

Biblioteksstyrelsen

J.nr. 324-3
10. juni 2002
LEA

Registrering af netpublikationer i Danmark

Rapport til Kulturministeriets arbejdsgruppe om bevaring af den virtuelle kulturarv

Indhold

1. Indledning

1.1 Resume

1.2 Anbefalinger

1.3 Nøglebegreber

2. Kulturarv

3. Hvorfor nationalbibliografi

3.1 Den internationale udvikling

3.2 Den danske nationalbibliografi

3.3 Pligtaflevering

3.4 Nationalbibliografi for netpublikationer og den danske forpligtelse

4. Hvordan gør vi

4.1 Opsporing, optagelseskriterier og gruppering

4.2 Omfang

4.3 Valg af registreringsformater for netpublikationer

5. Hvad foregår i øjeblikket

6. National strategi. Samspil lokalt og nationalt

6.1 Nationalbibliografi, samkatalog og genbrug

6.2 Fælles data og værktøjer

7. Afslutning

Bilag til rapporten

A. Eksempler på publikationstyper

B. Fælles data og værktøjer

1. Indledning

Formålet med denne rapport er at belyse omfanget af registreringsværdige netpublikationer på den danske del af Internettet og samtidig vise de registreringsaktiviteter, der foregår i dag.

Tanken med dette er at det skal danne baggrund for et udkast til en national strategi for registrering af danske netpublikationer, som hermed forelægges til drøftelse i Kulturministeriets arbejdsgruppe om bevaring af den virtuelle kulturarv.

Baggrunden for arbejdet er at Biblioteksstyrelsen og Dansk BiblioteksCenter medio 2001 afholdt et fælles seminar for ledelse og nøglemedarbejdere om brugen af metadata og registreringen af netpublikationer i Danmark. Der blev nedsat en fælles arbejdsgruppe med henblik på at udforme et udkast til en national strategi for registrering af danske netpublikationer. Sideløbende med dette forløb er det blevet klart at det rette forum for behandling af dette spørgsmål, er i forbindelse med Udredning 2 om bevaring af den virtuelle kulturarv, jf. Kommisorium for udredning vedrørende bevaring af kulturarven af 19. marts 2002.

1.1 Resume

Der hersker en udbredt opfattelse at Internettet rummer helt overvældende mængder af information hvilket i nogle sammenhænge er brugt som et argument for at man ikke kan foretage en systematisk registrering af netpublikationer. Det er dog samtidig en udbredt opfattelse at store mængder af ressourcerne på Internettet ikke er af værdi.

Derfor er der som en del af forarbejdet til denne rapport udført en statistisk undersøgelse af omfanget af registreringsværdige netpublikationer på den danske del af Internettet. Den viser at omfanget af sådanne publikationer er til at overskue: der er ca. 13.500 netpublikationer, hvor formentlig en meget stor del allerede er registreret, men ikke på koordineret vis. Der er en årlig tilvækst på ca. 4.000 stk. Disse tal er naturligvis behæftet med en statistisk usikkerhed og kan være enten lidt højere eller lidt lavere. Samtidig viser undersøgelsen at antallet af domæner stadig stiger, men at stigningstakten er stagnerende de sidst to år. Samtidig er ca. 60% af domænerne ikke taget i brug. D.v.s. at hverken antallet af domæner eller antallet af publikationer udvider sig eksponentielt.

Derfor bliver der med denne rapport lagt en model på bordet, som indebærer samarbejde mellem lokale biblioteker og centrale instanser om registrering af relevante netpublikationer med henblik på en sikring af nutidens og fremtidens adgang til kvaliteten på Internettet.

De statiske boglignende publikationer er allerede en del af nationalbibliografien, men de dynamiske er ligeså relevante ud fra den til enhver tid aktuelle brugsværdi af nationalbibliografien, nemlig at give borgerne adgang til information. Derfor

anbefalingen om at nationalbibliografien gøres mere dækkende på dette område i samklang med at det tilsvarende sker for bevaring af den virtuelle kulturarv.

Rapporten bygger ovenpå de erfaringer som DBC har opnået dels ved den nationalbibliografiske registrering af statistiske netpublikationer og dels ved andre registreringer, som i dag ikke er dækket af nationalbibliografiaftalen. De beskrevne metoder, udvælgelseskriterier, værktøjer m.v. er derfor en beskrivelse af dels en igangværende praksis og dels af et igangværende udviklingsarbejde.

1.2 anbefalinger

Arbejdsgruppen anbefaler følgende tiltag som en del af en national strategi for registrering af netpublikationer:

- At hullerne i den nuværende registrering af netpublikationer udfyldes, således at der sikres en løbende registrering af publikationer på Internettet af en vis værkhøjde.
- At Danmark vedstår sine internationale forpligtelser på registreringsområdet ved at nationalbibliografien udvides til også at dække de såkaldte dynamiske netpublikationer
- At nationalbibliografien for netpublikationer sikres ved en kombination af fælles registrering af publikationer på en række centrale sites udført af Dansk BiblioteksCenter og sikring af en mekanisme til indsamling af øvrige registreringer af netpublikationer uden for dette område udført af biblioteker og bibliotekernes forskellige fælles netjenester
- At etableringen af nationalbibliografi også for dynamiske netpublikationer bruges til at etablere et sammenhængende system, således at biblioteksvæsenets ressourcer anvendes rationelt
- At der sikres en sammenhæng mellem registrering og selektion til opbevaring ved at alle registrerede netpublikationer betragtes som så værdifulde at de skal bevares uanset mulige modeller med bevaring af statistiske udtrukne delmængder af det danske Internet.

1.3 Nøglebegreber

Der bliver i denne rapport anvendt en række begreber, som ikke nødvendigvis har en alment kendt betydning. I denne rapport er ordet netpublikation brugt på samme måde som i nedenstående citat fra *Bekendtgørelse om pligtaflevering af udgivne værker (nr 1041 af 17/12/1997)* for at adskille egentlige publikationer fra andet materiale på nettet.

Klip fra §1 i bekendtgørelsen:

Ved et værk forstås en afgrænset mængde af information, som må betragtes som en afsluttet og selvstændig enhed, jf. lovens § 1, stk. 2.

Et værk anses for udgivet, når eksemplarer af værket med ophavsmandens samtykke er bragt i handelen eller på anden måde spredt blandt almenheden, samt når der med ophavsmandens

samtykke tilgår almenheden meddelelse om, at eksemplarer af værket fremstilles og distribueres på bestilling, eller at værket er tilgængeligt i en database, hvorfra brugeren kan hjemtage et eksemplar, jf. lovens § 1, stk. 3 og 4.

Som udgivelse anses endvidere, jf. lovens § 1, stk. 5, eksemplarspredning blandt almenheden i medfør af bestemmelserne i kapitel 2 i lov om ophavsret.

Og for at definere dynamiske og statiske netpublikationer fra samme paragraf:

Som værker anses, jf. stk. 1, afgrænsede informationsmængder, der udgør afsluttede og selvstændige enheder. En løbende opdateret (dynamisk) enhed, f.eks. en hjemmeside på Internet, er således ikke afleveringspligtig som helhed, mens afsluttede (statiske) og selvstændige værker, der kan hjemtages, herunder fra en hjemmeside, er afleveringspligtige.

For *afsluttede (statiske) og selvstændige værker* er i rapporten brugt udtrykket *boglignende*.

2. Kulturarv

Der er en klar sammenhæng mellem at bevare den digitale kulturarv for eftertiden og at foretage registrering af netpublikationer. Men lige så lidt som på det trykte område er der tale om identiske opgaver. I runde tal bliver der årligt pligtafleveret en kvart million tryksager, hvoraf ca. 20.000 katalogiseres på nationalbibliografisk niveau og 8.000 på udvidet niveau – den såkaldte bibliotekskatalogisering.

Forholdene er helt tilsvarende for Internettet. Det er kun en lille delmængde som er egentlige publikationer og som derfor kan være berettiget til at blive registreret tilsvarende som for det trykte materiale.

Der foregår i dag en omfattende registrering – se afsnit 5 – af kvalitetsvurderede netpublikationer i bibliotekerne. Når nærværende rapport omhandler registrering af netpublikationer, er det således ikke et helt nyt projekt, men i forhold til bevaring af den digitale kulturarv handler det i høj grad om koordinering af igangværende initiativer.

Det vil næppe være muligt at have kopier af hele det danske Internet fra hvert eneste sekund året rundt. Derfor vil der under alle omstændigheder være tale om en selektiv bevaring. Kravene til en selektiv bevaring er naturligvis også en generel repræsentativitet, men det må også være nødvendigt at sikre at de egentlige publikationer bliver bevaret ikke blot som et tilfældigt udvalg, men også som et bevidst valg. Der skal derfor ske et samspil mellem aktiviteten 'bevaring' og aktiviteteten 'registrering' ved at alle registrerede publikationer bliver bevaret.

For de trykte publikationer gælder til en vis grad det omvendte, hvor bevaringen er et underlag for registreringen, idet Det Kongelige Bibliotek videregiver ikke nationalbibliografisk registreret materiale modtaget af pligtafleveringen til DBC.

Registreringen skal naturligvis også være med til at sikre adgangen til det registrerede materiale, hvad enten det alene findes i arkivet eller fortsat er tilgængeligt på det frie net. Et eksempel på dette er Det Administrative Biblioteks projekt startende denne sommer med opbevaring og formidling af statslige netpublikationer.

3. Hvorfor nationalbibliografi

Nationalbibliografien er rygraden i og grundlaget for al anden bibliografisk virksomhed. Nationalbibliografien er defineret som en fortegnelse over dokumenter der udgives eller er blevet udgivet i et land.

Nationalbibliografien er den nationale hukommelse. Der er i Danmark ikke andre end de nationalbibliografiske producenter, Dansk BiblioteksCenter og Det Kongelige Bibliotek der har til opgave at sikre at dansk produceret viden og oplevelse der er offentligt publiceret i trykt form eller på nettet, registreres og tilgængeliggøres. Både for nutiden men ikke mindst for eftertiden.

De værker / publikationer der produceres af danske virksomheder, myndigheder og kunstnere og som kun sendes ud på Internettet, bliver kun sikret hvis der er en instans der har til opgave at foretage den systematiske registrering der skal foregå kontinuerligt hvis alt hvad der skønnes væsentligt skal repræsenteres.

Der skal sikres søgemæssig adgang, dvs. at det er muligt at genfinde titler, forfattere og emner på grundlag af den bibliografiske registrering. Men også at sikre brugerne adgang eller tilgang til materialet. Dette understreger den naturlige sammenhæng mellem registrering og pligtaflevering.

Man kan ikke alene forlade sig på at selv gode søgemaskiner påtager sig den forpligtelse at arkivere og tilgængeliggøre dokumenter og internetsider. Formålet med søgemaskinen er et andet, nemlig at skabe adgang til de internetsider der på ethvert givet tidspunkt er tilgængelige.

De kulturbevarende institutioner har både interesse i og forpligtelse til at sikre adgangen til de kvalitetsdokumenter der ellers forsvinder fra Internettet.

Herudover ligger der i de internationale aftaler, vedtaget i IFLA og UNESCO-regi, den hensigtserklæring at der skal sikres brugerne adgang til den information og de dokumenter de har behov for, ikke bare nationalt men også internationalt.¹ Det ideelle grundlag for at det kan lade sig gøre, er at de enkelte nationer vedligeholder deres nationalbibliografi på grundlag af de vedtagne internationale anbefalinger.

Ambitionen er, sagt på en anden måde, at registrere al vidensproduktion og intellektuel produktion i tekst, lyd og billed som det kommer til udtryk inden for en nations geografisk afgrænsede område (ofte dækkende et sprogområde). Ofte er værker udgivet på andre sprog eller i andre lande med tilknytning til nationen også medtaget i nationalbibliografien.

¹ UAP: Universal Availability of Publications (IFLA-program)

3.1 Den internationale udvikling

De internationale rekommandationer og resolutioner fra UNESCO og IFLA vedrørende nationalbibliografi har grundlæggende været gennemgået og behandlet ved tre større konferencer efter 2. Verdenskrig.

I 1950 iværksatte Unesco det internationale bibliografiske koordineringsarbejde, og "*The Unesco Conference on the Improvement of Bibliographical Services*" udmøntedes i en anbefaling af at nationalbibliografien bør indeholde registreringer og indekseringer af en række dokumenttyper: en bogfortegnelse, en periodicaindex, fortegnelser over kort og atlas, over musikaler, av-materialer, offentlige tryksager, periodica, forlag, boghandlere og endelig en vejviser over biblioteker, lærde selskaber og andre informationsproducerende institutioner.

Det anbefaledes herudover at der etableres en nationalbibliografisk tjeneste, at nationalbiblioteket opbevarer dokumenter, og at nationerne lovgiver om pligtaflevering.

I 1977 fulgte UNESCO og IFLA op på anbefalingerne fra 1950, og efter konferencen dette år blev det "*Guidelines for the National Bibliographical Agency and the National Bibliography*" der satte den nye standard for nationalbibliografisk virksomhed.

Der blev i 1977 vedtaget retningslinier om at, nationalbibliografien har de praktiske formål at

- tilfredsstille nationale informationsbehov
- være værktøj til udvælgelse og anskaffelse af materiale
- være værktøj til litteratursøgning
- fungere som model for bibliotekernes katalogiseringer

Det blev samtidig understreget at de praktiske formål kræver et vist tempo i registreringen.

Mere overordnet lagde man desuden vægt på at nationalbibliografien har en historisk funktion, idet den afspejler landets sociale, kulturelle og økonomiske udvikling.

Ved den seneste opfølgning på anbefalingerne, på *International Conference on National Bibliographical Services* i København 1998, arrangeret af IFLA, blev anbefalingerne revideret med henvisning til udvikling af informationstjenester på Internettet og den dynamiske publikationsform der er blevet en realitet som følge af internet- og web-teknologiudvikling.

Den væsentligste ændring i forhold til de tidligere retningslinjer gælder nationalbibliografiens dækning af dokumenttyper der i 1998 udvidedes til en definition, der omfatter hele den nationale produktion, uanset medie, dvs. også netpublikationer.

Samtidig blev kravet om effektiv, hurtig registrering, samt at de bibliografiske indførsler i nationalbibliografien bør være baseret på internationalt anerkendte standarder understreget.

3.2 Den danske nationalbibliografi

Dansk bogfortegnelse

Dansk bogfortegnelse blev grundlagt af forlæggeren G.E.C. Gad i 1851. Den dækker den danske bogproduktion siden 1841. Gad stod for såvel udarbejdelse som udgivelse indtil 1945 hvor udarbejdelse og redaktion blev overtaget af Folkebibliotekernes Bibliografiske Kontor, det nuværende Dansk BiblioteksCenter.

Indtil 1976 stod Gads forlag stadig for udgivelsen som samtidig med at produktionen blev omlagt til edb også overgik til Bibliotekscentralen (nu Dansk BiblioteksCenter) således at udarbejdelse og udgivelse af *Dansk Bogfortegnelse* blev samlet.

Øvrige dele af nationalbibliografien

Bogudgivelser er som nævnt registreret i *Dansk Bogfortegnelse* fra 1841 og frem. I *Dansk Bogfortegnelse* er registreringsmetoden udvidet undervejs, men grundlæggende er indholdet danske bogudgivelser.

I løbet af 1900-tallet kommer en række fortegnelser over øvrige materialetyper til: 1915 *Dansk Tidsskriftsindex*, 1940 *Aviskronikindex*, (1979: sammensmeltes de to fortegnelser til *Dansk Artikelindex* og *Dansk Anmeldelsesindex*), 1931 *Dansk Musikfortegnelse*, 1950 *Dansk Tidsskriftsfortegnelse* (1970: ændret til *Dansk Periodicafortegnelse*), 1972 *Dansk Kortfortegnelse*, 1982 *Dansk Lydfortegnelse*, 1984 *Dansk Billedfortegnelse*, 1987 *Danske Musikoptyagelser*, 1996 *elektroniske publikationer i selvstændig fysisk form*, 1998 *Statistiske Netpublikationer*.

Tilsammen udgør disse fortegnelser den danske nationalbibliografi².

I 1991 indgik den danske stat ved Kulturministeriet aftale om udarbejdelse af nationalbibliografien med Dansk BiblioteksCenter.³

Fra 1983 har den samlede løbende nationalbibliografi været online tilgængelig i en fællesbase med lidt varierende indhold, fra 1994 kaldet DanBib, fra 2001 herudover offentlig tilgængelig på web-adressen www.bibliotek.dk.

² Den ajorførte liste over fortegnelser der indgår i nationalbibliografien, findes i danMARC2-formatet www.kat-format.dk

³ Nationalbibliografiaftalen

3.3 Pligtaflevering

Fra 1927 pålægger en pligtafleveringslov forlagene af aflevere bogtrykte værker til nationalbiblioteket.

Fra 1998 omfatter den danske pligtafleveringslov herudover pligtaflevering af audiovisuelle medier, fonogrammer og elektroniske publikationer i fysisk form samt statiske netpublikationer.

3.4 Nationalbibliografi for netpublikationer og den danske forpligtelse

Som det fremgår ovenfor er der i løbet af 1900-tallet iværksat produktion af fortegnelser over en lang række dokument- og materialetyper, mens der først i 1998 blev indført pligtaflevering af de nye medier.

I pligtafleveringsloven fra 1998 indgår at statiske internetdokumenter skal pligtafleveres. Det betyder i praksis at de anmeldes på en elektronisk blanket, samt at Det Kongelige Bibliotek og Statsbiblioteket arkiverer en udgave af værket på en separat server som skal sikre eftertiden adgang til værket, også når det eventuelt er ændret eller helt forsvundet fra det 'levende' net.

Vidensproduktionen på Internettet ændrer dog kontinuerligt karakter, og en sikring af den reelle produktion er ikke gjort med en registrering af de statiske publikationer alene.

Også indsamlings- og registeringsmetoder for de dynamiske netpublikationer må fastlægges, således at disse kan sikres for eftertiden. Ikke kun de bibliografiske data, men også reel adgang bør indgå i overvejelserne, såfremt man ønsker at leve op til den nationalbibliografiske forpligtelse.

Det her fremførte fører til den konklusion eller anbefaling at hvis Danmark skal kunne leve op til sine forpligtelser på det nationalbibliografiske område, må der tages skridt til en forsvarlig sikring også af den digitale kulturarv, herunder især den videns- og kulturproduktion der publiceres på Internettet. Der skal en egentlig systematisk overvågning til før man kan tale om en nationalbibliografi for de netbårne dokumenter.

Dette betones ligeledes i de internationale anbefalinger der opstiller det grundlæggende krav til nationalbibliografien at den på dækkende vis skal omfatte den løbende intellektuelle produktion, herunder også internetpublikationer. Dette vedrører ikke kun statiske afsluttede værker, men i lige så høj grad de dynamiske udgivelser der indeholder unik, nationalt produceret, information.

4. Hvordan gør vi

For publikationer på Internettet gælder der i sagens natur i nogle tilfælde andre forhold end for det trykte materiale – og for andre igen er der i praksis ingen forskel. Der er i begge tilfælde tale om en lille delmængde af samtlige publiceringer som falder inden for de nationalbibliografiske optagelseskriterier. Men hvor definitionerne for det trykte materiale har mange år bag sig, så er overførsel af den samme model til netområdet en relativ ny ting. Men i begge tilfælde er formålet helt overordnet at sikre adgang til materiale af værdi.

4.1 Opsporing, optagelseskriterier og gruppering

Hvordan afgøres om en given informationsmængde på nettet er en selvstændig publikation?

Et værk opfattes som en selvstændig netpublikation hvis værket kan siges at fungere som en logisk afgrænset informationsenhed der klart er adskilt fra anden information. Hertil kræves at 1) publikationen er udstyret med egen titel og 2) udgiver/forfatter rent teknisk/navigatorisk har gjort det nemt for læseren at 'færdes' inden for værkets rammer.

Hvordan afgøres om en netpublikation bør registreres i nationalbibliografien? For netpublikationer gælder de samme overordnede optagelseskriterier som for de traditionelle materialetyper hvilket i praksis betyder at kun en lille del af den samlede mængde eksisterende publikationer skal registreres. Optagelseskriterierne skal generelt sikre at det der optages, indeholder en vis informationsværdi.

Optagelseskriterierne for netpublikationer omfatter bl.a. følgende:

- Publikationer der har kommerciel, intern, meget lokal eller privat karakter, optages ikke⁴
- Publikationer skal have et omfang svarende til det omfangskriterium der gælder for trykte publikationers optagelse i Dansk Bogfortegnelse
- Publikationer hvis karakter sandsynliggør at de vil forsvinde fra nettet, optages ikke
- Publikationer der oprindeligt er kommet i anden form, f.eks. som bog i 1836, og som nu - i ubearbejdet form - er digitaliseret og lagt på nettet, optages ikke. Synspunktet er at registrering af nye originale værker skal prioriteres i den løbende nationalbibliografi.

Forskellen på de traditionelle materialetyper og netpublikationer er at den informationsmængde der bør frasorteres på nettet, er forholdsvis meget større. Det skyldes at det er nemmere og billigere at publicere på nettet, og at det generelle fravær af et redaktionelt filter har indflydelse på indholdet og kvaliteten af indholdet.

⁴ Se nærmere i Indoreg 1+2 med udførlige optagelseskriterier som kan ses her
<http://purl.dk/rapport/html.dk/>
<http://www.dbc.dk/udvikling/indoreg/>

Optagelseskriterierne for netpublikationer kan overtages fra de traditionelle materialetyper for så vidt angår de publikationer der ligner/afløser bøger og andre fastformsudgivelser.

Det er i praksis ikke så vanskeligt at vurdere bestandigheden af netpublikationer. Således blev der i 2001 ved URL-check af de 4.000 fortrinsvis nationalbibliografiske registreringer der indtil da var foretaget på DBC, kun konstateret ca. 20 (varigt) forsvundne netpublikationer. Der er således tale om en høj grad af stabilitet for registrerede netpublikationer.

Hvad angår hjemmesiden som publikationsform, er følgende retningslinjer DBC's praksis:

1. Optagelse af så overordnede enheder som muligt. Det gælder hjemmesider og databaser med et relativt homogent præg
2. Personlige hjemmesider og firmahjemmesider udelades som hovedregel.
3. Hjemmesider for offentlige institutioner udelukkes hvis de kun handler om institutionen selv eller kun virker på lokalt plan.

4.2 Omfang

Til grund for vurdering af omfanget af det årlige antal registreringsværdige netpublikationer er foretaget optælling ud fra en stikprøve på 2000 tilfældigt genererede subdomæner (second level domains) fra DK-domænet. Med ovenfor nævnte optagelseskriterier som redskab har fem DBC-medarbejdere i januar og april gennemgået en stikprøve på 2.000 (ca. 1/2 %) af de i alt 351.792 domæner fra hostmaster.dk.⁵

Forud for denne optælling lå dels en statistisk model og dels en forundersøgelse på et begrænset antal domæner. På baggrund heraf blev det vurderet at et grundlag på 2.000 tilfældigt genererede subdomæner, ville give et rimeligt fingerpeg på det samlede omfang af relevante registreringsenheder. Til sammenligning kan nævnes at folketingsmeningsmålinger som regel bygger på færre interview end 2.000 i forhold til en mere en ti gange så stor mængde af vælgere.

Selve optællingen foregik ved for hvert tilfældigt udtrukket subdomæne at søge på navnet med og uden foranstillet www, dvs. på <subdomæne>.dk og www.<subdomæne>.dk.

Undersøgelsen viste at en forbavsende stor del af subdomænerne ikke var aktive, dvs. de tilhørende websider var utilgængelige, ikke eksisterende eller meget mangelfulde (f.eks. reserveret til eventuel fremtidig brug). Der var i alt 1.178 ubrugte domæner af de 2.000. Heraf var 355 reserverede, 148 tomme eller under

⁵ Se nærmere i *Registrering af netpublikationer, estimering af mængder*. Rapport publiceres juni/juli 2002.

opbygning, 144 med automatisk viderestilling til anden adresse og endelig 531 som slet ikke kunne findes.

En særlig drilsk problemstilling er at flere subdomæner kan henvise til de samme sider. Dette kan foregå ved viderestilling fra det ene domæne til det andet, eller ved at flere subdomæner mappes sammen i nameserveren. I den udstrækning det har været muligt at gennemskue, er der korrigeret for denne fejlkilde i undersøgelsen.

Resultatet blev at 46 domæner havde et relevant registreringsværdigt indhold og i alt 76 publikationer blev udpeget som registreringsværdige. Heraf var 45 publikationer boglignende og den type har været en del af nationalbibliografien siden 1998. 31 dynamiske publikationer i form af databaser og hele hjemmesider blev fundet registreringsværdige. Den årlige tilvækst i subdomæner har de to sidste år været på godt 100.000, og tilvæksten i det første kvartal af 2002 er 27.200.

Holder denne undersøgelsestendens - og den er blevet bekræftet i to undersøgelser - betyder det at der aktuelt på .dk områdets 351.792 subdomæner (pr. 1. januar 2002) er der ca. 13.500 enheder registreringsværdige til nationalbibliografien, med en anslået årlig tilvækst på ca. 4.000 enheder. Disse tal er naturligvis behæftet med en statistisk usikkerhed og kan være enten lidt højere eller lavere. Hertil kommer supplerende enheder fra gamle domæner.

Den årlige tilvækst på boglignende/statiske netpublikationer som siden 1998 har været en del af nationalbibliografien, vil i den aktuelle situation være estimeret til ca. 2.500 enheder, mens tilvæksten af dynamiske netpublikationer som endnu ikke er omfattet af nationalbibliografien, estimeres til ca. 1.500.

I 2001 registrerede DBC 2.300 statiske netpublikationer, hvilket med en snusfornuftig (men ikke videnskabelig statistisk) vurdering indikerer at tallet på 2.500 er i underkanten. Det ville betyde at DBC's nuværende registrering, som bygger på pligtafleveringen og opsøgende arbejde, skulle dække mere end 90% af de statiske netpublikationer.

4.3 Valg af registreringsformater for netpublikationer

Nationalbibliografien for de traditionelle medier katalogiseres efter gældende katalogiseringsregler i danMARC2-formatet. Den praksis der har udviklet sig⁶ for netressourcer er imidlertid at disse i vid udstrækning registreres i Dublin Core-formatet, normalt forkortet DC.

En registrering i DC med en god emnebehandling og kontrollerede navne har for mange netressourcers vedkommende en lige så god funktionalitet som en enkel registrering i danMARC2.

⁶ hos Dansk BiblioteksCenter, i Danmarks Elektroniske Forskningsbibliotek og i de forskellige netjenester

Hovedforskellen er at der ikke følges strikse regler for hvorfra navne og forlag citeres ved registrering i Dublin Core. For de dynamiske netressourcer gælder at disse oplysninger ofte skifter plads og at der ikke er etableret en fælles praksis hos dataproducenterne.

Valg af danMARC2

Det anbefales at fortsætte med at registrere de statiske netpublikationer i danMARC2-formatet. De statiske netpublikationer er ofte parallelle udgaver af de publikationer der allerede i fastformsudgaver er registreret til nationalbibliografien. Nettidsskrifter med ISSN-numre skal fortsat registreres i danMARC2. Artikler fra disse tidsskrifter registreres indtil videre i danMARC2.

Der katalogiseres i henhold til Katalogiseringsreglernes kapitel 9 om elektroniske materialer⁷ med enkelte justeringer.

Valg af Dublin Core

Alle andre typer af netpublikationer anbefales forsøgsvis registreret i DC: databaser, hjemmesider, enkelte artikler/samlinger af artikler på hjemmesider, cd-tracks/skæringer, fragmenter fra hjemmesider.

⁷ http://www.kat-format.dk/kat_regl/Kat2.1f2.htm

5. Hvad foregår i øjeblikket

Selvom kun en mindre del af publikationerne på det danske Internet er omfattet af pligtafleveringsloven, sker der en ganske omfattende registrering af disse ressourcer i biblioteksregi. Behovet for at denne registrering foretages er stort, da internetkilder udgør en stadig større del af bibliotekernes formidlingsarbejde. Det er ikke afgørende om svaret på et spørgsmål kan findes i en bog, et tidsskrift eller i en publikation på nettet.

Udover registrering på linksamlinger mv. på det enkelte bibliotek, foregår der p.t. række projekter og aktiviteter, hvoraf kan nævnes:

Projekt/aktivitet	Antal registreringer
DanBib	
DBC's registrering af statistiske netpublikationer (en del af nationalbibliografien)	3.600
DBC testregistreringer af andet materiale	1.050
DBC indeksering af artikler i danske aviser og tidsskrifter	4.700
DBC indeksering af anmeldelser i danske aviser og tidsskrifter	2.200
Netkat: DBC registrerer på forslag af biblioteker	500
Indberetninger fra biblioteker	20.000
DanBib i alt	32.000
FNG – Folkebibliotekernes Netguide.	4.100
BibHit - Registrering af enkelte publikationer	8.100
DotBot – med særligt henblik på børn	1.300
Handelsskolernes Internetvejviser	1.350
DEF fagportaler:	
Bizigate - erhvervsøkonomisk portal	4.500
DVM - Det Virtuelle Musikbibliotek	4.000
Energiportalen – den energiteknologiske fagportal	200
Food-i - en vejviser til fødevarer og ernæring	550
Klinisk Information	800
TransPortalen	350

De forskellige tjenester har forskellige dækningsområder og optagelseskriterier, således at tallene ikke er direkte sammenlignelige. F.eks. rummer DVM mange kilder som ikke ville være optagelsesberettigede til nationalbibliografien.

Det skal understreges at der findes projekter/aktiviteter udover disse og at der stadig kommer flere til. Der er således fem nye fagportaler på vej i DEF-regi.

Der er endnu ikke etableret et egentligt samarbejde om disse registreringer, hvorfor der i et vist omfang er tale om overlap. De to første aktiviteter udført af

DBC dækker danske publikationer, mens de øvrige i forskelligt omfang dækker både danske og udenlandske kilder.

En del af bibliotekernes arbejde er at skabe adgang til bibliotekernes ressourcer, hvilket også sker ved skabelse af primært materiale. Eksempler er BogWeb med anmeldelser, artikler, nyheder, debat og klummer med relation til bøger, forfattere, læsning, andre medier, kultur og IT – samt ForfatterNet med omtale af knap 400 nulevende danske skønlitterære forfattere.

Det er derfor ikke muligt umiddelbart at give et tal for hvor stor en del af de registreringsværdige danske netpublikationer, som er dækket af disse 57.000 registreringer. En stor del er udenlandsk materiale ligesom der er tale om en vis mængde af dobbeltregistreringer. Men et forsigtigt skøn giver at en meget stor del af de registreringsværdige netpublikationer allerede er registreret.

Der er en række forskelle i de tekniske formater der anvendes. Derfor har Biblioteksstyrelsen i samarbejde med Dansk BiblioteksCenter fastlagt et udvekslingsformat for både beskrivende metadata og administrative metadata. Formålet er dels at sikre en samlet brug af de offentlige ressourcer der anvendes til registrering af netpublikationer og tilsvarende at udvide omfanget af genbrug, således at de offentlige ressourcer bruges mere rationelt.

6. National strategi. Samspil lokalt og nationalt

6.1 Nationalbibliografi, samkatalog og genbrug

Der bliver verdenen over ofret mange millioner Euro, dollars og også kroner på at gøre publikationer på Internettet tilgængelige for publikum. Dette gælder også i Danmark. Her tænkes ikke på de kommercielt baserede søgemaskiner, men primært på biblioteker m.v.'s arbejde med fagportaler og forskellige nationale løsninger såsom FNG, DEF fagportaler, BibHit og en række andre netjenester. Informationen på nettet er efterhånden blevet en afgørende del af enhver informationsformidling.

En afgørende forskel mellem trykte publikationer og netpublikationer er at den mulige tilgængelighed er en anden for netpublikationer. Mange trykte publikationer kommer kun i begrænset omfang til bibliotekernes kundskab. Hvis de pligtafløves indgår de i svært tilgængelige småtrykssamlinger, mens alle netpublikationer er umiddelbart tilgængelige. Den massive indeksering som de forskellige søgemaskiner står for, indebærer at publikationer uden for oplagte publiceringssteder bliver mulige at finde. Men omfanget af registreringer gør at de nemt forsvinder i mængden.

Det er en oplagt opgave for Biblioteksstyrelsen at medvirke til en national strategi til sikring af den mest rationelle registrering af registreringsværdige internetpublikationer. På den ene side er det en nationalbibliografisk opgave helt på linie med den nationalbibliografiske opgave med hensyn til trykt materiale - på den anden side medfører nettets natur at arbejdet må gribes an på en anden måde.

De statiske boglignende publikationer er allerede en del af nationalbibliografien, men de dynamiske er ligeså relevante ud fra den til enhver tid aktuelle brugsværdi af nationalbibliografien, nemlig at give borgerne adgang til information. Det indebærer at registreringen skal holdes ajour, og det gælder både de statiske og dynamiske netpublikationer. I begge tilfælde skal eksistens løbende kontrolleres, mens registreringen skal være på et beskrivelsesniveau, således at den normalt ikke løbende skal justeres. Ikke længere aktuelle registreringer vil dog fortsat have værdi ikke blot i forhold til materiale arkiveret af pligtafløversbibliotekerne, men også i sig selv som en oversigt over hvad der blev publiceret på nettet på et givet tidspunkt.

En løbende gennemgang af alle danske sites for at lokalisere registreringsværdigt materiale vil være uden sammenhæng med udbyttet og vil heller ikke være økonomisk forsvarlig. En systematisk nationalbibliografisk indsats i forhold til Internettet må derfor basere sig dels på identificering og løbende overvågning af publiceringstunge sites og dels på hvad der anmeldes frivilligt og som led i pligtafløringen. Håndteringen af dette materiale ligger i forlængelse af det arbejde som Dansk BiblioteksCenter i forvejen udfører for staten.

Da en sådan systematisk overvågning ikke kan dække hele det danske net, kan bibliotekerne i deres daglige informationsformidling ikke gå ud fra at dette er

tilfældet. Bibliotekerne anvender derfor mange andre værktøjer til at afsøge større dele af nettet og lokaliserer på denne måde mange relevante publikationer. Supplementet til den systematiske overvågning er derfor en løbende indsamling af de ikke-registrerede publikationer som bibliotekerne finder som led i det daglige arbejde. Der er p.t. to projekter som systematisk arbejder i denne sammenhæng: BibHit⁸ og Netkat⁹.

Med etablering af en egentlig nationalbibliografisk dækning som er tilgængelig i en fælles base, vil øvrige netjenester kunne arbejde mere målrettet i forhold til de dele af nettet, som ikke systematisk overvåges, og dobbeltregistreringer kan undgås.

6.2 Fælles data og værktøjer

Et samarbejde nationalt og lokalt bygger i første række på et fælles udvekslingsformat, som nu er defineret i et samarbejde mellem Biblioteksstyrelsen og DBC. Hertil kommer fælles værktøjer under udvikling i DBC's regi til understøttelse af en rationel arbejdsproces hvor dobbeltarbejde og dobbeltudvikling minimeres.

Kernen i de fælles værktøjer er fælles database - metadatapoolen - indeholdende alle de af biblioteksfællesskabet udarbejdede registreringer af netressourcer. Metadatapoolens rolle for netpublikationer kan til en vis grad sammenlignes med DanBibs for traditionelle medier. Årsagen til at Metadatapoolen fremstår som en selvstændig database og ikke bare som en del af DanBib, er primært valget af format.

Dataudveksling foregår helt analogt til DanBib. Hvor posterne i DanBib udveksles i danMARC2 indpakket i rammeformatet ISO 2709, sker udvekslingen af poster i metadatapoolen i Dublin Core (se afsnit 4.4). Denne analogi betyder at det meste af infrastrukturen fra dataudvekslingen med DanBib kan genbruges - såvel dataudveksling i form af filer som via Z39.50.

Hertil kommer værktøjer til høst og selektion (fremsøgning af relevante netsteder), herunder værktøjer som understøtter en traditionel overvågning af et udvalg af sites, men også værktøjer som ved brug af agentteknologi fremfinder relevante sites.

Endelig værktøjer til automatisk indeksering og klassifikation - ud fra indholdet af netpublikationerne - samt værktøjer til overvågning af netsteder (URL-check) og til genfindning (URN-administration).

⁸ Se nærmere: <http://www.bibhit.dk/info/kort.htm>

⁹ Se nærmere: <http://www.hilbib.dk/netkat.htm>

7. Afslutning

Nærværende rapport er udarbejdet på opdrag af Biblioteksstyrelsens og Dansk BiblioteksCenters kontaktudvalg og i udarbejdelsen har fra DBC deltaget Kirsten Waneck, Randi Diget Hansen og Tommy Schomacker og fra Biblioteksstyrelsen Charlotte C. Pedersen og Leif Andresen.

Bilag A

Eksempler på publikationstyper

Eksempler på nationalbibliografisk relevant indhold

Redegørelse om patientrettigheder i forbindelse med indførelse af Elektroniske Patient Journaler (EPJ) m.v. April 2001 –
www.sundhedsministeriet.dk/publika/epj2001/index1.htm
som tydeligt udgør en enhed, skønt værket ligger i en række enkelte filer.

Finansiell ordbog – <http://www.danskebank.dk/FinansiellOrdbog>
en database der afløser en trykt udgave

Se tre typiske - ret foranderlige - dynamiske publikationer:

Vejret de kommende dage - <http://www.dmi.dk/vejr/index.html>

Sundhedsvæsenets Patientklagenævn – <http://www.pkn.dk>

Vil du vide mere? – <http://www.vilduide.dk/forside.asp>

Selv om hele indholdene henholdsvis dele af indholdet udskiftes, falder de nye informationer stadig inden for den oprindelige emnebeskrivelse af publikationerne.

Eksempler på ikke relevant indhold (i forhold til nationalbibliografien)

Familien Lohmann-Hansens hjemmeside – <http://lohmann-hansen.dk/>
P. V. Mobiltoilet - <http://mobiltoiletter.dk/>
Disse sider er typiske eksempler på de mange sider der har indhold, men som falder udenfor optagelseskriterierne.

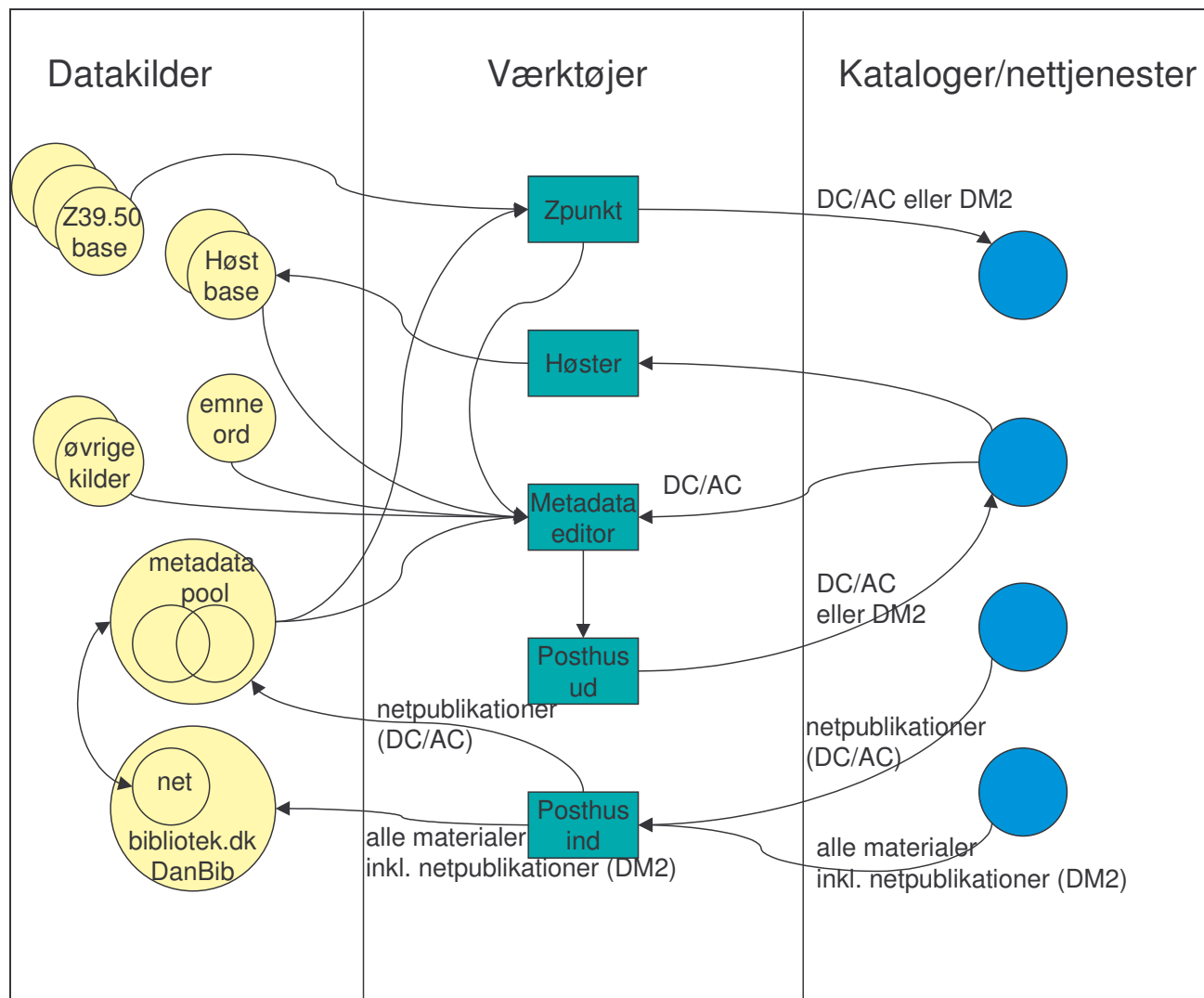
Se et typisk fragment

<http://competencehouse.dk>.

Fragmentet har en uprægnant titel ”Analyser og artikler”.

Bilag B

Fælles data og værktøjer



Diagrammet viser datakilder og værktøjer for opbygning og vedligeholdelse af metadata (linksamlinger) i bibliotekernes nettjenester. Hertil kommer (ikke viste) værktøjer til linkcheck, automatisk ajourføring og diverse hjælpefunktioner.

Metadatapoolen rummer ideelt set foreningsmængden af DBC's national-bibliografiske registreringer af netpublikationer og de metadata som rummes i bibliotekernes nettjenester og kataloger. Formatet er DC/AC (Dublin Core and Administrative Components).

Princippet er at poster i danMARC2 indlægges i DanBib/bibliotek.dk, og poster i DC/AC indlægges i metadatapoolen. Metadata som forefindes som danMARC2-poster i DanBib (og bibliotek.dk) vil være synlige ved opslag i metadatapoolen

enten virtuelt (Z39.50) eller fysisk kopieret fra DanBib. Disse poster er angivet som delmængden 'net' i DanBib/bibliotek.dk.

Metadatapoolen kan, som antydnet, afgrænses i en række delbaser, f.eks:

- alle poster som er relevante for bibliotek.dk
- alle poster fra et givet bibliotek
- DBC's nationalbibliografiske registreringer
- høstede poster som er automatisk indekserede.

Bibliotekerne og DBC vil kunne få adgang til metadata på en række måder, herunder:

- ved høstning af en specifik site
- via avanceret totalsøger som ved agentteknologi kan understøtte selektion
- via redskaber til automatisk indeksering og klassifikation
- via metadataeditoren er der adgang til metadata fra metadatapoolen, høstede data samt eventuelle eksterne Z39.50 baser
- metadataeditoren giver mulighed for at udbygge dataene, herunder genbrug af emnedata.

Ajourføring af lokalt katalog eller netjeneste kan ske på flere måder:

- download via Z39.50 (Zpunkt) i danMARC2 eller i DC/AC
- download via ftp (Posthuset) i danMARC2 eller i DC/AC
- automatisk ajourføring efter individuel profil
- samsøgning via Zpunkt (Z39.50).