) drøftet; workshop-konsortiet finder det interessant, men betragter det hørende til en senere fase af den nationale service.

Ansvar for servicen

Styrelsen for Bibliotek og Medier afholder en udbudsforretning med det formål at få etableret den nationale service. Som resultat af udbudsforretningen indgår Styrelsen for Bibliotek og Medier en kontrakt med den udvalgte entreprenør, som regulerer alle forhold vedrørende udvikling, drift, vedligeholdelse og videreudvikling af servicen for en tidsbegrænset periode på fx 5 år. Det er essentielt, at de bibliotekssites, som er tilsluttet servicen, sikres indflydelse på den fremadrettede udvikling. Der kan tænkes flere modeller for ansvarsfordelingen:

1. Det totale ansvar (teknisk, fagligt, juridisk, økonomisk, organisatorisk etc.) overdrages til entreprenøren, som etablerer et rådgivende udvalg, hvori bl.a. Styrelsen er repræsenteret.
2. Det "daglige" ansvar overdrages til entreprenøren. Rammerne for dette ansvar aftales på kvartalsvise møder mellem entreprenøren og Styrelsen. Ansvar for servicen ligger fortsat i Styrelsen, som etablerer et rådgivende udvalg.
3. Det tekniske ansvar overdrages til entreprenøren. Det økonomiske, juridiske, faglige og organisatoriske ansvar for servicen ligger fortsat i Styrelsen, som etablerer et rådgivende udvalg.
4. Det totale ansvar overdrages til et åbent konsortium; konsortiet er baseret på et sæt vedtægter som bl.a. giver biblioteker ret til at blive primære medlemmer. Alle beslutninger træffes af konsortiet.

Et flertal peger på denne løsning: Det totale ansvar overdrages til et konsortium, som er åbent for alle danske biblioteker og bibliotekssites.

Kommentar: De 4 modeller er dog ikke helt sammenlignelige; eksempelvis kan model 4 godt kombineres med model 2 eller model 3, idet konsortiet kan overdrage det daglige eller det tekniske ansvar til en entreprenør.

Tilslutning til servicen

Bibliotekskataloger, netbiblioteker, fagportaler og bibliotek.dk er primære sites, som har mulighed for at tilslutte sig servicen.

Skal der skelnes mellem forskellige tilslutningsformer for primære sites, altså om sitet fx er både leverandør og bruger, eller om sitet kun er bruger? Skal der skelnes mellem om sitet bruger webservice eller får dataleverancer?

Der indgås tilslutningsaftale mellem den nationale service og de enkelte primære sites som fastlægger samarbejdsvilkårene.

Skal andre offentlige og/eller kommercielle og private sites have adgang til den nationale service som webservice og/eller som dataleverance?

Primære brugere (sites):

Tilslutning til servicen og Betalingsparametre: Der er tilsyneladende lidt modstridende besvarelser på disse 2 spørgsmål: det anbefales at følge svaret for betalingsparametre, dvs. at der ikke skelnes mellem ydende og nydende sites, og at prismodel fastlægges ved en kombination af indbyggertal og sitets benyttelse.

Sekundære brugere (sites):

Ja, bidragyderne skal acceptere, at deres data både bruges i biblioteksites og på alle andre sites. Hvis et sekundært site bruger den nationale service i stor udstrækning bør den ansvarlige for servicen kunne opkræve betaling.

Upload af, ejerskab til og klager over data

Ved upload af brugerskabte data fra primære sites til den nationale service overdrages ejerskabet til den nationale service. Den nationale service har således mulighed for at slette (fx anstødelige data) og bevare data (fx hvis et primært site udtræder af samarbejdet) i egen database. Den nationale service kan indlægge filtre i data, som modtages fra primære sites.

Hvis det primære site giver brugeren mulighed for at ændre i egne data, skal sådanne ændringer automatisk overføres til den nationale service som rettelsesposter.

Klager over brugerskabte data behandles af det bibliotek mv. som er ansvarlig for upload af de pågældende data til den nationale service.

Brugere, der har oprettet data på et bibliotekssite, som leverer data til den nationale service, bevarer ejendomsretten til disse data. I en periode uden fælles login og profildata er der dog ingen garanti for, at brugeren kan rette eller slette egne data.

Betalingen vedrører selve servicen, hvorimod der ikke betales for data.

Finansiering af servicen

Udvikling og implementering: a) finansieres af Styrelsen? b) finansieres delvist (fx 50%) af Styrelsen; resten finansieres over driften, dvs. af de sites som bruger servicen? c) finansieres over driften, dvs. af de sites, som bruger servicen?

Drift og vedligeholdelse finansieres delvist ved: a) tilskud fra staten b) salg af brugerskabte data til andre c) reklamer og sponsorater – eller en kombination af disse?

Uanset disse evt. finansieringskilder anses det for nødvendigt med en årlig betaling fra primære sites, som bruger servicen. Hvad vil I betale for brugerskabte data på jeres site: a) under 25.000 kr./år ? b) mellem 25 og 50.000 kr./år? c) mere end 50.000 kr./år?

Betalingsparametre ved helt eller delvis finansiering via de primære sites: a) antal genererede brugerskabte data, b) antal viste brugerskabte data, c) antal indbyggere i kommunen (gælder ikke netbiblioteker) d) faste tariffer afhængigt af sitets størrelse/benyttelse – eller en kombination af disse?

Der er overvejende interesse for, at staten finansierer udvikling og implementering.

Der er overvejende interesse for, at staten giver underskudsgaranti for drift og vedligeholdelse for en periode på min. 4 år, dog højst 50% af omkostningerne, dvs. op til et beløb som svarer til den samlede betaling fra de bibliotekssites mv., som betaler for den nationale service.

Entreprenøren/konsortiet, der forestår servicen, har det samlede ansvar for drift og vedligeholdelse for minimum den periode, hvor staten yder tilbyder underskudsgaranti.

Særlige forhold som Styrelsen skal forholde sig til

Styrelsen opfordres til at udarbejde en kontrakt, som brugeren accepterer forud for aflevering af brugerskabte data. Kontrakten bør indeholde en beskrivelse af ejendoms- og brugsret til data samt rettigheder i forbindelse med sletning.

Styrelsen opfordres til at beskrive servicemål for den nationale service.

Styrelsen opfordres til at stille krav til entreprenøren om opbevaring af data, herunder håndtering af personhenførbare data.

Styrelsen opfordres til at afklare, om der kan ydes tilskud fra staten til den lokale implementering af servicen.

Styrelsen opfordres til at udarbejde krav til statistik.

Styrelsen opfordres til at overveje om der kan knyttes brugerskabte data til andre materialer end dem, som findes bibliotek.dk, fx netdokumenter.

Styrelsen opfordres til at publicere et forslag til en tidsplan for en national bsd-service bl.a. af hensyn til forberedende arbejde på de sites, som ønsker at implementere servicen.

Styrelsen opfordres til en politik mht. samme brugers flere brugeridentiteter.

Styrelsen opfordres til at definere servicen meget præcist aht. udbudsforretningen.

Brugsscenerier

Data skabt i dansk bibliotekssammenhæng (lokalkataloger, netbiblioteker, bibliotek.dk)

Brugeren vil se en eller anden form for data.

Kun de med grønt markerede processer udføres af den nationale service, resten er lokale opgaver.

Aktør	Start	Trin i processen	Slut	Kommentarer
Slutbruger i lokalkatalog	Ser på en post og vil gerne se andre brugers ratings, anmeldelser, tags, kommentarer m.v.	Klikker på link		Om data vises umiddelbart eller først efter et klik er en lokal afgørelse (eller en systemleverandørs afgørelse)
		Der laves en webserviceforespørgsel (GetData) på postens idnr. - med oplysning fra lokalsystemet om hvorvidt man ønsker data kun fra dette idnr eller ønsker det udfoldet til et værk		
		Webservicen udvider idnr til værk		

		hvis dette er specificeret i forespørgslen		
		Anmeldelser leveres retur på (et eller flere) idnr.		
		Data vises	Brugeren ser de ønskede data	

Hvilke typer brugerskabte data er interessante: ratings, anmeldelser, tags, kommentarer?

Der er størst interesse for ratings og anmeldelser, nogen interesse for kommentarer, tags og links til supplerende materiale.

Er det vigtigt at kunne identificere data som er skabt i en bestemt lokalkatalog/i et bestemt geografisk område/andre udvalgsriterier (køn, alder osv) - altså kun vise data fx fra Randers i sin lokale katalog? Eller "andre brugere på din alder har rated denne cd sådan"

Et lille flertal synes det er vigtigt. Evt. kan det være et tilbud, men ikke obligatorisk fra start. Senere vil brugerne måske være mere vant til at have mange persondata i deres profil. Meget få services med brugerskabte data forlanger i dag data som alder, køn, postnummer, uddannelsesinstitution.

Vær opmærksom på at dette stiller tilsvarende krav til hvor meget brugeren skal oplyse ved tilmelding til servicen. Vil servicen være attraktiv for brugeren hvis de skal svare på spørgsmål om sig selv? Skal sådanne data (om ønsket) trækkes fra lokalsystemets oplysninger om brugeren (her har man jo alder, køn og bopæl)?

Er brugeren i forvejen logget ind i en lokalkatalog når data skabes, kan data som køn, alder og postnummer gemmes. Brugeren skal godkende at disse data videregendes til dette formål.

Brugeren vil selv skabe data

I denne beskrivelse er der gået ud fra at brugeren opretter sig og identificerer sig i det lokale system.

Kun de med grønt markerede processer udføres af den nationale service, resten er lokale opgaver.

Aktør	Start	Trin i processen	Slut	Kommentarer
Slutbruger i lokalkatalog	Ser på en post og vil gerne rate,	Klikker på link a la "Fortæl hvad du		

	anmelde, tagge, kommentere m.v.	synes"		
		Om der er login eller ej er en lokal afgørelse		Se dog spørgsmål nedenfor
		Brugeren afgiver sine data - grænseflade bestemmes lokalt.		
		Der skal være adgang til en fælles beskrivelse af hvad data bliver brugt til og hvilke rettigheder brugeren har		
		Der laves en webservice- forespørgsel (createData) på postens idnr. med data på brugeren		Hvilket system der bruges til at identificere brugeren i den centrale base er ikke denne workshops hovedfokus. Men der skal bruges en identifikation som kan identificere brugeren selvom han bruger flere systemer, og som samtidig er ok i forhold til Datatilsynet. Og der skal i så høj grad som muligt bruges gennemsnitlige

				offentlige løsninger til single-sign-on.
		Webservicen gemmer data (om materialet og om brugeren) på materialets idnr		
		Der sendes besked om at data er gemt		
		Besked vises i lokal grænseflade eller sendes pr mail til brugeren	Brugeren ser beskeden	Om brugeren også ser sine data straks er en lokal beslutning.

Skal brugere kunne afgive data (ratings, tags osv - evt. kun nogle typer) anonymt?
 Afhænger det af, hvilke data, der er tale om? vil man f.eks. gerne kunne se kaldenavnet på brugeren, når det drejer sig om en kommentar eller en anmeldelse, hvorimod det betyder mindre med et tag eller en rating?

Er det vigtigt at brugeren også kan fjerne eller ændre sine data?

En nationalt bruger-id-løsning ville betyde at en bruger kan logge ind med samme login "alle steder". Men en sådan løsning er ikke klar nu.

Er det ok hvis det nationale system i første omgang IKKE har et fælles brugerid-system (dvs. data gemmes på det brugerid + biblioteksnummer som evt er med når der skabes data? - og en bruger af flere systemer kan ikke identificeres som én bruger).

Hvis forretningsmodellen vil indebære at data kan sælges eller viderestribueres, er det så noget brugeren (=dataskaber) skal sige ja til?

Brugeren SKAL identificere sig, men der er enighed om at en service kan fungere uden en national id-løsning, så én bruger kan have flere identiteter som så ikke nødvendigvis er knyttet sammen. Brugere skal kunne fjerne deres data (og dermed kan de ikke være anonymt afgivet).

Data skabt i andre sammenhænge

Andre steder end i danske biblioteker/bibliotekssites kan brugere skabe data om bøger, musik, film m.v. Fx LibraryThing, udenlandske bibliotekssites, boghandlersites fx saxo.com, imdb.com etc. Disse data skal kunne batch-uploades (en sådan service er ikke beskrevet endnu).

Skal den fælles database indeholde sådanne data (hvis de kan fås gratis/til en rimelig pris)?

Vil I betale ekstra for servicen hvis de indeholder data der er købt?

Skal det markeres på hvilket site sådanne data er skabt?

Et lille flertal vil gerne have sådanne data, men vil nok ikke betale for dem. Det skal kunne ses hvor data kommer fra.

Administrative funktioner

Anstødeligt indhold: Hvis en siteejer gøres opmærksom på brugerskabte data, som strider imod de regler der er for sitet, skal disse fjernes.

Hvordan skal det ske i praksis:

A. Det løses i første omgang i lokalt - om man giver brugeren adgang til at rapportere den slags i den lokale grænseflade er en lokal afgørelse. Siteejeren skal dernæst sende en slet-data til den centrale base, som sletter centralt.

Hvis der lokalt er gemt data, som er slettet i den centrale base, er det den lokale siteejers ansvar ikke at have opdateret sine data.

eller

B. Det løses ved en fælles udformet "Rapporter anstødeligt indhold"-knap (med tilhørende fælles disclaimer)

Der skal laves en fælles policy som brugeren skal acceptere første gang de laver brugerskabte data. Hvis der kun lagres data ét sted, slettes posterne så blot i den centrale base.

Søgemuligheder i den centrale base

Hvilke muligheder skal der være:

Man kan vælge at sige at alle ønsker om søgbarhed udover på idnr håndteres lokalt. Altså at man lokalt må lagre brugerskabte data og gøre dem søgbare på den måde man ønsker det.

Man kan også sige at den nationale database skal opbygges så man straks - eller senere - vil kunne foretage søgninger i brugerskabte data samtidig med søgninger i andre typer data. Det kan fx være relevant at søge i brugerskabte tags samtidig med søgning i almindelige bibliografiske data, at kunne sortere ud fra indhold i brugerskabte data (højest ratede titler først) samt en lang række af top-ét-eller-andet (mest ratede, mest anmeldte, nyest anmeldte osv).

Upload/download af større mængder poster: Både bibliotekssystemer og andre dataleverandører/databrugere kan have brug for at kunne up/downlade mange data/alle data. En sådan service er ikke beskrevet endnu. Upload skal være den samme service som for andre dataleverandører. Skal den nationale service kunne tilbyde både høstning af data og batch-upload?

Opdatering: Både de lokale systemer og den nationale service skal kunne levere opdateringsinformation til aftagere så de lokale og den nationale base kan opdateres fx med sletteposter.

Kommentarerne antyder at én national datalagring foretrækkes uden lokale delbaser, men der kan fx tænkes fællesbaser for ét bibliotekssystems brugerskabte data. Derfor ønskes diverse søgemuligheder i den centrale base. Hvordan data uploades er der ikke ultimative krav til.

Tekniske spørgsmål

Udgangspunktet for de tekniske spørgsmål er den løsning der er skitseret for brugerskabte data i bibliotek.dk:

Brugerskabte data til bibliotek.dk

Formålet med projekt 'B01c - Database over anmeldelser og ratings i bibliotek.dk' på udviklingsplanen for bibliotek.dk 2008 er at få etableret en service for brugerskabte data som kan bruges i forbindelse med bibliotek.dk og senere vil kunne udvides til brug i andre sammenhænge. Servicen og databasen udvikles således at den vil kunne udvides til brug i national sammenhæng.

Projektet indeholder følgende leverancer:

- Standardiseret beskrivelse af webservicen til oprettelse og hentning af brugerskabte data
- Database til brugerskabte data
- Webservice der benyttes i forbindelse med oprettelse af data via bibliotek.dk og ved visning af data i bibliotek.dk
- Administrationsmodul der håndterer overvågning af løsningen samt central sletning af brugerskabte data

Der er beskrevet følgende services til databasen over brugerskabte data:

- createData: en bruger skal kunne anmelde eller rate et materiale (en bibliografisk post) i bibliotek.dk brugergrænsefladen og få det gemt i databasen
- getData: denne service skal bruges af bibliotek.dk til at hente brugerskabte data om en given manifestation til visning i bibliotek.dk brugergrænsefladen. Den skal også bruges af administrator til at kunne hente data fx til sletning via administrationsmodul. På sigt vil servicen også kunne bruges af den bruger der har skabt data til at hente alle de data som han har skabt (men dette vil kræve brugeridentifikation i forbindelse med afgivelse af data (createData)).
- deleteData: en administrator skal kunne slette data i databasen, eller en slutbruger skal kunne slette egne data.

- updateData: data skal kunne rettes af en administrator eller den slutbruger der har skabt dem.

Kun createData og getData implementeres i første version på bibliotek.dk. deleteData håndteres i første omgang af et administrationsmodul, så det bliver kun administrator der vil kunne slette data og ikke brugeren der har oprettet data.

Databasen udvikles i første version til at kunne rumme ratings og reviews (anmeldelser). Databasen designes så den nemt kan udvides med fx tags og comments (og det samme gælder for webservicen der udvikles).

Alle deltagere i spørgeskemaundersøgelsen mener at de beskrevne services vil være tilstrækkelige for en national service. Der er dog et par kommentarer hvor der gives udtryk for at det først vil være muligt helt at afgøre når man mere erfaring med løsningen. Det har ikke været den tekniske del der har været i fokus i workshoppen (det er primært medtaget som baggrund), og der er ikke kommet kommentarer på bloggen hertil. Den tekniske beskrivelse er til høring blandt system-leverandørerne (deadline 22. september 2008), og herefter vil det være nemmere at komme med en konklusion omkring den tekniske løsning.

Sitet (navnet på det site), hvor data er skabt, skal dog under alle omstændigheder gemmes i den nationale base.

En WSDL er under udarbejdelse til en webservice til brugerskabte data.

Udkast kan ses her: http://www.danbib.dk/docs/ucd_1.wsdl (tilrettes løbende).
