

# Projektrapport

## Udvikling af forretningsmodel for udnyttelse af digitaliserede bøger

Statsbiblioteket  
Det Kongelige Bibliotek  
Aarhus Universitetsforlag  
Publizon A/S

Projektet er støttet af DEFF

## Indhold

Resume.....	3
Præsentation af projektet.....	4
Baggrund, målsætning og succeskriterier .....	4
Delprojekt 1: Forretningsmodel og jura.....	6
1A: Projektforløbet .....	6
Aftalestruktur .....	6
Rettighederne .....	7
Forretningsmodellen: .....	7
1. B: Resultater, konklusioner og ting der skal opmærksomhed på ved lignende projekter.....	8
Delprojekt 2. Digitaliseringen.....	9
2A: Projektforløbet .....	9
Hvad skal digitaliseres? .....	9
Digitaliseringen.....	10
2.B: Resultater, konklusioner og ting der skal opmærksomhed på ved lignende projekter.....	11
Delprojekt 3: Test af søgning.....	13
Test af søgning på Statsbiblioteket: .....	13
Test af søgning på Det Kongelige Bibliotek.....	13
Sammenfatning af søgetest: .....	14
Konklusion og perspektivering .....	15
Plan for formidling.....	16
Kontaktperson .....	17
Regnskab.....	18
Bilag:.....	19
Bilag 1 – Projektansøgning .....	20
Bilag 2: Aftale med Copydan.....	26
Bilag 4 – Aftale mellem projektet og Aarhus Universitetsforlag .....	30
Bilag 5 – Notat fra advokaten .....	32

## Resume

Projektet Dansk Google har haft et langt forløb med mange udfordringer undervejs. Projekttilsagn blev givet fra DEFF i december 2007, og først i efteråret 2011 er projektet nået til vejs ende.

Planen var digitalisere alle bogudgivelser fra Aarhus Universitetsforlag, for at benytte fuldteksterne som metadata i bibliotekernes løsninger til integreret søgning. Der var 3 selvstændige formål med dette:

- A. Forretningsmodel og jura: At afprøve en forretningsmodel som gør det muligt for både forlag og biblioteker at få glæde af de samme scanninger, og få afklaret en række juridiske problemstillinger i forhold til copyright
- B. Digitaliseringsprocessen: At teste selve digitaliseringsprocessen mellem bibliotek, forlag og scanningsleverandør
- C. Anvendelse af data til søgning: At få testet hvorvidt fuldtekstdata teknisk set kan integreres i bibliotekernes søgemaskiner, og få et første fingerpeg om nytteværdien for brugerne.

Oprindeligt var det hensigten at søge de juridiske problemstillinger løst uden aftalelicenser. Imidlertid kom der en ændring af ophavsretsloven i sommeren 2008, som netop gjorde det nemmere at indgå aftalelicenser. Det blev derfor besluttet at afprøve denne mulighed. Hermed startede et langt forløb hvor en aftale blev forhandlet på plads med Copydan i foråret 2009, for derefter at ligge til godkendelse i Kulturministeriet indtil maj 2010. Fra sommeren 2010 til sommeren 2011 er den praktiske del af projektet så blevet gennemført, med scanninger af ca. 260.000 sider, som efterfølgende er blevet testet i søgemaskinerne hos Det Kongelige Bibliotek og Statsbiblioteket.

Projektet forløb langt fra som tiltænkt, men trods det langstrakte forløb, kan vi konstatere at alle projektets mål er nået. Vi har ikke fundet løsninger som uden videre kan kopieres og bringes i anvendelse i massedigitaliseringsprojekter, men vi har høstet en række erfaringer, som vi håber kan blive nyttige for andre.

Projektgruppen  
November 2011

## Præsentation af projektet

### Baggrund, målsætning og succeskriterier

Dette er beskrevet i projektansøgningen:

*Projektet vil undersøge mulighederne for at etablere en juridisk, logistisk og forretningsmæssig sund basis for at udnytte fuldtekster i forbindelse med bibliotekernes søgegrænseflade. For at gøre dette vil projektet digitalisere alle bøger fra Århus Universitetsforlag og i forbindelse med dette vil omkostninger og det retlige spørgsmål blive afklaret*

#### **Formål**

*Projektets formål er gennem faktiske forsøg at etablere en forretningsmodel for at sikre, at danske fag- og forskningsbiblioteker kan få adgang til at udnytte data i forbindelse med søgninger – på linie med det, som sker i eksempelvis Google Print.*

*Forretningsmodellen baserer sig på privatretslige aftaler mellem forlag og leverandøren af de skannede bogsider på den ene side, og leverandøren af de skannede bogsider og bibliotekerne på den anden side, og projektet skal således også afprøve mulighederne for en model der ikke indebærer generelle aftalelicenser. Udviklingen af disse aftalemæssige rammer indgår som en del af projektet.*

*Endelig er det formålet, at fremme tilgængeligheden for forlagets udgivelser, således at forlagets forfattere opnår større synlighed både i biblioteksmæssig og kommerciel sammenhæng.*

#### **Mål**

*Målet med projektet er at skanne 300.000 bogsider fra forlaget, svarende til ca. 1300 titler. De skannede bogsider OCR behandles herefter og vil herefter kunne indekseres af bibliotekerne, således at titlerne kan fuld-tekst søges. Endvidere vil image-filerne blive arkiveret som en del af den løbende digitaliseringsopgave. Det er endvidere et mål for projektet at udvikle metoder til udarbejdelse af kataloglister der sikrer at værkeres rettigheder indehaves af forlaget, samt udarbejdelse af logistikplanlægning til bibliotekets leverance af de enkelte titler, da det er projektets hensigt at benytte bibliotekets samling som kilde til scanning.*

#### **Succeskriterier/resultater**

*At alle biblioteker, som måtte ønske det, har adgang til at udnytte fuldteksterne i deres indeksering og at de i forbindelse med fremvisning kan vise information, som af brugerne opleves som relevant.*

På grund af det lange forløb, og til dels på grund af de forhindringer der viste sig undervejs, er nogle dele af projektet faldet anderledes ud end først tænkt, men alle mål er nået på den ene eller

anden måde. I det følgende vil vi beskrive projektets forløb med et afsnit for hver af de tre hovedopgaver, og for hver del give A) En beskrivelse af projektførløbet B) Resultater, konklusioner og ting der skal opmærksomhed på ved lignende projekter.

## Delprojekt 1: Forretningsmodel og jura

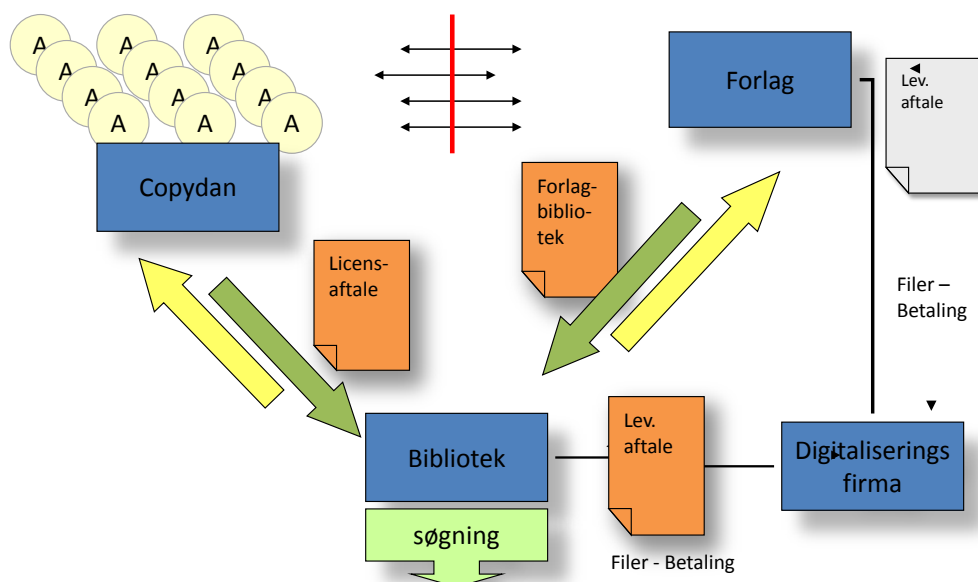
### 1A: Projektforløbet

En af de grundlæggende præmisser for denne del af projektet var at de digitaliserede filer skulle komme til gavn for flest muligt. Den oprindelige plan var at forholdet skulle reguleres ved ”privatretslige aftaler mellem forlag og leverandøren af de scannede bogsider på den ene side, og leverandøren af de scannede bogsider og bibliotekerne på den anden side.”

Da der imidlertid i sommeren 2008 blev vedtaget en ændring af ophavsretsloven, som indeholdt en ny paragraf, § 50 stk. 2, som gjorde det muligt at indgå en licensaftale med aftalelicensvirkning, dvs. at den også ville kunne omfatte rettighedshavere, som ikke er medlemmer af en rettighedshaverorganisation, hvorved rettighedsklareringen forenkles væsentligt, blev det besluttet at afprøve denne nye mulighed i stedet.

### Aftalestruktur

Det blev besluttet at indgå aftaler som skitseret i nedenstående model:



Aftale mellem Projektet og Copydan som repræsentant for autorrettighederne  
Aftale mellem Projektet og Aarhus Universitetsforlag som har udgiverrettighederne.  
Aftale mellem Projektet og Publizon om den konkrete digitalisering.

Det blev besluttet af aftalen mellem forlaget og Publizon om fremtidig udnyttelse af de digitaliserede filer – ”Lev. aftale” - ikke skulle udarbejdes i regi af projektet, men at projektet blot skulle sikre sig at de øvrige aftaler ikke forhindrer en sådan aftale. (Se bilag 5)

## **Rettighederne**

Der blev indledt forhandlinger med Copydan om indgåelse af en aftale som skulle klarere rettighederne i forhold til autorerne, dvs. forfattere, illustratører m.v. (bilag 2). Desuden blev der indgået en aftale med Aarhus Universitetsforlag (bilag 3) hvor forlaget giver sit samtykke til projektets anvendelse af deres udgivelser.

Aftalen med Copydan blev underskrevet af SB på vegne af projektet i maj 2009.

Aftalen omfatter digitaliseringen, og retten til at vise såkaldte snippets, små stumper af tekst, som indeholder søgeordet, som brugeren har anvendt. Det har ikke været drøftet med Copydan hvorvidt bibliotekerne havde haft ret til at digitalisere i medfør af Ophavsretslovens §16, så man kan ikke se aftalen som en stillingtagen til dette spørgsmål. Det viste sig umuligt at få forhandlet aftalen til at gælde alle medlemmer af DEFF. Copydan mente at det ville blive kompliceret, grænsende til det umulige, hvis alle forskningsbiblioteker, UC-biblioteker, gymnasiebiblioteker m.m. skulle have haft ret til at opbevare en kopi af de digitaliserede bøger. Efter rådføring med DEFF blev det derfor besluttet at acceptere en aftale som kun gælder for de to biblioteker i projektet. Aftalen er desuden tidsbegrænset, og skal genforhandles efter 2 år.

Et af formålene med projektet var at finde en forretningsmodel som gør det muligt for biblioteker at kunne bruge digitaliserede bøger til søgning. Det må man sige er lykkedes – i hvert fald for det konkrete projekt. Modellen med aftalelicens fungerer altså godt nok, men de konkrete vilkår i aftalen vil skulle tilpasses forholdene i hvert enkelt tilfælde. Såvel Copydan som bibliotekerne i projektet har da også eksplicit gjort opmærksom på at aftalen er ”uden præjudice” – at vilkår og priser ikke nødvendigvis kan overføres til andre projekter og aftaler.

Den utroligt lange sagsbehandlingstid kan formentlig og forhåbentlig tilskrives at det var helt nyt for alle parter at arbejde med aftalelicenserne. Man må gå ud fra at Copydans og Kulturministeriets godkendelse af fremtidige aftaler kan ske relativt hurtigt, og i så fald er det en velfungerende ordning, som også i mellemtiden er anvendt til flere andre aftaler.

## **Forretningsmodellen:**

I vores projekt stiller forlaget sine udgivelser til rådighed uden beregning. Copydan/autorerne får afregning som beskrevet ovenfor.

Fra starten har projektet været lagt sådan op at Publizon skulle scanne billigt mod til gengæld at kunne tjene penge på viderebearbejdning af de digitaliserede filer til de formål som forlaget måtte finde relevant på et senere tidspunkt. Der har været ideer fremme om udnyttelse, men endnu er ingen af dem blevet realiseret. Publizon har således endnu ikke tjent penge på sekundær udnyttelse, og har tabt penge på at deltage i projektet. Publizon ville, i tilfælde af fremtidige projekter, enten sikre sig fuld betaling for arbejdet, eller forhåndsftaler med forlaget om et vist minimumskøb til forud aftalte priser.

Man kan også stille spørgsmålstejn ved om denne model for finansiering overhovedet er anvendelig i praksis. Man bør nøje analysere hvilke parter der får hvilken specifik værdi ud af en digitalisering, og i højere grad indrette finansieringsmodellen herefter.

Forlaget har arbejdet med forskellige muligheder for at udnytte de digitaliserede bøger, men endnu ikke fundet en attraktiv model. Forlaget havde ved projektstart en forestilling om at andre akademiske forlag også ville digitalisere i stor stil, og at der kunne etableres en fælles platform, og dermed en markant afsætningskanal for akademiske e-bøger. Flere forlags deltagelse ville kunne nedbringe forlagets nødvendige investering i en platform til præsentation og evt. salg af det digitale materiale, og det større antal titler ville give mulighed for at tiltrække stor opmærksomhed fra potentielle købere. Og selv hvis der faktisk eksisterede sådanne afsætningskanaler, så er det jo også uafklaret om kunderne faktisk ville købe, At have bøgerne i digital form er altså ikke i sig selv nok til at skabe værdi for forlaget; der skal også være attraktive afsætningskanaler, som kan generere omsætning.

I løbet af projektets løbetid har forlaget besluttet at deltage i Google Book programmet, og lade Google søge i forlagets digitaliserede udgivelser. Men fordi forlaget ikke har filerne har de blot givet Google et eksemplar af alle de trykte bøger, og Google scanner dem så en gang til! Det er tankevækkende, at der måske oven i købet ikke er nogen af parterne der egentlig finder det uhensigtsmæssigt: For forlaget er det måske mere ukompliceret at levere 1000 bøger end at levere 1000 digitaliserede bøger. For Google er det måske også lettere at scanne bøgerne, og straks få det ønskede format, i stedet for at skulle bruge tid på at analysere de leverede digitale filer og konvertere til de formater som Google anvender.

Hvad angår prisen for digitalisering, så kan den måske faktisk komme ned på samme niveau som i dette projekt, også på almindelige markedsvilkår, hvis volumen er stort, og især hvis der kan benyttes eksemplarer stillet til rådighed af forlaget, som må skæres op.

## 1. B: Resultater, konklusioner og ting der skal opmærksomhed på ved lignende projekter

Det var meningen at teste forretningsmodellen, for at finde ud af om den kunne skabe værdi for alle de involverede parter:

Forlaget:

Har nu mulighed for at få adgang til alle filerne billigt, men i den anvendte model kun gennem digitaliseringsfirmaet. De digitaliserede bøger kunne genudnyttes til udgivelse som e-bøger, som book-on-demand eller andre former for udnyttelse, men det er ikke sket endnu.

Forlaget har bidraget med tid, bøger til opskæring og scanning, leverance af titellister fra økonomisystem, tjek af lister mhp. at tjekke at forlaget faktisk har udgiverrettighederne til de bøger der står på listen. Forlagets udbytte er at forlagets udgivelser og forfattere bliver mere synlige i de to bibliotekers søgesystemer, og dermed understøtter de forlagets formål: ”at fremme formidling af forskningsresultater og resultater af anden faglig virksomhed fra Aarhus Universitet”

Desuden har forlaget opnået en potentiel værdi som kan blive realiseret senere, såfremt der opstår bedre afsætningsmuligheder for e-bøger.

Digitaliseringsfirmaet:

Har fået betaling for at digitalisere, men har endnu ikke tjent penge på at viderebearbejde de digitaliserede bøger for forlaget. Firmaets konklusion er at fremtidige finansieringsmodeller i højere grad skal sikre en reel indtjening, også på kort sigt, enten ved fuld betaling af markedspris for arbejdet, eller ved forhåndsftaler med forlaget om udnyttelse.

Biblioteket:

Har fået digitaliserede filer som kan indgå i søgningen. Men værdien som middel til at forbedre søgningerne har vist sig at være svær at påvise – se afsnittet om test af søgning. Desuden har det krævet et meget stort arbejde at udarbejde lister, levere bøger, kvalitetssikre leverede digitale filer, integration i søgemaskinerne.

## Delprojekt 2. Digitaliseringen

### 2A: Projektforløbet

#### Hvad skal digitaliseres?

Hvad skal digitaliseres? Alle bøger fra Aarhus Universitetsforlag! Tilsyneladende enkle spørgsmål og formodet enkle svar. Men sådan var det ikke.

Projektet havde brug for en liste på ISBN-niveau, fordi både forlaget og Copydan har brug for at vide hvilke rettighedshavere der er knyttet til de udgivelser der indgår i projektet: Forlaget for at kunne tage stilling til om det er titler forlaget faktisk har udgiverrettighederne til, Copydan for at kunne viderebetale honorar.

Desuden skulle ISBN bruges som nøgle, således at digitaliserede filer kan matches med poster i bibliotekernes kataloger, sådan at brugernes søgninger kan lede dem hen til bøger, som bibliotekerne kan levere.

Statsbiblioteket gik i gang med at afprøve søgemetoder, for at kunne producere en liste. Der blev søgt på ISBN-nummerserier og forlagsnavn, herunder også adskillige varianter af forlagets navn. Men forlaget har også rettigheder til titler som har været udgivet under helt andre forlagsnavne og med andre ISBN-numre. Det kan være titler fra forlag som Aarhus Universitetsforlag siden har opkøbt eller er fusioneret med, eller enkelttitler som er udgivet af andre udgivere eller forlag, men som Aarhus Universitetsforlag efterfølgende har overtaget udgiverrettighederne til.

Statsbibliotekets katalog er desuden ikke primært struktureret efter ISBN, men efter titel. Det betyder at en post i katalogen kan referere til flere ISBN-numre, f.eks. grundbog og opgavebog, værker i flere bind med selvstændigt ISBN, udgivelser som har karakter af bog, men som er udgivet som et bind i en serie, evt. som tidsskrift.

Alt i alt viste det sig at være en kompliceret opgave at generere en komplet liste. For at supplere og krydstjekke bibliotekets liste blev den samkørt med oplysninger fra forlagets økonomi- og lagersystem.

Andre forhold der gjorde det svært at lave listen:

Forvirring som følge af ISBN-10 og ISBN-13

Flerbindsværker registeret som én post, med mange ISBN-numre som indgår i værket, må opsplittes til én post pr. ISBN.

Forlaget bruger ISBN som varenummer, dvs. at hvert unikt ISBN ikke nødvendigvis dækker en unik udgivelse. Et ISBN kan f.eks. dække over salgsvarianter som forskellig indbinding, sampak med andre ISBN osv.

Desuden dukkede spørgsmål op som f.eks. hvorvidt en ny trykt udgave skal scannes igen, og hvad med e-bogsudgave og papirbogsudgave?

Og hvordan identificerer man en unik bogudgivelse? Og hvad er det korrekte ISBN til en bog? Det som er trykt på bagsiden af bogen, det som er trykt i kolofonen, det som biblioteket har registreret i katalogen, det som Bibliotek.dk har registreret, det som forlaget har registreret? I langt de fleste tilfælde er alle disse forekomster af ISBN identisk for samme bog. Men undtagelserne forvirrer og tager utrolig lang tid at rede ud.

Vi ville gerne bruge poster fra Statsbibliotekets katalog, for at kunne køre lister ud med opstillingsnumre, for dermed at lette indsamlingen i magasinerne. Det var faktisk også meget nyttigt, men kostede meget ekstra arbejde med udarbejdelse af lister.

En konklusion her må være: Udarbejdelse af lister er besværligt ud fra Statsbibliotekets katalog, og udarbejdelse af (tilnærmelsesvist) *komplette* lister er utroligt tidskrævende. Hvis man vil lave tilsvarende projekter i større skala kan det være værd at overveje hvorvidt man skal forsøge at lave *komplette* lister, eller blot lave hurtige lister som fanger det meste. Skal slutdata kunne bruges til søgning, er det afgørende at bibliotekets katalog bruges som udgangspunkt. Hvis flere biblioteker skal benytte de digitaliserede bøger, skal man naturligvis også tage hensyn evt. forskelle mellem bibliotekerne. Hvis man bruger data fra forlagets systemer, eller registrerer nye data i forbindelse med scanningen, vil det være årsag til mange nye fejl i metadata, som betyder at de digitaliserede filer ikke matcher nogle af bibliotekskatalogens poster.

Til slut fik vi defineret en liste. Herefter fik forlaget listen, for at frasortere titler hvor der var tvivl om udgiverrettighederne, eller andre aftalemæssige tvivlsspørgsmål. Tilbage havde vi *masterlisten* – listen over de bøger der skulle digitaliseres. Forlaget afmærkede på listen alle de titler, som de havde på lager, og leverede et gratis eksemplar til Publizon, som kunne skæres op, og scannes i arkføder. Forlaget leverede desuden en række bøger digitalt, enten e-bøger eller pdf-filer, som havde været brugt til tryk. Disse konverterede Publizon til samme format som de scannede, sådan at de kunne indgå i søgematerialet.

## Digitaliseringen

Forlaget leverede en del bøger til scanningsfirmaet. Resten af bøgerne blev leveret af Statsbiblioteket. Digitaliseringsfirmaet blev oprettet som låner og fremfundne bøger blev scannet til udlån på sædvanlig vis af biblioteket. Ved indsamling af bøgerne i magasinet var det en fordel at vi havde trukket Statsbibliotekets opstillingsnumre med ud på vores lister. Listen blev sorteret efter opstilling før indsamling. Ved tilbagelevering blev bøgerne blot scannet og afleveret som normalt. Hele den praktiske logistik indebar selvfølgelig en smule arbejde for biblioteket, men udgjorde ikke nogen nævneværdig byrde i den store strøm af udlån.

Publizons produktion blev ikke styret ud fra masterlisten, men udelukkende ved at de selv registrerede de bøger, som blev leveret. Efter at Publizon havde leveret de digitale filer, blev det tydeligt at dette forhold gav anledning til mange uklarheder. Vi fik en liste fra Publizon over de leverede titler, men havde vi fået det, som vi havde bestilt? Det kostede en del blod, sved og tårer på begge side at få trådene redt ud, og undervejs dukkede der fejl frem, som var opstået ved den manuelle registrering.

Det er nødvendigt at man i et sådant projekt deler en fælles liste, evt. en database, for at undgå fejl og misforståelser. Så vidt muligt skal denne liste genereres på baggrund af bibliotekets katalog, dels for at sikre at fremtidige søgninger peger på materialer som biblioteket har, dels fordi bibliotekets katalogiseringspraksis formentlig betyder at bibliotekernes katalogers data er den bedst validerede version.

Desuden bør listens informationer danne grundlag for automatisk generering af filnavne og metadata i øvrigt. For Publizon spillede det ikke nogen større rolle om de brugte masterlisten eller selv registrerede bøgerne ved scanningen, men for biblioteket som modtager var det svært at danne sig et overblik over hvad der var leveret i forhold til hvad der var bestilt, og tilsvarende for forlaget, som har givet deres tilladelse til at titlerne på masterlisten kunne indgå i projektet.

Mange af problemerne omkring identitet, metadata og lister hænger faktisk sammen med at vi i projektet valgte at fravige den oprindelige plan KUN at bruge Statsbibliotekets bøger som udgangspunkt for digitaliseringen. Det er f.eks. også hvad man gør i det store norske digitaliseringsprojekt, Bokhylla.no. Vi valgte at bruge forlagets eksemplarer, som måtte skæres op, og allerede digitaliserede bøger fra forlaget – alt sammen fordi vi mente at det ville være billigere. Vi har ikke haft mulighed for at lave en egentlig cost-benefit analyse, men til fremtidige projekter må man nøje overveje om det faktisk giver en besparelse at blande forskellige typer af originaler.

Det Kongelige bibliotek fastsatte de tekniske specifikationer for de digitale filer: Scannede sider blev leveret i Tiff format, 600 x 600 dpi, og ocr-tekst blev leveret som xml-filer i TEI5 format.

Scanningskvaliteten var god, og ligeledes OCR-kvaliteten. Der er dog udfordringer med eksotiske sprog, og med bøger med blandede sprog. Det viste sig også at en hel del af Statsbibliotekets eksemplarer indeholdt brugernes noter, indstregninger osv. hvilket også i et vist omfang forstyrrede ocr-kvaliteten.

Til sidst havde vi ca. 260.000 digitaliserede bogsider!

## 2.B: Resultater, konklusioner og ting der skal opmærksomhed på ved lignende projekter

- Hvis ambitionen er at lave en *komplet* liste over forlagets udgivelser kræver det rigtig meget arbejde. I større projekter vil det være vigtigt at træffe et prioriteret valg af ambitionsniveau.
- Vigtigt at have en fælles liste eller database med titler, og at have skarpt fokus på datakvalitet.
- Undtagelser koster utroligt meget energi. Vigtigt at aftale en procedure for håndtering og afklaring.

- Scanningskvalitet og ocr-kvalitet er generelt god.

Til kommende projekter, som i stor skala skal digitalisere bøger, kan der hentes værdifulde erfaringer fra Bokhylla.no, som drives af det norske nationalbibliotek. De har bl.a. brugt et helt årsværk alene til at tilrette deres katalog og it-systemer, til at sikre konsistens mellem den fysiske bog, katalogposten og den digitaliserede tekst. Læs evt. mere på:

<http://www.nb.no/bokhylla/om/digitalisering-av-boeker-i-nasjonalbiblioteket>

## Delprojekt 3: Test af søgning

### Test af søgning på Statsbiblioteket:

I forbindelse med at vi fik adgang til fuldteksterne fra Århus Universitetsforlag har vi valgt at hæfte dem op på de eksemplarer (metadata) vi i forvejen havde i vort katalog. Vi har dermed indekseret fuldteksterne ind sammen med de relevante metadata og har gjort fuldteksterne søgbare.

Umiddelbart har det ikke udgjort noget teknisk problem, da Search på forhånd er tænkt til at kunne håndtere fuldtekster.

For brugeren opstår der dog nogle uhensigtsmæssigheder som vi ikke på forhånd havde været helt opmærksomme på.

Da vi indekserer hele teksten inkl. referencer, noter o.l. vil eksempelvis et forfatternavn i en reference få uforholdsmæssig megen vægt hvis der søges med et forfatternavn i det alm. søgefelt. Står forfatter XX i et eller flere referencer i en publikation vil en søgning på XX i den fulde tekst gøre at publikationen dukker op, rankeret relativt højt. Låneren kan derved få den opfattelse at XX er forfatter til det viste.

På samme måde vil fuldteksten til et værk rumme rigtigt mange ord, som måske ikke ligefrem er nøgleord for værket. Værket vil derfor let dukke op på resultatlisten hvis låneren søger på ord der er almindelige i fuldteksten. Da ordene ofte forekommer flere gange vil værket desuden kunne komme højt i rankeringen.

Da vi ikke har mulighed for at vise den indskannede fuldtekst, vil der ofte forekomme situationer hvor brugeren ikke kan forstå hvorfor en given post dukker op. Søgeordet/søgeordene forekommer tilsyneladende ikke i fuldposten (metadata), så hvorfor dukker det så op i søgeresultatet?

Især det biblioteksfaglige personale har fundet de ovennævnte konsekvenser meget irriterende og vi har derfor valgt at tildele de værker der har fået tilknyttet fuldtekster en ”negativ ranking” på forhånd således at de ikke rankeres så højt. De pågældende værker vil således, med mindre søgningen er ret præcis. Blive placeret ret langt nede i rankeringen.

Vi har (på grund af manglende understøttelse i vores søgemaskine) ikke haft mulighed for at vise snippets fra fuldteksten. Tilstedeværelse af sådanne ville have gjort problemet mindre, men næppe løst det.

### Test af søgning på Det Kongelige Bibliotek

Proceduren på Det Kongelige Bibliotek var:

- Et antal testposter blev indlæst i systemet ved hjælp af en File Splitter, som var blevet konfigureret i henhold til fuldtekstposternes dataformat
- De bibliografiske metadata blev normaliseret til Primo-posternes dedup-sektion

- Fuldtekstdata blev normaliseret til Primo-posternes search-sektion
- Der blev ikke normaliseret data til Primo-posternes display-sektion
- Efter indeksering kunne posterne søges frem på fuldtekstdata

Resultatet af forsøget var at

1. posterne kan fuldtekstindekseres i Primo, så fuldteksten er
  - a. søgbar
  - b. ikke vises i fuldt omfang
  - c. kun vises som kontekst for lokalisering af søgeord
2. deduplikering af fuldtekstposterne med eksisterende bogposter er pga. datagrundlaget ikke pålideligt nok til at denne metode kan anbefales

Konklusionen efter forsøget var derfor at vi vil anbefale en metode, hvor fuldtekstdata fra Publizon-posterne matches til eksisterende bogposter som berigelse af disse, hvorpå disse berigede poster fuldtekstindekseres

Metoden vil kræve en Enrichment Plug-in op på baggrund af eksisterende dokumentation fra Exlibris. Forsøg med denne metode er igangsat.

### **Sammenfatning af søgetest:**

Den samlede konklusion på test af søgning er at det teknisk set er muligt at integrere fuldteksterne. De begrænsede erfaringer med hvordan fuldteksterne påvirker kvaliteten af søgefunktionerne peger imidlertid på at fuldteksterne er svære at vægte korrekt, sådan at posterne med fuldtekst dukker op i søgeresultaterne i et omfang som giver mere værdi end støj.

En mulig forklaring kan være, at grundet at vi har så lidt fuldt digitaliseret materiale til rådighed, er det nødvendigt at give en relativt lille vægt på fuldtekstmaterialet for at søgninger på ord, som forekommer ofte i teksten ikke skal blive totalt domineret af bøger med fuldtekst. Dette forhold forventes at ændres når mere materiale er tilgængeligt for fuldtekstindeksering. Dette kan forklare, hvorfor Google Book Search fungerer godt hvorimod søgninger i dette blandede materiale med meget lidt i fuldtekstform, ikke fungerer så godt.

Muligheden for at vise en snippet synes meget anvendelig, men en mere kategorisk udtalelse vil kræve brugertest.

## Konklusion og perspektivering

Konklusionen i forhold til forretningsmodel og jura må være at aftalelicensen er en brugbar mulighed, når bare systemet der skal levere aftalen er på plads, og at det er muligt at etablere et samarbejde mellem forlag, biblioteker og digitaliseringsfirma, som kan producere brugbare digitaliserede filer. Der vil helt sikkert være brug for – og mulighed for – at forbedre mange aspekter i evt. kommende projekter.

Det viste sig at være en ret omfattende opgave at producere en (tilnærmelsesvis) komplet liste over forlagets udgivelser. Her må der formentlig prioriteres ned på komplement, hvis man vil gennemføre projekter i større skala. Det er nødvendigt at have fokus på kommunikationen mellem parterne, brug af fælles lister eller fælles database, og brug af denne til at producere metadata, for at minimere fejl undervejs. Selve scanningskvaliteten og kvaliteten af OCR må derimod siges at være temmelig god. Det er håndteringen af metadata der er en udfordring.

Siden dette projekt blev sat i gang er der sket en hel del andre ting omkring bibliotekernes søgning. Siden projektets start er Statsbibliotekets egne ca. 4.5 mio. poster med metadata blevet suppleret af flere nye kilder, bl.a. ca. 200 mio. artikler i fuldtekst, som Aarhus Universitet har købt adgang til, og desuden metadata for yderligere 700 mio. fuldtekstartikler, som udbydes af videnskabelige forlag. Bibliotekets integrerede søgning står derfor med den store udfordring at skulle give brugerne adgang til at søge med ét søgefelt a la Google, i enorme mængder af data, og stadig give et brugbart resultat til studerende og forskere.

Hvornår er det mere hensigtsmæssigt for brugeren at få en dansk fagbog blandt de første søgeresultater, end en udenlandsk bog eller tidsskriftsartikel? En søgning i Statsbibliotekets katalog på "Ludwig Wittgenstein" giver 10.485 poster. Hvordan hjælper vi bedst brugerne til at finde dem de har mest brug for? Introduktionen af indekserede danske fagbøger i fuldtekst kræver en tilpasning af søgemaskinernes algoritmer til ranking, og måske vil det først fungere godt, hvis den digitaliserede bog opmærkes yderligere, så der er mulighed for at prioritere vigtigheden af forekomster af ord efter om de f.eks. indgår i en kapiteloverskrift eller i en fodnote. Der er brug for nøjere undersøgelse af disse komplekse problemstillinger.

En anden ting der har rykket sig markant siden projektstart er anvendelsen af e-bøger. Specielt i USA har læsernes brugsmønster rykket sig voldsomt, sådan at e-bøger nu indtager en markant plads. Tendensen er også tydelig i Danmark, men meget mindre markant pt. I en akademisk sammenhæng er det dog svært at forestille sig at bøgerne ikke skulle følge tidsskriftartiklerne, og skifte til elektronisk form. Set i dét lys, er det måske ikke så interessant for forskningsbibliotekerne at bruge mange ressourcer på at få muligheden for at fremsøge papirbøger. Det virkeligt interessante ville være at kunne fremsøge bøger, og få adgang til at læse i dem med det samme, i digital form. Det virkeligt interessante digitaliseringsprojekt mellem forlag, øvrige rettighedshavere og biblioteker ville både omfatte søgning og tilgængeliggørelse i digital form.

## Plan for formidling

Projektrapport gøres tilgængelig på DEFF's website

Nyhed om projektets afslutning og afrapportering udsendes af DEFF

Rapport omtales på KB's website med link til projektrapporten på DEFF's website

Nyhed udsendes på forskbib-listen med link til rapporten

Vi vil skrive en kort artikel til Revy om erfaringerne

## Kontaktperson

Projektleder:  
Tonny Skovgård Jensen  
Statsbiblioteket  
Telefon: 40 41 77 23  
[tsj@statsbiblioteket.dk](mailto:tsj@statsbiblioteket.dk)

## Regnskab

Generelt har projektet holdt sig ret tæt til de budgetterede omkostninger. Der er tilføjet en omkostning på 2 x 10.000 kr. til Copydan, som ikke var budgetteret. Et næsten tilsvarende beløb er sparet på advokatrådgivningen. DEFF bevilgede 50.000 mindre end ansøgt, og dette beløb er sparet ved at digitalisere færre sider.

	Budget i ansøgning	Faktisk forbrug (excl. Moms)	Note
Udarbejdelse af kontrakt, advokat	50.000	26.853,42	50% af de samlede advokatomkostninger. Publizon har betalt resten.
Køb af digitaliseret tekst	300.000	258.184,00	
Honorar Copydan	0	20.231,00	10.000 + 10231
Møder og rejser SB		1.444,80	Går fra SB andel af bevilling
Projekt opgaver, Universitetsforlaget	50.000	50.000	Afregnes med forlaget, når restbevilling er modtaget.
Projekt opgaver, KB	100.000	100.000	Afregnes med KB når restbevilling er modtaget.
Projekt opgaver SB	100.000	100.000	Reduceres med det beløb som er brugt for meget.
Total forbrugt	600.000	556.713	

Brugt ud over bevilling  
- fratrækkes SB andel

6.713

### Til afregning fra DEFF:

Bevilget, DEFF		550.000,00	
Udbetalt aconto fra DEFF til SB		285.037,42	
<b>Restbevilling, til udbetaling fra DEFF ved aflæggelse af rapport og regnskab</b>		<b>264.962,58</b>	

## **Bilag:**

1. Projektansøgning
2. Projekttilsagn
3. Copydan aftale
4. Aftale mellem projektet og Aarhus Universitetsforlag
5. Notat fra advokaten om aftale mellem Aarhus Universitetsforlag og Publizon

## Bilag 1 – Projektansøgning

### DEFF ansøgning

#### Generelle oplysninger

Navn	Statsbiblioteket
Adresse	Universitetsparken, 8000 Århus C
CVR-nummer	10100682
Telefon	89 46 23 80
Fax	89 46 22 20
E-nostadresse	hcd@statsbiblioteket dk

#### Økonomisk/juridisk ansvarlig

Navn	Statsbiblioteket
Adresse	Universitetsparken, 8000 Århus C
CVR-nummer	10100682
Telefon	89 46 23 80

#### Projektitel

Udvikling af forretningsmodel for brug af data fra digitaliserede bøger

Tilskud søges til: **Projekt**

#### ***Samarbejdspartnere***

Det Kongelige Bibliotek ved Thomas Peter Sørensen,

Publizon ved Steffen Sørensen og

Århus Universitetsforlag ved Claes Hvidbak

#### ***Samlede projektudgifter***

Set fra projektet vil de samlede udgifter være 600.000 kr. idet meget store investeringer hos Publizon ikke er medtaget i budgettet men er erstattet af en købssum, som giver adgang for samtlige biblioteker i Danmark til at benytte data..

#### ***Ansøgt beløb***

600.000 kr.

## **Resume af projektet**

Projektet vil undersøge mulighederne for at etablere en juridisk, logistisk og forretningsmæssig sund basis for at udnytte fuldtekster i forbindelse med bibliotekernes søgegrænseflade. For at gøre dette vil projektet digitalisere alle bøger fra Århus Universitetsforlag og i forbindelse med dette vil omkostninger og det retlige spørgsmål blive afklaret.

## **Projektbeskrivelse og projektplan**

### **Baggrund**

De danske forskningsbiblioteker fokuserer i øjeblikket meget på udviklingen af funktionalitet til integrated search. Man ønsker ligesom hos Google og Amazon at tilbyde brugeren en nem adgang til al information gennem ét enkelt søgefelt. Bibliotekerne arbejder derfor med konstruktion af 5 indekser og intelligent søgning, så brugeren på baggrund af en simpel søgning får tilbudt lige netop de materialer, der efterspørges.

For at kunne tilbyde denne funktionalitet forsøger bibliotekerne at indhente så mange data som muligt om deres materialer. Det bedste er naturligvis at have adgang til den fulde tekst og kunne indeksere denne, sådan som Google gør med websider. Bibliotekerne arbejder på at få adgang til abstracts, indholdsfortegnelser og gerne den fulde tekst for de mange elektroniske tidsskrifter. Der er således tale om en øget synlighed og forbedret adgang til det udenlandske materiale, som i forvejen er let tilgængeligt, fordi det foreligger i digital form.

Set fra bibliotekernes perspektiv ville en digitalisering af især dansk faglitteratur således give en lang række fordele:

- Større synlighed - Dansk faglitteratur ville blive synligt på linje med den udenlandske litteratur
- Bedre søgning - Bibliotekerne kan tilbyde forbedrede søgemuligheder og præsentere kombinationer af materialer på nye måder
- Mere relevante resultater - Studerende og forskere vil spare tid ved litteratursøgninger
- Bedre materialeudvælgelse - Kvaliteten af studerendes opgaver vil øges ved, at de bliver mere opmærksomme på faglitteratur af højere kvalitet end netressourcer
- Bedre studie- og forskningsresultater – mulighed for søgning vil styrke arbejdet med selve materialet.
- Mere præcise referencer – søgningen vil give hurtigere identifikation af referencer og mulighed for at efterprøve henvisninger.

Overfor disse åbenbare fordele står de begrænsninger, som ophavsretten giver. Udgangspunktet i dansk ophavsret er, at den ophavsretlige beskyttelse først udløber 70 år efter ophavsmandens død. Det indebærer, at størstedelen af det relevante danske materiale er ophavsretligt beskyttet.

Bibliotekerne har med Ophavsretslovens § 16 en begrænset adgang til at digitalisere materiale. En mere omfattende digitalisering og on-demand tilrådighedsstillelse forudsætter indgåelse af aftaler med rettighedshaverne på eneretsbasis.

Det har hidtil vist sig at være vanskeligt at indgå aftaler om digitalisering med danske forlag. Det skyldes formentlig både det forhold, at forlagene i mange tilfælde ikke råder over de digitale rettigheder samt det forhold, at forlagene endnu ikke kan vurdere de fremtidige kommercielle perspektiver ved digitalt indhold og derfor er tøvende med hensyn til at indgå aftaler.

Spørgsmålet om de digitale rettigheder angår både de digitale rettigheder til allerede publiceret trykt materiale og de digitale rettigheder til fremtidige publikationer. Det er spørgsmål, der skal løses af forlag og forfattere.

## **Formål**

Projektets formål er gennem faktiske forsøg at etablere en forretningsmodel for at sikre, at danske fag- og forskningsbiblioteker kan få adgang til at udnytte data i forbindelse med søgninger – på linie med det, som sker i eksempelvis Google Print.

Forretningsmodellen baserer sig på privatretslige aftaler mellem forlag og leverandøren af de skannede bogsider på den ene side, og leverandøren af de scannede bogsider og bibliotekerne på den anden side, og projektet skal således også afprøve mulighederne for en model der ikke indebærer generelle aftalelicenser. Udviklingen af disse aftalemæssige rammer indgår som en del af projektet.

Endelig er det formålet, at fremme tilgængeligheden for forlagets udgivelser, således at forlagets forfattere opnår større synlighed både i biblioteksmæssig og kommerciel sammenhæng.

## **Mål**

Målet med projektet er at skanne 300.000 bogsider fra forlaget, svarende til ca. 1300 titler. De skannede bogsider OCR behandles herefter og vil herefter kunne indekseres af bibliotekerne, således at titlerne kan fuld-tekst søges. Endvidere vil image-filerne blive arkiveret som en del af den løbende digitaliseringsopgave.

Det er endvidere et mål for projektet at udvikle metoder til udarbejdelse af kataloglister der sikrer at værkernes rettigheder indehaves af forlaget, samt udarbejdelse af logistikplanlægning til bibliotekets leverance af de enkelte titler, da det er projektets hensigt at benytte bibliotekets samling som kilde til scanning.

## Målgruppe

Målgruppen for servicen er alle biblioteker, som arbejder med at udnytte tekster til etablering af bedre navigationsmuligheder og specielt bedre støtte i forbindelse med brugerens beslutning om et givent søgeresultat er relevant..

## Hovedaktiviteter

### Planlægning og aftalerammer

- Udvikling af aftalemæssige rammer
- Planlægning og indsamling af data til katalogliste
- Udarbejdelse af logistikplan til bibliotekets levering af 1300 bøger til skanning
- aftale om leverancens omfang – tekst, billeder, ToC, etc.

### Tekniske vurderinger

Udvikling af leveringsformat til indeksering

Udvikling og vurdering af den automatiserede OCR kvalitet. Her er det specielt forhold som sammenblanding af sprog i værkerne der vil give udfordringer.

Vurdering og udvikling af Image-filer til arkivering. Der benyttes SLR fotos som med 10,2 Mpixel som udgangspunkt.

Vurdering og udvikling af evt. efterbehandling af image-filer.

### Praktiske aktiviteter

- Planlægning og implementering af boglevering til ekstern leverandør (Publizon)
- Scanning og efterbehandling af image
- Tilbagelevering af bøger til bibliotek
- Levering af tekstmateriale til indeksering og image-filer til arkivering.
- aftale om leverancens omfang
- Integration af data i Primo og Summa
- Afholdelse af seminar ved projektets afslutning.

## Succeskriterier/resultater

At alle biblioteker, som måtte ønske det, har adgang til at udnytte fuldteksterne i deres indeksering og at de i forbindelse med fremvisning kan vise information, som af brugerne opleves som relevant.

## Organisationsplan (styring)

Projektet vil have en styregruppe med deltagelse af de fire aktører. Styregruppens ansvar er at sikre, at projektet lever op til mål og at aftaler og forretningsmodel udvikles på en måde, som tilgodeser såvel den private investor og bibliotekerne

Den daglige ledelse vil ske som følger:

Forhandling med forlag og udkast til aftalegrundlag: Publizon

Logistik i forbindelse med levering: SB

Sikring af formater og initiativer i forbindelse med langtidsbevaring: KB

## Ressourceplan

Der er estimeret med 2 mandmåneder til hhv. KB og SB til håndtering af de praktiske opgaver i forhold til projektet, herunder en delvis finansiering af integration af data i forhold til Primo og Summa.

Der er estimeret med 1 mandmåned til Universitetsforlaget til dækning af deres omkostninger i forhold til levering af bøger, aftaler, diskussion om visning osv.

## Tidsplan

Projektet vil operere med en forventet indskanning på 200 bøger pr. uge.

December-januar: udarbejdelse af aftaler

Januar: planlægning af skanneprocessen

Februar- juni: indskanning

September-November: Brug af data i forbindelse med Integrated search.

## Erfaringsindsamling og informationssøgning

Projektet følger ideen fra Google og Microsoft, hvor der etableres et public-private partnerskab, som hvis det skal være en succes skal berige alle deltagere. Erfaringerne fra Google ses som en succes af de deltagende partnere.

## Evalueringsplan

Evalueringen har to faser, hvor kun den første er inddraget her.

Første fase er en sikring af, at materialet er brugbart i forbindelse med integrated search. Denne del vil blive evalueret i forbindelse med Summa og Primo. Anden fase vil se på en brucevaluering for at fastslå, om en søgning baseret på tilstedeværelsen af disse ekstra data beriger deres oplevelse af søgeprocessen således at materiale bedre fremfindes.

Kun første fase er med i denne ansøgning.

### **Formidling af resultater**

Resultatet vil blive formidlet gennem en artikel i DF-Revy og gennem et seminar ved projektets afslutning.

### ***Budget***

Udarbejdelse af kontrakt: 50.000 kr (50% af 100.000 kr,)

Køb af OCR tekst til anvendelse ved integreret søgning og aflevering af billedfiler til KB og SB (til intern lagring): 300.000 kr.

Analyse af arbejdsgang og gennemførelse: 150.000 (1 mdr. til hver af KB, SB og Universitetsforlaget)

Test af data i forhold til integrated search produktet: 1 mandmåned til KB og SB.

Samlet: 500.000 kr.

## Bilag 2: Aftale med Copydan

Statsbiblioteket  
Universitetsparken  
8000 Århus C

Og

Foreningerne Copydan Tekst & Node og Billedkunst  
Østerfælled Torv 10  
2100 København Ø  
(i det følgende betegnet Copydan)

Har dags dato indgået følgende:

### **Forsøgsaftale om digital eksemplarfremsstilling af ophavsretligt beskyttet materiale**

#### Præambel:

1. Formålet med denne aftale er at give brugeren adgang til at indscanne værker fra Aarhus Universitetsforlag til brug for bogsøgning. Aftalen har aftalelicensvirkning for autorerne jf. ophavsretslovens § 50, stk. 2. Statsbiblioteket er projektleder for et projekt under DEFF, Danmarks Elektroniske Forsknings og Fagbiblioteker, som har til formål at afprøve modeller for scanning og bogsøgning som beskrevet i nærværende aftale. Brugeren defineres i denne aftale som Statsbiblioteket samt Det Kongelige Bibliotek. Statsbiblioteket er som projektleder forpligtet til at sikre at bestemmelserne i nærværende aftale overholdes.
2. Copydan Tekst & Node er fællesorganisation for samtlige kategorier af rettighedshavere, undtagen billedautorer, som kan indgå aftale med aftalelicensvirkning jf. ophavsretslovens § 50, stk. 2. Foreningen er godkendt af kulturministeriet til at indgå aftalen jf. ophavsretslovens § 50, stk. 4.
3. Copydan Billedkunst er den godkendte fællesorganisation som i henhold til individuelle fuldmagter og ophavsretslovens § 24a, kan licensere gengivelse af offentliggjorte kunstværker. Ved billeder forstås kunstværker, tegninger, fotografier, fotografier af billedkunst og kunsthåndværk, grafik mv. Foreningen er godkendt af kulturministeriet til at indgå aftalen jf. ophavsretslovens § 50, stk. 4.

På denne baggrund har parterne indgået nedenstående aftale:

#### **§ 1 Aftalens omfang**

Aftalen giver brugeren ret til digital eksemplarfremsstilling af ophavsretligt beskyttet materiale udgivet af Aarhus Universitetsforlag, som aftalt mellem Statsbiblioteket og Aarhus Universitetsforlag (jf. bilag). Digitaliseringen sker med henblik på udvikling af en funktionalitet til ”integreret søgning” og tilgængeliggørelse af de af brugeren anvendte søgeord med omgivende tekst.

#### Stk. 2

Aftalen omfatter:

- Indscanning af værker lagret i TIFF billedformat (600x600 DPI)
- OCR scanning af TIFF filerne til tekst-filer.
- Tilgængeliggørelse af korte uddrag.
- Lagring af fuldtekst.

#### Stk. 3

Det indscannede materiale skal lagres i en password-beskyttet database som kun brugerens ansatte har adgang til. Slutbrugere må således ikke have adgang til databasen. Værkerne gøres alene til genstand for indeksering og søgning samt visning af korte uddrag fra værkerne i forbindelse med søgeresultatet. ”Korte uddrag” defineres som 30 – 50 ord eller 3 – 5 linjer.

#### Stk. 4

Aftalen omfatter ikke eksemplarfremsstilling i de tilfælde hvor ophavsmanden overfor de aftalesluttende parter har nedlagt forbud mod værkets udnyttelse jf. § 50, stk. 2 sidste pkt.

### **§ 2 Indhentelse af særskilt samtykke**

Såfremt brugeren ønsker at anvende andre udnyttelsesformer der falder uden for denne aftale, skal brugeren indhente særskilt samtykke hertil og aftale vilkår herfor. Copydan vil så vidt muligt være behjælpelig med at indhente samtykke fra rettighedshaverne.

### **§ 3 vederlag**

Der skal i henhold til nærværende aftale betales et årligt vederlag fastsat til 10.000 kr. i forsøgsperioden.

Vederlaget reguleres hvert år pr. 1. januar på grundlag af udviklingen i lønindekset for den private sektor for 3. kvartal i forhold til 3. kvartal det foregående år.

#### Stk. 2

Betalingsfristen er 30 dage fra datoen for fremsendelse af faktura (faktureringsdatoen). Ved overskridelse af betalingsfristen betales morarente i henhold til renteloven.

### **§ 4 Indberetning**

Brugeren skal efter nærmere aftale med Copydan hvert år inden den 31. december fremsende en opdateret liste til Copydan over indscannede værker med angivelse af bibliografiske oplysninger i form af ISBN og oplysninger om titel, forfatter, udgiver og eventuel billedkunstner (illustrator el. lign.) til brug for Copydan’s fordeling af de indkomne vederlag.

### **§ 5 Øvrige vilkår**

Brugeren kan anvende 3.mand som fremmed medhjælp for teknisk at muliggøre den ovennævnte eksemplarfremsstilling og tilgængeliggørelse. Brugeren er ansvarlig for, at den eksemplarfremsstilling og tilgængeliggørelse af ophavsretligt beskyttet værker, brugeren foretager

eller lader andre foretage, overholder bestemmelserne i denne aftale. Brugeren er tillige ansvarlig for, at eksemplarfremsættelse, lagring og tilgængeliggørelse foregår i et teknisk sikkert miljø. Brugere er som statslige institutioner underlagt bestemmelserne i DS 484 om informationssikkerhed.

#### Stk. 2

Copydan er forpligtet til at friholde brugeren for ethvert krav, som måtte blive fremsat af danske eller udenlandske rettighedshavere, på grundlag af den eksemplarfremsættelse, som brugeren har foretaget eller ladet andre foretage i henhold til denne aftale, forudsat eksemplarfremsættelsen er sket i overensstemmelse med vilkårene i denne aftale

#### Stk. 3

Eventuelle tvister mellem parterne vedrørende denne aftale skal i første omgang søges bilagt ved forhandling. Hvis forhandling ikke fører til en løsning, kan hver af parterne indbringe tvisten for de almindelige domstole.

### **§ 6 Aftalens varighed**

Aftalen træder i kraft den 1. april 2009 og gælder i en 2-årig periode. Der er tale om en forsøgsaftale som er uden præjudice for eventuelle fremtidige aftaler mellem parterne, herunder det fastsatte vederlag.

#### Stk. 2.

Senest seks måneder inden aftalens ophør skal parterne i fællesskab forhandle en ny aftale, herunder størrelsen af vederlaget med udgangspunkt i erfaringerne med forsøgsaftalen.

#### Stk. 3.

Såfremt parterne ikke har indgået ny aftale inden udløbet af de 2 år, kan nærværende aftale forlænges, hvis parterne er enige herom.

#### Stk. 4.

I tilfælde af væsentlig misligholdelse kan aftalen opsiges skriftligt med 14 dages varsel.

#### Stk. 5

Ved aftalens ophør bortfalder såvel brugerens som Copydans rettigheder og forpligtigelser.

Dato: \_\_\_\_\_

For:

Statsbiblioteket:

---

Copydan Tekst & Node &  
Copydan Billedkunst:

---

## Bilag 3 – Aftale mellem projektet og Aarhus Universitetsforlag

Mellem

Aarhus Universitetsforlag  
Langelandsgade 177  
8200 Århus N  
(I det følgende kaldet Forlaget)

og

Statsbiblioteket  
Universitetsparken  
8000 Århus C

Aftales følgende:

### Præambel

Statsbiblioteket er projektleder for et projekt under DEFF, som skal afprøve en forretningsmodel for scanning af bøger med henblik på at lagre den fulde tekst, gøre det muligt for forskningsbibliotekernes brugere at søge i fuldteksten samt at præsentere søgeresultater som korte uddrag fra fuldteksten hvori søgeordet indgår.

Brugerne af det indscannede materiale defineres i denne aftale som Statsbiblioteket og de øvrige forskningsbiblioteker som er medlem af DEFF.

Aarhus Universitetsforlag er en erhvervsdrivende fond, der er oprettet i 1985 med det formål "at fremme formidling af forskningsresultater og resultater af anden faglig virksomhed fra Aarhus Universitet".

På denne baggrund indgås dags dato følgende

### **Aftale om benyttelse af ophavsretligt beskyttet materiale udgivet af Aarhus Universitetsforlag**

1.

I det omfang Forlaget har bemyndigelse til det, giver Forlaget Bruger en non eksklusiv ret til at gennemføre de i Bilag 1, Aftale mellem Statsbiblioteket og Copydan, beskrevne aktiviteter. Aftalen dækker samtlige udgivelser fra Forlaget, skønsmæssigt opgjort til 1300 titler.

2.

Bilag 1, Aftale mellem Statsbiblioteket og Copydan, varetager forfatterens og billedkunstnerens rettigheder, med Copydan som repræsentant for forfattere og billedkunstnere. Nærværende aftale dækker således kun Forlagets rettigheder i forhold til den beskrevne udnyttelse, og medfører i øvrigt ingen indskrænkninger i Forlagets rettigheder.

3.

Nærværende aftale træder i kraft 1.2.2009, og gælder uden tidsbegrænsning. Efter 5 år kan begge parter dog opsige aftalen med 6 måneders varsel.

4.

Statsbiblioteket indgår aftale med Publizon A/S om levering af scannede bøger i digital form. Evt. aftaler mellem Publizon A/S og Forlaget er Statsbiblioteket uvedkommende.

5.

Statsbiblioteket, som projektleder på DEFF-projektet, beskrevet i Bilag 2, afholder omkostninger til indscanning m.m. Forlaget modtager vederlag for sin deltagelse i projektet, som beskrevet i bilag 2, men modtager ikke herudover vederlag for udnyttelsen.

Sted og dato:

\_\_\_\_\_  
For Statsbiblioteket

\_\_\_\_\_  
For Aarhus Universitetsforlag

## **Bilag 4 - Notat fra advokaten**

Se næste side.

Sagsnr. 24239-0

NOTAT OM AFTALE MELLEM PUBLIZON A/S OG AARHUS  
UNIVERSITETSFORLAG VEDRØRENDE RETTEN TIL  
INDSCANNING AF VÆRKER OG LAGRING AF INDSCANNEDE  
FILER

## 1. **Beskrivelse af problemstillingen**

Publizon A/S (herefter benævnt "Publizon") har bedt Bender von Haller Dragsted om at udarbejde et notat vedrørende en aftale mellem Publizon og Aarhus Universitetsforlag (herefter benævnt Forlaget) om indscanning af Forlagets værker og lagring disse indscannede filer.

En eventuel trepartsaftale med Statsbiblioteket behandles ikke i dette notat.

Baggrunden for dette notat er, at de danske forskningsbiblioteker er i gang med udvikling af en funktionalitet, der muliggør integrated search, således at bibliotekerne kan tilbyde brugerne adgang til alt information gennem ét søgefelt.

På denne baggrund er der iværksat et digitaliseringsprojekt, der involverer Statsbiblioteket, Det Kongelige Bibliotek, Publizon, Forlaget og Copydan, der repræsenterer rettighedshaverne.

På nuværende tidspunkt er der udarbejdet en forsøgsaftale mellem Statsbiblioteket og Copydan, der giver Statsbiblioteket adgang til at indscanne værker fra Forlaget til brug for bogsøgning. Denne aftale forventes underskrevet i nærmeste fremtid.

Forlaget og Publizon ønsker som alternativ til forsøgsaftalen mellem Statsbiblioteket og Copydan at få belyst muligheden for at indgå en aftale mellem Forlaget og Publizon om indscanning og lagring af Forlagets værker.

## 2. **Scanning til lager**

De juridiske problemstillinger i forbindelse med en eventuel aftale mellem Forlaget og Publizon afhænger af den benyttelse af de indscannede og lagrede værker, som parterne ønsker at foretage.

I det omfang Publizon fungerer som eksternt lager i forbindelse med opbevaring af de elektroniske indscanninger for Forlaget, vurderer Bender von Haller Dragsted, at der ikke umiddelbart er nogle problemer forbundet hermed, eftersom det ophavsretligt beskyttede materiale ikke herved tilgængeliggøres for almenheden.

Publizon kan derfor efter Bender von Haller Dragsteds mening uden yderligere overvejelser indgå en aftale med Forlaget, hvis der udelukkende er tale om, at Publizon skal fungere som ekstern elektronisk lagerfunktion.

## 3. **Anden brug**

Såfremt Forlaget og Publizon efter, der er foretaget indscanning til lager, ønsker at benytte sig af det indscannede materiale til eksempelvis e-books eller books-on-demand, rejser denne brug spørgsmålet omkring afklaring af ophavsrettigheder.

Forlaget skal i denne forbindelse sikre sig, at der er sket en clearing af ophavsrettighederne med rettighedshaverne – forfatterne og andre rettighedshavere eksempelvis illustratører – således at udnyttelsen gennem sådan aftale er gjort lovlig, eller at rettighederne på anden vis er clearret eksempelvis gennem en selvstændig Copydan-aftale.

I denne sammenhæng er det relevant at understrege, at en overdragelse af retten til at udnytte et værk i et bestemt henseende som hovedregel ikke omfatter retten til at udnytte værket i andre henseender. Det vil sige, at Forlaget som udgangspunkt skal sikre sig, at de har ret til at udnytte det pågældende værk i en elektronisk indscannet udgave.

Opsummerende er det derfor efter Bender von Haller Dragsteds opfattelse vigtigt at være opmærksom på, at alle rettigheder er afklaret i forbindelse med anden udnyttelse end den eksterne lagerfunktion, herunder udnyttelse i form af e-books eller books-on-demand.

#### **4. Kopi til Statsbiblioteket**

Ligeledes er det relevant at se på, hvordan Publizon står i forhold til Statsbiblioteket i kraft af forsøgsaftalen mellem Statsbiblioteket og Copydan.

Det er i forsøgsaftalen mellem Statsbiblioteket og Copydan forudsat, at Statsbiblioteket kan benytte fremmed teknisk medhjælp i forbindelse med blandt andet indscanning. Publizon kan derfor af Statsbiblioteket hyres til at foretage indscanninger.

Det er derfor interessant, hvorvidt Publizon kan videregive en kopi af det allerede til Forlagets brug elektronisk indscannede værk til Statsbiblioteket, såfremt Statsbiblioteket anmoder om en kopi af et bestemt værk og henviser til de relevante bestemmelser i forsøgsaftalen mellem Statsbiblioteket og Copydan.

Retstillingen inden for dette område er uklar, eftersom der ikke foreligger litteratur eller domspraksis, som kan belyse området. Det er dog Bender von Haller Dragsteds opfattelse, at en sådan videregivelse vil være lovlig.

Der henvises i den forbindelse til, at rettighederne i forhold til rettighedshaverne allerede er clearret i kraft af den aftale, Statsbiblioteket har indgået med Copydan om indscanning af værker, og at de indscannede kopier, som opbevares hos Publizon, er lovlige.

Det er således Bender von Haller Dragsteds vurdering, at det i forhold til videregivelsessituationen ikke er væsentligt, om der er tale om en lovlig kopi (selve indscanningen) af et værk eller en lovlig kopi af indscanningen (en kopi af kopien) af værket, eftersom rettighedsgrundlaget i begge tilfælde er afklaret i

henholdsvis aftalen mellem Publizon og Forlaget og forsøgsaftalen mellem Statsbiblioteket og Copydan.

Derudover vil en betragtning omkring ressourcospild være relevant, såfremt Publizon på ny skulle foretage indscanning af værker, Publizon allerede i kraft af aftalen Publizon og Forlaget imellem har liggende i elektronisk udgave.

#### **5. Efter udløbet af forsøgsaftalen mellem Statsbiblioteket og Copydan**

Afslutningsvis skal det bemærkes, at det også er et uafklaret spørgsmål, hvad der skal ske med de indscannede værker, når prøveperioden på de 2 år udløber.

Der er ikke taget stilling til dette i forsøgsaftalen mellem Statsbiblioteket og Copydan, og der er således en risiko for, at Statsbiblioteket skal fjerne det indscannede materiale, når forsøgsaftalen mellem Statsbiblioteket og Copydan udløber.

For Publizons vedkommende vil der dog ikke være en sådan risiko, eftersom Publizon i kraft af aftalen med Forlaget og Forlagets aftale med rettighedshaverne har ret til at opbevare den elektroniske kopi, som er lavet efter aftale med Forlaget.

Hvis Publizon A/S har spørgsmål eller ønsker en uddybning af et eller flere punkter i notatet, står vi naturligvis fortsat til rådighed.

Skanderborg, den 7. maj 2009  
Bender von Haller Dragsted

Nis Peter Dall  
Advokat