

Rapport fra Arbejdsgruppen vedr.
nationalbibliografisk optagelse og registrering
af netressourcer (udkast)

Styrelsen for Bibliotek og Medier, juni 2008

Indholdsfortegnelse

1. Arbejdsgruppe og formål	4
1.1 Arbejdsgruppens kommissorium	4
1.2 Baggrund	5
1.21 Scenario 1 (Danica-indeks)	6
1.22 Scenario 2 (Producent-genererede metadata)	6
1.23 Scenario 3 (Den selektive registrering og formidling)	7
1.3 Samspil mellem scenarierne i den anbefalede strategi	7
1.4 Nationalbibliografi	7
1.41 Lovgrundlaget	8
1.42 Definition af Nationalbibliografien	8
1.5 Dansk Nationalbibliografis relation til pligtaflevering	9
1.51 Dansk Bogfortegnelse	9
1.52 Dansk tidsskriftfortegnelse/Dansk periodicafortegnelse	9
1.53 Kulturbevaring	10
1.54 Statisk/dynamisk	10
1.6 Registrering af statiske og dynamiske ressourcer 1996-	11
1.7 Registreringsmængder	11
1.8 Formidling	12
2. Model: Selektiv nationalbibliografisk registrering af internetressourcer	16
2.1 Hvorfor operere med en selektiv model?	16
2.2 Adgang	16
2.3 Hvad bør medtages i nationalbibliografien?	17
2.31 Materialetyper	17
2.32 Selektionskriterier	17
2.33 Fuldstændighed	18
2.4 Registreringsmetoder	19
2.5 Hvad med det der ikke registreres?	19
2.6 Konsekvens	19
3. Model: Fuldstændig nationalbibliografisk registrering af internetressourcer	20
3.1 Hvorfor operere med en fuldstændig model?	20
3.2 Kulturarvsvinklen	20
3.3 Princip for formidling	20
3.4 Metode for berigelse	21
3.5 Princip for selektion til berigelse	21
3.6 Konsekvenser i forhold til nuværende optagelsespraksis	22
4. Model: Selektiv nationalbibliografisk registrering af internetressourcer	23
4.1 Modargumenter	23
4.11 Generelt om registrering af internetressourcer og brugeradfærd	23
4.12 Specielt til den foreslåede selektive model	23
5. Model: Fuldstændig nationalbibliografisk registrering af internetressourcer	25
5.1 Modargumenter	25
5.11 Fuldstændighedsprincippet	25
5.12 Fritekstsøgning	25
5.13 Producentens egne metadata	25
5.14 Automatisk berigelse og automatisk klassifikation	26
5.15 Figuren	26

5.16 Konsekvenser i forhold til nuværende optagelsespraksis	26
5.17 Adgang	26
5.18 Brugerne	27
6. Sammenfatning	28
7. Anbefaling	28
8. IFLA	29
9. Litteraturliste	30

1. Arbejdsgruppe og formål

Bibliografisk Råd nedsatte i 2007 en arbejdsgruppe, som opfølgning på anbefalinger i rapporten *Dansk strategi for registrering og formidling af internetressourcer* (DSRI) fra 2005, med følgende medlemmer:

Eva Fønss-Jørgensen, Statsbiblioteket
Hanne Hørl Hansen, DBC
Grethe Jacobsen, Det Kongelige Bibliotek
Eva Sønderstrup-Andersen, Styrelsen for Bibliotek og Medier (sekretær)
Bodil Wöhnert, Esbjerg Kommunes Biblioteker

I DSRI rapporten anbefales det:

"[...] at der arbejdes videre med afdækning af registreringsbehov og selektionskriterier for dynamiske Internetressourcer og andre materialer end tekster på nettet, herunder om en selektion af registreringsværdige ressourcer kan ske i et samspil mellem decentrale og centrale aktører."
(DSRI 2005, s. 53, punkt 5)

og at

Biblioteksstyrelsen hurtigst muligt nedsætter en eller flere arbejdsgrupper, der skal:

1. *Undersøge rationale i den nuværende optagelsespraksis for netressourcer i Nationalbibliografien – specielt med henblik på at analysere værdien for brugerne af den strikte skelnen mellem statiske og dynamiske ressourcer*
2. *Undersøge muligheder for at øge mængden af 'gode' metadata hos producenterne og udbyderne af netressourcer. Dette kan for eksempel ske gennem promovning, udvikling af værktøjer samt indgåelse af aftaler med producenter, distributører og andre.*

(DSRI 2005, s. 55)

Gruppen har afholdt tre heldagsmøder og to halvdagsmøder. Forud for møderne er der udarbejdet skriftligt materiale.

1.1 Arbejdsgruppens kommissorium

Kommissoriet blev udarbejdet i september 2007 på baggrund af ovenstående punkt 1 og lyder:

Arbejdsgruppen skal udvikle anbefalinger i forhold til Nationalbibliografien vedr. optagelse og registrering af netressourcer.

I disse anbefalinger skal følgende facetter indgå:

- *Afgrænsning af Nationalbibliografien*
 - *Udgangspunkt vil være brugsscenerier (hvordan og i hvilke sammenhænge vil brugerne anvende Nationalbibliografien og de nationalbibliografiske poster?)*

- *Selektionskriterier*
 - *Vurdering og prioritering af forskellige typer af netressourcer på baggrund af deres kompleksitet og sammensætning, samt af deres tilgængelighed og anvendelse (brugerbehov)*
- *Registreringsniveau(er)*
 - *Vurdering af nødvendige registreringsniveau(er) givet netressourcernes kompleksitet og sammensætning, samt deres tilgængelighed og anvendelse (brugerbehov)*
- *Samarbejder/Partnerskaber med henblik på at optimere producent metadata*
 - *Offentlig*
 - *Privat*

Metode:

I IFLA-regi arbejdes der med at få færdiggjort *Guidelines for National Bibliographies in the digital age* (GNBDA). Arbejdsgruppen skal afprøve og give input til det videre arbejde med disse guidelines – med særligt fokus på de elementer, der omhandler optagelse/selection og registrering af netressourcer.

1.2 Baggrund

I rapporten ”Dansk strategi for registrering og formidling af internetressourcer” (DSRI, 2005) konkluderes det (summeret op):

- *at formålet med en registrering af internetressourcer ”er at formidle disse ressourcer på den optimale måde til slutbrugeren, som primært er den danske borger”*
- *at registrering og formidling af ressourcerne til enhver tid skal ske så rationelt som muligt under hensyntagen til de forskellige brugerbehov*
- *at brugerbehov ikke er entydige, idet man i ”såvel forskningsmæssig som kulturarvmæssig sammenhæng vægter fuldstændighed tungere end brugervenlighed og nem tilgængelighed, hvorimod aktualitet, overskuelighed og nem tilgængelighed vejer tungt i informationsformidling til samtiden”*
- *at der skal tages højde for teknologiske begrænsninger (søgemaskiner understøtter alene høstning og indeksering af tekster – ikke andre materialetyper, søgemaskiner understøtter pt. ikke struktureret søgning og er mangelfuld mht. emnesøgning) og de juridiske begrænsninger (persondatalov, ophavsretslov mv.)*

(DSRI, 2005 s. 53)

DSRI anbefaler på baggrund af dette, at der udformes en strategi for national registrering af netressourcer, der ”skal understøtte de differentierede brugerbehov og tage højde for aktuelle juridiske og teknologiske barrierer.” I rapporten er skitseret tre scenarier, som i kombination bør danne

grundlag for den anbefalede strategi. De tre scenarier refereres her med en opdatering af situationen (maj 2008).

1.21 Scenario 1 (Danica-indeks)

Dette scenario er baseret på en indeksering af det til Netarkivet indsamlede internetmateriale. Scenariet tilgodeser kulturarvsaspektet men kan også tænkes anvendt som alternativ til det levende net, når en tidligere registreret URL konstateres utilgængelig. På grund af juridiske barrierer blev scenariet på rapportens udgivelsestidspunkt vurderet ikke realiserbart, da der ikke er almen adgang til Netarkivet.

Status 2008: De juridiske barrierer er fortsat ikke fjernet, og der er pt. kun adgang til arkivet for forskere. Netarkivet har iværksat et arbejde, der ved juridisk ekspertise afdækker forholdet mellem persondataloven og de alment samfundsmæssige krav om fri adgang til information for at fremme en ændring i lovgivningen på området.

1.22 Scenario 2 (Producent-genererede metadata)

Dette scenario skulle give adgang til internetressourcer, dels via søgemaskiner, der søger i metadata, dels via bibliotekskataloger, såfremt metadata bliver høstet og konverteret til MARC-poster. URL som del af sådanne metadata vil være anvendelig, så længe materialet er knyttet til den oplyste URL på det levende net. En eventuel kobling af metadata og en arkiv-URL eller URN ville gøre det muligt at give permanent adgang til en ressource, enten på internettet eller i Netarkivet. Dog vil adgangen til Netarkivet være omfattet af de samme juridiske barrierer som anført i scenario 1.

Status 2008: DSRI anbefalede, at der blev nedsat en arbejdsgruppe, der skulle arbejde på at øge mængden af gode metadata fra producent- og udbyderside (jfr. ovenstående). Af nærværende arbejdsgruppes kommissorium fremgår det, at gruppen også skal overveje samarbejder og partnerskaber, der kan "*optimere producent metadata*".

Pt. arbejdes der i regi af Dansk Standard S24 (Information og dokumentation) på en vejledning vedrørende blivende netpubliceret information ("*Den artige netpublikation*"). Vejledningen vil indeholde råd til udgivere i almindelighed om hvilke metoder, der bør tages i anvendelse, hvis netinformationen skal kunne genfindes på længere sigt. Vejledningen vil således give råd om berigelse med metadata og forklare, hvorfor metadata i det hele taget bør tilføres blivende netpubliceret materiale, herunder en redegørelse for forholdet til det nationale netarkiv. Det kan ikke vides, hvornår dette arbejde afsluttes, om indholdet vil tilgodese andre materialetyper end tekster (f.eks. film og musik) tilstrækkeligt og endeligt, om retningslinjerne vil vinde udbredelse.

Hvad angår producentgenererede metadata findes disse online til alt materiale, da man ikke kan publicere noget på en server uden et minimum af oplysninger om sig selv og sit materiale. Omfanget og kvaliteten af tilknyttede metadata er dog meget svingende. Ofte er kun titlen tilknyttet som metadata. Det er til trods herfor arbejdsgruppens opfattelse, at der er stigende forståelse hos producenterne for nødvendigheden af metadata, ikke mindst af ophavsretlige grunde. DOI systemet (Digital Object Identifier), der refererer til metadata, og som kan bruges som udgangspunkt for hentning af metadata, er således blevet tildelt 33 mio. objekter pr. maj 2008. DOI bruges især til overvågning af rettigheder¹.

¹ <http://www.doi.org/>

Der har hidtil manglet redskaber i biblioteksverdenen, der kan udnytte gode metadata, og som ikke findes som MARC-poster i bibliotekskataloger. Med lanceringen af SUMMA (udviklet af Statsbiblioteket) og PRIMO (udviklet af Ex Libris for Det Kongelige Bibliotek) er der kommet fokus på integreret søgning ("integrated search") og dermed på, at metadata ikke nødvendigvis skal være lagrede som MARC-poster for at kunne præsenteres for brugeren i sammenhæng med data fra bibliotekskatalogens traditionelle MARC-poster. En anden mulighed, der også praktiseres, er at tage eventuelle metadata og konvertere disse til MARC-poster, så de kan nå sammen med katalogens øvrige poster ("federated search").

Der arbejdes således fra forskelligt hold på at understøtte scenario 2, men der mangler stadig en national strategi og dermed en tidshorizont for implementeringen af dette scenario. Målet er således ikke umiddelbart i sigte.

1.23 Scenario 3 (Den selektive registrering og formidling)

Dette scenario lægger op til en selektiv registrering af netressourcer i katalogen, således at man sikrer skabelsen af metadata på de mest relevante netmaterialer og dermed øger formidlingsmulighederne for de udvalgte netressourcer. Registreringen af de udvalgte netressourcer kan være baseret på en automatisk eller en manuel metode.

Status 2008: Scenario 3 var på rapportens tilblivelsestidspunkt allerede delvist realiseret, idet der på enkelte materialeområder optages statiske netpublikationer i nationalbibliografien. Disse metadata anvendes af lokale bibliotekskataloger og nationale tjenester som bibliotek.dk og øger mulighederne for fælles formidling af traditionelle biblioteksmaterialer og (udvalgte) netressourcer. På nuværende tidspunkt vil det være muligt at overføre samme praksis for udvalgte dynamiske ressourcer.

1.3 Samspil mellem scenarierne i den anbefalede strategi

I DSRI fremgår det, at ændringer i de enkelte scenariers realiserbarhed løbende må give anledning til at tilrette og omlægge de tiltag, man har sat i gang:

"[...] balancen mellem de tre scenarier i strategien kan ændre sig over tid efterhånden som udgivere selv etablerer stringente metadata og efterhånden som de teknologiske og juridiske begrænsninger ved scenario 1 nedbrydes"

(DSRI 2005 s. 53)

Perspektivet er dog uklart, da problematikken i forhold til netarkivet ikke umiddelbart står til at løse.

1.4 Nationalbibliografi

Da nærværende rapport ifølge kommissoriet skal "udvikle anbefalinger i forhold til Nationalbibliografien", følger her en kort oversigt over den danske nationalbibliografi og forholdet til biblioteksvirksomhed og pligtaflevering.

1.41 Lovgrundlaget

Lovgrundlaget for den danske nationalbibliografi er fæstet i:

Lov om biblioteksvirksomhed nr. 340 af 17. maj 2000

- *§ 16. Staten afholder udgiften til udarbejdelse af nationalbibliografien, der indgår i den nationale fælles bibliotekskatalog*
- *Stk. 2. Staten bidrager til, at den nationale fælles bibliotekskatalog i videst muligt omfang stilles til rådighed for enhver ved adgang via Internet*

Lov om pligtaflevering af offentliggjort materiale nr. 1439 af 22. december 2004

- *§ 6. Udgiveren er forpligtet til efter påkrav fra pligtafleveringsinstitutionen at give sådanne oplysninger, som er nødvendige for korrekt registrering af værket i Nationalbibliografien*

1.42 Definition af Nationalbibliografien

En nationalbibliografi skal ifølge internationale standarder tjene to formål:

- *Genfinding/formidling: Alt registreres med henblik på at kunne genfinde og formidle et lands samlede nationale output*
- *Kulturbevaring: Alt registreres ud fra et ønske om at have en beskrivelse af et lands samlede nationale output*

Den nyeste internationale definition fra 1998 lyder som følger:

- *Nationalbibliografien bør omfatte den løbende nationale publicerings-virksomhed, og hvor det lader sig gøre, tillige sørge for retrospektiv dækning. Om nødvendigt bør selektionskriterier defineres og publiceres af den nationalbibliografiske institution.*
- *Nationalbibliografien bør indeholde registreringer af materiale på alle de sprog og/eller i alle de alfabeter, som anvendes i den pågældende stat, og - hvor det er muligt - bør de sprog og/eller alfabeter, som publikationerne oprindeligt er udkommet på, indgå i registreringerne*

(Rekommandationer fra International Conference on National Bibliographic Services, København, november 1998²)

For tiden arbejder IFLA's Working Group on Guidelines for National Bibliographies med en definition, der lyder:

“The Working Group finds that the coverage of the national bibliography is the essential thing and it should be the whole national output in the understanding of all different types of publications, but it should not necessarily be exhaustive. There should be pragmatic formal selection criteria. In the

² <http://www.ifla.org/VI/3/icnbs/fina-dk.pdf>

same way all the types of publications are not necessarily going to be catalogued and formatted at the same level and following the same rules. The registrations can be made by different people (cataloguers, producers, etc., in different places) or automatic/semiautomatic systems following different sets of rules depending on user needs and technology. This model could be used because the technology today allows it. It would be wonderful if all the data were published in the same database, but it might not be necessary. It is meanwhile a recommendation that the data could be retrieved from the same interface. It is still important that there is a central institution/the national bibliographic agency to make/coordinate rules and selection criteria, etc.”

(“Guidelines for national bibliographies in the electronic age”, Draft 2007 s. 5)

I dette forslag ligger en erkendelse af, at den traditionelle nationalbibliografi fremover må opgive ambitionen om fuldstændighed og i stedet være selektiv. Samtidig åbner definitionen for den mulighed, at nationalbibliografien i stigende grad vil kunne - og bør - spille sammen med de teknologiske redskaber, der udvikles til søgning på nettet i bred forstand.

1.5 Dansk Nationalbibliografis relation til pligtaflevering

Pligtaflevering til Det Kongelige Bibliotek har eksisteret siden 1697, og siden 1781 med det formål at opbygge en national samling af trykt materiale (bøger, tidsskrifter, aviser, kort, billeder etc.) produceret i landet. Det modtagne materiale blev indtil 1960 registreret i kataloger inddelt i emner. Selve registreringen er sparsom: forfatters efternavn, forbogstaver, titel, udgiver/redaktør (for tidsskrifter), udgivelsessted og -år, bindtal og opstillingssignatur.

Siden pligtafleveringsloven af 1902 er der blevet skelnet mellem småtryk og bøger/tidsskrifter/aviser. Småtryk defineres som udgivelser uden forfatternavn, udgivet af en korporation, og behandlede korporationens aktuelle interne anliggender. Det betød, at fra 1902 blev en stadig større del af de pligtafleverede tryksager ikke længere registreret i Det Kongelige Biblioteks kataloger og dermed ikke i bogfortegnelsen.

1.51 Dansk Bogfortegnelse

Denne ældste og største del af den danske nationalbibliografi blev fra 1841-1939 (dækningsår) udarbejdet af Det Kongelige Bibliotek i samarbejde med Gads forlag. Derefter overgik udarbejdelsen først til Bibliotekstilsynet, senere Bibliotekscentralen og nu DBC. Siden 2004 har DBC for Det Kongelige Bibliotek registreret de af bibliotekets monografier, der også optages i Dansk Bogfortegnelse.

1.52 Dansk tidsskriftfortegnelse/Dansk periodicafortegnelse

DPF 1949 er siden 1960 udarbejdet af Det Kongelige Bibliotek og udgivet i samarbejde med DBC. Den er udarbejdet udelukkende på grundlag af de pligtafleverede tidsskrifter og årspublikationer. Småtryksperiodika optages ikke i DPF, hvilket i praksis betyder, at ca. trefjerdedele af de pligtafleverede tidsskrifter ikke er registreret i Nationalbibliografien.

Nationalbibliografien er i løbet af 1900-tallet blevet udvidet med andre fortegnelser, som dækker både pligtafleverede materialer, f.eks. artikelindeks, kort- og musikfortegnelser, og materialer, der indtil 1998 ikke var omfattet af pligtaflevering, f.eks. lyd- og billedfortegnelser. For alle fortegnelser gælder det, at der selekteres ifølge formelle optagelseskriterier blandt alle udgivne materialer, så kun udvalgte værker optages i nationalbibliografien og dermed får en post på nationalbibliografisk

niveau. Ingen af nationalbibliografiens dele kommer længere i trykt form og posterne, der udgør disse fortegnelser, skal derfor søges i DanBib³.

1.53 Kulturbevaring

Med lov om pligtaflevering af 1997 (Lov nr. 423 om pligtaflevering af udgivne værker af 10. juni 1997 § 1 stk. 4) blev genstanden, der skulle pligtaflleveres, ændret fra bogtrykte materialer til udgivne værker, herunder værker der var tilgængelige i en database. For at kunne implementere dette i forholdet til det nye og hastigt udviklende internet, blev der i den ledsagende bekendtgørelse (bekendtgørelse nr. 1041 om pligtaflevering af udgivne værker af 17. december 1997, § 1 stk. 2, 9) skelnet mellem statiske og dynamiske enheder på internettet, hvoraf kun de statiske blev anset for værker, der var omfattet af pligtafleveringsloven. Denne skelnen var en praktisk foranstaltning forårsaget af de juridiske og teknologiske betingelser i efteråret 1997. Af bemærkningerne til lovforslaget fremgår det, at loven bør revideres i takt med den teknologiske udvikling, hvilket også skete med vedtagelsen af lov nr. 1439 om pligtaflevering af offentliggjort materiale af 22. december 2004⁴. En væsentlig ændring i 2004-loven var, at afleveringskriteriet gik fra at være baseret på udgivne værker til offentliggjorte værker. Herved skete der en meget omfattende udvidelse af pligtafleveringens dækningsområde, idet nu også dynamiske værker på internettet (dvs. hele .dk-domænet samt den del af internettet, der retter sig mod et dansk publikum), danske radio/tv-udsendelser og film blev afleveringspligtige. Der skete i den forbindelse ingen ændring af Nationalbibliografiens optagelseskriterier, som derfor fortsat alene omfatter de ovennævnte fysiske værker og de statiske internetdokumenter. Kort sagt dækker den nationalbibliografiske registrering i dag kun en meget lille del af de værker, som indsamles via pligtafleveringsloven.

1.54 Statisk/dynamisk

Der er bred enighed i arbejdsgruppen om at ophæve sondringen mellem statiske og dynamiske publikationer, da denne skelnen er foretaget af praktiske årsager og ikke ud fra relevans. Internettet er i sin natur dynamisk, hvilket betyder, at en skelnen mellem statiske og dynamiske ressourcer vil være en kunstig afgrænsning. Dette er bl.a. også erkendt i pligtafleveringssammenhæng.

³ En beskrivelse af hele nationalbibliografien samt et essay (1997) af Mona Madsen findes på Det Kongelige Biblioteks hjemmeside (<http://www.kb.dk/da/kb/service/nationalbibliografi/>) og en beskrivelse af de dele som DBC producerer på DBCs hjemmeside: http://www.dbc.dk/top/top_natbib_mv.html

⁴ Denne lov skulle revideres i 2008/09, men revisionen er udsat til 2010/11

1.6 Registrering af statiske og dynamiske ressourcer 1996-

DBC har siden 1996 registreret internetpublikationer, først som forsøg (1996-1999) og siden på driftsbasis. For statiske tekstmaterialer følges Optagelseskriterierne for Dansk bogfortegnelse og Dansk periodicafortegnelse 5. udgave (1998):

- *Elektronisk lagrede publikationer i selvstændig fysisk form, f.eks. disketter, cd-rom, samt statiske netpublikationer optages (2.3. Elektronisk lagrede publikationer)*
- *Periodica (se def. under 5.2) opfattes også i denne sammenhæng som statiske netpublikationer (Forord til 5. udgave)*

Udover optagelseskriterierne, er der en række retningslinjer for optagelse nævnt i Indoreg 2-rapporten fra 1999⁵:

- *Entitet: Enkelt fil eller en samling af en række filer. Logisk afgrænset med egen titel. I registreringsammenhæng "et værk". Selvstændigt publiceret*
- *Andre afgrænsninger af hensyn til ressourcer mv.: Så overordnede enheder som muligt*
- *Optagelse af nye originale værker prioriteres i forhold til netudgaver af eksisterende værker*
- *Begrænset optagelse af personlige hjemmesider og selvstændige enkeltpublikationer herpå*
- *"Bestandighedskriterie"*
- *Omfang for tekst på >20000 typografiske enheder*

1.7 Registreringsmængder

En søgning i maj 2008 på monografiske netdokumenter gav 16.733 hits. Heraf er 4782 suppleret med, den af primært folkebibliotekerne finansierede, bibliotekskatalogisering (emneord, indholdsbeskrivende noter mv.). Der køres ugentligt url-tjek på de registrerede materialer.

Dynamiske netpublikationer er ikke omfattet af nationalbibliografien. Som bibliotekskatalogisering har DBC gennem årene på forsøgsbasis registreret en række af disse primært i Dublin Core, men også enkelte i MARC. Dublin Core-poster er herefter konverteret til MARC og indgår i DanBib. En søgning i januar gav 1.682 poster. Registreringen i Dublin Core omfatter ikke navneformskontrol, men der tildeles kontrollerede emneord, ligesom der laves indholdsbeskrivende noter. Der laves med jævne mellemrum url-tjek.

Dansk periodicafortegnelse begyndte i midten af 90'erne at optage online danske periodika, men kun publikationer som svarede til den danske definition af periodika. I 2003 udkom den nye version af ISSN Manual, som udvidede periodikabegrebet til også at omfatte "continuing resources". Dette blev indarbejdet i den nye ISSN-standard (ISO 3297) fra 2007, hvorefter følgende udgivelser omfattet af standarden bør tildeles ISSN:

⁵ <http://gammel.dbc.dk/udvikl/indoreg/>

- ”Serials”, der defineres som ”continuing resource issued in a succession of discrete issues or parts, usually bearing numbering, that has no predetermined conclusion”, hvilket svarer til den danske definition af ”periodika”
- ”Continuing resources”, der defineres som ”publication, in any medium, that is issued over time with no predetermined conclusion and made available to the public”, med en yderligere forklaring: “Such a publication is usually issued in successive or integrating issues which generally have numerical and/or chronological designations ... Continuing resources include serials such as newspapers, periodicals, journals, magazines, etc., and ongoing integrating resources such as loose-leaf publications that are continually updated and Web sites that are continually updated”

ISSN Danmark begyndte i 2003 at tildele ISSN til danske databaser og web steder og samtidig registrere disse til Dansk periodikafortegnelse. Der tildeles ikke ISSN til institutionelle og personlige hjemmesider eller weblogs. En søgning i april 2008 gav 1.303 poster i Dansk periodikafortegnelse. ISSN Manual er igen under revision, og der er nedsat en arbejdsgruppe, som skal komme med nogle retningslinjer for tildeling af ISSN til online publikationer ud fra en pragmatisk synsvinkel.

Trods disse udvidelser i den nationalbibliografiske dækning, er der stadig stor forskel på, hvad der registreres i nationalbibliografien, og hvad der er omfattet af pligtafleveringsloven som kulturarv.

1.8 Formidling

Ifølge Informationsordbogen⁶ er formidling:

”[...] en aktivitet, der fungerer som forbindende mellemed. I sin bredeste betydning al medieret kommunikation, inkl. sprogligt medieret og litterært medieret kommunikation. I denne betydning er alle biblioteker og databaseværter, museer, forlag, tidsskrifter osv. formidlingsinstitutioner. I en snævrere betydning mediering fra specialiserede faggrupper til den bredere offentlighed, dvs. omfattende f.eks. oversættelse, popularisering og undervisning. I denne snævre forståelse er folkebiblioteker, men ikke forskningsbiblioteker formidlingsinstitutioner.”

Denne dobbelttydige definition rammer lige ned i en helt væsentlig problemstilling for arbejdsgruppen:

Er formidling af internetressourcer ensbetydende med at stille disse til rådighed efter en berigelse baseret på automatiske rutiner?

Eller

Er formidling af internetressourcer ensbetydende med at stille disse til rådighed efter en selektion (f.eks. baseret på den statslige informationspolitik, Lov om Biblioteksvirksomhed og katalogisering) og en nationalbibliografisk registrering?

Nærværende rapport vil præsentere to modeller, kulturarvsmodellen og den selektive registreringsmodel, der begge forudsætter, at der sker en bearbejdning af det indsamlede netmateriale af hensyn

⁶ <http://www.informationsordbogen.dk/concept.php?cid=2468>

til formidlingen. De to modeller adskiller sig ved forskellig opfattelse af formålet med nationalbibliografien; ved metode, omfang af dækning samt kvaliteten af metadata.

I arbejdsgruppen har der været enighed om, at en åbning af Netarkivet er en god ide, - dog ud fra forskellige motiver. Netarkivet kan enten ses som fundamentet for formidling af netressourcer eller som et arkiv, som man henter ressourcer fra, der er forsvundet på det levende net (netarkivet ajourføres ikke dagligt). For at kunne udnytte Netarkivet og dets indhold til disse formål kræves det dog, at Netarkivet bliver alment tilgængeligt. Formålet er således at gøre det klart, at det er nødvendigt med en politisk stillingtagen til hvilken model, der skal sættes på.

1.9 Brugeradfærd / brugerbehov og registreringsniveau

I rapporten DSRI kortlægges brugerbehov og funktionalitet i relation til kompleksitet og sammensætning samt tilgængelighed og benyttelse. Med udgangspunkt i IFLA's FRBR - Functional Requirements for Bibliographic Records (IFLA, 1998) beskrives, hvad brugeren fundamentalt set behøver en bibliografisk beskrivelse til, og rapporten sætter disse krav i relation til søgning af internetressourcer. Rapporten beskriver ligeledes forskellige brugerkategorier og deres informationsbehov (DSRI, 2005 s. 17-20).

Rapporten behandler desuden Dublin Core (DC) standarden med henvisning til, at DC er det mest dominerende regelsæt af forenklede regler. Desuden refereres konklusioner fra et Library of Congress-projekt fra februar 2005, hvor det fastslås, at der ikke er forbindelse mellem udvikling af programmer til dannelse af digitale ressourcer og udvikling af teknikker til genfindning. Automatisk dannelse af metadata bør ifølge DSRI tilstræbes – men med mulighed for tilføjelser - f. eks. i forbindelse med emnebeskrivelser (DSRI, 2005 s. 10-12).

Søgbarhed og genfindning på internettet har siden 2005 udviklet sig markant. Der ses stigende kommerciel og prestigemæssig interesse i at kunne søges og rangordnes højt på de store søgemaskiner, hvilket har haft den konsekvens, at dataleverandører og webredaktører har øget opmærksomheden på dette. Google har efterhånden udviklet flere niveauer af søgefunktioner og er gået i samarbejde med bibliotekakataloger som OCLC og bibliotek.dk, ligesom de sammen med biblioteker verden over digitaliserer ”verdenskulturarven”.

For producenten handler det således om at kunne søges og blive synlig. Producenten vil typisk vægte de metadata, som det vides, at de store søgemaskiner anvender. For brugeren handler det om at søge, udvælge og få adgang til ressourcen. Kort sagt: der skal ”varedeklareres” på en måde, så brugerne finder relevante information i forhold til deres behov. En varedeklaration lavet af hensyn til søgemaskiner er ikke nødvendigvis identisk med en varedeklaration lavet ud fra brugernes behov.

Den generelle samfundsudvikling går i retning af øget offentlig og administrativ kommunikation med og dokumentation for borgeren (borger.dk m.m.). Denne udvikling er bl.a. et resultat af den statslige informationspolitik, som den er formuleret i Info-samfundet år 2000 (Dybkjær Rapporten, 1994) og Forskningsministeriets betænkning Information til Tiden (Betænkning nr. 1342, 1997). Der vil derfor formentlig være behov for flere niveauer af dokumentation fra den dybdeborende til det fagligt professionelle niveau contra ”light”-udgaven med hurtigt overblik til den brede brugergruppe.

En nyligt foretaget inddeling af brugere findes i Alsted-klassificering af brugertyper i forhold til forudsætninger og barrierer set i relation til IKT-kompetencer samt informationsbehov (Alsted

Research, 2005). Inddelingen har bl.a. dannet baggrund for udvikling af personas til brug for borger.dk (Den Digitale Taskforce, 2006).

10 % af brugerne er typisk professionelle brugere – herunder forskere og bibliotekarer/ informationsformidlere, mens en langt større del af brugerne er den almindelige borger, herunder studerende. Det giver anledning til overvejelser om målgruppe i relation til udvælgelse, præsentation og tilgængelighed.

Wiggins (2007) har, som led i IFLA's GNBDA, opregnet en række indsamlingskategorier, som kan sammenfattes i følgende fem kategorier (ikke udtømmende):

- Forskningsmæssig interesse
 - Ressourcens intellektuelle og forskningsmæssige niveau, herunder forskningsmæssige værdi for fremtiden
 - Ressourcens repræsentationsværdi som samleobjekt med værdi for fremtidige forskere

- Generel befolkningsmæssig interesse
 - Ressourcens nuværende interesse set i forhold til forskellige og unikke perspektiver på landets sociale, kulturelle og politiske landskab
 - Ressourcens værdi set i relation til mangfoldigheden i befolknings sammensætningen, herunder ressourcens afspejling af minoriteter, undergrundsaktiviteter og forskellige opfattelser af landets historiske, kulturelle og sociale begivenheder m.m.

- Fysiske og tekniske forhold
 - Ressourcens tekniske forudsætninger for fremtidig brug (brugbarhed i forhold til navigation, funktionalitet, pålidelighed osv.)
 - Ressourcens metadata, muligheden for berigende associerede metadata
 - Ressourcens repræsentationsværdi som tidssvarende format

- Tilgængelighed
 - Unik ressource hvor indholdet udelukkende er tilgængeligt via webben
 - Ressourcens værdi set i forhold til den foreliggende analoge Nationalbibliografi
 - Sandsynligheden for at ressourcen bliver gemt andetsteds, herunder risikovurdering af om ressourcen ellers forsvinder for samtiden og fremtiden

- Særlige forhold
 - Ophavets særlige status og renommé

Behovet vil gå i retning af ønske om overblik suppleret med konkrete faktuelle oplysninger og muligheden for uddybende og uerkendt relevant information. Desuden vil der være et ønske om tilgængelighed af den registrerede netressource.

For den ikke-professionelle bruger vil overskuelighed være af betydning. Interessen vil samle sig om, hvorvidt en publikation er tilgængelig i analog og/eller elektronisk form. Dublin Core forekommer at være absolut minimum for et registreringsniveau – og, som allerede nævnt i DSRI-rapporten, synes hjemmesider i nogen grad at have taget DC til sig, bl.a. fordi DC giver basis for høstning på søgemaskinerne.

Man kan i dag forestille sig, at links vil blive udnyttet i udstrakt grad til videre beskrivelse – og at brugerne vil foretrække andre informationskilder end de traditionelle bibliotekskataloger. Spørgsmålet i denne sammenhæng er derfor, om ressourcerne skal bruges på at udarbejde anbefalinger jf. beskrivelsesstandarder med henblik på at sikre formidling af udvalgte centrale netressourcer eller, om man skal koncentrere sig om principper for udvælgelse af datakilder og sikring af deres bestandighed på internettet i det omfang, der er tale om national kulturarv.

2. Model: Selektiv nationalbibliografisk registrering af internetressourcer

2.1 Hvorfor operere med en selektiv model?

Internettet har gjort det uproblematisk og let at publicere. Den store mængde af materialer på nettet gør, at der netop her er behov for selektion med henblik på biblioteksmæssig brug. Ikke alt er relevant i nationalbibliografisk sammenhæng, og en vis grad af informationsværdi bør være til stede. Selektionen bør foretages i nationalbibliografisk regi, idet nationalbibliografien indgår som en væsentlig del af bibliotekernes fælleskatalog. Selektionen vil både være til gavn for folke- og forskningsbiblioteker og ikke mindst for borgerne som slutbrugere.

Selektion i forhold til hensynet til bibliotekernes formidling er i nationalbibliografisk sammenhæng ikke et nyt tiltag, bl.a. har man på artikelområdet altid været selektiv i forhold til, hvilke danske periodika der blev indekseret. IFLA's GNBDA lægger ligeledes op til selektion og til yderligere differentiering imellem det nationalbibliografisk registrerede ved hjælp af forskellige niveauer for registrering. Det er således ikke kun i Danmark, men også på internationalt plan, der tænkes i selektion.

Nationalbibliografien har ikke kun til opgave at registrere enkelte dokumenter. En helt central opgave er at skabe basale sammenhænge imellem registrerede dokumenter, hvilket bl.a. fremgår af de krav, der er opstillet i FRBR. I praksis er valg af niveauet for registrering i sidste ende også en økonomisk vurdering af, hvad der er midler og ressourcer til. Denne vurdering bør bygge på en faglig prioritering af materialerne, sådan som kommissoriet for nærværende arbejdsgruppe afspejler.

IFLA's GNBDA opererer med tre materialekategorier i nationalbibliografisk sammenhæng. Det anbefales, at materialer med stor betydning registreres på autoritativt niveau ("access points are controlled"). Registreringen erkendes at være dyr, men besparelsen, i form af rationelt genbrug, opvejer dette. Som eksempler på "stor betydning" nævnes materiale, der tidligere ville have haft en traditionel publicering (surrogat for eller fortsættelse af tidligere trykt ressource) og ressourcer med forskningsmæssig værdi.

I den modsatte ende af betydningspektret finder man ressourcer af personlig eller forbigående karakter, som nu og fremover må skønnes at have lav interesse, og som oftest er uden eller med få metadata. Disse er ikke af nationalbibliografisk interesse. Imellem disse to punkter ligger en gruppe materialer af en vis interesse, der niveaumæssigt med hensyn til metadata eksempelvis spænder fra informative DC-metadata til mindre fine og ukurante metadata.

2.2 Adgang

Et væsentligt aspekt ved registreringen af netressourcer er adgangen til disse. bibliotek.dk er i dag en central indgang for mange brugere. Her præsenteres brugeren for en værkvisning, der f.eks. angiver trykt version, netversion, lydversion, netlydversion mv. af samme materiale. Adgang angives for gratispublikationers vedkommende gennem URL'er og for betalingsbelagte publikationers vedkommende gennem lokaliseringer. Her kan det også anføres, om der er fjernadgang til materialet. Begge dele er betinget af bibliografiske poster på materialet.

Brugere på nettet mangler generelt en basal viden om internettet, internettets funktioner og mekanismer samt evnen til at kvalitetsvurdere internetressourcer. Samtidig er det de færreste der har et klart defineret informationsbehov og søgestrategi når de søger (British Library & JISC, 2008).

For brugerens oplevelse af værkpræsentation og adgang er eksistensen af bibliografiske poster og lokaliseringer derfor et meget vigtigt element, ligesom de bibliografiske poster via DanBib umiddelbart kan genbruges i den lokale katalog.

For gratisressourcer, der er selekteret til nationalbibliografien, ville et åbent netarkiv og links hertil være en ideel situation som tilbagefaldsmulighed, hvis ressourcen ikke længere findes på den oprindelige URL. Da dette pt. ikke er opnåeligt, bør løsninger med automatiserede URL-tjek og brug af linkresolving løbende justeres i takt med den teknologiske udvikling. URL-tjek foretages i dag af de nationalbibliografisk registrerede ressourcer. Det er vigtigt for brugeroplevelsen ikke at have for mange URL'er på dokumentniveau, som ikke er aktive længere, da det vil opfattes som støj.

2.3 Hvad bør medtages i nationalbibliografien?

2.31 Materialetyper

I dag medtages statiske publikationer på bog-, lyd- og filmområdet. Artikler og musik på nettet er ikke omfattet af de nationalbibliografiske optagelseskriterier. Nationalbibliografien bør omfatte alle materialetyper på nettet. F.eks. må musikområdet forventes at bevæge sig fra fysisk publicering til netpublicering, og den bibliografiske behandling med hensyn til standardtitler på klassiske værker er central i bibliotekernes formidling af musik.

På artikelområdet har der allerede været afvist indekseværdige periodika – udelukkende fordi deres udgivelsesform var elektronisk. De afviste periodika udgives ikke i fysisk form og indekseres derfor slet ikke. Såfremt den selektive model vælges, bør materialeområderne gennemgås med hensyn til optagelseskriterier. F.eks. ville det på musik- og filmområdet give god mening at prioritere materialer, der kan distribueres i biblioteksregi frem for materialer, som kun kan distribueres i privat regi.

2.32 Selektionskriterier

I IFLA's GNBDA nævnes det, at materialer, der er surrogat for eller fortsættelse af tidligere trykt ressource, og ressourcer med forskningsmæssig værdi, må betragtes som havende stor betydning i nationalbibliografisk sammenhæng. Når der ses bort fra ophævelse af differentieringen imellem statiske og dynamiske publikationer er der stort sammenfald mellem disse principper, og hvad der i praksis i dag registreres nationalbibliografisk på netområdet vedrørende bøger.

Optagelseskriterierne for netpublikationer omfatter bl.a. følgende:

- Publikationer der har kommerciel, intern, meget lokal eller privat karakter, optages ikke
- 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.

De almindelige optagelseskriterier for Dansk Bogfortegnelse er herudover også gældende.

DBC har på forsøgsbasis og udenfor nationalbibliografien udarbejdet løse registreringer og tematisk sammensatte pakker af registreringer på hjemmesider, linksamlinger og databaser. Der produceres i Dublin Core, men konverteres efterfølgende til danMARC2 - netop for at sikre optimalt samspil med andre data i centrale og lokale kataloger. De sider, der registreres, er typisk sider med et homogent præg og et højt informationsniveau – ikke personlige hjemmesider, firmahjemmesider og kommercielle sider.

Hvis der ændres i optagelsespraksis, således at man ophører med at skelne mellem statiske og dynamiske publikationer, må nogle af denne type registreringer forventes at falde ind under de nationalbibliografiske optagelseskriterier. F.eks. ville en database som Verdens-dyr.dk være relevant i nationalbibliografisk sammenhæng.

Som eksempler på Dublin Core-registrerede sider kan nævnes:

- *videnskab.dk* (internetportal med en samlet indgang til information om forskning i Danmark - oprettet på foranledning af Tænk tanken vedrørende forståelse for forskning)
- *Artikler og rapporter - rygning.com* (materialesamling fra Kræftens Bekæmpelse for formidlere over for børn og unge)
- *Den sociale kvalitetsdatabase* (Amtsrådsforeningen)

Denne type af materialer er ikke kun snævert folkebiblioteksrettede men afspejler forskellige informationer om det offentlige rum til brug for oplysning og uddannelse. Siderne har nødvendigvis ikke samme bestandighed på nettet som de hidtil nationalbibliografisk registrerede.

Et forslag kunne her være, at informationssider af en vis tyngde produceret af offentlige myndigheder og institutioner, politiske partier og interesseorganisationer blev registreret i nationalbibliografien, evt. på et reduceret niveau, sådan som det også er tilfældet med DBCs Dublin Core poster i dag.

2.33 Fuldstændighed

IFLA's GNBDA åbner mulighed for, at det kan være et legitimt mål for nationalbibliografien på netområdet at søge at være repræsentativ frem for fuldstændig. Dette kan være en fornuftig og realiserbar økonomisk indgangsvinkel til registreringen på netområdet. Pt. registreres netpublikationer dels på baggrund af udgiveres og andres indberetninger, dels på baggrund af oplysninger fra det tilsvarende fysiske materiale. Dette bør kombineres med automatisk selekterede forslag til registreringsemner.

Den repræsentative model giver også mulighed for at opsætte rammer for omfanget af antal registreringer. Særligt på den type af publikationer, der ligger i mellemlaget, kan dette være af betydning. Omvendt er det også værd at bemærke, at der i de undersøgelser DBC har foretaget i 2002⁷ og 2004⁸ gøres op med myten om, at mængden af registreringsværdige publikationer vokser eks-

⁷ <http://gammel.dbc.dk/nationalbibliografi/domaenedk.htm>

⁸ <http://gammel.dbc.dk/nationalbibliografi/domaene.htm>

plosivt. Både i 2002 og 2004 sås en tilvækst på ca. 4000 netpublikationer indenfor de nuværende optagelseskriterier for nationalbibliografien.

2.4 Registreringsmetoder

Redskaberne til registrering er principielt irrelevante set ud fra en nationalbibliografisk synsvinkel, så længe der er et fornuftigt forhold mellem pris og kvalitet mv. Valget af en selektiv model er dog ikke ensbetydende med registrering i gammeldags forstand fremover. Der skal løbende arbejdes på udvikling af redskaber.

DBC har på Dublin Core-poster arbejdet med ripning af metadata fra internetsider til registreringen. I projektet "Assisteret indeksering"⁹ (Diget Hansen, 2007) har der været arbejdet med software, der ud fra metadata danner forslag til den bibliografiske beskrivelse. Samtidig foretages en gennemgang af den fulde tekst op imod definerede algoritmer. Resultatet er et samlet forslag til emneord til posten. Dette projekt er i dag i drift. Det er klart, at der arbejdes hen imod at udbrede sådanne redskaber.

2.5 Hvad med det der ikke registreres?

Der er ikke noget i vejen for, at en selektiv nationalbibliografisk registrering kan suppleres af søgninger udenfor egentlige bibliografiske poster f.eks. ned i netarkivets indekser. Den selektive model udelukker ikke nødvendigvis det frasorterede, men har til formål at sikre en tilstrækkelig nationalbibliografisk, og biblioteksmæssig funktionel, registrering af det mest betydningsfulde. Det er ikke en nationalbibliografisk opgave og i bibliotekernes interesse at agere som konkurrent til de store søgemaskiner. Det må være en nationalbibliografisk opgave at gøre det nemt for brugerne at finde relevant information.

2.6 Konsekvens

Konsekvensen af at vælge den selektive model vil betyde en udvidelse af nationalbibliografien. Omfanget og de økonomiske konsekvenser heraf afhænger af flere faktorer, bl.a. en endelig fastlæggelse af selektionskriterier for de enkelte materialetyper evt. kombineret med fastlagte rammer for registreringsmængder, når målet ikke længere er fuldstændighed. På nogle områder må det formodes, at netudgaver vil fortrænge tilsvarende fysiske udgaver, eksempelvis på musikområdet.

⁹ DBC gennemførte i 2007 forsøg med automatisk genererede poster inklusive emneord ud fra strukturerede fuldtekster på artikelområdet. I modsætning til automatisk indeksering finder der ved assisteret indeksering en slutkontrol sted. Forsøget overgår til drift i 2008

3. Model: Fuldstændig nationalbibliografisk registrering af internetressourcer

3.1 Hvorfor operere med en fuldstændig model?

Udgangspunktet er taget fra Kulturministeriets rapport *Udredning om bevaring af kulturarven*, 2003:

”I betragtning af de meget store mængder materiale, der udgives på internettet, og de meget store omkostninger, en sædvanlig nationalbibliografisk registrering ville medføre, og i betragtning af de andre søgemuligheder, der er til rådighed, finder udvalget, ud fra en samlet vurdering, at en sædvanlig nationalbibliografisk registrering ikke bør foretages for internetmateriale.”

(Udredning om bevaring af kulturarven, 2003 s. 149)

3.2 Kulturarvsvinklen

1. Nettets informationer kan søges og fremfindes umiddelbart – i modsætning til information i værker i fysisk form
2. Hvis nationalbibliografien skal omfatte ”den løbende nationale publicerings-virksomhed¹⁰”, kan dette mål opnås (jfr. og figur 1) via en indeksering af det materiale, der findes i Netarkivet
3. Nettets ”værker” er foranderlige, også selvom en netressource er udsprunget af et fysisk værk, hvorfor sædvanlige katalogiserings- og klassifikationsregler og -rutiner, udviklet til fysiske værker, ikke er tilstrækkelige
4. Katalogiseringseksperter bør bruges til at udvikle automatiske rutiner til berigelse af de producent-genererede metadata, så ressourcen kan sættes ind i en sammenhæng med relevante ressourcer og derved formidles mere præcist

3.3 Princip for formidling

Formidlingen af den indsamlede kulturarv på nettet sker på grundlag af

- fritekstsøgning (eksisterende, kommercielle søgemaskiner)
- producentens egne metadata, der automatisk beriges i det omfang, der er mulighed for det
- automatisk klassifikation (tildelte emneord ved hjælp af algoritmer)

¹⁰ Rekommandationer fra International Conference on National Bibliographic Services, København, november 1998

3.4 Metode for berigelse

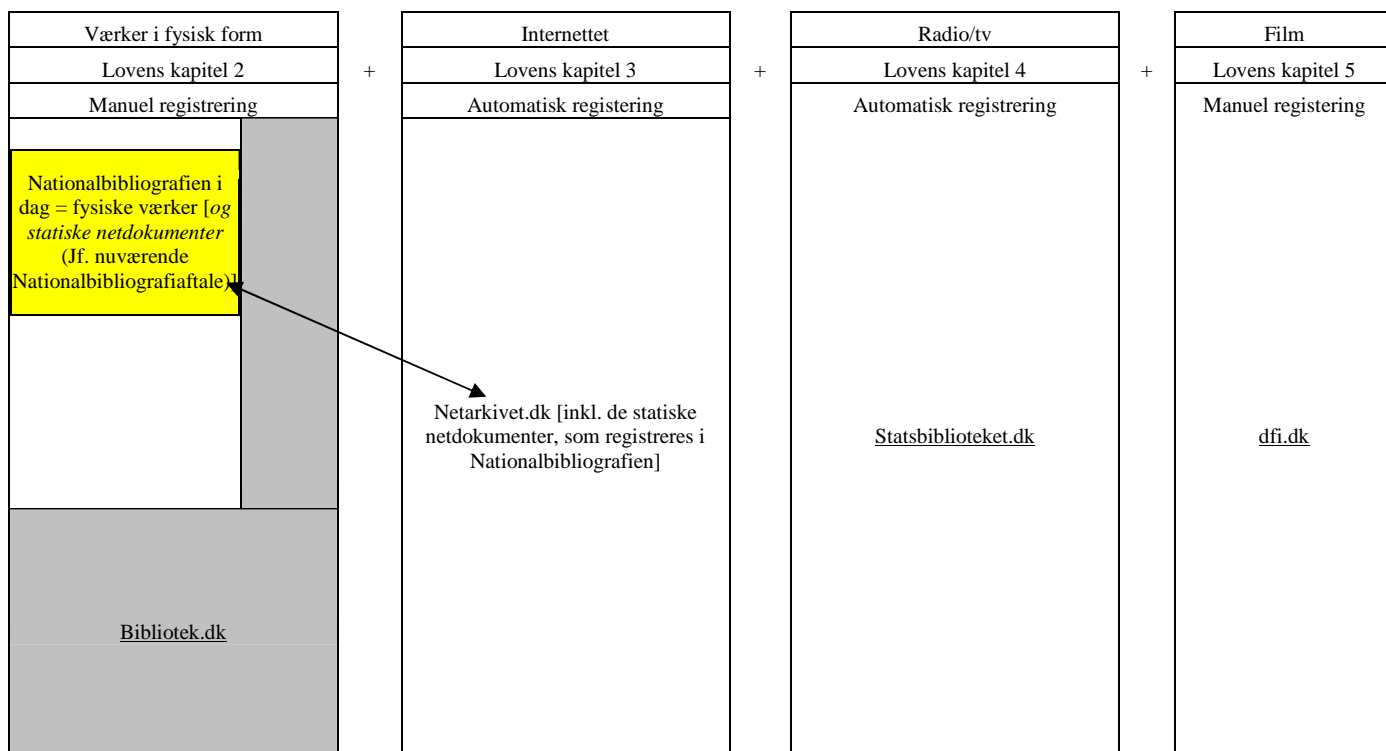
Metadata høstes, og ved hjælp af automatiske rutiner beriges disse ved at tilføje emneord og klassifikationsnotationer, forbedre ophavsoplysninger o.l.

3.5 Princip for selektion til berigelse

Selektion af de ressourcer, der skal have tildelt berigede metadata, sker hos producenten af netmateriale igennem hans/hendes tildeling af metadata. Jo bedre metadata producenten genererer, jo bedre berigelser kan foretages. Nogle af redskaberne til berigelse af metadata i Netarkivet findes allerede i form af den dokumentation, der produceres af Netarkivet i forbindelse med høstninger, især de selektive høstninger og begivenhedshøstninger (f.eks. opdeling af web-steder i kategorier i forbindelse med indsamling af valg).

Det skal understreges, at redskaber, der udvikles til berigelse af metadata, **også** kan bruges på det aktive net, på trods af at adgangen til Netarkivet pt. er begrænset til forskere. Hvis berigelsesredskaber kobles sammen med en national - og generelt anvendt - URN model kan adgangen til en ressource sikres, idet man ved hjælp af URN'en vil kunne finde ressourcen i Netarkivet, hvis den ikke længere er tilgængelig på det aktive net.

Figur 1: Pligtaflevering og nationalbibliografi i dag (2008):



3.6 Konsekvenser i forhold til nuværende optagelsespraksis

- Nationalbibliografibegrebet udvides til også at omfatte Netrivets danicaindeks. Det betyder i praksis, at den nuværende nationalbibliografis optagelseskriterier fremover alene skal fokusere på at registrere værker i fysisk form (hvorfor produktet måske derfor bør skifte navn til "Nationalbibliografi for værker i fysisk form")
- Ca. 2.600 netpublikationer, der pt. årligt registreres for bibliotekerne, vil ikke længere skulle registreres, men findes ved hjælp af fritekstsøgninger eller fokuserede søgninger ved hjælp af berigede metadata
- Ca. 450 periodiske netressourcer vil ikke længere skulle registreres. De skal i stedet findes via fritekstsøgninger eller fokuserede søgninger ved hjælp af berigede metadata, med mindre de har fået tildelt et ISSN og derfor skal registreres til ISSN basen
- De midler, der pt. anvendes på registrering af netressourcer i nationalbibliografien, allokeres til udvikling af automatiske rutiner for berigelse.

4. Model: Selektiv nationalbibliografisk registrering af internetressourcer

4.1 Modargumenter

Følgende afsnit indeholder argumenter som taler imod den selektive model.

4.11 Generelt om registrering af internetressourcer og brugeradfærd

1. Online materialer har medfødte metadata, der er tilgængelige for alle ved hjælp af en søgemaskine i modsætning til et titelblad, cd-omslag o. lign. på et fysisk værk, der kun er tilgængelige for den, der holder værket i hånden. Der er derfor ingen grund til at registrere online materiale med redskaber og metoder beregnet for fysiske værker. Sagt på anden vis: fremtidens problemer kan ikke løses med fortidens redskaber
2. Udviklingen af metoder til automatisk indeksering og hurtig og effektiv søgning i internetressourcer går rivende stærkt, såvel nationalt som internationalt. De ekstra økonomiske midler, der evt. måtte være til rådighed med henblik på at gøre Nationalbibliografien tidssvarende, bør anvendes til videreudvikling af disse automatiske metoder frem for at anvende dem til mere eller mindre 'håndholdt' registrering af et lille subjektivt udvalg af netressourcer
3. Alle nyere undersøgelser viser, at allerede nutidens brugere flittigt benytter det levende net til informationssøgning, og at søgning i bibliotekskataloger kun anses som værende én blandt andre muligheder for at finde relevant information. Sagt på anden vis: bibliotekskatalogen er ikke 'verdens navle' for brugerne i deres informationssøgning.

4.12 Specielt til den foreslåede selektive model

4. Der sættes lighedstegn mellem nationalbibliografi og bibliotekskatalogisering, dvs. der lægges kvalitative, men ikke formelle kriterier til grund for selektion til nationalbibliografien. Dette er et paradigmeskift i forhold til det hidtidige nationalbibliografibegreb, hvilket er betænkeligt. Nationalbibliografien skal ikke afspejle et subjektivt udvalg af materialer, men give et objektivt billede af et lands samlede produktion over tid. Et begreb som "registreringsværdigt" og "værker af en vis tyngde" bør derfor slet ikke anvendes i denne sammenhæng
5. Forslag vedrørende hvilke internetressourcer, der skal selekteres til registrering, er præget af det fysiske værksbegreb, og ikke af hvordan internettet ser ud i dag, og hvor det er på vej hen. Derfor er forslagene ikke langtidsholdbare
6. Online materialer er dynamiske, så en vurdering med henblik på selektion til bibliotekskatalogisering vil kun gælde for en given dato. Et "registreringsværdigt" netsted af i dag kan nemt blive morgendagens "ikke-registreringsværdige" og ikke-vedligeholdte netsted med irrelevante og/eller forældede data. De nævnte materialekategorier (fra "materialer med stor

betydning” til ”ressourcer af personlig eller forbigående karakter”) er alt for subjektive og forankrede i situationen her og nu, og vil kun være gældende på kort sigt

7. Misforholdet mellem antallet af bibliotekskatalogiserede online ressourcer og omfanget af nyttigt online materiale vil konstant øges, mens udbyttet af investering i selektion og traditionel registrering af online materialer konstant vil mindskes
8. Adgang til Netarkivet er ganske vist begrænset, men der er ikke tale om et 'dødt' arkiv, idet der er adgang for forskere. Det er specielt forhold vedr. Persondataloven, der pt. hindrer adgang for ikke-forskere, men Kulturministeriet har netop bevilget midler til Netarkivet med henblik på at afklare og på sigt løse de juridiske problemer, således at der kan gives adgang for en bredere kreds.
9. Den selektive model knytter sig til IFLA's forslag til guidelines for nationalbibliografisk registrering, som på den ene side siger, at nationalbibliografien bør omfatte "*hele det nationale output i form af alle typer af publikationer*", men på den anden side samtidig siger, at nationalbibliografien "*ikke behøver at være udtømmende, men at der bør anlægges "pragmatiske selektionskriterier"*", idet det ikke er muligt at registrere alt. Det kan undre, at verdens førende biblioteksorganisation i sine bestræbelser på at gøre biblioteksstandarder tidsvarende anlægger et så snævert – og i realiteten elitært - perspektiv på nationalbibliografi-begrebet, når det nu er muligt med dagens og morgendagens teknologi at kunne redefinere nationalbibliografiens genstandsfelt til automatisk at registrere så godt som det samlede internet i et land.

5. Model: Fuldstændig nationalbibliografisk registrering af internetressourcer

5.1 Modargumenter

Følgende afsnit indeholder argumenter som taler imod den selektive model.

5.11 Fuldstændighedsprincippet

Udvalget bag rapporten *Udredning om bevaring af kulturarven* er nedsat ud fra et kulturarvsaspekt. Den danske nationalbibliografi er indtil nu ikke kun blevet udarbejdet af kulturarvshensyn men i høj grad også af hensyn til rationalet i det samlede biblioteksvæsen (genbrug, infrastruktur mv.). Nationalbibliografien er da også primært lovfæstet i Lov om biblioteksvirksomhed.

Med kulturarvsmodellen vægtes fuldstændighed. Løsningen betyder i praksis, at det der opnås er en funktionalitet svarende til det de store søgemaskiner allerede leverer, men afgrænset til dansk netmateriale og som noget nyt suppleret med automatiske berigelser.

Det publiceringsfilter, der eksisterer på det fysiske område, eksisterer ikke på netområdet – alt kan nemt lægges ud på nettet af alle. Det gør selektion til et regulært behov på nettet. Sangeren Peter Gabriel, der har åbnet nettjenester på musik- og filmområdet, formulerer det således:

”I første omgang gav internettet os mulighed for at finde alt indhold - når som helst og hvor som helst. Men i øjeblikket, hvor mange af os er ved at drukne i valg, så har vi brug for redskaber til at foretage gode valg”

(Gabriel, 2008)

Det er en oplagt opgave for nationalbibliografien at påtage sig selektionsopgaven og lade de store søgemaskiner tage sig af fuldstændigheden søgemæssigt. Hvis en bruger ønsker fuldstændighed, må man formode, at det primært vil være total fuldstændighed og ikke kun begrænset til danske forhold.

5.12 Fritekstsøgning

Denne funktionalitet fås allerede via søgemaskinerne. Ren fritekstsøgning opfylder på ingen måde de krav, FRBR opstiller som basiskrav til nationalbibliografi. Fritekstsøgning understøtter kun tekster - andre materialetyper såsom musik og film giver ikke mulighed for denne type søgning.

5.13 Producentens egne metadata

Det er oplagt, at der er fokus på metadata fra producenterne side, men i hvor høj grad dette har betydet markant udbredelse og så væsentlige forbedringer af metadata, at man kan basere nationalbibliografien herpå, betvivles. Udgiveres incitament til at udforme metadata er primært at blive fundet via de store søgemaskiner – ikke nødvendigvis at lave gode metadata i bibliografisk forstand. Google ignorerer f.eks. metadata med keywords. Bureauer der rådgiver om udformning af hjemmesider er meget opmærksomme på dette. På webbureauets Nam Nam Designs hjemmeside¹¹ ses følgende anvisning:

¹¹ <http://www.namnamdesign.dk/artikler/51/hvordan-virker-google-hvad-er-sogemaskineoptimering>

Når du optimerer til Google skal du bl.a. være opmærksom på følgende ting:

- *meta keywords er værdiløse*
- *udformning af title-tags*
- *linkpopularitet*
- *indhold*

Tidligere var en top 20 placering på Google nogelunde sikker, hvis du plottede 20-30 gode meta-keywords ind i din html-kode. Det gælder ikke længere! Google ignorerer meta-keywords, og det er derfor ikke her du skal bruge dine kræfter.

(Nam Nam Design, 2004)

5.14 Automatisk berigelse og automatisk klassifikation

(Tildele emneord ved hjælp af algoritmer, klassifikationsnotationer, forbedre ophavsangivelser o. l.)

Om noget skal beriges og et eventuelt omfang heraf, foreslås afgjort ud fra, om der i forvejen er metadata til materialet. En sådan selektion virker usystematisk og tilfældig. Materialer, der ikke indeholder brugbare metadata, mistes.

Kvaliteten af automatisk berigelse betvivles. Det beskrevne forslag mangler konkretiseringer med hensyn til hvilket niveau, man præcis kan vente sig af registreringen. At et materiale er foranderligt i sin levetid, er ikke et nyt fænomen. Det har altid været gældende for periodika. Dertil kommer, at den foreslåede automatik til registrering pt. ikke er eksisterende. Udviklingen af sådanne redskaber må være at betragte som et forskningsprojekt på internationalt plan og ikke en driftsopgave for den danske nationalbibliografi.

5.15 Figuren

Figuren afspejler den sammensmeltning af biblioteks- og arkivområdet, der findes på nettet. Bibliotekernes primære rolle er ikke en arkivfunktion men at være biblioteker i henhold til Lov om biblioteksvirksomhed. Figuren viser, at der allerede nu i 2008 finder en automatisk registrering sted. Der er vel nærmere tale om en høstning med efterfølgende automatisk genererede indekser og ikke en egentlig registrering.

5.16 Konsekvenser i forhold til nuværende optagelsespraksis

ISSN-tildelte periodika foreslås fortsat registreret til ISSN-basen. Motivet herfor må være en forpligtelse i forhold til ISSN-standarden. Konsekvensen må i så fald være den samme for ISBN-tildelte publikationer, hvor tilsvarende forpligtelser gør sig gældende.

5.17 Adgang

Så længe der ikke er adgang til netarkivet, vil der være et stadigt voksende antal døde links i registreringer set i forhold til det levende net.

Hvis Netarkivet åbnes, vil materialer, der ikke findes på det levende net, mere indgå i søgninger fra Netarkivet. Meget af det, der dør på nettet, er netop uaktuelt historisk materiale, der ikke har bred interesse længere, og det vil dermed virke som støj ved søgning.

Med hensyn til betalingsbelagte ressourcer er det i dag muligt for brugerne, via bibliotekernes lokaliseringer på de bibliografiske poster, at få adgang til eller oplysninger om adgang direkte fra bibliotek.dk. Kulturarvsmodellen understøtter ikke dette. Dette vil sige, at det ikke længere vil være muligt samlet at se fra hvilke biblioteker, man kan få adgang til et bestemt materiale på nettet og samtidig få præsenteret alternative materialetyper. Eksempel: *”hvorfra kan jeg som bruger låne en netlydbog, og hvis dette ikke er muligt, kan jeg så låne den i fysisk form?”*

5.18 Brugerne

Fokus for kulturarvsmodellen er fuldstændigheden. Formidlingsprincippet i modellen bliver for bredt, og dermed er der ikke tale om et alternativ til de almindelige søgemaskiner. Modellen skaber ikke den overskuelighed for brugeren, som må være et af målene for en nationalbibliografi.

6. Sammenfatning

Nærværende rapport afspejler nødvendigheden af, at der bliver taget stilling til, om hovedformålet med Nationalbibliografien på netområdet fremover skal være:

- et redskab, der primært anvendes til kulturarvsbevaring (Lov om pligtaflevering af offentliggjort materiale).
- et redskab, der primært anvendes til at dokumentere og understøtte formidling af relevante ressourcer til den almene offentlighed (Lov om Biblioteksvirksomhed)

Forskellene ligger i fuldstændighedsprincip contra brugerorienteret selektion, i vurderingen af hvad der er tilstrækkelig formidling, i vurderingen af hvad der er tilstrækkeligt nationalbibliografisk niveau samt i holdninger til adgang, herunder samspil med eksisterende infrastruktur.

Ved den fuldstændige model lægges der vægt på, at der gives adgang til samtlige internetressourcer. Dette skal gøres ved høstning af metadata. Dertil skal der udvikles redskaber til berigelse af metadata samt hjælp til automatiske rutiner i forhold til tilføjelse af emneord m.m.

Ved den selektive model lægges der vægt på, at de store mængder af materialer på nettet gør det nødvendigt at foretage selektion med henblik på biblioteksmæssigt brug. Der skal skabes basale sammenhænge mellem registrerede dokumenter, skabes overskuelighed for brugerne, og brugeren skal ikke mindst understøttes i udvælgelse. Nationalbibliografisk selektion sanktioneres i IFLA's GNBDA, Draft 2007.

Det har ikke været muligt for arbejdsgruppen at blive enige om hvilken model, der primært skal arbejdes hen imod. Som en følge heraf har det ikke været muligt at bruge kommissoriet, som det var tiltænkt.

7. anbefaling

En forudsætning for videre arbejde er derfor, at der tages stilling til **hvilken** af de to modeller, der skal efterstræbes og dermed underbygges, dvs. om kulturarvshensyn eller formidling skal vægtes højest.

8. IFLA

Gennem kommissoriet var arbejdsgruppen samtidig blevet bedt om at set på, hvor anvendelig Guidelines for National Bibliographies in the Digital Age (GNBDA), Draft 2007 var i dette arbejde.

Alle gruppens medlemmer har læst rapporten, hvorefter den er blevet diskuteret. Dette har afstedkommet følgende generelle kommentarer:

- Det er tydeligt, at rapporten henvender sig til en bred landekreds på forskellige udviklingsstadier. Dette betyder, at man i Danmark, på de fleste punkter, allerede er i gang
- Der fokuseres meget på tekst i traditionel form
- Guidelines kommer generelt godt rundt og kan bruges som huskeliste til, om alle aspekter er taget i betragtning. Der savnes dog en mere direkte konkret anbefaling i forhold til, hvad der skal registreres og på hvilken måde, særligt i forhold til det elektroniske område
- Brugere bliver opremset, men deres behov prioriteres ikke indbyrdes
- Samlet set er der mere tale om et katalog over muligheder/aspekter frem for guidelines.

9. Litteraturliste

Alsted Research (2005). *Digital kommunikation mellem det offentlige og borgerne*. Videnskabsministeriet, IT-og Telestyrelsen, Finansministeriet, Den Digitale Taskforce, Amtsrådsforeningen, Kommunernes Landsforening og Københavns Kommune. Lokaliseret d. 10. juni 2008 på www: <http://www.videnskabsministeriet.dk/site/forside/publikationer/2005/digital-kommunikation-mellem-det-offentlige-og-borgerne>

Bekendtgørelse af lov om ophavsret. Bekendtgørelse nr. 763 af 30. juni 2006. Lokaliseret d. 10. juni 2008 på www: <https://www.retsinformation.dk/Forms/R0710.aspx?id=12014>

Biblioteksstyrelsen (2005). *Dansk strategi for registrering og formidling af internetressourcer. Rapport udarbejdet for Nationalbibliografisk Udvalg*. Lokaliseret d. 10. juni 2008 på www: http://www.bs.dk/publikationer/rapporter_andre/netressourcer/index.htm

British Library & JISC (Joint Information Systems Committee) (2008). *Information behaviour of the researcher of the future*. Lokaliseret d. 10. juni 2008 på www: http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf

Dansk bibliotekscenter (1999). *INDOREG 2. Registrering af statiske og dynamiske publikationer samt elektroniske tidsskriftartikler*. Lokaliseret d. 10. juni 2008 på www: <http://gammel.dbc.dk/udvikl/indoreg/framepage.htm>

Dansk bibliotekscenter (2002). *domæne.dk og nationalbibliografi? Registrering af netpublikationer, estimering af mængder 2002*. Lokaliseret d. 10. juni 2008 på www: <http://gammel.dbc.dk/nationalbibliografi/domaenedk.htm>

Dansk bibliotekscenter (2004). *domæne.dk og nationalbibliografien 2004*. Registrering af netpublikationer, estimering af mængder 2004. Lokaliseret d. 10. juni 2008 på www: <http://gammel.dbc.dk/nationalbibliografi/domaene.htm>

DBC (1998). *Optagelseskriterier for bøger og periodica, 5. udgave*. Lokaliseret d. 10. juni 2008 på www: http://www.dbc.dk/om_dbc/nationale_opgaver/nationalbibliografi/nationalbibliografi/optagelseskriterier/optagelseskriterier-for-dansk-bogfortegnelse-og

Diget Hansen, Randi (2007). *Assisteret indeksering på DBC*. Temadag Danmarks Biblioteksskole 4. oktober 2007. Lokaliseret d. 10. juni 2008 på www: <http://www.db.dk/kurser/temadage/e2007/udviklingstendenser/randi%20hansen%20ppt.ppt#283,1>, Assisteret indeksering på DBC

Den Digitale Taskforce (2006). *Personas til den fællesoffentlige borgerportal*. Lokaliseret d. 10. juni 2008 på www: <http://borger.dk/lib/docs/Personasrapport%20v%201.1.pdf>

Forskningsministeriet (1994). *Dybkjær Rapporten. Informationssamfundet år 2000*. Lokaliseret d. 10. juni 2008 på www: <http://videnskabsministeriet.dk/site/forside/publikationer/1996/dybkjaer-rapporten-info-samfundet-aar-2000>

Forskningsministeriet (1997). *Information til tiden – Betænkning nr. 1342, 1997*. Lokaliseret d. 10. juni 2008 på www: <http://videnskabsministeriet.dk/site/forside/publikationer/1997/information-til-tiden-betaenkning-nr-1342>

IFLA (1998). *Functional Requirements for Bibliographic Records*. Lokaliseret d. 10. juni 2008 på www: <http://www.ifla.org/VII/s13/frbr/frbr.pdf>

IFLA (1998). *Rekommendationer fra International Conference on National Bibliographic Service*. København. Lokaliseret d. 10. juni 2008 på www: <http://www.ifla.org/VI/3/icnbs/fina-dk.pdf>

Guidelines for National Bibliographies in the Electronic Age. *IFLA Working Group on Guidelines for National Bibliographies (Draft 2007)*.

Informationsordbogen - ordbog for informationshåndtering, bog og bibliotek. Lokaliseret d. 10. juni på www: <http://www.informationsordbogen.dk/>

The International DOI Foundation. *The DOI[®] System*. Lokaliseret d. 10. juni på www: <http://www.doi.org/>

International Organization for Standardization (2007). ISO 3297:2007. *Information and documentation - International standard serial number (ISSN)*. Lokaliseret d. 10. juni 2008 på www: http://www.iso.org/iso/catalogue_detail?csnumber=39601

Kulturministeriet (2003). *Udredning om bevaring af kulturarven*. Lokaliseret d. 10. juni 2008 på www: http://www.kum.dk/graphics/kum/downloads/Publikationer/Bevaring_af_Kulturarven.pdf

Lov om biblioteksvirksomhed. Lov nr. 340 af 17. maj 2000. Lokaliseret d. 10. juni 2008 på www: <https://www.retsinformation.dk/Forms/R0710.aspx?id=11770>

Lov om behandling af personoplysninger. Lov nr. 429 af 31. maj 2000. Lokaliseret d. 10. juni 2008 på www: <https://www.retsinformation.dk/Forms/R0710.aspx?id=828>

Lov om offentlige Bibliotekers Ret til Frieksemplarer af Tryksager m.m. 2. maj 1902. Genoptrykt som aktstykke nr. 47 i: Horstbøll, H. & Lauridsen, J.T. (red.) (1998). *Den trykte kulturarv: Pligtaflevering gennem 300 år*, 16. København: Det Kongelige Bibliotek, Statsbiblioteket & Museum Tusulanum

Lov om pligtaflevering af offentliggjort materiale. Lov nr. 1439 af 22. december 2004. Lokaliseret d. 10. juni 2008 på www: <http://www.pligtaflevering.dk/loven/index.htm>

Lov om pligtaflevering af udgivne værker. Lov nr. 423 af 10. juni 1997. Lokaliseret d. 10. juni 2008 på www: <https://www.retsinformation.dk/Forms/R0710.aspx?id=85005>

Nam Nam Design (2004). *Hvordan virker Google? Hvad er søgemaskineoptimering?* Lokaliseret d. 10. juni 2008 på www: <http://www.namnamdesign.dk/artikler/51/hvordan-virker-google-hvad-er-sogemaskineoptimering>

Wiggins, B (2007). Selection Principles. I: IFLA Working Group on Guidelines for National Bibliographies (Draft 2007). *Guidelines for National Bibliographies in the Electronic Age*

Villum Jensen, Kasper (2008). *Filter skal skaffe bedre film og musik på nettet*. Interview med Peter Gabriel. Lokaliseret d. 10. juni 2008 på www:

http://www.comon.dk/news/filter.skalkaffe.bedrefilm.og.musik.paa nettet_36358.html