



IT- og Telestyrelsen

Ministeriet for Videnskab
Teknologi og Udvikling

Unikke Identifikatorer til digitale objekter

Standard

IT- og Telestyrelsen, København den 6. december 2006

Standardisering

Kolofon:

Titel: Unikke Identifikatorer til digitale objekter

Denne standard kan frit anvendes af alle. Citeres fra standarden i andre publikationer til offentligheden, skal angives korrekt kildehenvisning.

Denne standard er udarbejdet af arbejdsgruppen vedrørende unikke identifikatorer til digitale objekter og godkendt på Datastandardiseringskomitéens møde d. 6 december 2006.

Kontaktperson

Chefkonsulent Palle Aagaard, IT og Telestyrelsen

Mail-adresse: paa@itst.dk

Telefon: 33 37 92 67

URN: UUID: 44dce900-85d8-11db-b606-0800200c9a66

Forord	4
Resume	5
1. Universelle, unikke og vedvarende identifikatorer	6
2. Baggrund.....	6
2.1 Hvad udgør et digitalt objekt?.....	7
2.2 Brug af identifikatorer	8
2.3 Vedvarenhed – også kaldet persistens.....	9
2.4 Hovedtyper af digitale objekter.....	10
2.5 Hvorledes skal et nationalt identifikatorsystem udformes	11
2.6 Metadata og identifikatorer	12
3. Beskrivelse af anbefalet identifikatorsystem	12
3.1 Uniform Resource Name (URN)	13
3.2 Universal Unique Identifier (UUID)	14
3.2.1 Eksempel på generering af tidsbaseret UUID	15
3.2.2 Anvendelsen af UUID som identifikatorer	16
4. Standardregler for universel, unik og vedvarende identifikation ...	17
4.1 Navngivning og tildeling	17
4.2 Administrativt set-up og praksisregler	18
4.2.1 Opmærkning af universel, unik og vedvarende identifikator	18
5. Resolution og resolver teknikker	21
6. Læs mere.....	23
Bilag 1. Arbejdsgrundlag. Fælles identifikatorer for digitale objekter	24
Bilag 2. UUID mere detaljeret	26

Forord

I regi af ”Dansk Standard S-24 Information og Dokumentation” udviklede en mindre arbejdsgruppe i starten af 2004 en indledende rapport om et fælles nationalt koncept for universelle, unikke og vedvarende identifikatorer til digitale objekter.

([http://www.oio.dk/files/Identifikatorer til digitale objekter - DS24 okt 2003-tilpasset.doc](http://www.oio.dk/files/Identifikatorer_til_digitale_objekter_-_DS24_okt_2003-tilpasset.doc))

Et sådant identifikatorsystem skal anvendes til identifikation af digitale objekter, som kan være web-sider, netpublikationer, elektroniske dokumenter, billed-filer, data-filer, geo-data etc. Universelle, unikke og vedvarende identifikatorer vil gøre det lettere fx at henvise til og kommunikere om et givent digitalt objekt. Det vil også gøre det lettere at håndtere data om givne digitale objekter. Det påtænkte identifikatorsystem skal således i princippet kunne anvendes til al identificering af digitale objekter, er ikke begrænset til bestemte sektorer og kan således også anvendes i både den offentlige og private sektor.

Dansk Standard S 24 (DS S24) tilsluttede sig som helhed den oprindelige arbejdsgruppes synspunkter, og den nationale XML-komité, i regi af Videnskabsministeriet, gav på sit august-møde 2004 også sin tilslutning til synspunkterne. XML-komiteen anbefalede, at der i første omgang via en workshop for en bredere kreds blev arbejdet videre med sagen. På en workshop den 18. november 2004 i IT- og Telestyrelsen blev således nedsat en bredere arbejdsgruppe med det formål at arbejde videre med et fælles nationalt koncept på området. Denne bredere arbejdsgruppe har arbejdet siden årsskiftet 2004/2005 .

Efter at have indkredset og indledende behandlet problemstillingerne i første halvår af 2005 har arbejdsgruppen efterfølgende, og som planlagt allerede fra arbejdets start, siden efteråret 2005 fået bistand af konsulenter fra firmaet Devoteam Fischer & Lorenz, der derefter fortsatte og videreudbyggede arbejdsgruppens hidtidige arbejde med rapporten ”Universelle, unikke og vedvarende identifikatorer for digitale objekter. 17. marts 2006”

På baggrund af ovennævnte rapport har arbejdsgruppen udarbejdet sine anbefalinger; men har dog anbefalet anvendelsen af Uniform Resource Names frem for eXtensible Resource Identifier.

Arbejdsgruppens medlemmer og gruppens arbejdsgrundlag kan ses i bilag 1.

Det skal tilføjes, at arbejdsgruppen har lagt størst vægt på at få udformet et sæt af holdbare metoder til identificering af digitale objekter og i mindre grad har lagt vægt på forskellige helt konkrete anvendelser, som fx brug af resolvere til hjælp ved kulturarvens bevarelse. Dette fordi det undervejs i arbejdsgruppens arbejde blev skønnet, at fx resolveranvendelse i høj grad er et organisatorisk/administrativt anliggende, som i sin udmøntning kræver helt særlig forpligtende opbakning. Arbejdsgruppen er dog af den opfattelse, at de foreliggende anbefalinger kan indgå i et eventuelt fortsat arbejde/projekt vedrørende resolvere for internetpubliceret materiale. Arbejdsgruppen anbefaler tillige, at der i denne sammenhæng bliver oprettet et særligt helt overordnet namespace (”NID”) med tilhørende regelsæt.

Resume

Denne standard anbefaler en universel, unik og vedvarende identifikation af digitale objekter. Et digitalt objekt er her ”**information i form af talkoder**”. Det kan være digitale tekster, billeder, lyd, information, datafiler (fx XML-filer) etc. Herunder hører digitale repræsentationer af fysiske objekter, fx personer eller geografiske steder. En identifikator er **en sekvens af tegn, der muliggør identificering**.

Flere muligheder for præsentation af identifikatorer er vurderet, og valget er faldet på URN (Uniform Resource Names). URN'er har som formål at virke som lokations-uafhængige identifikatorer af objekter.

Der er allerede forskellige identifikatorsystemer i brug til forskellige typer af objekter (fx til digitale bøger), men der er ikke tradition for brug af fælles systemer for alle typer af digitale objekter. For de tilfælde hvor der ikke allerede anvendes et fælles system, anbefales brugen af Universal Unique Identifier (UUID).

Identifikatorer kan således **fx** se ud som:

- urn:uuid:6e8bc430-9c3a-11d9-9669-0800200c9a66
- urn:oasis:member:A00024-x
- urn:isbn:1-56592- 262-x

hvor eksemplerne ”uuid”, ”isbn” og ”oasis” repræsenterer ”NID”, og hvert af de efterstillede led repræsenterer ”NSS”, der fastlægges af ejeren af namespace (NID).

1. Universelle, unikke og vedvarende identifikatorer

En identifikator opfattes her som en sekvens af tegn, der muliggør identificering; og en universel, unik og vedvarende identifikator i denne forbindelse er en blivende entydig reference til et digitalt objekt.

Universel betyder, at identifikatoren er anvendelig til alle objekter indenfor alle domæner og emner. **Unik** betyder, at identifikatoren er entydig, eller der eksisterer "kun en af slagsen". En given identifikator knyttes således kun til ét objekt. At identifikatoren er **vedvarende** betyder, at den er blivende, permanent eller persistent. Med andre ord når en given identifikator er knyttet til et objekt, er den gældende så længe, at objektet eksisterer.

Tanken om et identifikatorsystem er ikke ny. Et identifikatorsystem, som vi kender i Danmark, er Det Centrale Personregister CPR. CPR-nummeret identificerer entydigt hvert individ i Danmark. Et andet eksempel på et identifikatorsystem er fra bøgernes verden, hvor man arbejder med International Standard Bog Nummer ISBN. ISBN nummeret er det, som gør det muligt til enhver tid entydigt at identificere en bestemt publikation.

Et konkret problem inden for IT-verdenen er problematikken om unikke nøgler, der entydigt kan identificere fx filer, brugere eller transaktioner i et IT-system. Der eksisterer fx ikke en fælles standard for sådanne unikke nøgler eller identifikatorer.

Når man udformer IT-systemer, har man ofte defineret sine egne unikke nøgler til brug for systemet. En unik nøgle i et givent IT-system vil således være unik i dette ene IT-system, men ikke nødvendigvis i andre systemer. Universelle, unikke og vedvarende identifikatorer tilbyder en standardmetode for unikke nøgler. Ved at bruge identifikatoren som unik nøgle vil nøglen være universel unik på tværs af IT-systemer.

Et andet konkret problem på fx Internettet er, at der eksisterer mange digitale objekter, og antallet vokser. Lige nu er Uniform Resource Locator (URL) den mest udbredte identifikator brugt til at identificere digitale objekter på Internettet, og det betyder, at man linker ved hjælp af URL. URL er ikke egnet til livslang identifikation af et objekt, for URL giver udelukkende den nuværende placering for objektet eller en kopi af objektet. Objektet som URLen peger på, kan man uden videre ændre, flytte eller slette, og et objekt kan således i sin levetid have mange forskellige URLer. Det gør det umuligt at identificere digitale objekter unikt med mindre, at det fx som visse netpublikationer har et ISBN nummer.

Det er derfor nødvendigt at have unikke, universelle og vedvarende identifikatorer (nøgler) for alle typer af digitale objekter.

2. Baggrund

I de følgende afsnit præsenteres definitionen af, hvad et digitalt objekt er. Det er disse objekter, som den universelle, unikke og vedvarende identifikator skal gælde for.

2.1 Hvad udgør et digitalt objekt?

Universelle, unikke og vedvarende identifikatorer kan knyttes til alle former for digitale objekter. Hvad er så et digitalt objekt i denne forbindelse? Der er mange forskellige måder at definere dette, f. eks. at det er elektronisk information, der kan lagres i en elektronisk fil. Informationsordbogen har denne definition: *"Informationsbærende genstand på digital form. Bruges oftest om genstande, som er tilgængelige via Internet. Genstandene kan fx være dokumenter, lydoptagelser, video eller en kombination af sådanne former."*¹

I denne standard er følgende definition valgt:

Information i form af talkoder

Det kan være digitale tekster, billeder, lyd, information, datafiler (fx XML-filer) etc. Herunder hører digitale repræsentationer af fysiske objekter, fx geografiske steder. Man kan skelne mellem:

- a) Et enkeltstående statisk objekt
- b) Et objekt der består af to eller flere statiske objekter
- c) Et enkeltstående dynamisk objekt
- d) Et objekt der består af to eller flere dynamiske objekter

Digitale objekter kan være statiske enkeltstående objekter, fx et dokument hvis egenskaber fastlægges en gang for alle.

Et enkeltstående objekt er relativt nemt at håndtere, mens et dynamisk objekt og et objekt, der består af flere objekter, fx en hjemmeside, er sværere. Ved flere samhørende objekter skal man beslutte, om hvert af objekterne, der udgør et samlet objekt, skal tildeles hver en identifikator, og hvordan man i øvrigt vil angive, at de hører sammen.

Ved et dynamisk objekt, forstået på den måde at præsentationen af informationen er baseret på kontinuert opdatering, skal man beslutte, hvor meget ændring der kan tolereres, før der er tale om et nyt objekt. Som eksempel kan anføres en internetpubliceret vejrudsigt, hvor vejrudsigten for et geografisk område skifter meget ofte. I et sådant tilfælde er det digitale objekt både at opfatte som selve vejrudsigten, der således bør identificeres selvstændigt, og hver af de enkelte opdateringer af indholdet, som kan have hver sin identifikator. Med andre ord er dynamiske objekter, objekter hvor indholdet ændrer sig over tid - og kan godt tilknyttes en unik identifikator.

Med udgangspunkt i definitionen af et digitalt objekt er et objekt selv de mindste data-filer. Fx kan en netpublikation (der samlet set er ét digitalt objekt) bestå af flere sammenknyttede eller sammenlængede filer eller kan være dannet ud fra felter i en database. Derfor kan der i teorien være knyttet flere identifikatorer til samme fil, der hver peger på et digitalt objekt. Ligeledes kan en hjemmeside bestå af flere enkelte objekter. Igen bør hjemmesiden have en identifikator for sig selv, og hvert objekt på hjemmesiden kan få en identifikator, hvis de skal/kan bruges i andre sammenhæng. Identifikatorerne bør i denne situation referere til hinanden enten via identifikatoren eller øvrige metadata tilknyttet identifikatoren eller objektet.

¹ Online: Udgiverselskabet Informationsordbogen: Informationsordbogen - ordbog for informationshåndtering, bog og bibliotek [Set d. 2. november 2005]

Der er således flere situationer, hvor det er svært at afgrænse det digitale objekt. Dette er problemstillinger, som ikke kan afgøres endeligt, men som man må tage stilling til i det konkrete arbejde med universelle, unikke og vedvarende identifikatorer. Det afhænger i høj grad af den formodede anvendelse af objektet, hvilket bedst kan vurderes af ejeren af objektet.

Man kan dog så vidt mulig gå ud fra, at ethvert objekt, der bliver stillet til rådighed offentligt, og som vil blive anvendt, citeret eller linket til af andre objekter, har brug for en identifikator.

2.2 Brug af identifikatorer

Identifikatorer relaterer sig til forskellige objekter, og alle disse objekter påvirker kravet til identifikatoren. Ønsket er en identifikator, som holder hele objektets "liv". En væsentlig egenskab ved et objekts identifikator skal derfor være, at når den en gang er blevet tilknyttet et objekt, forbliver den identifikator for dette objekt, og man kan ikke anvende den til identifikation af andre objekter. Præcis som et CPR-nummer, en nummerplade til bilen eller et ISBN-nummer.

Der er fx brug for identifikatorer til²:

- Hjemmesider/portaler og dele deraf
- E-publikationer
- Læringsobjekter
- Sager
- Dokumenter
- Medieobjekter
- Datafiler, herunder digitale repræsentationer af fysiske objekter
- Metadatasæt
- Billeder
- Lyd
- Noder
- Musik
- Bøger
- Tidsskrifter

2.2.1 Universel, unik og vedvarende identifikator i ESDH

Et eksempel på en problemstilling er digitale objekter (typisk et dokument) i et Elektronisk Sags- og Dokumenthåndteringssystem (fremover ESDH system), som man ønsker at offentliggøre på en hjemmeside. I det øjeblik går objektet fra at være næsten usynligt uden for den pågældende organisation til at være offentligt tilgængeligt, og objektet vil få brug for at være identificeret ved en universel, unik og vedvarende identifikator. URL'en, som vil identificere objektet, er ofte ikke tilstrækkeligt vedvarende.

Objektet (typisk et dokument) kan også have behov for at blive unikt og vedvarende identificeret i det øjeblik, hvor en sagsbehandler i en organisation sender objektet til en anden organisation. Her kan objektet blive udvekslet frem og tilbage i flere versioner, og hver gang ønsker man at identificere objektet unikt, universelt og vedvarende. Hermed kan man bl.a. undgå at registrere det samme objekt flere gange.

² Nogle af objekttyperne har allerede et default identifikatorsystem fx bøger (ISBN) og tidsskrifter (International Standard Serial Number, ISSN).

2.2.2 Udveksling af objekter

I det hele taget kan en identifikator spille en rolle hver gang et digitalt objekt sendes fra et sted til et andet, idet det vil forenkle kommunikation vedrørende det digitale objekt. Hvis der ikke oprettes et fælles system, eller der skabes enighed om fælles retningslinier, så vil man typisk være henvist til, at hver ejer af et digitalt objekt opererer med sit lokalt definerede identifikatorsystem eller metadata. Når der ikke findes et universelt system – en fælles standard – er der risiko for, at forskellige digitale objekter påføres enslydende identifikatorer.

2.2.3 Skanning af gamle kort eller luftfotografier

En identifikator kan også anvendes ved skanning af gamle kort eller luftfotos. Ved at tilføje en identifikator ved skanningen, kan man senere genfinde det digitale billede ved hjælp af en metadatabeskrivelse. Denne metadatabeskrivelse vil være indgangen til en katalog for kort eller luftfotos. I kataloget er det fysiske sted, som fx er fotograferet, repræsenteret vha. koordinater, og billedet er repræsenteret af metadata. Den universelle, unikke og vedvarende identifikator er således koblingen mellem de to dele, og kan således anvendes i forbindelse med søgning efter et specielt billede.

2.2.4 Internetpubliceret materiale

I forbindelse med internetpubliceret materiale er URL den mest udbredte identifikator, selvom en URL i virkeligheden alene er en lokalisering. URL'er er som nævnt tidligere ikke er egnet til livslang identifikation af et objekt, da URL'en for et givent objekt kan ændre sig i objektets levetid. Objektet, som URL'en peger på, kan man uden videre ændre, flytte eller slette, og et objekt kan således i sin levetid have mange forskellige URL'er. Det er derfor nødvendigt også at have en unik identifikator tilknyttet internetpublicerede digitale objekter, idet disse identifikatorer kan anvendes i såkaldte resolvere (se afsnit 5), som sammenknytter objektets vedvarende identifikator med objektets øjeblikkelige placering (URL). Dette betyder, at med kendt identifikator vil et objekt altid kunne lokaliseres (genfindes). Det bør nævnes, at ovenstående betragtninger gælder, uanset om adgangen til objektet er frit tilgængeligt, eller adgangen er password-beskyttet.

2.3 Vedvarende – også kaldet persistens

At en identifikator er vedvarende, betyder i denne sammenhæng, at den vedvarende knyttes til et digitalt objekt. Når objektet bliver flyttet fra et sted til et andet, har man dermed stadig mulighed for at finde det vha. identifikatoren. Vedvarende handler således om:

At identifikatoren og objektet skal være vedvarende forbundet, således at en vedvarende henvisning til objektet er sikret.

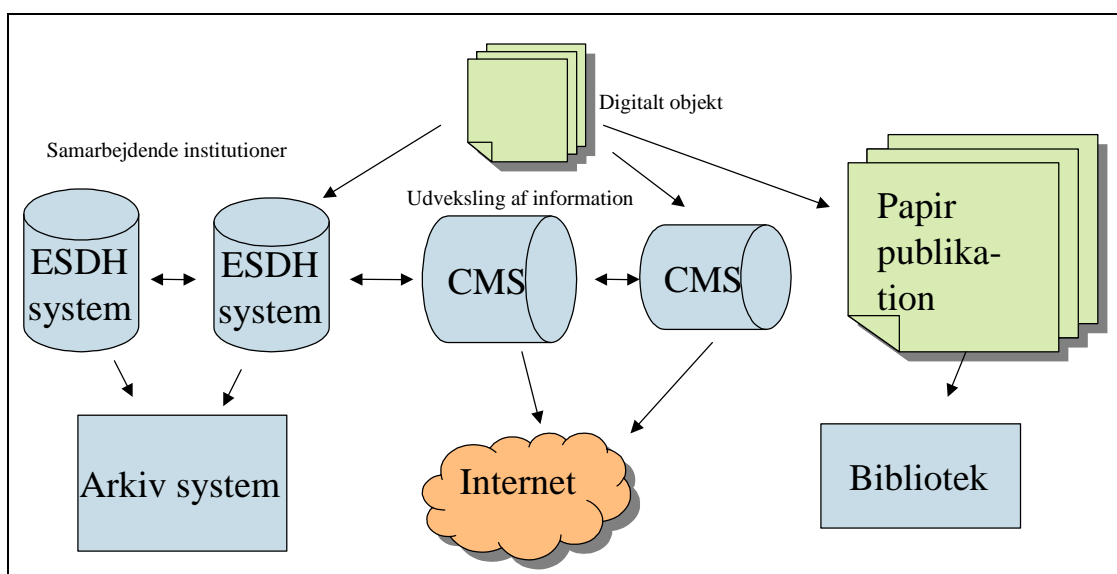
Den vedvarende del af en identifikator viser en bestemt udnyttelse af at indføre dem. Dermed handler det ikke kun om at identificere objekter, men også om at håndtere lokalisering og at sikre adgangen til digitale objekter.

Vedvarende vil være et spørgsmål om politikker og tilslutning til det valgte identifikatorsystem. Der skal være nogen (typisk en organisation), der ønsker at sikre, at et objekt vedvarende kan identificeres og findes. Den samme organisation skal være villige til at arbejde for at sikre vedvarende, dvs. opdatere sammenhængen mellem objektet og dets lokalisering fx under brug af såkaldt resolver (se afsnit 5). Med andre ord er en af forudsætningerne for vedvarende, fx at der

opbygges en organisation, at der fx anvendes resolver-teknikker, samt at dette ”set-up” vedligeholdes/administreres.

Vedvarende er fx især relevant for henvisninger til videnskabelig litteratur, rapporter, vejledninger mv. samt for at linke til netpublikationer. I forhold til netpublikationer er en universel, unik og vedvarende identifikator særlig interessant, da det ofte er et problem, at netpublikationer skifter URL og dermed bryder alle links til objektet.

Et digitalt objekt forbliver i sin levetid næppe på samme lokation, og objektet flytter sandsynligvis også mellem forskellige systemer. Et objekt (i dette tilfælde et dokument) kan fx blive udvekslet mellem to ESDH systemer eller fra et ESDH system og til et Content Management System, CMS. Dokumentet kan måske flyttes fra et CMS til et andet CMS eller til offentliggørelse på Internettet. Figuren nedenfor illustrerer det digitale objekts potentielle ”vandring” mellem systemer.



2.4 Hovedtyper af digitale objekter

For at indkredse nogle problemstillinger vedrørende digitale objekter og tildeling af universel, unik og vedvarende identifikator, er det hensigtsmæssigt at skelne mellem to tænkte situationer.

- Det digitale objekt har ”værkshøjde”
- Det digitale objekt har ikke ”værkshøjde”

Det bør nævnes, at ”værkshøjde”-begrebet er taget fra ophavsretsloven og her udelukkende anvendes som praktisk term for at beskrive to forskellige håndteringssituationer vedrørende digitale objekter.

Digitale objekter med ”værkshøjde” vil være af en sådan karakter, at objektets ejer er parat til at investere væsentlig tid til håndtering og behandling af det digitale objekt. Det kan fx være elektroniske bøger, love, musik og billeder, hvor ejeren investerer en del ressourcer på frembringelsen samt distribution og evt. salg. Derfor vil digitale objekter med værkshøjde have en stor chance for at blive tildelt en identifikator og andre metadata, og der er en ligeledes chance for,

at ejeren vil sørge for at opdatere objektets lokation i (d)en resolver-service, som anvendes til at finde objektet.

Digitale objekter i denne forbindelse uden ”værkshøjde” er alle andre digitale objekter. Ejeren af objektet har oftest ikke en salgsmæssig interesse i objektet, og der er en mindre chance for, at objekterne ikke tildeles en identifikator og en endnu mindre chance for, at henvisninger til lokation opdateres.

Det er endvidere karakteristisk, at digitale objekter med ”værkshøjde” antalmæssigt er relativt begrænset, hvorimod antallet af digitale objekter uden ”værkshøjde” må betragtes som meget stort og derfor uden tradition for, at man vil investere væsentlig tid på håndteringen af disse.

Ud fra et praktisk synspunkt er det således overkommeligt også at håndtere og bruge tid på digitale objekter med ”værkshøjde” (for det gør man i forvejen), hvad angår tildeling af identifikatorer. Hvorimod den samme situation antageligvis ikke gør sig gældende for digitale objekter uden ”værkshøjde”.

Derfor må en universel, unik og vedvarende identifikator og det tilhørende ”administrative set-up” indrettes med henblik på størst effektivitet. Der skal være mindst håndteringsmæssigt besvær ved at sammenknytte identifikatorer, digitale objekter og deres lokation. Dette vil sige, at tildeling af identifikatorer til digitale objekter skal foregå så automatisk som muligt og kun i de tilfælde, hvor det digitale objekt har ”værkshøjde”, må der evt. være mere komplicerede arbejdsprocesser involveret.

2.5 Hvorledes skal et nationalt identifikatorsystem udformes

For at udforme et nationalt identifikatorsystem er det nødvendigt at opstille en række krav, som systemet må opfylde. Det er minimumskrav, som et identifikatorsystem skal opfylde.

Identifikatorerne skal:

- være universelt unikke.
- kun anvendes én gang, dvs. de må ikke duplikeres uanset antal identiteter og tid.
- både kunne tildeles lokalt af objektets ejer eller organisation og kunne tildeles automatisk maskingenereret.
- være vedvarende (stabile/blivende) uanset arbitrære forandringer i objektets attributter, fx navneskift eller skift fra en organisation til en anden³.
- ikke nødvendigvis registreres centralt, da det er omkostningsfuldt økonomisk og administrativt.
- ikke nødvendigvis være betydningsbærende.
- kunne bruges af alle – private og erhvervsliv.
- være omkostningsfrie eller meget billige.

³ Det er arbejdsgruppens opfattelse, at namespace for identifikatorer (se også senere) ikke bør organiseres på organisationsniveau, men bør organiseres funktionsstyret fx på fagligt domæne niveau.

2.6 Metadata og identifikatorer

En identifikator vil ofte være at betragte som et metadata-element i forhold til et digitalt objekt. Tænkes på en tekst, som det digitale objekt, vil en identifikator tilknyttet teksten (informationsindholdet) kunne være ét metadata-element blandt flere metadata-elementer.

Der findes mange forskellige metadata standarder/anbefalinger til forskellige formål, hvor disse formål har at gøre med fx typen af information (tekster, billeder, sager, sagsbehandlingsdokumenter, lyd, geo-data etc.) eller fx håndteringen af de digitale objekter. Der kan nævnes Dublin Core, FESD-datamodel, MoReq, DanMARC2 etc. Antallet af forskellige metadata anbefalinger, der anvendes, er meget høj og varierer fra fagligt domæne til fagligt domæne.

Karakteristisk for de fleste metadata standarder/anbefalinger er, at disse også indeholder et element, der anvendes til identificering af den information, metadata-sættet beskriver. Fx knytter Dublin Core sig til tekster og har næsten udelukkende til formål at tilvejebringe mulighed for genfindning. Dublin Core har fx metadata-elementerne (kun en lille delmængde):

Title
Creator
Subject
Identifier

Af disse metadata-elementer, er det ”Identifier”, der indeholder (unik) identifikator, ”subject” indeholder emnedata, som beskriver de emner, objektet (teksten) handler om etc. Et andet eksempel på en metadata-profil, hvor lignende forhold gør sig gældende, kan ses på side 17, hvor et lille udsnit af FESD-datamodellen er gengivet.

Da andre metadata-elementer indeholder forskellig information om objekterne, er det således ikke nødvendigt, at en identifikator er betydningsbærende eller indeholde information i øvrigt. Nødvendig yderligere information kan/bør således tilknyttes via metadata standarder, som de nu engang er udformet til den type af information objekt ejeren har med at gøre. Det kan i forlængelse heraf også nævnes, at det næppe er muligt at udforme en betydningsbærende identifikator, som kan indeholde al nødvendig meta-information for alle informationstyper og ”domæner”.

Endelig skal tilføjes, at har objekt-ejeren at gøre med digitale objekter i form af netop tekster, dokumenter, billeder, lyd o.l., vil det næsten altid være nødvendigt at tilknytte yderligere meta-information til objekterne end blot en identifikator.

3. Beskrivelse af anbefalet identifikatorsystem

Fra arbejdsgruppens start har det kendte og lidt ældre Uniform Resource Names (URN) været en mulighed i forbindelse med præsentation af identifikatorer. Men undervejs i arbejdet blev arbejdsgruppen opmærksom på eXtensible Resource Identifier (XRI) fra OASIS som en mere tidssvarende og mere udbygget metode. Arbejdsgruppen har studeret XRI i detalje, men XRI er dog – mod forventning – ikke blevet en vedtaget standard i OASIS regi, og en afgørelse af dette kan have lange og ukendte udsigter, hvorfor arbejdsgruppen ikke på nuværende tidspunkt kan anbefale XRI.

Derfor anbefaler arbejdsgruppen URN'er, som er både kendte og afprøvede, også fordi XRI'er kan opfattes som en videreudvikling af URN'er. Dvs. hvis XRI på et tidspunkt vedtages som standard, kan URN beskrevne identifikatorer indeholdes i XRI'er.

3.1 Uniform Resource Name (URN)

URN er udviklet af The Internet Engineering Task Force (IETF) i erkendelse af de problemer, der er knyttet til at bruge de lokationsafhængige URL'er som identifikatorer. IETF begyndte arbejdet med URN i 1996, og beskrivelsen af URN er dækket af flere RFC dokumenter. URN og URL udgør tilsammen sættet af en identifikator kendt som Uniform Resource Identifier URI. URN er en åben, ikke-proprietær standard, der tilbyder en identifikator uafhængig af placering.

URN består af 3 dele:

- URN sekvensen "URN:", der identificerer, at det er en URN.
- Namespace Identifier (NID), der identificerer den domæansvarlige instans. Det er et unikt **registreret** navn for det valgte identifikatorsystem, som Namespace Specific String i URN syntaksen indeholder.
- Namespace Specific String (NSS), som identificerer det konkrete objekt, dvs. identifikatoren.

$\langle \text{URN} \rangle := \text{"urn:"} \langle \text{NID} \rangle \text{"} \langle \text{NSS} \rangle$
--

NID tildeles af navnemyndigheden Internet Assigned Numbers Authority (IANA), og NSS varetages af den namespace ansvarlige for NID. Regler for benyttet syntax kan ses i RFC 2141 – "URN syntax". Det kan dog her nævnes, at URN samt NID benævnelserne ikke er "case-specifikke", og at NSS regler bestemmes af den ansvarlige for NID.

Af "IANA Registry of URN Namespaces" [RFC 2141 og RFC 3406 senest opdateret februar 2006] fremgår, at der p.t. findes 28 registrerede namespaces, der kan anvendes i forbindelse med URN'er. Heraf fx UUID, ISBN og ISSN.

Definitioner og mekanismer for dannelse af URN namespaces er beskrevet i RFC 3406. Dette dokument bygger på to nøgleprincipper. Tildeling af URN er en styret proces (dvs. en URN skal følge reglerne for det specifikke namespace), og rummet af URN namespaces er styret.

I URN kan registrerede (NID) identifikatorer indsættes uafhængigt af hvilke identifikatorer, der er på tale. Der kan fx indsættes ISSN, ISBN, UUID eller andre nummersystemer som identifikator i URN.

- urn:ietf:rfc:3187 – Brug af IETF Namespace
- urn:isbn:1-56592-262-x – Brug af bognummersystem namespace
- urn:uuid:6e8bc430-9c3a-11d9-9669-0800200c9a66 – Brug af uuid (se senere)

Det bør videre nævnes, at der eksisterer også et enkelt eksempel på et nationalt namespace, NZL, for New Zealand.

Ønsker en organisation eller tilsvarende eget NID, kan dette muligvis opnås efter ansøgning. Der findes procedurer, som skal overholdes. Formelt set skal ansøgningen udformes som en RFC, som eventuelt kan bygge på en formular, der findes i den tidligere nævnte RFC 3406. Ansøgningen skal beskrive det ansøgte URN namespace – formål, forslag til NID, struktur for NSS, henvisning til standarder og anden dokumentation samt overvejelser vedrørende tildeling af URN, resolution, valideringsregler etc.

Ansøgningen behandles af IETF (the Internet Engineering Task Force), som lægger vægt på en række forhold, som omfatter såvel tekniske som politiske og organisatoriske aspekter. De tekniske aspekter omfatter blandt andet, at objekterne skal være tilgængelige via Internettet. De politiske og organisatoriske aspekter omfatter behovet for og fordelene ved at oprette det pågældende namespace samt ansøgerinstitutionens stabilitet og kompetence.

Institutionen skal have en vis stabilitet, eller det skal være afklaret, hvordan det pågældende namespace kan sikres, hvis institutionen nedlægges. Institutionen skal tillige kunne sikre, at tildelingen af URN'er indenfor det pågældende domæne foregår korrekt, herunder at de ikke genbruges, og at persistens sikres.

Det skal endelig tilføjes vedrørende URN'er, at der ikke er defineret standard resolver-teknikker (teknikker til sammenhæng mellem identifikator og lokation fx URL) til denne syntaks, men der er eksperimenteret med software i praksis. Eksempelvis har Folketinget en demo-resolver, der er baseret på URN syntaksen

3.2 Universal Unique Identifier (UUID)

UUID betyder Universal Unique Identifier⁴. Man omtaler også identifikatoren som Globally Unique Identifier (GUID eller GUID). UUID er en identifikator, der kan fungere som en del af den universelle, unikke og vedvarende identifikator for et digitalt objekt.

UUID startede som en de facto standard og er i dag defineret i standarden ISO/IEC 9834-8:2005 af International Telecommunication Union.⁵ Herudover beskriver IETF dokumentet RFC 4122 UUID og teknikken bag. UUID er ikke en brugervendt identifikator, dvs. den er ikke læsbar for mennesker, men for maskiner.

Der findes flere varianter/versioner af UUID - dvs. flere måder at generere en UUID på jf. Bilag 2. Et eksempel på en UUID i version 1 som hexadecimal streng er:

e76af270-4c63-11da-8cd6-0800200c9a66

Det skal nævnes, at alle versioner er mulige at anvende for at generere en UUID.

Karakteristika for UUID er:

- En universel, unik identifikator
- Kræver ikke central registrering

⁴ Syntaksen for brug af UUIDen er beskrevet i: P. Leach, M. Mealling, and R. Salz, "A Universally Unique Identifier (UUID) URN Namespace", RFC 4122, July 2005 [Online <ftp://ftp.rfc-editor.org/in-notes/rfc4122.txt>, set d. 11. November 2005]

⁵ Information technology – Open Systems Interconnection – Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 Object Identifier components. ISO/IEC 9834-8:2005

- Kan bruges af alle
- Kan genereres uden Internetforbindelse
- On-demand generering af UUID kan ske fuldt ud automatisk
- Ikke performancekrævende at generere
- Næsten ingen omkostninger forbundet
- Kan ikke resolves alene og har derfor brug for syntaks som URN.

Set i forhold til kravene til en identifikator (jf. tidligere), så opfylder UUID dem alle.

3.2.1 Eksempel på generering af tidsbaseret UUID

En UUID kan blive genereret lokalt på ens computer og vil være universel unik. Det der sikrer, at en tidsbaseret UUID er universel unik, er kombinationen af:

- Hardware adressen (MAC adressen)
- Tidsstempet
- Tilfældighedsprincipper

Hardwareadressen kommer fra det første netværkskort hos den vært (host), som genererer UUIDen. Eftersom denne adresse for hvert netværkskort er tildelt af én global autoritet, er den garanteret at være unik. Det er vigtigt at understrege, at metoden til at generere UUID ikke er bundet til det fysiske netværkskort. Så det er ikke et problem for UUID, hvis netværkskortet ikke fungerer korrekt.

Med kombinationen af de tre listede elementer, er det meget usandsynligt, at to digitale objekter bliver tildelt den samme UUID. Som et eksempel kan man forestille sig, at den samme person på samme netværkskort i præcis samme sekund opretter to objekter, der får tildelt en UUID. Her vil tilfældighedsprincipperne sikre, at UUIDen er unik.

Et problem med UUID, som eksperter har identificeret, er lande, der bruger sommertid/vintertid. Problemet er dog meget lille, og det er ikke vurderet som havende en reel betydning for identifikatoren.

UUID bliver automatisk generet af computeren, og det betyder, at der ikke skal være en central administrativ enhed, som skal tildele identifikatoren. Der er ydermere frit tilgængelige UUID generatorer på Internettet⁶, og man vil kunne lave en webservice, der kan stille UUID til rådighed. På den måde er det også sikret, at man kan tildele fx databaseposter en UUID.

Opsummeret er fordelene ved at anvende UUID, at det er en gratis løsning, som er tilgængelig for alle og kan tilmed fungere automatisk uden behov for administrative ressourcer. Identifikatoren har også den fordel, at den ikke kun henvender sig til en type af digitale objekter, men kan bruges til mange forskellige.

⁶ Fx kan man finde en på hjemmesiden for International Telecommunication Union [Online: <http://www.itu.int/itu-t/asn1/uuid.html>] set d. 5. januar 2006]

3.2.2 Anvendelsen af UUID som identifikatorer

UUID er en kendt teknik, som bliver anvendt i flere sammenhænge. Flere dokumenthåndteringssystemer anvender også UUID som universel, unik identifikator. Derfor er det tydeligt, at tendensen i Danmark fremover er at bruge UUID mere og mere.

Et eksempel på anvendelsen af UUID er således, at i forbindelse med databaseposter (som indeholder metadata, dvs. beskriver digitale objekter etc.) vil der altid være behov for automatisk generering af identifikatorer – eller nøgler. Sådanne identifikatorer knyttes typisk til hvert sæt af samhörrende data/metadata (entiteter). De mest udbredte relationsdatabasesystemer tilbyder endvidere UUID-metoden til at generere identifikatorer eller nøgler.

Da digitale objekter i meget høj grad opbevares eller håndteres af databasesystemer, fx knyttet til ESDH eller CMS, er det hensigtsmæssigt at bibeholde identifikatoreren, der allerede i databasen automatisk tilknyttes det digitale objekt.

I et ESDH system, hvor databasen lagrer oplysninger/metadata om et dokument, udgør disse oplysninger/metadata en post i databasen, og til denne post er knyttet en identifikator/nøgle. Nedenstående er en forkortet metadataprofil knyttet til et dokument taget fra FESD-standardiseringen:

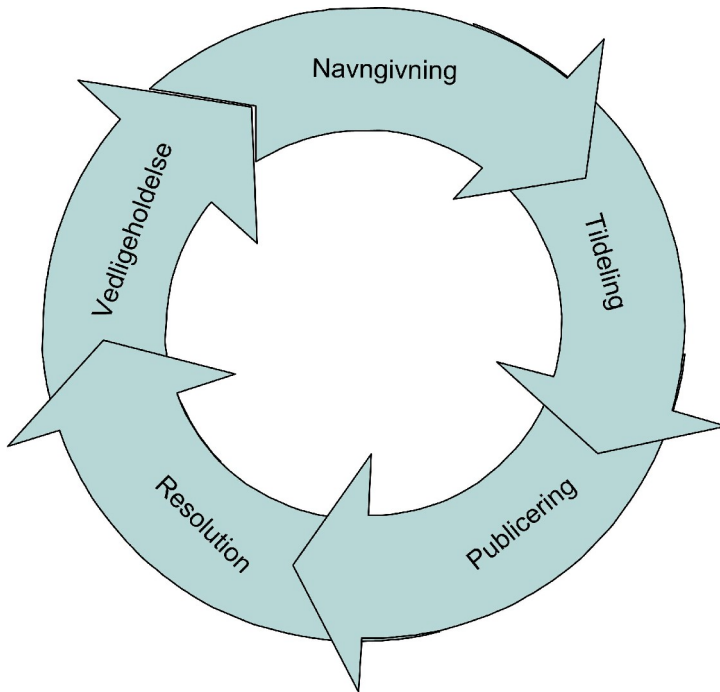
DokumentID
DokumentTitel
DokumentUdarbejdetAf
DokumentDato

DokumentID kan således udfyldes med en identifikator/nøgle baseret på UUID, og den konkrete identifikator følger det pågældende dokument ”fra fødsel til død”.

For en internetpubliceret tekst (digitalt objekt) baseret på et CMS, gør sig lignende forhold gældende. De enkelte tekstdele er ofte felter i en post, og til sådan en post knyttes også en identifikator.

4. Standardregler for universel, unik og vedvarende identifikation

Standardreglerne for identificering med universelle, unikke og vedvarende identifikatorer tager udgangspunkt i livscyklusen for identifikatoren nedenfor.



Livscyklusen er iterativ, dvs. når et objekt får en identifikator, så er det første skridt Navngivning. Cyklusen slutter ikke nødvendigvis ved Vedligeholdelse, men fortsætter, idet objektet kan blive til en ny version og for nogen dermed et nyt objekt. Så skal den nye versions identifikator navngives osv.

4.1 Navngivning og tildeling

Alle digitale objekter bør tildeles en identifikator. Praksis for tildeling, herunder praksis i forbindelse med versionering, afgøres af objekt ejeren under hensyntagen til gældende regler for domænet, hvor objektet håndteres.

- En identifikator kan tildeles et digitalt objekt af objektets ejer eller af den, som udfører opgaven for objekt ejeren. Andre kan tildele objektet en identifikator, hvis objektet ikke har en i forvejen.
- Til ”indpakning” eller præsentation af identifikatoren anvendes URN jf. tidlige afsnit.
- Hvad angår NID og NSS i URN identifikatorstrengen, kan ejeren af det digitale objekt selv beslutte, hvilken der skal anvendes - fx UUID, ISBN eller andet kendt system registreret hos IANA Registry of URN Namespaces. Dog anbefales brug af en version af UUID, såfremt der ikke allerede findes et de facto identifikatorsystem for den pågældende type af digitale objekter, eller andre forhold/analyser tilsiger, at UUID ikke er den optimale løsning.
- En version af UUID kan anvendes som unikke nøgler i forbindelse med poster i databaser. UUID kan således indgå i analysearbejdet i forbindelse med oprettelse af nye databaser, eller når der alligevel sker omlægning af gamle systemer. Brugen af UUID (eller andre unikke identifikatorsystemer) er først og fremmest relevant i forbindelse med digitale objekter, der udveksles mellem systemer eller offentliggøres.

- Der kan anvendes flere identifikatorer knyttet til samme digitale objekt, hvis det er identifikatorer fra forskellige namespaces (NID). Det betyder, at samme digitale objekt fx kan identificeres ved både UUID og et andet identifikatorsystem.
- Én universel, unik og vedvarende identifikator kan kun anvendes til ét objekt, og det må aldrig bruges til at identificere et andet objekt.
- Det anbefales, at kopier af et digitalt objekt (dvs. forskellige lokaliseringer af objektet) har samme identifikator som original-objektet. Dette gælder på det tidspunkt, hvor man skaber original objektet og objekt kopier.*

* Med ”kopi” forstås ”nøjagtig kopi” af originalen (indholdsmæssigt som formatmæssigt). Sådanne kopier opfattes som helt identiske, hvorfor de også skal have samme identifikator, idet fx sammenligninger af (og referencer til) digitale objekter ellers bliver meningsløse.

4.2 Administrativt set-up og praksisregler

Objektet skal som udgangspunkt have tildelt en identifikator i det system, hvor objektet er skabt. Samtidig kan man registrere andre metadata om objektet. I de tilfælde, hvor ophavsmanden ikke ønsker at offentliggøre objektet, så er der i princippet kun behov for at tildele objektet fx en UUID eller en anden universel, unik og vedvarende identifikator. I det øjeblik man ønsker at offentliggøre, fx udveksle objektet med andre interne eller eksterne systemer, så er der behov for URN syntaksen.

Arbejdsgruppen vil pege på følgende udgangspunkter for et administrativt set-up:

Hvad den universelle, unikke og vedvarende identifikator skal identificere

Identifikatoren skal identificere digitale objekter, jf. definition og beskrivelser i tidligere afsnit. Identifikatorer kan tildeles såvel dynamiske som statiske objekter.

Hvem tildeler den universelle, unikke og vedvarende identifikator

Det digitale objekt skal så vidt muligt tildeles en identifikator af den person eller organisation, der har skabt det, med mindre der er fastlagt regler gennem lovgivningen.

Hvordan man genererer en universel, unik og vedvarende identifikator

Ønsker man at anvende UUID, så kan enhver pc genere en UUID. Umiddelbart kræver det et udviklingsværktøj installeret på pc'en. Endvidere findes forskellige UUID generatorer på nettet

Hvorledes håndteres kopier og versionering

Først og fremmest skal konstateres, at nøjagtige kopier ikke er nye versioner. Kopier af et digitalt objekt bibeholder det originale objekts identifikator (jf. ovenstående anbefalinger). Forskellige versioner af ellers samme objekt tildeles forskellige identifikatorer. Kriterier for, hvornår et objekt befinder sig i en ny version, afgøres af objekt-ejeren. I visse tilfælde kan fx ændringen af blot en karakter i det digitale objekt give anledning til, at der er tale om en ny version. Er det digitale objekt derimod fx en mange sider stor elektronisk rapport, kan det være hensigtsmæssigt/praktisk, at der er mere brede rammer for, hvornår der foreligger en ny version.

4.2.1 Opmærkning af universel, unik og vedvarende identifikator

Der findes i princippet to forskellige måder at tilknytte identifikatorer til digitale objekter. Tilknytningen kan foregå ved, at identifikatoren indgår i en metadatabeskrivelse – evt. er metadatabeskrivelsen - af det digitale objekt; en metadatabeskrivelse der som selvstændig fil

følger det digitale objekt fra ”vugge til grav”. Den principielt anden måde at tilknytte en identifikator til et digital objekt på, er ved at inkorporere identifikatoren i det digitale objekt, hvorved identifikatoren så automatisk følger det digitale objekt fra ”vugge til grav” – metoden kan også kaldes indlejring. Disse to metoder kan også udtrykkes som følger:

- Følgefil (tilknytningsfil (metadata) med identifikator som ”følger” det digitale objekt)
- Indlejring (identifikator opmærket i det digitale objekt eller er del af det digitale objekt)

I ESDH-verdenen forgår tilknytning på den først beskrevne måde, idet metadata om dokumenter etc. registreres i en database med permanent henvisning til dokument mv. Denne metode er nødvendiggjort af vanskelighederne – især historisk – ved at indlejre metadata i fx tekstbehandlingsdokumenter.

Databaseposter har typisk allerede identifikatorer /nøgler inkorporeret i databaseposten – parallelt til muligheder vedrørende XML. Denne situation ligner det, der kendes fra XHTML og HTML verdenen – hvis de digitale objekter udtrykkes ved hjælp af disse – idet disse opmærkningsmetoder giver mulighed for indlejring af metadata – og dermed også identifikatorer.

I det hele taget må det betragtes som hensigtsmæssigt at identifikatorer – og andre metadata – er indlejret i de digitale objekter. Det gør det lettere at bevare sammenknytning mellem objekt og identifikator, og da digitale objekter skabt med de nævnte kendte (åbne standarder) opmærkningsprog er åbent maskinlæsbare (fx for høsterbotter), er det også lettere at registrere (fx for søgemaskiner) og dermed udnytte sammenhæng mellem identifikator og digitalt objekt.

I det følgende gives en kort gennemgang af nogle muligheder for indlejring af metadata/identifikatorer, hvis det digitale objekter er formuleret med kendte opmærkningsprog.

Såfremt der ønskes en XML-opmærkning af identifikatoren fx i forbindelse med udveksling af data, kan en opmærkning se ud som følger for en instans:

```
...  
<identifier>  
  urn:NID:NSS  
</identifier>  
...
```

Eller som konkret default eksempel:

```
...  
<identifier>  
  urn:uuid:6e8bc430-9c3a-11d9-9669-0800200c9a66  
</identifier>  
...
```

I forskellige sammenhænge kan der være særlige måder at udtrykke identifikatorer på. I Dublin Core sammenhæng – dvs. metadata især knyttet til tekst-ressourcer – udtrykkes identifikatorer i det tilhørende Dublin Core namespace (ved angivelse af prefix dc) . Dvs. ovennævnte eksempler vil – som instans – se ud som følger for en netpublikation med ISBN-nummer:

```
<dc:identifiser>
  urn:isbn:1-56592-262-x
</dc:identifiser>
```

En tilsvarende HTML eller XHTML indlejring vil kunne have følgende udseende også udtrykt i et Dublin Core namespace:

```
<META NAME= "DC.Identifiser" CONTENT="urn:isbn:1-56592-262-x">
```

På tilsvarende måde udtrykker FESD-standardiseringen identifikatorer ved "UniqueIdentifier" af typen "fesd:UniqueIdentifierType" (således også i eget namespace), hvilket fører til følgende instans eksempel for udveksling af ESDH-data under anvendelse af UUID:

```
...
<UniqueIdentifier>
  urn:uuid:6e9bc430-9c3b-11d9-6438-0800200c9a54
</UniqueIdentifier>
...
```

5. Resolution og resolver teknikker

En central problemstilling er sammenknytning af identifikatorer og de tilhørende digitale objekter.

For at løse denne problemstilling anvendes ofte teknikker, som går under betegnelsen ”**resolvere**”. I princippet er resolvere simple registre, der sammenknytter identifikator og det digitale objekts lokation.

URL’er, som man ofte bruger til både at navngive og lokalisere objekter på internettet, er modtagelig overfor organisatoriske ændringer og dermed ikke vedvarende. Et digitalt objekt kan i sin livscyklus have mange URL’er. Identifikator og lokalisering er her adskilt, og derfor er der behov for en teknik, som lokaliserer ressourcen via dens identifikator. Denne teknik betegnes på engelsk ”resolution”.

I internet-publiceringssammenhæng anvendes til tider sådanne resolvere for at sammenknytte identifikator (fx URN) og URL for objektet. Vedligeholdes/opdateres sammenhængende oplysninger om identifikator og lokation (URN og URL) vil et givent digitalt objekt i princippet altid kunne genfindes (genfinding ved søgning kræver dog typisk også andre registrerede metadata end identifikator knyttet til det digitale objekt, idet søgeren sjældent kender den tilknyttede identifikator).

Manuel vedligeholdelse/opdatering er vanskelig under dynamiske forhold. Man kan således ønske, at man kunne kræve af resolver-miljøet, at det automatisk opdaterer links til objekternes placering. Altså at det udfører en slags udvidet linktjek, hvor den udover at tjekke linket, også kan finde objektets nye placering og registrere det. Sådant mulighed findes ikke i almindeligt kendte resolvere. Sammenhængen mellem identifikator og lokalisering afhænger således i stedet af mennesker og deres ønske om at sikre, at man altid kan finde objektets fysiske placering. Det fører tilbage til diskussionen om objekter med og uden værkshøjde – eller med andre ord om objektet har så stor ”værdi”, at man vil ofre ressourcer på manuel vedligeholdelse/opdatering.

Der er flere projekter, som arbejder med resolvere. Se bl.a. rapporten ”*Universelle, unikke og vedvarende identifikatorer for digitale objekter. 17 marts 2006*” for en beskrivelse af nationale og internationale projekter.

I den forbindelse skal især fremhæves Folketingets administration, som har ønsket at publicere Folketingets dokumenter bredere end blot gennem hjemmesiden www.ft.dk og dermed stille information til rådighed for informationsforbrugerne gennem andre kanaler end internetsider. Man har derfor valgt at gennemføre et pilotprojekt, der afdækker mulighederne for og omfanget af den indsats, der skal investeres i at publicere dokumenter ved brug af URN navngivning.

Folketingets administration har derfor fået udviklet en datatransformationsprocedure, der flytter data ud i en database på websiden, en simpel URN vedtagelse og en webside, der kan opløse URN referencen til et dokument (dvs. en resolver) og vise det med sammenhængende URI links til relaterede dokumenter (resultat). Resolver teknikker i tilknytning til URN’er er således ikke ukendt nationalt set.

Hvad angår en eventuel fælles central resolver eller et system af resolvere, kan de foreliggende anbefalinger indgå i et eventuelt fortsat arbejde/projekt vedrørende sådanne resolvere og internetpubliceret materiale. I det kommende arbejde/projekt bør indgå konkret stillingtagen til nødvendige resolvermetadata samt administrative og organisatoriske procedurer i forbindelse med

vedligehold af lokaliseringsoplysninger og resolvermetadata; ligesom interfaces og tilgængelighedsforhold bør behandles.

Det skal endelig tilføjes, at i forbindelse med et fortsat arbejde/projekt anbefales det, at der i denne sammenhæng bliver oprettet et særligt helt overordnet dansk namespace ("NID") med et regelsæt baseret på disse anbefalinger, og som fx er målrettet resolveranvendelse.

6. Læs mere

1. "Universelle, unikke og vedvarende identifikatorer for digitale objekter". 17 marts 2006. Devoteam Fisher & Lorenz
[<http://www.oio.dk/dataudveksling/fora?o=e578246f33d9e637dd4c844fffbec123>]

URN

2. RFC 2141 "URN Syntax" R. Moats, 1997 [<http://www.ietf.org/rfc/rfc2141.txt> Set 16 oktober 2006]
3. RFC 2611 "URN Namespace Definition Mechanisms" L. Daigle m. fl., June 1999
[<http://www.ietf.org/rfc/rfc2611.txt> Set 16 oktober 2006]
4. RFC 3406 "Uniform Resource Names (URN) Namespace Definition Mechanisms" L. Daigle m.fl. October 2002 [<http://www.ietf.org/rfc/rfc3406.txt> Set 16 oktober 2006]

UUID

5. Information technology – Open Systems Interconnection – Procedures for the operation of OSI Registration Authorities: Generation and registration of Universally Unique Identifiers (UUIDs) and their use as ASN.1 Object Identifier components. ISO/IEC 9834-8:2005
6. P. Leach, M. Mealling, and R. Salz, "A Universally Unique Identifier (UUID) URN Namespace", RFC 4122, July 2005 [Online <ftp://ftp.rfc-editor.org/in-notes/rfc4122.txt>, set d.16 oktober 2006]
7. UUID-generator kan man finde på hjemmesiden for International Telecommunication Union [Online: <http://www.itu.int/itu-t/asn1/uuid.html> set d.16 oktober 2006]

URI og IRI

8. T. Berners-Lee, R. Fielding, and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, January 2005 [Online <http://www.ietf.org/rfc/rfc3986.txt> set d. 9. november 2005]
9. M. Duerst and M. Suignard, "Internationalized Resource Identifiers (IRIs)", January 2005 [Online <http://www.ietf.org/rfc/rfc3987.txt> set d.16 oktober 2006]
10. Naming and addressing: URIs, URLs. W3C [<http://www.w3.org/Addressing/> set 16 oktober 2006]

Bilag 1. Arbejdsgrundlag. Fælles identifikatorer for digitale objekter

I regi af Dansk Standard S 24 har en arbejdsgruppe [i 2004] udviklet nogle indledende tanker og ideer om et fælles nationalt koncept for unikke identifikatorer til digitale objekter – som hovedregel til digital information.

Det påtænkte identifikatorsystem skal anvendes til identifikation af digitale objekter, som kan være web-sider, netpublikationer, elektroniske dokumenter, billed-filer, geo-data etc.

Unikke identifikatorer vil gøre det lettere fx at henvise til og kommunikere om et givent digitalt objekt. Unik identificering vil også fx gøre det lettere at håndtere data om givne digitale objekter.

På en workshop d.18 november 2004 i IT- og Telestyrelsen er nedsat denne arbejdsgruppe for identifikatorer med det formål at arbejde videre med et fælles nationalt koncept på området.

Arbejdsgruppen for identifikatorer skal tage udgangspunkt i notatet fra DS S 24 arbejdsgruppen samt konklusioner på workshoppen d.18. november 2004, og arbejdsgruppen skal orientere sig om kendt internationalt arbejde på området.

Formålet med arbejdsgruppen for identifikatorer er at udarbejde et konkret forslag til – og regelsæt for – hvorledes et nationalt koncept for fælles unikke identifikatorer til digitale objekter kan udmøntes.

Arbejdsgruppen skal bl.a. beskæftige sig med og beskrive følgende:

- Formål med de fælles identifikator, herunder hvad udgør et digitalt objekt, som kan tildeles en fælles national identifikator, og hvilke hovedtyper findes af digitale objekter.
- Syntaks og semantik. Hvorledes skal et nationalt fælles identifikatorsystem udformes samt tilhørende regler for syntaks-mæssig beskrivelse og stabilitet over tid. I denne forbindelse skal der tages stilling til muligheder for anvendelse af URN'er, og der skal tages stilling til sameksistens af forskellige identifikatorsystemer.
- Administrativt set-up. Hvorledes skal tildeling af identifikatorer til digitale objekter foregå, herunder stillingtagen til namespace'es, administrativt set-up og evt. central administration/overvågning samt mulighed for web-service ved den praktiske håndtering. Det er i denne forbindelse også ønskeligt med estimater for eventuelle omkostninger i forhold til det, der foreslås.
- Praksisregler og vejledning herunder beskrivelsen af mindst én use-case.
- Forholdet til og muligheden for persistens ("persistent identifiers") med tilhørende teknikker fx resolver-teknikker.
- Internationale tiltag på området.
- Forhold til aflevering til rigsarkivet samt pligtaflevering afdækkes.

- Sammenhæng til autentifikation (af det digitale objekt) samt ophavsrettigheder til det digitale objekt.

Det anbefales, at arbejdet udformes på en måde således, at resultatet er direkte brugsorienteret og handlingsanvisende.

Deltagelse i arbejdsgruppen er frivillig. Arbejdsgruppen for identifikatorer afleverer sine resultater i form af et notat eller tilsvarende til Dansk Standard S 24 senest december 2005. Resultaterne vil også blive forelagt den nationale XML-komité (senere Datastandardiseringskomité).

Arbejdsgruppen for unikke identifikatorer består af:

Adrian Price, Danmarks Veterinær- og Jordbrugsbibliotek
Grethe Jacobsen, Det Kongelige Bibliotek
Flemming Nissen, Kort og Matrikelstyrelsen
Jytte Hansen, Dansk BiblioteksCenter
Marianne Alenius, Museum Tusulanum Epress og Eforlag
Anders Egeberg, ScanJour
Søren Broberg Nielsen, Civilstyrelsen
Knud Erik Petersen, Folketingets Administration
Leif Andresen, Biblioteksstyrelsen
Tommy Schomacker, Dansk BiblioteksCenter as
Palle Aagaard, IT- og Telestyrelsen

Bilag 2. UUID mere detaljeret

Da teknikken er veldokumenteret, vil det følgende primært gennemgå UUID i kontekst af den universelle, unikke og vedvarende identifikator.

En UUID er basalt set et 128 bit nummer, som er opdelt i 6 felter, der er med til at definere forskellige typer af UUIDer. De 6 felter er:

- a) "TimeLow" feltet
- b) "TimeMid" feltet
- c) "VersionAndTimeHigh" feltet
- d) "VariantAndClockSeqHigh" feltet
- e) "ClockSeqLow" feltet
- f) "Node" feltet

Felterne Variant (VariantAndClockSeqHigh) og Version (VersionAndTimeHigh) er dem, som bestemmer identifikatorens struktur og den måde, som den konstrueres på.

Variantfeltet

Variantfeltet kan udfyldes med 3 typer af varianter, der beskriver strukturen for UUID. Variantfeltet fortolker den øvrige del af UUIDen. Fortolkningen afhænger af hvilken variant, der er valgt.

De 3 varianter til Variantfeltet er:

- **0 NCS** (Denne variant er reserveret til Network Computing System NCS for at sikre bagudkompatibilitet)
- **10** (Denne variant er den, som er defineret i ISO/ IEC 9834-8:2005)
- **110 Microsoft** (Denne variant er reserveret til Microsoft for at sikre bagudkompatibilitet)

Versionsfeltet

Feltet for version beskriver, hvordan UUID er genereret, om den fx er generet på baggrund af tid.

De 5 versioner af UUID er:

- **Version 1: Tidsbaseret.** Baseret på tidspunkt og MAC-adresse. Den klassiske version af UUID, dannet ud fra 60-bit tidsstempel en 14-bit clocksequence og en 48-bit MAC adresse. MAC adressen kan enten være den ægte fysiske netværksadresse (NIC) eller en tilfældig tildelt adresse.
- **Version 2: Distributed Computing Environment DCE baseret.**
- **Version 3: Navnebaseret MD5-hash⁷.** Denne version af UUID er dannet på baggrund af en 128-bit MD5-hash af meddelelsen og en 128-bits namespace-UUID samt et navn af vilkårlig længde.
- **Version 4: Randomiseret eller pseudo-randomiseret.** Versionen af UUID dannes på baggrund af 128-bit tilfældig data
- **Version 5: Navnebaseret SHA-1⁸.** UUID dannes på baggrund af 160-bit SHA-1 hash af objektet, samt en 128-bits namespace UUID og et navn af vilkårlig længde.

⁷ Står for Message Digest algorithm 5

⁸ Står for Secure Hash Algorithm 1