

# Permanente identifikatorer for digitale objekter

Uddrag af DEF rapporten:

Kortlægning af standardiseringsmæssige tiltag og behov for samme

Af Adrian Price og Kasper Løvschall

## Adrian Price

Bibliotekskonsulent  
Danmarks Veterinær- og Jord-  
brugsbibliotek  
ap@kvl.dk

## Kasper Løvschall

Udviklingskonsulent  
Aalborg Universitetsbibliotek  
kl@aub.aau.dk



## Permanente identifikatorer ("persistent identifiers")

For at sikre at objekter i et arkiv bliver ved med at være tilgængelige, er det afgørende, at man kan garantere en entydig og vedvarende identifikation og adressering af objektet. En Uniform Resource Locator (URL) er den mest udbredte metode i dag, hvor en ressources URL bruges til både at identificere (navngive) ressourcen, angive hvor ressourcen befinder sig, objektets adresse, samt den protokol som skal anvendes til at få adgang til ressourcen (f.eks. http, ftp osv.).

At en URL kombinerer både identifikation og adressering, anses for at være URL'ens svaghed: Hvis et objekt flyttes til en ny adresse, f.eks. en anden server, vil objektet tildeles en ny URL, som samtidig er en ny identifikation, og den vedvarende adgang til objektet vil blive svækket. Der er behov for et system, som klart og vedvarende adskiller identifikation/navngivning og adressering/lokalisering. Et objekt skal tildeles en entydig identifikator, et navn, som objektet beholder "hele livet". Til denne identifikator kan kobles objektets aktuelle adresse og i tilfælde af, at objektet flyttes, kan den nye adresse kobles til den uændrede identifikator, og dermed fastholdes adgang over tid. Som erstatning for det nuværende URL-baserede lokaliseringssystem arbejdes der på at udvikle og indføre metoder og systemer, som bedre vil sikre fremtidig adgang til netressourcer. Der udvikles på metoder og systemer til indførelsen af det, som

oftest betegnes "persistent identifiers" på engelsk, permanente identifikatorer (PI).

Indførelse af et system, som bedre vil sikre adgang til objekter over tid, er ikke kun vigtigt for "isolerede objekter", men er også af betydning for linkning mellem objekter: f.eks. fra en reference i en artikel til selve artiklen, styring af adgang til ressourcer i forbindelse med betaling for benyttelse og brug af objekter i flere forskellige sammenhænge, f.eks. i e-lærings miljøer<sup>1</sup>. En sikring af adgang til netressourcer er således et afgørende element i opbygning af en "digital infrastruktur".

## Metoder og systemer til permanente identifikatorer

I løbet af de sidste cirka ti år er der blevet arbejdet på forskellige metoder og systemer, som har til formål at sikre vedvarende adgang til e-ressourcer. Udgangspunkt for de fleste er erkendelsen af, at det er afgørende, at navngivning og lokalisering holdes skarpt adskilt. Det vil sige, at uafhængigt af hinanden skal det fastlægges, hvordan et objekt skal identificeres ved et entydigt og standardiseret navn, og dernæst skal det fastlægges, hvordan dette navn skal anvendes i lokalisering af objektet. At navngivning og lokalisering er adskilt og uafhængig betyder, at man til enhver tid kan ændre lokalisering af objektet uden at svække adgang.

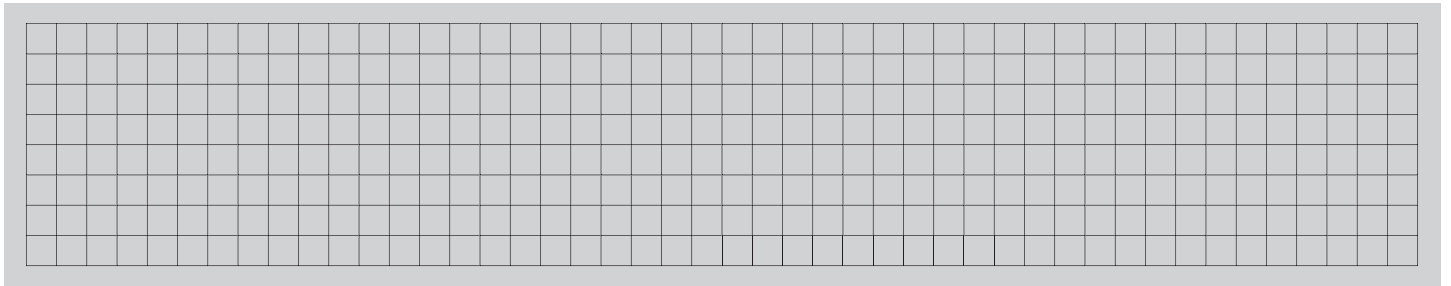
Det er også vigtigt at holde navngivning og adressering adskilt, da infrastrukturen for vedvarende adressering (lokalt og globalt) endnu ikke er fuldt udbygget

eller afgjort. Til indførelse af PI'er er der navngivningskonventioner til udformning af identifikator (en meget udbredt konvention er f.eks. Universal Resource Name (URN), se næste afsnit), og der findes softwaresystemer, som anvender standardiserede navngivningskonventioner, og som kan implementeres til adressering af digitale objekter (f.eks. Handle systemet, se næste afsnit). Der findes metoder og systemer til etablering af infrastrukturen.

## Resolution

På grund af adskillelsen mellem identifikation og adressering er der behov for en mekanisme, som giver adgang til ressourcen via dens identifikator. Denne mekanisme betegnes på engelsk "resolution". DNS (Domain Name System) er sådan en "resolution" mekanisme, som via et domænenavn, f.eks. www.bibliotek.dk, giver adgang ("resolves") til en bestemt vært med et bestemt IP nummer.

Der er endnu ikke implementeret en global resolution service for andre protokoller end til almindelige URL'er (hostnames), så PI'er i dag skal kobles til en almindelig URL, som så anvender DNS-systemet til at give adgang til objektet. Fordi DNS systemet er så udbredt, bruger de fleste PI systemer i dag http-protokolens redirection til at omstille til objektets aktuelle lokale. Denne situation forventes ikke at ændre sig på kort sigt, og det vil også blive svært at ændre på længere sigt, da DNS systemet er så udbredt.



Der skal også nævnes et andet svagt punkt, dog af mindre betydning: almindelige internetbrowsere, hvor kun "almindelige" URLs kan anvendes i dag. Det vil sige, for at anvende et PI i en browser, skal identifikatoren kobles til en almindelig URL, hvor der så anvendes http-protokol til at "resolve" objektet til en adresse. At ændre denne browserbegrænsning, hvor der vil være mulighed for at anvende andre resolutionsprotokoller, vil kræve enighed om en ændring i browsere, som skal gennemføres af softwareproducenterne. Det forudsætter, at man er blevet enig om hvordan en "resolutionsprotokol" skal fungere – og formentlig på globalt plan. (En midlertidig løsning vha. en browser-plugin, hvor plug-in sørger for at oversætte til http, er mulig og er blevet afprøvet, men denne løsning er uholdbar i længden).

PI området deler sig i konventioner for navngivning af objekter alene og konventioner for navngivning og systemer for resolution af disse objekt-ID'er til deres lokalisering. I det følgende beskrives kort de muligheder, som findes i dag, og i fig. 1 nedenfor er en oversigt over disse konventioner og systemer. Disse konventioner og systemer beskrives i de efterfølgende afsnit.

#### Uniform Resource Name – URN [URN]

**Figur 1. Navngivningskonventioner og PI systemer**

| Navngivningskonventioner | "Actionable-systemer" |
|--------------------------|-----------------------|
| URN                      |                       |
| HANDLE                   |                       |
| DOI                      |                       |
|                          | PURL                  |
|                          | POI                   |
| ARK                      |                       |

\* I et "actionabelt system" er resolvingmekanismen (lokalisering) specificeret som en integreret del af systemet



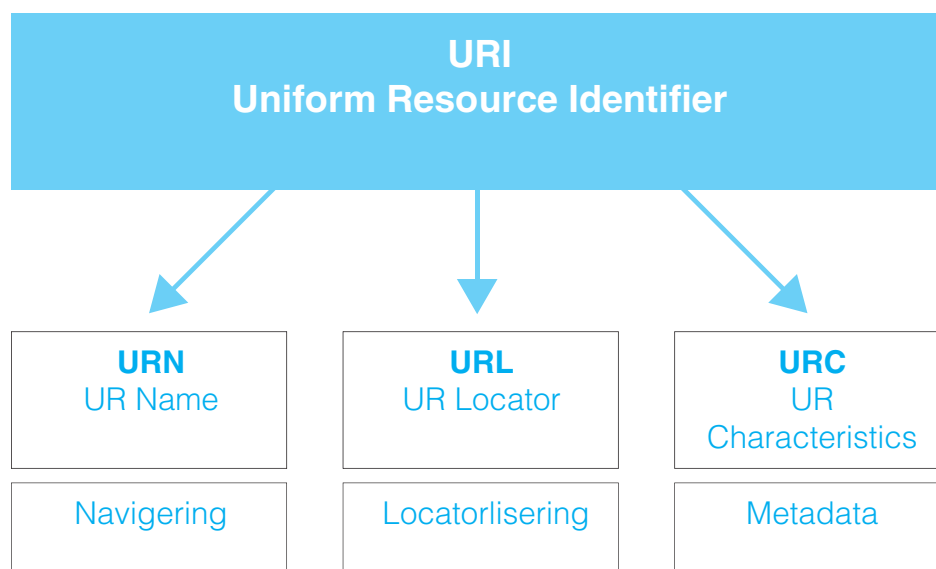
Uniform Resource Name – URN – er en navngivningskonvention udviklet af Internet Engineering Task Force Network Working Group og fremsat i 1992. URN er nok den mest udbredte konvention for navngivning af ressourcer og anvendes af en række andre systemer.

Sammenhæng for URN er det "klassiske", men på nuværende tidspunkt ukomplette (eller snarere fragmenterede) udgangspunkt:

Figuren foroven præsenteres her, fordi "grundsystematikken" bag URI stadig er vigtig at holde for øje. Navngivning, lokalisering og metadata (de tre nederste elementer i figuren) udgør stadig grundpillerne i PI'er, selv om billedet i dag er mere fragmenteret og ukomplet.

En Uniform Resource Identifier (URI) [URI] specificerer en syntaks og semantik for identificering og for lokalisering af en ressource. En ressource tildeles en formel identifikation via et URN [URN], og ressourcen lokaliseres (adresseres) via en specificeret mekanisme (eller protokol) i en URL [URL]. (De mest udbredte protokoller er f.eks. http, ftp osv.)

Fordi et URN kun identificerer (navngiver) et objekt, skal der kobles en lokaliseringsmekanisme til, dvs. der kræves en såkaldt "resolver", som kobler objektets URN til ressourcens adresse. Der har været flere forsøg på at udvikle et resolution system til URN, sidst i 2001 med Dynamic Delegation Discovery System (DDDS), hvor det eksisterende Domain Name Service (DNS) foreslås anvendt til



**Figur 2. URI og dens grundlæggende karakteristika**

at lokalisere ”resolvers” med information om ressourcerne. DDDS systemet er ikke slået igennem.

#### URN syntaks

Hvis man ser på opbygning af et URN, så består et URN af følgende elementer: urn:nid:nss hvor ”nid” er den registrerede Namespace Identifier, og ”nss” er Namespace Specific String, objektets identifikator, som tildeles af NID. NID registreres hos Internet Assigned Numbers Authority [IANA].

Et eksempel på et URN, hvor NID er det registrerede namespace ”NBN”:

**urn:nbn:fi-fe19981001**

„nbn“ – National Bibliography Number – er den registrerede Namespace Identifier (NID) som definerer syntaksen: urn:nbn:<ISO country code>-<assigned string>. I ovenstående eksempel er: fe19981001 den Namespace Specific String (NSS), som tildeles ressourcen, i dette tilfælde af det finske nationale bibliotek. Det er også muligt at definere ”sub-namespaces” efter NID, og i et URN er de adskilt med et ”:”.

NBN namespace blev oprindeligt udviklet og registreret af det finske nationalbibliotek og siden accepteret af CDNL – Committee of Directors of National Libraries – som et URN-baseret system, som kunne anvendes af nationale biblioteker.

URN specifikation fastlægger syntaksen i opbygningen af et URN, men tager f.eks. ikke stilling til, hvordan de specifikke objektidentifikatorer (NSS) udformes: det bestemmes af NID. Der arbejdes på metoder til f.eks. at anvende ISBN- og

ISSN-identifikatorer i et URN.

#### Relevante URN/NBN aktiviteter

**Die Deutsche Bibliothek (DDB)**, som siden 1998 har koordineret en central indsamling og formidling af tyske online-afhandlinger har, som en del af ”Dissertation Online”-projektet [DISSONLINE], etableret et nationalt ”URN management system”, hvor tildeling og vedligeholdelse af permanente identifikatorer er en af hovedingredienserne [DISSPI]. DDB anvender URNs og har siden september 2001 haft ansvar for URNs tildelt under namespace nbn:de. Til dato er der registreret ca. 14.000 URNs for afhandlinger på vegne af 29 deltagende biblioteker.

DDB’s ”URN management system” har mange interessante elementer. De deltagende biblioteker har adgang til et websystem til tildeling og vedligeholdelse af URNs for de enkelte ressourcer. Ressourcerne tilmeldes DDB systemet, som sørger for en efterfølgende høstning af ressourcen. Kernen er selvfølgelig selve URN ==> URL resolving tjeneste, som giver adgang til ressourcen, men der udføres også regelmæssig linkcheck (daglige) og ”MD5 check sum” kontrol af filerne udføres en gang månedligt.

De enkelte URNs tildeles decentralt af de deltagende universitetsbiblioteker. Selve URN er udformet som et hierarkisk system, hvor det institutionelle tilhørsforhold også er afspejlet:

urn:nbn:de:[designation of library associations]:[official notation of university libraries]-[production

number][P]

hvor ”production number” er et lokalt tildelt løbenummer og ”P” et check-ciffer kalkuleret ud fra hele URN. Check-ciffer kalkuleres og tilføjes URN, når et bibliotek anmelder en ressource i DDBs URN management system. Som det fremgår af ovenstående URN, er der et stærkt ”semantisk” element i udformning af en URN-streng, som kan være problematisk. Vi vender tilbage til dette spørgsmål i afsnittet med anbefalinger.

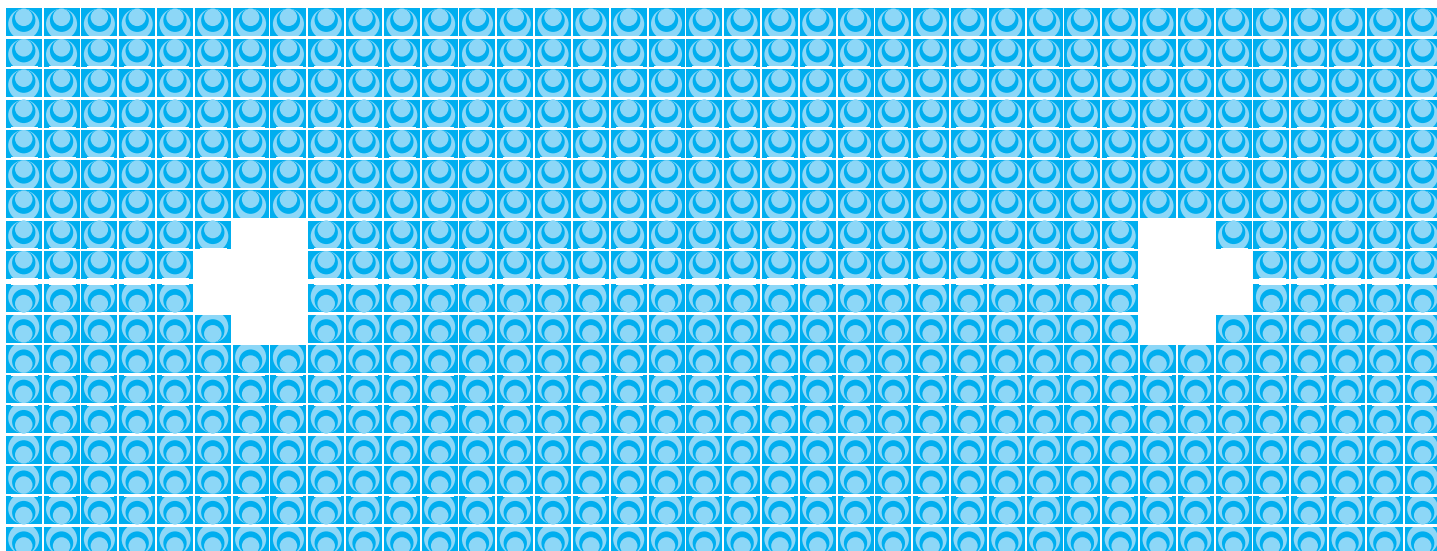
DDBs system kan anvendes med 1:1 relationer, hvor ét URN peger mod ét URL og 1:n relationer, hvor ét URN peger mod flere URLs.

URN resolver system drives af Bibliothekservice-Zentrum Baden-Württemberg (BSZ), udviklet under CARMEN-AP4 projekt, som afsluttedes i 2002. Systemet videreudvikles i 2002-2004 i EPICUR-projekt. [EPICUR]

#### Et nordisk URN resolver service projekt

har fået støtte fra NORDINFO til etablering af en nordisk URN resolver service. [NORDISKURN] Denne URN resolver service vil anvende NBN namespace og forventes at blive taget i brug primo 2005. Projektet udføres af de fem nordiske lande med Uppsala Universitetsbibliotek, Enheten for digital publicering, som projektleder. Der er planer om at anvende open source software, som er blevet anvendt af Kungliga Biblioteket i Sverige siden 2002. Projektperioden er 01.01.04 – 31.12.04.

Første del af projektet vil være skabelse af selve URN resolver service. Med tildeling af færre ressourcer end forventet,



vil projektet søge andre midler til et senere forløb til f.eks. udvikling og indførelse af metoder til flytning af ”arkivpakker” mellem arkiver, anvendelse af METS mm.

Ifølge projektets reviderede tidsplan vil en kravspecifikation være færdig ultimo juli 2004 og implementering af den URN: NBN resolver service påbegyndes primo november 2004.

### En arbejdsgruppe under Dansk

**Standard** blev nedsat i april 2003 til at foretage nogle indledende undersøgelser vedrørende permanente identifikatorer. I arbejdsgruppens rapport [DSPI] fra november 2003 anbefales det (efter en kort gennemgang af de eksisterende modeller), at arbejdet fortsætter med fastlæggelse af den nødvendige infrastruktur for et system baseret på URN navngivningsmodel. Det bemærkes, at arbejdsgruppen ikke kun beskæftigede sig med en infrastruktur under ”nbn namespace”, men tog et bredere perspektiv. Arbejdsgruppens anbefalinger vedrører også kun URN som navngivningskonvention og ikke systemet, som skal bruges til ”resolving” af identifikatorer til objekternes lokalisering.

### Handle System – [HANDLE]

Handle-systemet er udviklet af Corporation for National Research Initiatives (CNRI) og bygger på det arbejde, som er blevet udført til at fastlægge den infrastruktur, som er nødvendig for opbygning af digitale biblioteker/samlinger. (Se f.eks. [Kahn]). Handle-systemet er et globalt, distribueret system for navngivning, vedligeholdelse og adressering (”resolving”)

af identifikatorer for digitale objekter. Disse ”identifikatorer” kaldes ”handles”. [HANDLE]

Systemet er et distribueret system, som består af en Global Handle Registry, hvortil er registreret alle Local Handle Services. Identifikatorer tildeles af de registrerede ”naming authorities”. ”Handles” som tildeles ressourcerne, består af et prefix, som er den navngivende myndighed og et suffix, det tildelte unikke navn (”unique local name”). Prefix og suffix separeres af ”/”. Med Handle-systemet tildeles digitale ressourcer unikke identifikatorer (”handles”), som bruges til at lokalisere og få adgang til ressourcen via en ”resolver-service”.

Udover at være et system for navngivning og lokalisering af ressourcer er det samtidig et software-system, som frit kan downloades.

Handle-systemet er meget udbredt. Der er tildelt cirka 1000 prefixes, hvoraf cirka 700 anvendes af Digital Object Identifier (DOI) System under International DOI Foundation (IDF) (se et følgende afsnit). CNRI står bag udvikling og vedligeholdelse af Global Handle Registry, og registrering er gratis. CNRI forudser, at det i fremtiden sandsynligvis vil blive nødvendigt at indføre betaling for denne registrering af de lokale myndigheder og for vedligeholdelse af systemet.

CNRI tilbyder også en gratis Public Local Handle Service (LHS), hvor alle kan tildele resource-handles, og hvor disse handles kan kobles til et URL. Denne LHS er så ”naming authority” 1030, og brugerne af denne service får tildelt en

”sub-naming authority”, f.eks. 1030.34 (som så er prefix i identifikatoren). Efter man har oprettet sine handles, kan de blive ”resolved” ved at anvende CNRI’s proxyserver eller en plug-in, udviklet af CNRI. (For mere information se <http://hdl.handle.net/4263537/4090>).

### Digital Object Identifiers – [DOI]

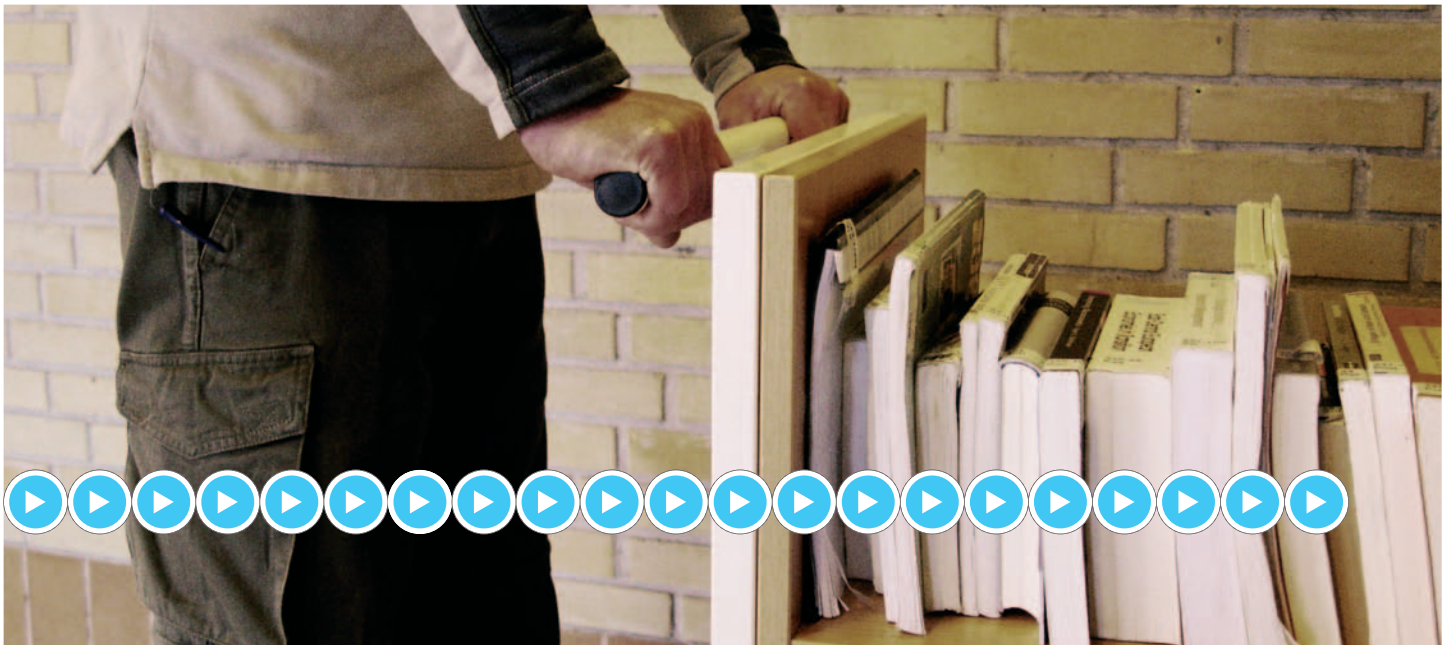
Digital Object Identifiers (DOI) er en specifik implementering af Handle-systemet og var oprindeligt rettet mod forlagsbranchens bekymringer over den ophavsretslige beskyttelse af digitale ressourcer. DOI blev introduceret i 1997.

Som i Handle-systemet består et DOI af et prefix, som identificerer den organisation, som er medlem af The International DOI Foundation (IDF) [IDF], efterfulgt af et ”/”, efterfulgt af et suffix, som er den identifikator, der får tildelt en ressource. Prefixet er den udstedende organisation, og det er denne organisation, som bestemmer, hvordan suffixet (objektets ID) udformes. Det kan f.eks. være et ISBN eller et helt andet identifikationssystem.

Et eksempel: ”10.1045/february2003-featured.collection” er et DOI. ”10.1045” er den registrerede organisation (D-Lib Magazine) og ”february2003-featured.collection” er identifikator for en artikel i februar 2003 nummer med titlen ”Flora and Fauna of the Great Lakes Region”. Ved at anvende IDF ”resolution service” (<http://dx.doi.org/10.1045/february2003-featured.collection>) får vi adgang til artiklen via URL <http://www.dlib.org/dlib/february03/02featured-collection.html>.

IDFs ”resolution service” anvender





Handle-systemet, udviklet af CNRI.

IDF beskriver DOI systemet, som bestående af fire komponenter [PASKIN]: (1) en nummererings- (identifikations) komponent, som er en implementering af URI/URN, (2) en beskrivelseskomponent af objektet, som er blevet identificeret. Til "beskrivelsen" af objektet bruges metadata baseret på <indec>-systemet. [INDECS] Indecs (Interoperability of data in e-commerce systems) er metadata til ophavsretslig beskyttelse af digitale objekter. (3) En "resolution" komponent, som giver adgang til det identificerede objekt og, som allerede nævnt, er baseret på Handle-systemet og (4) "policy" komponent, er den organisation, som står bag, og som sørger for finansiering og udvikling af DOI systemet.

IDF og National Information Standards Organisation (NISO) har samarbejdet om publicering af en standard (ANSI/NISO Z39.84-2000) for DOI. [DOI/NISO]

Forretningsmodellen bag DOI-systemet er en række "registration agencies" som betaler et årligt medlemskab af IDF, den non-profit organisation, som står bag DOI-systemet. Disse "registration agencies" kan så udstede objekt identifikatorer for en pris, som de selv fastsætter. For hver DOI, som udstedes af en registration agency, betales en afgift til IDF, som er pt. cirka fire amerikanske cent.

Som beskrevet, udsprang DOI systemet af forlagsbranchen, men modellen i sig selv forudsætter ikke nødvendigvis en kommerciel forbindelse mellem en informationsbruger og en ressource. Medlemmerne af The International DOI Foundation (IDF) repræsenterer mange

forskellige typer informationsformidlere bestående af kommercielle virksomheder, biblioteker, organisationer osv.

For en kort gennemgang af en række identifikatorsystemer henvises til en værdifuld rapport udarbejdet af The Stationary Office (TSO) [TSO]. Rapporten koncentrerer sig om og anbefaler DOI-systemet for anvendelse i den engelske "JISC community" (Joint Information Services Community), og i rapporten er eksempler for anvendelse af DOI i forskellige sektorer såsom forlagsbranchen, fagportaler, universiteter osv. Herunder også eksempler på hvordan DOI's tilknyttede metadatamodel (<indec> - interoperability of data in ecommerce systems) kan anvendes.

I juni 2004 har TSO meddelt, at de nu vil udstede DOIs gratis til at fremme anvendelsen, især i det engelske "e-Government" regi. (TSOFREE) Finansiering af TSO som "registration agency" vil være i form af en årlig betaling til TSO.

En af de mest velkendte implementeringer af DOI-systemet er CrossRef, en organisation etableret af videnskabelige udgivere, som anvender DOIs til at linke mellem referencer i artikler. Som medlem af CrossRef kan man få sine tidsskriftartikler registreret med en DOI og tilhørende metadata i databasen (som administreres af CNRI). Tilsvarende kan en udgiver forsyne citationer i artikler med outbound links, til de referencer som allerede findes i "DOI databasen".

#### **Persistent Uniform Resource Locator**

#### **– PURL [PURL]**

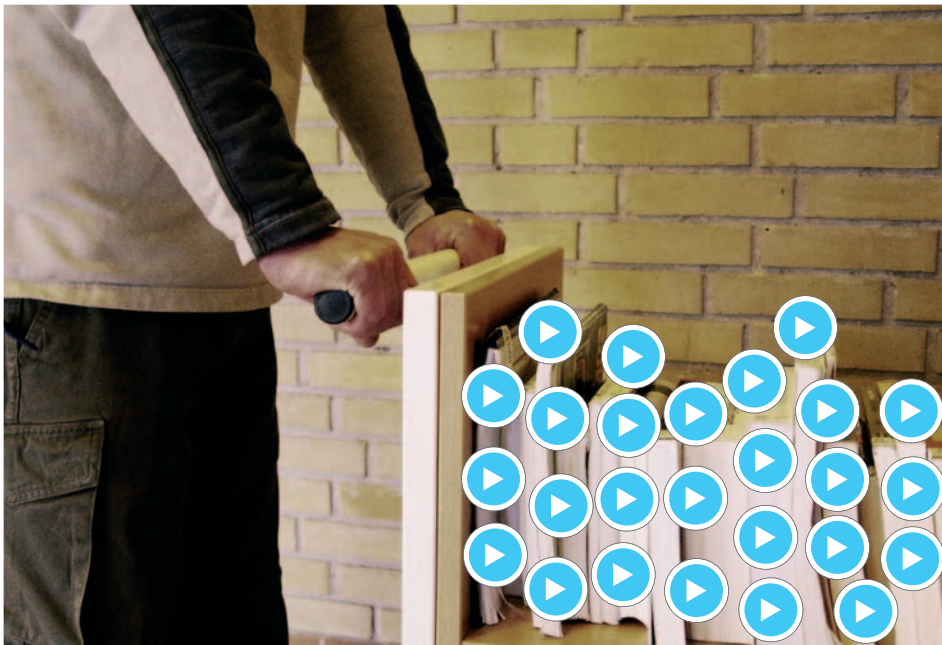
En Persistent Uniform Resource Locator (PURL) er en "almindelig" URL, som bruger http-protokollens redirection til at slå den rigtige URL op i en PURL registreringservice. Objektets "rigtige" URL bliver derefter returneret til brugerens browser, som så giver adgang til objektet. En PURL kan kun håndtere 1:1 relationer, én PURL lokaliserer kun ét objekt. PURL systemet blev introduceret af OCLC i 1995, og softwaren kan downloades fra PURL website.

PURL systemet består således kun af en resolver service og ingen konventioner til at skabe standardisering/best practice i navngivning af objekter.

Pr. maj 2004 er der blevet udstedt ca. 600.000 PURLs og resolved ca. 86.000.000 PURLs .

En PURL service er en mellem-løsning, dvs. overvejende rettet mod adressering/ lokalisering af objekter og uden en stringent objektidentifikation. Som "mellem-løsning" er PURL blevet anvendt, mens URN-systemet har været under udvikling. Den danske PURL service (<http://purl.dk/>) oprettet af DBC, ophørte i starten af 2004.

Der er for nylig blevet fremsat en specifikation for en såkaldt POI – PURL-based Object Identifier – for anvendelse med Open Archives Initiative (OAI) og OAI identifier format i OAI repositorer. [POI] En POI er en URI som anvender OAI identifier som namespace-identifier, f.eks.: <http://purl.org/poi/abcd.dk/doc12345.5>, hvor "abcd.dk" er OAI identifier. Der er også et sæt guidelines for, hvordan POIs kan blive "resolved". [POIRES] PURL



software anbefales til at resolve POIs.

### Archival Resource Key (ARK) – [ARK]

ARK-systemet er et eksperimentelt system, udviklet ud fra den erkendelse, at adgang til digitalt materiale på længere sigt ikke kun garanteres ved at indføre nogle tekniske foranstaltninger, som sikrer digitale objekters integritet og ved at sørge for en infrastruktur til navngivning og adressering af objekterne. Der er endnu et led, som er afgørende: det organisatoriske. Kun en institution med den nødvendige og tilstrækkelige organisatoriske infrastruktur, kan i sidste ende sikre adgang over tid.

ARK identifikator er en unik URL til en bestemt ressource. En "ARK URL" linker til tre ting: (1) metadata tilknyttet et digitalt objekt (2) de(n) fil(er) som udgør objektet og (3) et "commitment statement" fra den institution, som har ansvar for objektet, dvs. den institution som skal sikre adgang til og bevaring af objektet.

ARK syntaksen er:

[<http://NMAH/jark:/NAAN/Name>

hvor NMAH er Name Mapping Authority Hostport, NAAN er Name Assigning Authority Number, et unikt nummer for den instans, som har navngivet objektet, med Name. NMAH, en almindelig URL, kan og vil formentlig ændres og hører således ikke til objektets identifikation, kun lokalisering (på et givent tidspunkt). Det, som følger efter "ark:" label, er objektets permanente identifikator. Når et NMAH for et objekt ændres, skal det aktuelle NMAH findes i NAAN registret (som pt. findes på US National Library of Medicine).

ARK systemet bruger http redirection, og derfor kan almindelige browsere

bruges. Der er pt. ti institutioner, som er registreret som Name Assigning Authority.

Som nævnt ovenfor kan en ARK identifikator returnere objektets tilknyttede metadata og en "commitment statement". Det gøres ved at tilføje hhv. et "?" eller to "??" efter Name. Det metadataformat, som anvendes nu, er et simpelt "label-colon-value"format, Electronic Resource Citations (ERCs), som tager udgangspunkt i Dublin Core.

ARK systemet er udviklet af US National Library of Medicine og er under afprøvning på University of California Digital Library (<http://www.cdlib.org/>). "A founding principle of the ARK is that persistence is purely a matter of service, and is neither inherent in an object nor conferred on it by a particular naming syntax. The best an identifier can do is lead users to those services." [KUNZE]

### Konklusioner og anbefalinger

Som et vigtigt led i sikring af adgang til digitale ressourcer og genbrug af ressourcer på tværs af anvendelsesområder er det afgørende, at der indføres et system, som vil sikre standardiseret navngivning af ressourcer uafhængig af ressourcernes lokalisering, for at opnå vedvarende og global adgang.

Et entydigt navn vil i sig selv ikke give et objekt status som "bevaring sikret for eftertiden", og at separere navnet fra lokalisering vil heller ikke give status som "adgang sikret for eftertiden". Den eneste entitet, som kan garantere denne status, er den instans som til enhver tid har erklæret sig som "objektets kustode". Kravene må

være, at et objekt kan (1) identificeres entydigt, og det ligger i navngivning og metadata, (2) lokaliseres til enhver tid, og (3) at der findes en bevaringsstrategi for en ressource.

Med baggrund i ovenstående gennemgang af de eksisterende konventioner, systemer og mål, kan der opstilles et sæt anbefalinger, som kan indgå i etablering af et permanent identifikator system.

### 1. Det er vigtigt, at adskillelse af navn og lokaliseringsmekanisme fastholdes

Der vil formentlig eksistere flere "konkurrerende" lokaliseringssystemer side om side, så det er vigtigt, at de forskellige objekt-ID'er kan anvendes i de forskellige lokaliseringssystemer, også hvis systemerne skulle skiftes ud, og hvis to systemer skulle anvendes parallelt. (Og der vil formentlig på sigt være behov for et globalt resolutionssystem for forskellige navngivningskonventioner og forskellige resolutionssystemer.)

### 2. URN anbefales som navngivningskonvention

Det anbefales, at man anvender URN navngivningskonventioner og med brug af nbn namespace og evt. flere sub-namespaces (efter nbn:dk).

Der skal nedfældes regler for citationer af URNs, da et URN i sig selv ikke indeholder lokaliseringsmekanismen, og de kan pt. ikke anvendes i browsere uden brug af plug-in.

### 3. Semantik i objektnavngivning

Der er en del diskussion (men ingen entydig konsensus) om, hvorvidt objektnavne skal være helt uden semantik, eller om de til en vis grad kan indeholde et semantisk element. Der synes dog at være konsensus om, at navne bør være forholdsvis neutrale og ikke udelukkende skal bestå af semantiske elementer i form af f.eks. flere institutionsnavne eller en hierarkisk opbygning (f.eks. geografisk eller institutionelle osv.). Det er en fordel, at de er genkendelige og kan "læses" af mennesker. At anvende en resources automatisk genererede MD5 checksum som ID i et URN (f.eks. urn:dbb-2c3ee69s23ecf567ec7yt5re33ace234), som anvendes f.eks. i Finland, synes at være overdrevent neutralt. (Men det kan være fornuftigt at anvende MD5 checksum i andre sammenhænge). Herudover kan der være en fordel i, at en identifikator indeholder mulighed for branding.



Vi foreslår, at man anvender en biblioteks- eller en universitetsforkortelse som første led i objekt ID efterfulgt af et neutralt løbenummer (som f.eks. kan bestå af et årstal efterfulgt af et løbenummer). Biblioteks- eller universitetsforkortelserne (som selvfølgelig skal tildeles entydigt) er knyttet til institutioner, som kan være flygtige og ikke altid entydige, men her henviser de ikke til en aktuell lokalisering, ansvar eller permanens men til objektets oprindelse. Herudover er de genkendelige, har en mnemonisk karakter og indeholder et element af branding.

En anden mulighed er anvendelsen af International Standard Identifier for Libraries and Related Organizations (ISIL) identifikatorer [ISIL]ii. ISIL identifikatorer er alfanumerisk og består af en landekode (specificeret i ISO-3166-1) som præfiks og en "library identifier" som suffiks. Mellem præfiks og suffiks er et "-". ISIL-koden er på maks. 16 tegn og kan variere i længde.

Et automatisk genereret check-ciffer som en del af objekt-ID (som i det tyske DDB projekt) kan anbefales. Og ligesom det tyske DDB system, kunne MD5 checksum og check-ciffer tildeles i et URN management system, når man lokalt registrerer sin ressource-ID.

Inden man anbefaler standardisering af navngivning af objekter, skal man undersøge, om en standardisering er mulig på tværs af lokale systemer.

I et eventuelt kommende follow-up projekt vedr. PI'er kan der fremsættes regler for navngivning. Guidelines for navngivning vil også være at foretrække frem for frit slag.

**4. Ressourcerne skal navngives lokalt**  
ID'er skal tildeles ressourcerne lokalt i henhold til de krav eller guidelines, som udformes.

**5. Relationer til andre instanser/sektorer**  
Da det ikke kun er biblioteker, som har interesse i at etablere en PI-service, er det vigtigt, at biblioteker indgår i et samarbejde med andre interessenter om etablering af en URN-resolver service og om udformning af URN navngivningskonventioner, dvs. systematisering af de relevante namespaces og objekt-ID'er. På denne måde vil man sikre en "organisatorisk stabilitet", som er afgørende.

**6. Service versus "bare" resolution system**

Permanens (adgang til digitale ressourcer over tid) er ikke kun et spørgsmål om teknik men i afgørende grad et spørgsmål om institutioner. (Gen)anvendelse af en voksende mængde digitale ressourcer på tværs af sektorer er ikke bare et spørgsmål om at lokalisere, men også om at kunne identificere og vælge. Permanens og effektiv anvendelse er begge afhængige af robuste services – ikke kun systemer. Det vil formentlig være afgørende, at system(er) indeholder services, som fremmer andet end "bare" lokalisering. Disse services kunne f.eks. være:

- Objektets metadata gøres tilgængelige med henblik på genfindning og udvælgelse.
  - En ID kan adressere til flere ressourcer (f.eks. flere udgaver af samme værk).
  - Der knyttes MD5 check-sum til ressourcerne.
  - Der knyttes et check-ciffer som en del af identifikatorer.
  - Der foretages logning af benyttelse mhp. statistik.
- Hvis man opererer med et system bestående af centrale og decentrale arkiver, f.eks. i forbindelse med pligtaflevering til nationalbiblioteker fra universitetsbiblioteker, vil følgende services være af værdi:
- Regelmæssigt linkcheck af lokale ressourcer.
  - Et system for lokal management af tildelte identifikatorer.
  - Automatisk høstning af ressourcer i forbindelse med f.eks. pligtaflevering.

### 7. Valg af resolution service

Groft sagt er der i dag tre muligheder for etablering af en resolution service: 1. Man anvender eller udvikler egen resolution software, baseret på urn (som f.eks. i det fælles nordiske NORDINFO projekt). 2. Der etableres en DOI registration service i Danmark og 3. Man anvender PURL software (centralt eller decentralt).

Afgørende er, at det system, man vælger, kan tilbyde de nødvendige services, som påpeget i pkt. 6. Det kan kun (1) en egen udviklet resolution service (f.eks. NORDINFO projekt, hvis projektet har den ambition), og (2) DOI, som er udbredt og har den nødvendige organisation bag. DOI anvendes i mange forskellige sektorer: biblioteker, centraladministration, erhvervsvirksomheder osv. Og da permanente identifikatorer har interesse i forskellige sektorer, kunne det være en fordel at etablere et lokaliseringssystem,

som i sig selv ikke var tilknyttet én bestemt sektor, f.eks. det Handle-baserede DOI system.

DOI er samtidig ikke bare en resolution service, og der er f.eks. et velfunderet metadata-system tilknyttet. Systemet er også robust i form af sin internationalt udbredte anvendelse. Systemet forudsætter oprettelse af en lokal "registration agency", som vil have økonomiske forpligtelser til IDF. Dette vil selvfølgelig medføre en betaling for de udstedte DOIs – et eller andet sted i systemet.

Da overdrivelse ind imellem kan understrege en påstand, kan en service, som udelukkende består af resolution, etableres ved anvendelse af PURL software, som vil være en billig løsning!

### 8. Bevaring af digitale ressourcer

Det, som giver persistent identifikatorer mening, er ikke kun et spørgsmål om teknik eller overholdelse af internationale navngivningskonventioner og -standarder. Et afgørende aspekt er, at teknikken og standarder varetages og udføres i en organisatorisk infrastruktur, og som dermed bliver en vigtig del af stabiliteten.

I forbindelse med kommende aktiviteter vedrørende bevaring, kan det anbefales, at man undersøger hvordan bevaringsstrategier, erklæringer eller garantier kan tilknyttes et objekts metadata, herunder også hvordan f.eks. "mekaniske" MD5 checksum eller lignende metoder kan gøres eksplicit for brugerne.

### Note

Der er for nylig blevet nedsat en arbejdsgruppe med IT og Telestyrelsen som den koordinerende part, som skal forsøge at nå til enighed om indførelse af fælles regler for udformning af identifikatorer for digitale objekter. "Fælles" regler betyder, at arbejdsgruppen skal forsøge at udforme regler for navngivning af identifikatorer på tværs af meget forskellige sektorer. Det forventes, at arbejdsgruppen vil fremlægge et forslag i første halvdel af 2005. Arbejdsgruppen har på nuværende tidspunkt ikke implementering af en resolver service på dagsorden, så denne opgave venter.



### Referencer

[ARK]: *Archival Resource Key*, [www.cdlib.org/inside/diglib/ark/](http://www.cdlib.org/inside/diglib/ark/), University of California Digital Library

[DISSFORMAT]: *Präferenzregelung für die Archivierung elektronischer Publikationen*, [www.ddb.de/wir/pdf/paerferenzregelung.pdf](http://www.ddb.de/wir/pdf/paerferenzregelung.pdf), 2000

[DISSONLINE]: *Digitale Dissertationen im Internet*, [www.dissonline.de/](http://www.dissonline.de/)

[DISSPI]: *Persistent identifier*, [www.persistent-identifier.de/](http://www.persistent-identifier.de/), Die Deutsche Bibliothek

[DOI]: *Digital Object Identifiers*, [www.doi.org/](http://www.doi.org/), The International DOI Foundation (IDF).

[DOI/NISO]: *Syntax for the Digital Object Identifier*, [www.niso.org/standards/resources/Z39-84-2000.pdf](http://www.niso.org/standards/resources/Z39-84-2000.pdf), 2000

[DSPI]: *Identifikatorer til digitale objekter* (Bilag til XML-komit møde, s. 3–11), [www.oio.dk/files/Materiale\\_2\\_sending\\_til\\_mødet\\_den\\_11.\\_august\\_04.pdf](http://www.oio.dk/files/Materiale_2_sending_til_mødet_den_11._august_04.pdf)

[EPICUR]: *Enhancement of Persistent Identifier Services - Comprehensive Method for unequivocal Resource Identification (EPICUR)*, (urn:nbn:de:1111-2003121811), <http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:1111-2003121811>

[HANDLE]: *Handle System*, <http://www.handle.net/in-dex.html>, Corporation for National Research Initiatives

[IANA]: *Internet Assigned Numbers Authority*, [www.iana.org/](http://www.iana.org/)

[IDF]: *The International DOI Foundation*, [www.doi.org/](http://www.doi.org/)

[INDECS]: *Interoperability of data in e-commerce systems*, [www.indecs.org/](http://www.indecs.org/)

[ISIL]: *International Standard Identifier for Libraries and Related Organizations (ISIL)*, [www.bs.dk/isil/](http://www.bs.dk/isil/)

[KAHN]: Kahn Robert, Wilensky, Robert: *A Framework for Distributed Digital Object Services*, <http://hdl.handle.net/4263537/5001>

[KUNZE]: *Towards Electronic Persistence Using ARK Identifiers*, [www.cdlib.org/inside/diglib/ark/arkcdl.pdf](http://www.cdlib.org/inside/diglib/ark/arkcdl.pdf), California Digital Library, University of California

[NLA-PI]: *Persistent Identification Systems*, [www.nla.gov.au/initiatives/persistence/PIcontents.html](http://www.nla.gov.au/initiatives/persistence/PIcontents.html), 2001

[NORDISKURN]: *Garanterad dokumentåtkomst nu och i framtiden*, <http://epc.ub.uu.se/niwiki/pmwiki.php>

[OAI]: *Open Archives Initiative*, [www.openarchives.org/](http://www.openarchives.org/)

[PASKIN]: Norman Paskin, DOI: *A 2003 Progress Report*, doi:<http://dx.doi.org/10.1045/june2003-paskin>, International DOI Foundation, 2003.

[POI]: *The PURL-based Object Identifier (POI)*, [www.ukoln.ac.uk/distributed-systems/poi/](http://www.ukoln.ac.uk/distributed-systems/poi/), 2004

[POIRES]: *POI resolver guidelines*, [www.ukoln.ac.uk/distributed-systems/poi/resolver-guidelines/](http://www.ukoln.ac.uk/distributed-systems/poi/resolver-guidelines/), 2004

[PURL]: *Persistent Uniform Resource Locator*, [www.purl.org/](http://www.purl.org/)

[TSO]: *Digital Object Identifiers for publishers and the e-learning Community, A Report for the JISC from TSO*, [www.jisc.ac.uk/index.cfm?name=project\\_tso](http://www.jisc.ac.uk/index.cfm?name=project_tso), 2003

[TSOFREE]: *Free digital object identifiers pave way to linking public information*, [www.tsoid.com/downloads/DOI%20free%20V.4.2%20Final.pdf](http://www.tsoid.com/downloads/DOI%20free%20V.4.2%20Final.pdf), 2004

[URI]: *Uniform Resource Identifiers (URI)*, [www.ietf.org/rfc/rfc2396.txt](http://www.ietf.org/rfc/rfc2396.txt)

[URN]: *Functional Requirements for Uniform Resource Names*, [www.ietf.org/rfc/rfc1737.txt](http://www.ietf.org/rfc/rfc1737.txt)  
URN Syntax, [www.ietf.org/rfc/rfc2141.txt](http://www.ietf.org/rfc/rfc2141.txt)  
URN Namespace Definition Mechanisms, [www.ietf.org/rfc/rfc2611.txt](http://www.ietf.org/rfc/rfc2611.txt)

i For en speciel rapport om permanente identifikatorer og e-læring se [TSO].

ii Anvendelse af ISIL er under diskussion af Dublin Core Collection Description Working Group

## DF Revy Kort nyt

### Høringssvar vedrørende Bekendtgørelse til Lov om Biblioteksvirksomhed

Den 15. november sendte kulturminister Brian Mikkelsen et udkast til en ny biblioteks bekendtgørelse til høring. Kulturministeriet ønsker med ændringen af bekendtgørelsen at indføre en karenstid på fire måneder, inden bibliotekerne må udlåne nyudgivne musik-cd'er.

Fra Biblioteksparaplyen, som blandt andet omfatter Danmarks Forskningsbiblioteksforening, er der kommet følgende reaktion, som er sendt til Kulturministeriet:

#### Biblioteksparaplyen:

Bibliotekarforbundet  
Bibliotekslederforeningen  
Danmarks Biblioteksforening  
Danmarks Forskningsbiblioteksforening  
Danmarks Skolebibliotekarere  
HK Kommunal Biblioteksudvalget  
HK-STAT Danmark. TR-Kollegiet for de Statslige Biblioteker  
Kommunernes Skolebiblioteksforening

### Vedr. Bekendtgørelse til Lov om Biblioteksvirksomhed. Høring november 2004

Den danske Biblioteksparaply, der samler de otte nationale biblioteksforeninger og -organisationer, ønsker i forbindelse med ovennævnte høring at gøre indsigelse mod to forhold af principiel karakter, som man finder stærkt problematiske.

#### Bekendtgørelsen

1. Denne Bekendtgørelse vil gribe afgørende ind i de grundlæggende intentioner, der ligger i Lov om biblioteksvirksomhed, nemlig de ny og fremadrettede bestemmelser omkring ligestilling af medierne og det hævdede krav om aktualitet i materialeudbudet. Bekendtgørelsen ses derfor ikke som en uddybning og fortolkning af biblioteksloven men som et instrument, som vil ændre loven på centrale punkter, der indgik som forudsætninger i selve lovvedtagelsen.

2. Eftersom Folketinget og Folketingets Kulturudvalg ikke hører i forbindelse med gennemførelsen af en bekendtgørelse, bør Bekendtgørelsen ikke gå imod lovens bestemmelser og intentioner. Folketingets og dermed borgernes indflydelse på lovgivningen indskrænkes væsentligt, hvis en bekendtgørelse er i modstrid med den vedtagne lov.

#### Høringsprocessen

3. Eftersom der er tale om en bekendtgørelse med et særdeles kontroversielt indhold, jf. pkt. 1 og 2, er det dybt beklageligt at bekendtgørelsen kun er udsendt til høring i en begrænset kreds blandt interessenterne.

4. Ligeledes er det beklageligt, at høringsfristen kun har været to uger. Bekendtgørelsen til biblioteksloven er et særdeles vigtigt redskab i forvaltningen af loven og dermed i arbejdet med at betjene borgerne efter lovens intentioner. De organisationer, som er høringsberettigede, har krav på tid til at udarbejde deres høringssvar, herunder at konsultere deres bagland.

En bekendtgørelse med et så kontroversielt indhold bør således ikke blot behandles som en hurtig og formel ekspeditionssag. Ved ovennævnte procedure i høringsfasen har Kulturministeriet tilsidesat såvel Folketinget som de interesseorganisationer, som repræsenterer borgerne. Vi anbefaler derfor, at sagen genovervejes.

**Med venlig hilsen  
Charlotte Rasmussen  
Formand for Biblioteksparaplyen**