

**domæne.dk og nationalbibliografien 2004**  
registrering af netpublikationer, estimering af mængder

**Dansk BiblioteksCenter**  
2004

[ISBN 87-552-2588-8](#)



## INDHOLD:

### [Resumé](#)

### [Dansk BiblioteksCenters arbejde med netpublikationer:](#)

#### [Optagelseskriterier](#)

#### [Vurdering af omfang](#)

#### [Valg af registreringsformater](#)

### [Overblik over udførte registreringer af netpublikationer i Danmark](#)

### [Nærmere gennemgang af tællingen:](#)

#### [Datagrundlag](#)

#### [Arbejdsmetode](#)

#### [Resultat](#)

#### [Sammenligning med første tælling og relevans af fremtidige tællinger](#)

#### [Bilag 1. Brugerrapport fra stikprøveprogrammet](#)

#### [Bilag 2. Sammenligning af undersøgelserne i 2002 og 2004](#)

## Resumé

### ***Hvor mange registreringsværdige publikationer i 2004?***

Formålet med denne rapport er at belyse omfanget af registreringsværdige netpublikationer i 2004 på den danske del af internettet.

### ***Myten***

At internettet rummer helt overvældende mængder af information er i forskellige sammenhænge brugt som argument for at det er en uoverskuelig opgave at registrere netpublikationer.

### ***Undersøgelserne***

Vi valgte derfor i 2002 at lave en statistisk undersøgelse af omfanget af registreringsværdige netpublikationer på den danske del af internettet. Netpublikationerne skulle opfylde optagelseskriterierne til nationalbibliografien.

I 2002 bestod det statistiske materiale af 2000 domæner. Vi har i foråret 2004 lavet endnu en statistisk undersøgelse hvor omfanget af domæner er øget til 5000. Undersøgelsen fra 2002 kan ses her<sup>1</sup>

### ***Resultatet***

Undersøgelsen viser at omfanget af netpublikationer der opfylder optagelseskriterierne til nationalbibliografien, fortsat er til at overskue: 2002-undersøgelsens bud på en årlig tilvækst: ca. 4000 netpublikationer hvor formentlig en meget stor del allerede er registreret, er blevet bekræftet. Tallet er naturligvis behæftet med en statistisk usikkerhed og kan være enten lidt højere eller lidt lavere.

## Dansk BiblioteksCenters arbejde med netpublikationer

Dansk Bibliotekscenter har siden 1996 arbejdet med at udvikle et koncept for nationalbibliografi, omfattende netpublikationer.

Det gælder opsporing, optagelseskriterier, registreringsniveau og formater.

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 publikationer 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. I begge tilfælde er formålet helt overordnet at sikre adgang til materiale af værdi.

---

<sup>1</sup> Se rapporten her:  
<http://www.dbc.dk/nationalbibliografi/domaenedk.htm>

## Optagelseskriterier

### **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. Det betyder i praksis 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.

### **Optagelseskriterier**

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

Optagelseskriterierne for netpublikationer omfatter bl.a. følgende:

- Publikationer der har kommerciel, intern, meget lokal eller privat karakter, optages ikke<sup>2</sup>
- 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.

### **Mængder**

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.

### **Bestandighed**

Siden vi påbegyndte registreringen af internetpublikationer, har vi frem til medio maj 2004 registreret 8328 netpublikationer hvoraf størstedelen falder ind under nationalbibliografien. I samme periode har vi kun konstateret 398 (varigt) forsvundne netpublikationer. Der er tale om en høj grad af stabilitet for registrerede netpublikationer idet de forsvundne publikationer kun udgør 4,7% af det samlede antal registreringer.

---

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

Idag modtager vi ugentligt en automatisk produceret liste over URLer der ikke virker. Denne liste gennemgås manuelt: en lang række flyttede publikationer opspores hvorefter registreringerne rettes til. Denne arbejdsgang bliver overflødiggjort hvis der etableres en URN-løsning eller en automatisering af URL-tjekket.

### ***Specielt om hjemmesider***

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 institutionerne kun virker på lokalt plan.

## **Vurdering af omfang**

### ***5000 domæner er gennemgået***

Til grund for vurdering af omfanget af det årlige antal registreringsværdige netpublikationer er foretaget optælling ud fra en stikprøve på 5000 tilfældigt genererede domæner (second level domains) fra DK-domænet. Med ovenfor nævnte optagelseskriterier som redskab har seks DBC-medarbejdere i marts 2004 gennemgået en stikprøve på 5000 (ca. 1 %) af de i alt 500.146 domæner fra hostmaster.dk.

### ***Den statistiske model***

Forud for denne optælling lå dels en statistisk model og dels en forundersøgelse på et begrænset antal domæner. Den første undersøgelse som vi foretog i foråret 2002 byggede på 2000 domæner, men da vi efterfølgende har udviklet et langt bedre og hurtigere arbejdsredskab<sup>3</sup> til arbejdet, har vi i nærværende undersøgelse udvidet stikprøven til 5000 domæner.

### ***Masser af utilgængelige websider***

Også denne undersøgelse afslørede at en forbavsende stor del af domænerne ikke var aktive, dvs. de tilhørende websider var utilgængelige, ikke eksisterende eller meget mangelfulde (f.eks. reserveret til eventuel fremtidig brug). I alt 3919 domæner blev frasorteret maskinelt. De resterende 1081 domæner blev gennemgået manuelt.

### ***Resultatet***

Det endelige resultat blev at 43 domæner havde et relevant registreringsværdigt indhold og i alt 349 publikationer blev udpeget som registreringsværdige. Heraf var 313 publikationer boglignende. Den type har været en del af nationalbibliografien siden 1998. 37 dynamiske publikationer i form af databaser og hele hjemmesider blev fundet registreringsværdige. Den årlige tilvækst i domæner har de to sidste år tilsammen været på 116.418

---

<sup>3</sup> Læs om programmet på s. 8

Holder denne undersøgelsestendens - og den er blevet bekræftet i to undersøgelser - betyder det at der aktuelt på .dk områdets 500.146 domæner (pr. 1. februar 2004) er ialt ca. 34.840 enheder registreringsværdige til nationalbibliografien, med en anslået årlig tilvækst på ca. 4000 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.

### ***Nationalbibliografiske registreringer i 2003***

I 2003 registrerede DBC 2406 statiske netpublikationer, hvilket med en snusfornuftig (men ikke videnskabelig statistisk) vurdering indikerer at tallet på ca. 4000 i årlig tilvækst er realistisk.

### **Valg af registreringsformater**

Nationalbibliografien for de traditionelle medier katalogiseres efter gældende katalogiseringsregler i danMARC2-formatet. Den praksis der har udviklet sig<sup>4</sup> 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.

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 materialer<sup>5</sup> med enkelte justeringer.

### ***Valg af Dublin Core***

Alle andre typer af netpublikationer anbefales forsøgsvis registreret i DC:

F.eks. databaser, hjemmesider, enkelte artikler/samlinger af artikler på hjemmesider og fragmenter fra hjemmesider.

---

<sup>4</sup> Hos Dansk BiblioteksCenter, i Danmarks Elektroniske Forskningsbibliotek og i de forskellige netjenester

<sup>5</sup> [http://www.kat-format.dk/kat\\_regl/Kat2.1f2.htm](http://www.kat-format.dk/kat_regl/Kat2.1f2.htm)

## Overblik over udførte registreringer af netpublikationer i Danmark

Overblik over hvad der er registreret pr. primo maj 2004








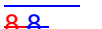
Projekt/aktivitet	Antal registreringer
<b>DBC – udelukkende danske ressourcer</b>	
DBC's registrering af statiske netpublikationer (en del af nationalbibliografien)	7.659
DBC testregistreringer af andet materiale (emnepakker mm i Dublin Core)	1.649
DBC indeksering af artikler i danske aviser og tidsskrifter	14.128
DBC indeksering af anmeldelser i danske aviser og tidsskrifter	4.823
DBC registreringer i dM2 bl.a. på forslag af biblioteker (Netkat mm)	669

## Nærmere gennemgang af tællingen

### Datagrundlag

Stikprøven bestod af 5000 af de i alt 500.146 domæner i DK-domænet pr 1. februar 2004.

Den årlige tilvækst i antal domæner er:

	1. —❖— 1997:	33.523
	2. —❖— 1998:	34.229
	3. —❖— 1999:	65.820
	4. —❖— 2000:	107.419
	5. —❖— 2001:	103.065
	6. —❖— 2002:	43.882
	7. —❖— 2003:	72.536
	8. —❖—	

### Arbejdsmetode

Programmet som dannede rammen for vores arbejde er udarbejdet af Martin Krøis<sup>6</sup>.

Via programmet udtages en stikprøve på den ønskede størrelse – i dette tilfælde 5000 domæner.

Programmets indbyggede robot frasorterede dernæst automatisk de domæner i stikprøven som ifølge statuskode<sup>7</sup> var uden indhold eller på anden måde utilgængelige.

Dernæst udpegede vi domæner uden reelt indhold (typisk webhoteller) og satte robotten til at frasortere denne type ved sammenligning.

Det betød at der automatisk blev frasorteret 3.534 domæner således at den manuelle gennemgang blev begrænset til 1.466 domæner.

Når programmet præsenterede os for et domæne med indhold, vurderede vi som det første om indholdet var relevant i registreringsøjemed.

<sup>6</sup> Ansat i DBC som programmør i 2. halvår af 2003

<sup>7</sup> Se oversigt over anerkendte statuskoder i

<http://www.faqs.org/rfcs/rfc2616.html>



Såfremt indholdet var relevant var næste skridt at afgøre hvorvidt det var statisk, dynamisk, en kombination eller om hele domænet skulle registreres.

Afslutningsvis har programmet et rapporteringsmodul som samler samtlige brugeres resultater i en overskuelig oversigt. Se evt. bilag 1.

## Resultat

5000 domæner er omfattet af undersøgelsen.

Ud af de 5000 domæner fandt vi

1081 med indhold

43 af disse domæner med

registreringværdigt indhold:

6 domæner med statiske boglignende publikationer:

313 registreringsværdige publikationer

1 domæne indeholdt en dynamisk publikation

35 domæner var relevante til en selvstændig registrering af hjemmesiden

(Af disse er seks domæner allerede registreret)

43 relevante domæner af 5000 = 0,86%

350 relevante registreringsenheder fra 43 relevante domæner = ca. 8,1 relevante registreringsenheder pr. relevant domæne.

En fremskrivning af disse resultater ville sige at ca. 4300 domæner (0,86% af 500.146) ville være relevante.

Og disse ville indeholde i alt ca. 34.840 relevante registreringsenheder.

De ovenfor udførte tællinger og beregninger har jo været rettet mod *det samlede nuværende net* og siger således ikke noget om tilvækst.

Hvis vi fx antager en tilvækst på ca. 60.000 domæner årligt, ender vi med en tilvækst på godt 4000 publikationer pr. år.

Det virker som om de 5000 domæner vi har gennemgået giver et realistisk billede af den danske del af nettet.

Som forventet var en meget stor mængde erhvervsdomæner uden relevant indhold.

Der er desuden stadig en betydelig mængde tomme, reserverede eller utilgængelige domæner – mere end 78 %.

## Sammenligning med første tælling og relevans af fremtidige tællinger

Stikprøven er vokset fra 0,6% i 2002 til 1% i 2004.

I 2004 er næsten 78,4% uden reelt indhold, mod 63,4% i 2002. Er denne forøgelse et tegn på at nettet "sander til"?

I 2004 har 0,86% af domænerne relevant indhold, mod 1,75% i 2002.

Til gengæld fandt vi i gennemsnit flere registreringsværdige publikationer pr. domæne i 2004 – nemlig 8,1 mod 1,7 i 2002. Forklaringen er at stikprøven i 2004 indeholdt to domæner med hhv. 188 og 80 registreringsværdige publikationer. Netop fundet af domæner med relativt store mængder registreringsrelevant indhold kan forskyde resultatet fra stikprøve til stikprøve.

## **Bilag 1:** **Brugerrapport fra stikprøveprogrammet**

### **Antal frasorterede domæner i alt 3919, herunder:**

Antal frasorterede domæner via statuskoder: 411  
Antal manuelt frasorterede domæner: 385  
Antal frasorterede domæner via sammenligning efter manuel udpegning: 3123

### **Antal domæner med indhold i alt 1081, herunder:**

Antal domæner med ikke-relevant indhold: 1038  
Antal domæner med relevant indhold: 43  
Herunder:  
Antal domæner med statiske boglignende publikationer: 6  
Antal domæner med fortrinsvis dynamiske publikationer: 1  
Antal domæner hvor hele domænet bliver registreret: 35  
Antal domæner med fragmenter: 0  
Antal domæner med en kombination af ovenstående: 1

### **Statiske boglignende publikationer (A):**

Domains antal med 1 publikationer: 1  
Domains antal med 8 publikationer: 1  
Domains antal med 14 publikationer: 1  
Domains antal med 17 publikationer: 1  
Domains antal med 80 publikationer: 1  
Domains antal med 188 publikationer: 1

### **Fortrinsvis dynamiske publikationer (B):**

Domains antal med 1 publikationer: 1

### **Hele hjemmesiden (C):**

Domains antal med 1 publikationer: 35

### **Kombination af A,B,C,D:**

Domæner med 5 statiske boglignende publikationer og hele hjemmesiden: 1

## Bilag 2: Sammenligning

	2004	2004	2004	2002	2002	2002
Subdomænestørrelse	<b>500.146</b>			<b>351.792</b>		
I alt behandlet	<b>5000</b>	<b>1%</b>		<b>2000</b>	<b>0,6%</b>	
Antal frasorterede domæner i alt	<b>3919</b>	<b>78,4%</b>		<b>1268</b>	<b>63,4%</b>	
Herunder:						
Antal frasorterede domæner via statuskoder	411					
Antal manuelt frasorterede domæner	385			1268 <sup>8</sup>		
Antal frasorterede domæner via sammenligning efter manuel udpegning	3123					
Antal domæner med indhold i alt	<b>1081</b>	<b>21,6%</b>		<b>732</b>	<b>36,6%</b>	
Herunder:						
Antal domæner med ikke-relevant indhold	1038					
<i>Antal domæner med relevant indhold</i>	<b>43</b>	<b>0,86%</b>		<b>35</b>	<b>1,75%</b>	
Herunder:						
Antal domæner med statiske boglignende publikationer	6					
Antal domæner med fortrinsvis dynamiske publikationer	1					
Antal domæner hvor hele domænet bliver registreret	35					
Antal domæner med fragmenter	0					
Antal domæner med en kombination af ovenstående	1					
Herunder:						
Statiske boglignende publikationer (A)						
Antal domæner med 1 publikation	1					
Antal domæner med 8 publikationer	1					
Antal domæner med 14 publikationer	1					
Antal domæner med 17 publikationer	1					
Antal domæner med 80 publikationer	1					
Antal domæner med 188 publikationer	1					
<b>Antal statiske publikationer i alt (A+5 fra kombination)</b>	<b>313</b>			<b>45</b>		
Fortrinsvis dynamiske publikationer (B)						
Antal domæner med 1 publikation	1					
Hele hjemmesiden (C)						
Antal domæner med 1 publikationer (dvs. selve hjemmesiden)	35					
<b>Antal dynamiske publikationer i alt (B,C+1 fra kombination)</b>	<b>37</b>			<b>31</b>		
Kombination af A,B,C,D						

<sup>8</sup> Der var kun manuel sortering i 2002.

Domæne med 5 statiske boglignende publikationer og hele hjemmesiden	1				
<b>Registreringsværdige publikationer i alt</b>	<b>350</b>		<b>8,1 publ.pr.dom.</b>	<b>76</b>	<b>1,7 publ.pr.dom.</b>