

Sven Winnell:

Om metataggar på Internet

den 8 augusti 2000, reviderad 020428. Några kompletteringar om XML, RDF od. Kompletterad 020517 med Statskontorets vägledning om XML och metadata.

Det finns miljontals Webbsidor på Internet och antalet stiger snabbt. Hur ska man kunna hitta bland dem alla? Det finns ett stort antal organisationer och företag som tillhandahåller kataloger över sidor och sökanordningar av olika slag.

Problemet med att hitta dokument är inte nytt. Biblioteken har haft det sedan länge. Biblioteken använder sig av klassifikationssystem och nyckelord/ämnesord. Det finns i världen ett stort antal olika klassifikationssystem. Länderna har vanligen nationalbibliografier där de klassificerar landets trycksaker. I Sverige används i nationalbibliografen SAB-systemet som tillkom omkring 1920 och är tämligen udda i förhållande till system som används i andra länder. På många håll använda system är amerikanen Deweys system (DDC) som tillkom 1876 och den på detta senare uppbyggda Universella decimalklassifikationen (UDK). De här nämnda bibliotekssystemen har organisationer som vartefter vidareutvecklar systemen och utger nya versioner av dem.

Problemet med att hålla ordning på dokument är inte nytt men har fått en ny dimension i och med tillkomsten av datortekniken och Internet.

Webbsidorna på Internet bör förses med "metataggar" som gör det möjligt att katalogisera webbsidor och använda sökmotorer på ett effektivt sätt.

Sökning

Sökning i LIBRIS :

<http://www.libris.kb.se/avancerad.sokning.html>

Avancerad sökning:

Valfria sökord:

Titel/ord i titel: (Alla titlar. Artiklar i. Recensioner av)

Författare/medarbetare:

Institution/kongress:

Ämnesord (databas):

Klassifikation :

Förlag/förlagsort:

ISBN/ISSN m.m.:

Begränsa sökfrågan med:

Utgivningsår:

Bibliotekskoder:

Språk: (Ingen begränsning. övers. från svenska. Många språk att välja bland.)

Publikationstyp: (Ingen begränsning. Tidskrifter/årböcker. Tidningar. Skriftserier. Avhandlingar.

Kongresser. Offentligt tryck. Festskrifter. Artiklar. Recensioner. Ej artiklar/recensioner.

Barn & Ungdom. Romaner (svenskt). Dramatik (svenskt). Noveller (svenskt). Poesi (svenskt). Förlagets beskrivning.)

Medietyp: (Ingen begränsning. Musikalier .Ljudupptagningar. Kartor. Film/Video. Mikrofilm/fiche.

Bilder. Affischer. Digitaliserade bilder. Digitaliserade affischer. Elektroniska resurser.Nätpublicerade tidskrifter.)

Ett nytt LIBRIS Websök beräknas vara klart våren 2002. Nuvarande Websök uppdateras inte med poster registrerade efter 23 jan 2002.Senast ändrad 5 februari 2002 av infomaster@libris.kb.se, LIBRIS/Kungl.biblioteket

Sökning i Svesök.

<http://www.svesok.kb.se/main.html>

Svesök består av två databaser, dels drygt ~~5,5~~ 15 miljoner svenska webblesurser, insamlade av en robot inom projekt Kulturarw3, dels av idag ca ~~3400~~ 4550 utvalda länkar, ordnade inom nedanstående ämnesområden.

OBS! Innehållet i de två databaserna är insamlade på olika sätt. I Svesöks länkkatalog granskas och katalogiseras webblesurser efter urval. Den andra innehåller allt som finns på den svenska delen av Internet och automatindexeras utan redaktionell granskning. Om man därför väljer att söka direkt i sökfältet ovan så får man visserligen en större mängd träffar men man riskerar att även få träffar som kan uppfattas som irrelevanta. Vill man undvika detta bör man gå till "Avancerad sökning" och där, tillsammans med sin sökning, klicka i rutan "Sök endast i Svesöks länkkatalog".

Bibliotek, uppslagsverk och länksamlingar.

http://www.svesok.kb.se/Bibliotek_uppslagsverk_och_lansamlingar/

Datorer, Internet och IT.

http://www.svesok.kb.se/Datorer_Internet_och_IT/

Ekonomi, samhälle och utbildning.

http://www.svesok.kb.se/Ekonomi_samhalle_och_utbildning/

Geografi, trafik och turism.

http://www.svesok.kb.se/Geografi_trafik_och_turism/

Hem, sport och fritid.

http://www.svesok.kb.se/Hem_sport_och_fritid/

Historia, filosofi och religion.

http://www.svesok.kb.se/Historia_filosofi_och_religion/

Konst, kultur och språk.

http://www.svesok.kb.se/Konst_kultur_och_sprak/

Massmedia, nyheter och evenemang.

http://www.svesok.kb.se/Massmedia_nyheter_och_evenemang/

Medicin, hälsa och psykologi.

http://www.svesok.kb.se/Medicin_halsa_och_psykologi/

Natur, teknik och matematik.

http://www.svesok.kb.se/Natur_teknik_och_matematik/

Sökning i Svesök.

<http://www.svesok.kb.se/search/short.html>

Avancerad sökning

Här söker du både i Svesöks länkkatalog och bland alla svenska webbsidor samtidigt. Mer information om sökingångarna får du genom att klicka på respektive sökingång.

Söktips

Valfria sökord:

Titel/ord i titel:

Författare/textansvarig:

Utgivare:

Ämnesord:

Fler sökingångar

Sök endast i Svesöks länkkatalog

Visa endast titel

Sökning i SAFARI (svensk forskning).

~~<http://safari.hsv.se/>~~ <http://safari.vr.se/>

SAFARI hjälper dig att hitta svensk forskning på Internet. Alla sökträffar leder till information från universitet, högskolor och andra forskande myndigheter.

Enkel sökning:

Målgrupp: (public/ allmänheten, industry/ näringslivet, scientific/ forskarsamhället, students/ högskolestuderande, alla kategorier.)

Forskande myndighet: (stort antal att välja på):

"">Alla forskande myndigheter

"bth.se">Blekinge Tekniska Högskola<

chalmers.se">Chalmers Tekniska Högskola<

"foa.se">Försvarets forskningsanstalt<

"gu.se">Göteborgs universitet<

"hhs.se">Handelshögskolan i Stockholm<

"hkr.se">Högskolan Kristianstad<

"his.se">Högskolan Skövde<

"hb.se">Högskolan i Borås<

"hig.se">Högskolan i Gävle<

"hh.se">Högskolan i Halmstad<

"hj.se">Högskolan i Jönköping<

"hik.se">Högskolan i Kalmar<

"irf.se">Institutet för rymdfysik<

"kau.se">Karlstads universitet<

"ki.se">Karolinska institutet<

"kfb.se">KommunikationsForskningsBeredningen<

"kth.se">Kungliga Tekniska Högskolan<

"liu.se">Linköpings universitet<

"luth.se">Luleå Tekniska Universitet<

"lth.se">Lunds Tekniska Högskola<

"lu.se">Lunds universitet<

"lhs.se">Lärarhögskolan i Stockholm<

"mah.se">Malmö högskola<

"mh.se">Mitthögskolan<

"mdh.se">Mälardalens högskola<

"polar.se">Polarforskningssekretariatet<

"snsb.se">Rymdstyrelsen<

"socforsk.se">Socialvetenskapliga forskningsrådet<

"su.se">Stockholms universitet<

"sp.se">Sveriges Provnings- och Forskningsinstitut<

"slu.se">Sveriges lantbruksuniversitet<

"smhi.se">Sveriges meteorologiska och hydrologiskainstitut<

"sh.se">Södertörns högskola<

"umu.se">Umeå universitet<

"uu.se">Uppsala universitet<

"vxu.se">Växjö universitet<

"oru.se">Örebro universitet<

Fritextsökning.

Ämnesvis sökning:

Biomedicinska vetenskaper,

<http://safari.vr.se/browse/structure/B000.html>.sv

Humaniora, <http://safari.vr.se/browse/structure/H000.html>.sv

Naturvetenskap, <http://safari.vr.se/browse/structure/P000.html>.sv

Rättsvetenskap och juridik,

<http://safari.vr.se/browse/structure/S110.html>.sv

Samhällsvetenskap,

<http://safari.vr.se/browse/structure/S000.html>.sv

Teknik, <http://safari.vr.se/browse/structure/T000.html>.sv

Svesöks sökmaskin.

Utdrag ur <http://www.svesok.kb.se/info/index.html>

"av [Anders Sundström](#), Sunstone Systems AB, ver 2, 1998-10-20.

Introduktion

Svesök består huvudsakligen av tre delar: metadatagenerator, söktjänst och katalogtjänst.

Metadatageneratorn används i två syften, dels för att anmäla en länk till länkkatalogen och därmed (indirekt) till söktjänsten, dels för att skapa metadata till dokument och därmed också (indirekt) i söktjänsten via roboten.

Söktjänsten används för att återsöka dels katalogtjänsten, dels hela den svenska delen av Internet.

Katalogtjänsten används för att beskriva utvalda webbsidor (enligt Svesöks [urvalskriterier](#)). Länkarna finns i en hierarkisk ämnesstruktur som gör det möjligt för användaren att navigera bland länkarna.

Systemet består av ett antal olika samverkande delar som tillsammans bildar Svesök. Syftet med detta dokument är att översiktligt beskriva dessa delar så att den intresserade kan bilda sig en uppfattning om hur Svesök fungerar och på så vis kunna få större nytta av systemet.

Denna tekniska beskrivning vänder sig både till användare som söker information med hjälp av Svesök och till användare som vill sprida information med hjälp av Svesök. Både i det förra och det senare fallet blir resultatet bättre om användaren vet hur systemet samlar in och sprider information och hur metadata används för att ge ett förbättrat resultat.

Systemöversikt

Svesök består av följande delar:

- Metadatagenerator
- Redaktionell databas (SOLID)
- Redaktionell applikation
- Insamlingsrobot
- Metadataextraherare
- Sökmotor (BRS/SEARCH)
- Sökapplikation
- Länkkatalogs applikation

Hårdvara

Svesök körs på en Sun Enterprise 4000 UNIX-server som är placerad på LIBRIS, Kungl. biblioteket.

Mjukvara

Applikationen har utvecklats i PERL och C av [Dansk Internet Selskab](#).

Databashanterare för katalogtjänsten är relationsdatabasen SOLID. Databashanterare för söktjänsten är fritextdatabasen [BRS/ SEARCH](#) från [Dataware Technologies](#) levererad av [Sunstone Systems](#).

Svesök nyttjar den insamlingsrobot som används av [Kulturarw3](#) på Kungl. biblioteket. Roboten har ursprungligen utvecklats av [NetLab](#) i Lund.

Metadata

[Metadata](#) är data om data, dvs information som beskriver en resurs. Detta dokument har t.ex. metadata i form av titel och författare.

Svesöks metadatagenerator erbjuder vem som helst att med hjälp av ett [registreringsformulär](#) skapa metadata för att beskriva ett dokument. Användaren fyller i formuläret och generatorn skapar metadata i HTML-format som användaren kan klippa in i ett eget HTML-doku-

ment för att publicera på Internet.

Metadata skapas enligt [Dublin Core](#)-formatet som får allt större internationell spridning. Elementet titel heter t.ex. DC.Title och representeras i HTML som:

`<META NAME="DC.Title" CONTENT="Svesök - Teknisk beskrivning">`.

När Svesöks robot (och andra robotar) kommer till dokumentet kan detta indexerats effektivare genom att roboten kan dra nytta av det metadata som användaren lagt i dokumentet.

Om användaren önskar kan dokumentet samtidigt som metadata skapas också anmälas som en kandidat till Svesöks katalogtjänst. Anmälningarna bearbetas av Svesöks redaktion och länkar som uppfyller Svesöks urvalskriterier tas in i katalogen.

Länkkatalogens struktur baseras på Svesöks interna ämnesstruktur. För detta används inte metadata i dokumenten annat än för presentation av den bibliografiska beskrivningen.

I söktjänsten erbjuds både fältsökning och relevansrankning som baseras på metadata från dokumenten. Söktjänsten indexerar metadata enligt Dublin Core-formatet, men även metadata från andra metadata-system tas omhand genom att mappas till motsvarande Dublin Core-element.

Sökmotor

Svesöks sökmotor erbjuder användaren fritext- och fältsökning på dels ämnessidor och länkar i Svesöks katalogtjänst, dels HTML-sidor som finns på den "svenska" delen av Internet, insamlade av roboten. Med svenska menas webbsidor med adress som slutar på ".se", webbservrar med ".com", ".org", ".net" och ".edu" som är registrerade med svensk postadress samt svenska sidor registrerade under .nu.

Sökmotorn som används är BRS/Search, samma system som används för [LIBRIS Websök](#). Denlöpande texten är sökbar i fritext och

metadata är sökbar både i fält och i fritext.

Vid sökning i fritext används relevansrankning för att de mest relevanta länkarna ska presenteras först. Metadatafälten är indelade i två grupper där dokument med träff i den första fältgruppen rankas högre än träffar i den andra fältgruppen. Därefter rankas dokument med träff enbart i fritext. Inom varje grupp visas i första hand ämnessidor i Svesök, i andra hand länkar i Svesöks katalog och i tredje hand länkar insamlade av roboten. Sammantaget innebär detta att dokument med träff i metadata rankas högre samt att dokument från Svesöks länkkatalog rankas högre än dokument som hämtats in av roboten.

Fältsökning möjliggör mer precis sökning även om användaren måste vara medveten om att flertalet dokument i Svesök (till skillnad från t.ex. LIBRIS Websök) helt saknar metadata och därmed missas vid fältsökning. även vid fältsökning presenteras ämnessidor i Svesök först, följt av länkar i katalogen och därefter länkar från roboten. Här sker ingen relevansrankning.

Det är inte bara DC-metadata som används vid fältsökning och relevansrankning utan även andra typer av metadata. Detta för att i så hög grad som möjligt utnyttja de metadata som överhuvudtaget finns i HTML-dokumentet som roboten samlat in.

Databas

Databasen består av följande delar:

Ämnessidor i Svesök

Länkar i katalogen

Länkar insamlade av roboten

Ämnessidorna har visst metadata, är sökbara i fritext och även delvis med fältsökning. Länkarna i katalogen har metadata skapat/ kontrollerat av redaktionen och är sökbara såväl med fritextsökning som med fältsökning.

Länkar insamlade av roboten kan ha metadata och är då sökbara såväl med fritextsökning som medfältsökning. De kan också helt sakna metadata och därför endast vara sökbara med fritextsökning. Detta innebär bl.a. att användaren vid fältsökning helt missar dokument insamlade av roboten vilka inte har metadata.

Insamlingsrobot

Svesök använder sig av Kulturarw3s robot som förutom att samla in svenska webbsidor för att bevaras åt eftervärlden sedan en tid också levererar alla HTML-sidor till Svesök.

Roboten löper runt den "svenska" delen av Internet i syfte att bevara den information som publiceras där för eftervärlden. Roboten genomser varje dokument som laddats ner och eventuella URL:er som påträffas läggs in i en kö som successivt betas av.

Eftersom det tar roboten ett antal veckor att ta sig "ett varv" runt den svenska delen av Internet kan det ta en tid innan ett nytt dokument blir tillgängligt i Svesök.

Alla textsidor som roboten hittar förs kontinuerligt över till Svesök där de tas omhand och läggs in i fritextdatabasen. Eventuellt metadata i HTML-dokumentet görs fältsökbart på samma sätt som metadata som registrerats i katalogen.

Teknisk projektorganisation

Svesök har utvecklats av [Dansk Internet Selskap](#). Projektledare från DIS är [Jakob Sand](#) och tekniskt ansvarig [Brian Peter Thorsbro](#).

Kulturarw3 levererar textsidor till Svesök. Ansvarig från Kulturarw3 är [Allan Arvidson](#).

[Sunstone Systems](#) har ansvarat för övergripande teknisk samordning inom projektet samt installation av fritextdatabashanteraren BRS/SEARCH, specifikation av Dublin Core-element, databasstruk

tur och relevansrankning. Teknisk rådgivare från Sunstone är [Anders Sundström](#).

Från LIBRIS-avdelningen, Kungl. biblioteket medverkar [Kai Pöykiö](#) och [Martin Malmsten](#).

Skapad: 1998-05-01, **senast uppdaterad:** 2000-08-10, **e-post:** svesok@kb.se"

KB och nationalbibliografin

Kungl.Bibliotekets verksamhet med nationalbibliografin behandlas på "Nationalbibliografin. Monografienheten (NBm)" (Uppdaterad ~~000331~~ 020212).

<http://www.kb.se/Nbm/nbmstart.htm>"Monografienhetens huvudsakliga uppgifter är att katalogisera, klassificera och sätta ämnesord på svenska monografier för databasen LIBRIS. Vid enheten sammanställs Sveriges nationalbibliografi Svensk bokförteckning."

KB, Metadata, Dublin Core

Utdrag från Kungl.Biblioteket, KB, sida om Metadata och Dublin Core (Uppdaterad 000426).
<http://www.kb.se/bus/dc/metadata.htm>

"Vad är metadata?"

Med metadata menas data om data, dvs en beskrivning eller en utsaga om något slags data. Prefixet meta- har länge använts på det här sättet inom filosofin, särskilt på logikens område. I den meningen innehåller t.ex. en katalog eller en bibliografi metadata om de böcker eller handskrifter den förtecknar.

Idag används termen ofta inom IT-området. Där betyder metadata en beskrivning av en webbsida, som ligger på sidan själv. Det är som om man fäste ett katalogkort direkt på det elektroniska dokumentet.

Varför behövs metadata?

Ett problem med Internet är att det är svårt att hitta det man söker med de sökmaskiner som finns idag. Man får alltför många träffar. Söker man ett träd, får man hela skogen som svar, och det är svårt att hitta det rätta trädet bland alla de andra. Sökmotorerna på Internet söker i dokumentets hela text. Resultatet blir oftast långa listor med dålig rangordning. Sidor som verkligen behandlar det sökta ämnet, blandas med sådana där det nämns i en tabell, parentes eller fotnot. Ofta är det först framåt sidan tio i listan man hittar något som liknar det man söker.

Man skulle kunna tänka sig att biblioteken och då särskilt nationalbiblioteken tog på sig att katalogisera webbsidorna så att de blev lätta att finna. Men det är orealistiskt. KB katalogiserar omkring 15.000 svenska böcker per år, medan de svenska webbsidornas antal i dag är

flera miljoner. Därför vore det bra om de som producerar sidorna själva beskrev dokumenten med de viktigaste uppgifterna om upphovsman, ämne etc. Man har i årtal diskuterat behovet av metadata i Internetvärlden. Men man har hittills haft svårt att ena sig om formerna, utöver själva html-taggen <META>.

Kungl. biblioteket stöder Dublin Core!

Allt fler aktörer håller nu på att försöka enas kring ett av de olika förslag till metadata som diskuterats, Dublin Core, och de element som ska ingå. Därför stöder KB Dublin Core. Du kan redan i dag sätta metadata på dina sidor, och på så sätt göra dem lättare att hitta. Använd något av formulären på sidan "Skapa metadata". De metadata som skapas kan Du kopiera och klistra in i ditt HTML-dokument mellan HEAD-taggar <HEAD></HEAD>."

Dublin Core behandlas på Kungl.Bibliotekets sida
"Referenser till information om metadata och Dublin Core" (Uppdaterad 000426).
<http://www.kb.se/bus/dc/referenser.htm>

Där står bl a:

"Dublin Core Metadata Initiative (DCMI) är den organisation som administrerar arbetet med och utvecklingen av Dublin Core. Vill man fördjupa sig ordentligt i ämnet är deras hemsida en bra utgångspunkt. <http://purl.org/DC/>

*** Nordic Metadata Projects** - ett samnordiskt metadataprojekt grundat av NORDINFO. Härfinns också många länkar att gå vidare från. <http://hul.helsinki.fi/meta/index.html>

Kungl.biblioteket har en avdelning,"BUS", som bl a har hand om bibliotekets verksamheter rörande Dublin Core. Se Kungl.Bibliotekets sida"Enheten för bibliografisk utveckling och samordning (BUS)" (Uppdaterad 000516 010927). <http://www.kb.se/bus/bus.htm>

Där finns hänvisningar till bl a BUS-nytt, BUS sidor om Metadata och Dublin Core, SAB:s klassifikationssystem (rättelser, ändringar och tillämpningar), SAB:s klassifikationssystem (NB:s rubrikregister till Svensk bokförteckning, Svenska ämnesord (databas), Svesök, BUS sidor om ämnesord.

a) Kungl.Biblioteket, KB, behandlar Dublin Core på sin sida"Dublin Core metadata-element, Version 1.1: Referensmodell" (Uppdaterad 000529 000928).
http://www.kb.se/bus/dc/DC_SV1_1.HTM

Dublin Cores metataggar består för närvarande av 15"element", 15 olika typer av uppgifter. Det är inte alldeles klart vad som menas med de 15 elementen, det förekommer olika uttolkningar.Tolkningar finns dessutom på bl a följande sidor b)-d)

b) Dublin Core Metadata Element Set: Ett svenskt referensdokument.

<http://www.sics.se/%7Epreben/DC/dcref-swe.html>

"Detta är en preliminär version av översättningen av Dublin Core till Svenska. Senast uppdaterat: 1999-02-28".

c) Sid 16-18 i"Metadata enligt Dublin Core",
<http://www.kult.lu.se/bivil/mag99.html> [online och pdf]: tillämpningar och konsekvenser i de svenska kvalitetssöktjänsterna SAFARI, Svenska miljönätet och Svesök (Uppsatsen daterad 29 juni 99).Med" en sammanfattande redogörelse för de femton DC-elementen. Vissa av dem är fortfarande under utveckling och det är inte helt klart hur de i framtiden kommer att användas".

d)"Dublin Core Metadata Template" på
<http://www.lub.lu.se/cgi-bin/nmdc.pl>

(Uppdaterad 980317): Provided by the "Nordic Metadata Project".

Utdrag från Kungl.Biblioteket, KB, sida "Skapa metadata". (Uppdaterad 000426).
<http://www.kb.se/bus/dc/skapametadata.htm>

" Det enklaste sättet att skapa metadata på är att använda en s.k. metadata-generator. En sådan genererar färdig metadata utifrån ett formulär. Man fyller själv i formuläret, skickar iväg det och får tillbaka en färdig"metadatapost" som man sen kan klistra in i sitt eget dokument. Den ska placeras någonstans mellan html-taggar <HEAD> </HEAD>. Formulär hittar man bl.a. hos följande tjänster:

Nordic Metadata Projects (<http://www.lub.lu.se/cgi-bin/nmdc.pl>) ... Formuläret är lätt att läsa, innehåller boxar för samtliga Dublin Core-element, har hjälptexter samt popup-menyer för val av språk, ämnesordslistor m.m....

Svesök (<http://www.svesok.kb.se/user/submit/>) Denna DC-generator på Kungl. biblioteket används framför allt när man vill anmäla sin sida till Svesöks länkkatalog över svenska webbsidor. Men man måste inte platsa i länkkatalogen för att få använda generatoren, det går alldeles utmärkt att bara välja skapa metadata och att sedan kopiera och klistra in den i sitt eget dokument.

SAFARI (<http://safari.hsv.se/metadata/?lang=sv>) För Högskoleverkets SAFARI gäller samma som för Svesök ovan, man kan använda generatoren bara för att skapa metadata och klistra in i sitt dokument. Formuläret är dock speciellt anpassat för institutioner och organisationer som vill anmäla sidor om pågående forskning."

Hjälpmedel för metadata

Nordic Metadata Project

Metadatagenerator: Utdrag från "Dublin Core Metadata Template" på (<http://www.lub.lu.se/cgi-bin/nmdc.pl>) (Uppdaterad 980317):

" This service is provided by the "Nordic Metadata Project" in order to assure good support for the creation of Dublin Core metadata to the Nordic "Net-publisher" community. If you use the metadata created by this form and follow our examples, term lists and recommendations, your HTML documents will carry high quality metadata.....

1 TITLE of the resource to be described

Alternative title (Titles other than main title)

2 CREATOR (Name of the person or organization primarily responsible for creating the intellectual content) Creator name. Personal name: "Last, First" .Corporate name.Creator's (Email) address.

3 SUBJECT: Keywords (Your own keywords describing the topic of the resource, one per box)

3 SUBJECT: Controlled vocabulary (Keywords from established schemes, one per box) (Se särskild lista nedan)

3 SUBJECT: Classification (Notations for the resource, one per box) (Se särskild lista nedan)

4 DESCRIPTION (Abstract, content description)

5 PUBLISHER (University department, corporate entity etc.) Publisher's (Email) address

6 CONTRIBUTOR (Name of significant contributors other than

the creator) Contributors namePersonal name: "Last, First"Corporate name

7 DATE (Date associated with the creation or availability of the resource) 1996-09-23 (ISO 8601).

19960923 (ANSI.X3.30-1985).

Mon, 23 Sep 1996 (RFC 822)

8 TYPE (Category of the resource) (Se särskild lista)

9 FORMAT (Data representation of the resource, MIME type) (Se särskild lista)

10 IDENTIFIER: URL (Location of the document. Start with 'http://')

10 IDENTIFIER (String or number used to uniquely identify the resource described by this metadata) URN, ISBN, ISSN, SICI, FPI
If you think of your resource as a Finnish, Norwegian or Swedish publication, we are authorized to give you a unique (URN) number here: Get a URN

11 SOURCE (Unique string or number for a printed or digital work from which this resource is derived) Free text, URL, URN, ISBN, ISSN

12 LANGUAGE of the content of the resource described (lång lista med nära 200 språk att välja mellan).

13 RELATION (Relationship to other resources) Free text, URL, URN, ISBN

14 COVERAGE (Spatial and/or temporal characteristics of the resource) Free text, Place name, Period name

15 RIGHTS (Link to a copyright notice etc.) Free text, URL, URN....."

Dublin Core Metadata Template på
(<http://www.lub.lu.se/cgi-bin/nmdc.pl>) :
Alternativ som kan väljas:

Dublin Core Metadata Template på

(<http://www.lub.lu.se/cgi-bin/nmdc.pl>) :

Alternativ som kan väljas:

3 SUBJECT: Controlled vocabulary (Keywords from established schemes)

Library of Congress Subject Headings
General Finnish Subject Headings
General Finnish Subject Headings (in Swedish)
Finnish Geographical SH
Finnish Music Thesaurus
Finnish Legal SH
Finnish Thes. of the Social Sciences
Religion Indexes SH (FI)
Finnish Forestry SH
Finnish Agricultural SH
Index Terms for Education and Psych. (FI)
SAB ämnesordslista (Sweden)
HUMORD (Norw. Thes. for Arts and Humanities)
NTUB (Norw. Thes. for Techn. Sciences)
DBC SH (Denmark)
LANDSLYK (Nat. Library Iceland)
KERFEFN (Icelandic Systematic Thes.)
Medical Subject Headings (MeSH)
Art & Architecture Thesaurus
Union List of Artist's Names
LC Thes. for Graphic Materials I: Subject Terms
TGM2: Genre and Physical Characteristic Terms
HASSET: Humanities and Social Science Electronic Thesaurus
RCHME/English Heritage Thesaurus
OECD Macrothesaurus
EUDISED

SCIMP/SCANP Thesaurus
Legislative IV
Global Legal INEURODICAUTOMERIC
Ei Thesaurus
AGROVOC
Thesaurus of Geographic Names
NASA Thesaurus
Astronomy Thesaurus
Index Nominum Genericorum (Plantarum), ING
Mammal Species

3 SUBJECT: Classification (Notations for the resource)

Dewey Decimal Classification
YKL (Finnish Public Libr. Class.)
SAB Class. (Sweden)
MOS Class.
DK5 (Danish Decimal Class.)
Library of Congress Class.
Universal Decimal Classification
Colon Classification
Regensburger Verbundklassifikation
National Library of Medicine Class.
Journal of Economic Literature Classification
Engineering Information Class.
Mathematics Subject Classification
ACM Computing Classification System
Physics and Astronomy Classification Scheme
Iconclass
Social History & Industrial Classification

8 TYPE (Category of the resource)

Text
Text.Abstract
Text.Advertisement
Text.Article

Text.Correspondence
Text.Correspondence.Discussion
Text.Correspondence.Email
Text.Correspondence.Letter
Text.Correspondence.Postcard
Text.Dictionary
Text.Form
Text.Homepage
Text.Homepage.Organizational
Text.Homepage.Personal
Text.Index
Text.Manuscript
Text.Minutes
Text.Monograph
Text.Pamphlet
Text.Poem
Text.Proceedings
Text.Promotion
Text.Serial
Text.Serial.Journal
Text.Serial.Magazine
Text.Serial.Newsletter
Text.Serial.Newspaper
Text.TechReport
Text.Thesis
Text.Thesis.Doctoral
Text.Thesis.Masters
Image
Image.Moving
Image.Moving.Animation
Image.Moving.Film
Image.Photograph
Image.Graphic
Sound
Sound.Ambient
Sound.Effect
Sound.Music
Sound.Narration
Sound.Speech
Software
Software.Executable
Software.Source
Data
Data.Numeric

Data.Spatial
Data.Spectral
Data.Statistical
Data.Structured-Text
Interactive
Interactive.Chat
Interactive.Games
Interactive.Multimedia
Interactive.VR

9 FORMAT (Data representation of the resource, MIME type)

text/html (.htm, .html)
application/msword (.doc)
application/pdf (.pdf)
application/postscript(.ai,.eps,.ps)
application/powerpoint (.ppt)
application/rtf(.rtf)
application/wordperfect5.1
application/x-latex (.latex)
image/gif (.gif)
image/jpeg (.jpeg, .jpe, .jpg)
text/plain (.txt)
text/x-sgml (.sgml, .sgm)

Metadatagenerator:

Registrera webbsidor i Svesök

<http://www.svesok.kb.se/user/submit/> Där sök långt
formulär, leder till

<http://www.svesok.kb.se/user/submit/urlsearch.pl> (Senast
uppdaterad 000120)

Fyll i relevanta fält:

Nätadress:	(URL) (Ange sidans/webbplatsens URL)
Standardnummer:	
Titel:	(Ange sidans/webbplatsens titel)
Alternativ titel:	(Ange om titeln på skärmsidan är en annan än dokumentets titel)
Författare/textansvarig:	(Ange Person och/eller organisation som är ansvarig för sidans innehåll. Skriv: Efternamn, Förnamn eller organisationens namn) (ex. Hansson, Peter)
Medarbetare:	(ex. Hansson, Peter)
Utgivare:	(Ange namn på den som har publicerat dokumentet på Internet)
Utgivarens e-post:	
Kontrollerade ämnesord:	(Ange ett ord eller begrepp per fält. Orden ska hämtas ur Svenska ämnesord.)
Egna ämnesord:	(Ange ett ord eller begrepp per fält)
Sammanfattning:	(Beskriv dokumentet med 3-5 rader)
Önskemål om placering i Svesöks ämnesstruktur:	(Ange önskemål om var sidan/webbplatsen ska placeras i Svesöks ämnesstruktur)
Datum:	(Använd enligt:1998-05-23)
Datum då sidan skapades:	

Datum då sidan senast uppdaterades:

Om sidan uppdateras, hur ofta görs det:

Om sidan är temporär, när skall den i så fall tas bort ur Svesök:

Typ av resurs:

(Välj det alternativ som passar bäst)

Format:

Källa:

Språk:

Relation:

Täckning:

Rättigheter:

Anmärkning:

Adressuppgifter:

(Välj det eller de alternativ som passar bäst för att beskriva sidan/webbplatsen)

(Välj det eller de format som används på sidan/webbplatsen)

(Om sidan kommer från ett större tryckt eller digitalt verk, ange källans titel,URL,ISSN etc)

(Välj det eller de språk som används huvudsakligen i dokumentet)

(Om sidan hör samman med andra dokument eller är en del av en webbplats, ange titel, URL, ISBN)

(Ange om innehållet är begränsat till en viss tidsperiod eller geografiskt område)

(Ange här vad som gäller för användningen av din sida/ webbplats med en kort fras eller en URL till egen upphovsrätt-text)

(Eventuella meddelanden till Svesöks redaktion)

Vill du ha nyheter om Svesök skickade till dig via e-post?

Namn: E-postadress:

**Förhandsgranska registreringen.
Skapa metadata. Registrera och skapa metadata. Återställ**

Metadatagenerator:

Registrera webbsidor i SAFARI

<http://safari.vr.se/metadata/safari.pl?lang=sv>

(Registreringshjälp se <http://safari.vr.se/metadata/reghelp.html>)

(SAFARI använder CERIF klassifikationssystem)

TITEL på den källa/resurs som ska registreras.

FÖRFATTARE/TEXTANSVARIG (Namn på den person eller organisation som är huvudansvarig för källans intellektuella innehåll. Ange Efternamn, Förnamn för personer.)

Författarens (Email) adress

URL (Denna källas www-adress. Börja med <http://>)

ÄMNE: Klassifikation CERIF (Nyckelord ur CERIF som beskriver källan. OBS! Ange bara ett ämnesord per fält). CERIF verktyg (med vars hjälp man kan få fram och föra in önskade ämnesord från CERIF)

ÄMNE: Nyckelord (Fritt valda nyckelord som beskriver källan. Du kan använda nyckelord separerade med komma ','))

Målgrupp (Ange vilken av följande målgrupper som denna resurs i första hand vänder sig till)

BESKRIVNING (Sammanfattning av resursens innehåll)

UTGIVARE (Institution, avdelning eller motsvarande inom organisationen. Om organisationen är en myndighet, ange även myndighetens namn (avdelning/myndighet))

Förläggarens (Email) adress

DATUM (Bäst före datum)

TYP (Resursens kategori/typ). TYPverktyg (med vars hjälp man kan få fram och föra in önskad typ)

SPRÅK i vilket resursens innehåll är skrivet

TÄCKNING: geografisk (Geografisk täckning för resursen). TÄCKNINGSS verktyg (med vars hjälp man kan ange länder).

Oöppnade DC-fält (som kan öppnas och komplettera ovanstående)

Alternativ titel (Titel skild från huvudtiteln)

ÄMNE: Vokabulär (Nyckelord från en etablerad vokabulär, ett per fält)

ÄMNE: Klassifikation (Klassificeringskoder efter etablerade system, en per fält)

MEDFÖRFATTARE (Namn för medförfattare, undantaget huvudförfattaren)

DATUM (Datum kopplat till skapandet eller tillgängliggörandet av resursen)

FORMAT (MIME-typ för resursen)

KÄLLA (Unik identifierare för källan till den beskrivna resursen)

BEROENDEN (Beroende/förhållande till andra resurser)

TÄCKNING (Rums och/eller tidsmässig täckning för resursen)

RÄTTIGHETER (Länk till en copyright text, etc.)

Copyright © 1998 Kungl. biblioteket. Skapad: 1998-05-12. Senast uppdaterad: 2000-01-20

Resurstyper för Safari:

Beskrivning av personlig expertis

Kortfattad beskrivning av projekt

Omfattande beskrivning av projekt

Textdokument:

Artiklar o korta texter, t ex essäer, noveller

Avhandlingar o examensarbeten i allmänhet

Beskrivning av kompetens och expertis

Proceedings.

Sammanfattning eller beskrivning av objekt

Ett kort dokument som ger allmän

information om ett ämne

Hemsidor

Informationsmaterial om produkter, tjänster eller organisation

Manuskript, opublicerat verk

Meddelanden till personer eller grupper

Monografi eller bokliknande verk

Periodiska utkommande resurser

Poesi. Dikt eller diktsamlingar

Preprint av ett arbete

Protokoll ell anteckningar fr möten

Rapport

Samling av bidrag till en konferens.

En sammanfattande förteckning

Andra resurstyper:

Bilder, animationer och andra visuella medier

Olika former av data

Interaktiva resurser för en eller flera användare.

Olika former av ljudinspelningar

Programvaror i allmänhet

Allmänt schema för metataggar.

(DC står för Dublin Core) Numren på metataggarna ingår ej i taggarna. Taggarna kan innehålla följande. Efter "content=" ska det stå det man vill ha inskrivet.

<HTML><HEAD> <TITLE></TITLE>

1 Titel <meta name="DC.Title" content=" ">

1 <meta name="DC.Title" LANG="en" content=" "> (Engelsk titel)

1 <meta name="DC.Title.Alternative1" content=" ">

1 <meta name="DC.Title.Alternative2" content=" ">

2 Upp- <meta name="DC.Creator" content=" ">

2 hovsman <meta name="DC.Creator.PersonalName" content=" ">

2 <meta name="DC.Creator.PersonalName.Address" content=" "> (Epostadress)

2 <meta name="DC.Creator.CorporateName" content=" ">

2 <meta name="DC.Creator.CorporateName.Address" content=" "> (Epostadress)

3 Ämne <meta name="DC.Subject" content=" ">

3 <meta name="DC.Subject" LANG="en" content=" "> (på engelska)

3 <meta name="DC.Subject.Keyword" content=" ">

3 <meta name="DC.Subject.Keyword" content=" ">

3 <meta name="DC.Subject.Keywords" content=" ">

3 <meta name="DC.Subject.ControlledVocabulary" content="(scheme=SABAO) ">

3 <meta name="DC.Subject.ControlledVocabulary" content="(scheme=SABAO) ">

3 <meta name="DC.Subject.Classification" content="(scheme=SAB) ">

3 <meta name="DC.Subject.Classification" content="(scheme=UDK) ">

3 <meta name="DC.Subject.Classification" content="(scheme=DDC) ">

3 <meta name="DC.Subject.Classification" content="(scheme=-) ">

4 Beskriv- <meta name="DC.Description" content=" ">

4 ning <meta name="DC.Description" LANG="en" content=" "> (på engelska)

5 Utgivare <meta name="DC.Publisher" content=" ">

5 <meta name="DC.Publisher.PersonalName" content=" ">

5 <meta name="DC.Publisher.PersonalName.Address" content=" "> (Epostadress)

5 <meta name="DC.Publisher.CorporateName" content=" ">

5 <meta name="DC.Publisher.CorporateName.Address" content=" "> (Epostadress)

6 Med- <meta name="DC.Contributor " content=" ">

6 arbetare <meta name="DC.Contributor.PersonalName" content=" ">

6 <meta name="DC.Contributor.PersonalName.Address" content=" "> (Epostadress)

6 <meta name="DC.Contributor.PersonalName2" content=" ">

6 <metaname="DC.Contributor.PersonalName2.Address" content=" "> (Epostadress)

6 <meta name="DC.Contributor.CorporateName" content=" ">

6 <meta name="DC.Contributor.CorporateName.Address" content=" "> (Epostadress)

7 Datum <meta name="DC.Date" content="(scheme=ISO8601) 2000-07-01">

7	<meta name="DC.Date.Created" content="(scheme=ISO8601) 1999-06-01">
7	<meta name="DC.Date.Modified" content="(scheme=ISO8601) 1999-09-09">
8 Typ	<meta name="DC.Type" content=" (scheme=-) ? ">
8	<meta name="DC.Type" content=" (scheme=)Text.Homepage.Personal ">
8	<meta name="DC.Type" content=" (scheme=)Text.Homepage.Organizational ">
9 Format	<meta name="DC.Format" content=" (scheme=MIME)text/html ">
10 Identifier-	<meta name="DC.Identifier" content=" (scheme=URL) ">
10 kator	<meta name="DC.Identifier" content=" (scheme=URN) ">
10	<meta name="DC.Identifier" content=" (scheme=ISBN) "> (URL. URN. ISBN. ISSN. SICI. FPI .)
11 Källa	<meta name="DC.Source" content=" (scheme=-) - "> (Free text. URL. URN. ISBN. ISSN.)
12 Språk	<meta name="DC.Language" content=" (scheme=ISO639-1)sv">
13 Relation	<meta name="DC.Relation" content=" (scheme=-)- "> (Free text . URL. URN. ISBN. ISSN.)
14 Täck-	<meta name="DC.Coverage" content=" "> (Free text)
14 ning	<meta name="DC.Coverage.Place name" content=" (scheme=-) ">

14 <meta name="DC.Coverage.Period name" content=" (scheme=-) ">

15 Rättig- <meta name="DC.Rights" content=" ">
heter (Link to: Free text. URL. URN.)
</HEAD> <BODY></BODY></HTML>

*"scheme=" anger den standard man använder när man anger "content"

* Lang anger språkkod för ett visst metadataelement, dvs på vilket språk dess värde är uttryckt

Exempel: <META NAME="DC.Subject.Keyword"LANG="sv" CONTENT=" ". Används i samband med Title, Subject, Description.

* SAFARI har en tagg"målgrupp", men den ingår inte i Dublin Core-taggar.

* När man anger ämnesord bör man helst ange bara ett ord per tagg, men SAFARI tillåter flera ordåtskilda av komma när det gäller fritt valda ämnesord.

* I Svesök bör man använda SABs ämnesord. SAB ämnesordslista (Sweden)= SABAO. KB-BUS har ca 30 000 ämnesord i databas: <http://www.arnesord.kb.se/>

* SABs ämnesrebriker (enligt Webb-sida på Internet 5 juli 2000 : "Rubrikregister till Svensk bokförteckning"

<http://www.kb.se/Nbm/rubrikregister.htm>) finns i"En jämförelse mellan tre klassifikationssystem: SW, UDK och SAB" på <http://w1.861.telia.com/~u86105430/omr102.html>

* CERIFs ämnen finns under SAFARI på <http://w1.861.telia.com/~u86105430/omr102.html>

Övrigt.

Det finns på Internet många frågetecken, åsikter och kompletteringar kring metataggar och Dublin Core. Här ovan har tagits med det som bedömts viktigast i en översikt som avses ge en introduktion till problemen kring sökning på Internet. I område 102 i "Samhällsplaneringens problem" har tagits in redovisningar av klassifikationssystem och hänvisningar till kataloger och söksystem på Internet.

XML, RDF o d

Metadata för medicinskt bruk: RDF od. 22 sidor.

<http://www.kib.ki.se/kib/project/rdf/report/rdf-report-2000-08-21.pdf>

Användning av Resource Description Framework för medicinska resurser på Internet. John Parnefjord, Ulf Kronman och Elisabet Fri-gell. Karolinska Institutets Bibliotek. Version 1.1.21 augusti 2000. Rapporten har skrivits med stöd från BIBSAM inom ramen för Karolinska Institutets Biblioteks åtaganden som ansvarsbibliotek inom medicin och omvårdnad. Mer information om rapporten och Karolinska Institutets Biblioteks aktiviteter kring RDF finns på bibliotekets webbplats på adressen <http://www.kib.ki.se/kib/project/rdf/>.

“Sammanfattning

Internet har länge varit i behov av en mer organiserad beskrivning av sina informationsresurser. Ett flertal standarder för metadata har tagits fram, men inte fått någon större spridning, eftersom de varit för begränsade eller specialiserade. Webbkonsortiet W3C har tagit fram ett standardiserat ramverk för metadata vid namn Resource Description Framework - RDF. RDF är en tillämpning av XML – Extensible Markup Language – och ser ut att ha alla förutsättningar för att kunna bli en framtida enhetlig standard för metadatahantering på Internet.

RDF är ett utbyggbart ramverk där man samtidigt kan använda flera olika system för metadata. Ramverket är så konstruerat att metadata-system kan utvecklas och användas utan central kontroll och koordinering. Utbyggbarheten hos RDF kommer sannolikt att säkra

framtiden för metadatastandarden Dublin Core, eftersom RDF tillåter att organisationer lägger till sina egna system för specialiserade metadata utan att förändra Dublin Core.

RDF är ännu ofärdig och oprövat. Redskap för produktion och tolkning av RDF saknas. För den som står i begrepp att skapa ett eget metadata-system är det ändå rekommendabelt att så snart som möjligt börja använda sig av RDF.”

“Vad är metadata?

Prefixet *meta* brukar i allmänhet användas för att beteckna att man talar om något – att man höjer abstraktionsnivån och försöker beskriva eller förklara något. Med den definitionen betyder metadata att man använder data för att beskriva eller förklara innebörden av vissa andra data. Metadata är således data med *information om data*.

Metadata är något vi använder dagligen, oftast utan att tänka på det. Ta följande exempel: siffrorna 6453782 betyder troligen ingenting för dig, men om vi tillför lite metainformation genom att skriva *Tel: 6453782* blir det genast klart för de flesta av oss att siffrorna ska tolkas som ett telefonnummer. Datat *Tel:* har alltså givit datat 6453782 en innebörd, en tolkning. Man kan alltså säga att metadata hjälper oss att omvandla data till meningsfull *information*.

Ett traditionellt bibliotek är ett annat bra exempel på användning av metadata. Bibliotekskatalogen är ju inget annat än en stor samling metadata över de böcker och tidskrifter som finns i biblioteket. Utan uppgifterna i katalogen skulle det bli mycket svårt och tidsödande att hitta det man söker.

Behövs metadata på Internet?

Med över två miljarder webbsidor är Internet idag en ovärderlig källa till information. Det är dock ofta svårt att hitta det man söker ute på nätet. Det vanligaste sättet att navigera på nätet är att söka hos de web-

bplatser som samlar in webbsidor och erbjuder dem i söktjänster med namn som till exempel AltaVista, HotBot, Evreka, EuroSeek och Fast Search. De program som samlar in dokumenten från nätet och gör dem sökbara läser av och förtecknar vartenda ord i dokumenten. Det innebär att man oftast kan hitta ett dokument utan att några metadata behöver tillföras, till skillnad från en traditionell bibliotekskatalog.

Bristen på strukturerad metainformation i de stora söktjänsterna leder dock ofta till att man får onödigt många sökträffar och låg precision i sin sökning. Om jag till exempel använder ordet *Strindberg* för att försöka hitta texter av Strindberg i söktjänsten Evreka får jag massvis med förslag på sidor. Eftersom August Strindberg är en berömd författare har det skrivits mycket *om* honom. Däremot är det mer sparsamt med sidor med texter där Strindberg själv är *författare*. Om jag skulle kunna tillföra lite metainformation till min sökning som sade att jag bara är intresserad av sidor med texter där Strindberg är författare skulle situationen genast bli bättre. En sådan sökning förutsätter dock att den som har publicerat en text av Strindberg på Internet också har tillfört den metainformationen till resursen, något som tyvärr är mycket ovanligt idag.

Vartefter mängden dokument på Internet ökar kommer alltså modellen med att söka fritt blandalla orden i dokumenten alltså att bli allt mindre hållbar. För att söktjänsterna ska kunna erbjuda förbättrad sökning med hjälp av metainformation krävs då att dokumenten förses med metadata enligt någon fastställd standard.

Metoder för metadata på Internet

Det har tagits ett flertal initiativ till metadatastandarder för dokument på Internet. Mångfalden visar att det finns ett stort behov av förbättrade beskrivningar av nätets resurser. Den rika floran har dock i sig hittills utgjort ett hot mot den praktiska användningen av metadata-beskrivningar. Splittringen har lett till att det har saknats stöd både för

produktion av metadata i publiceringsredskapen och för användning av dem i sökredskapen. Det är inte *flera* standarder vi behöver utan *en* standard som majoriteten kan ansluta sig till.

De två vanligaste metoderna för att foga metadata till webbsidor är idag HTMLs inbyggda *META*-tagg och biblioteksvärldens initiativ *Dublin Core* (DC). Vi börjar med en kort beskrivning av dessa metoder och en diskussion av problemen med dem. Vi går sedan över till att beskriva det som ser ut att kunna bli den framtida samordnande standarden för metadata, nämligen RDF - Resource Description Framework.

HTML META

I standardkodningen av webbdokument finns ett grundläggande stöd för metainformation i dokumentets sidhuvud. Metainformationen utgörs av två olika HTML-taggar – *TITLE* och *META* – som beskriver dokumentets titel, innehåll, upphovsman och några nyckelord. Dessa metadata presenteras inte på själva webbsidan, utan är enbart avsedda att vara tillgängliga för sökning. Söktjänster som stödjer dessa metadata använder dem för att förbättra sökresultat och för presentation i träfflistor. Exempel på sådana tjänster är AltaVista, Evreka, HotBot och InfoSeek.

TITLE är den enda metainformation som är "obligatorisk" enligt HTML-standard. Övrig metainformation läggs till med hjälp av HTML-taggen META, som används för att ange både författare, beskrivning och nyckelord.

En beskrivning av en webbsida skulle med HTML META kunna se ut på följande vis:

```
<HEAD>
```

```
<TITLE>Metadata för medicinskt bruk</TITLE>
```

```
<META NAME="author" CONTENT="John Parnefjord, Ulf Kron-  
man, Elisabet Frigell">
```

```
<META NAME="description" CONTENT="Metadata för medicinskt bruk">
<META NAME="keywords" CONTENT="Metadata, Resource Description Framework, RDF">
</HEAD>
```

Av exemplet framgår att HTML META har begränsade möjligheter till att beskriva ett dokument uttömmande. Här saknas till exempel uppgifter om utgivare, upphovsrätt och publiceringsdatum. Standarden säger heller inget om hur innehållet i taggarna ska vara utformat.

En undersökning gjord av tidskriften Nature [Nature] visar att en tredjedel av dokumenten på webben har metadata i form av HTML META. I undersökningen återfanns dock 123 olika varianter på HTML META, varav flertalet inte användes för att beskriva själva innehållet på sidorna, utan för interna administrativa ändamål.

Problemen med HTML META kan således sammanfattas med:

- * För yttlig beskrivning av dokumenten
- * Ej standardiserad utformning av innehållet i beskrivningarna
- * Avsaknad av utbyggbarhet

Till försvar för HTML META kan man dock anföra att det är en fungerande och väl spridd standard som stöds av många söktjänster.

Dublin Core

För att avhjälpa bristerna i HTML META togs år 1995 ett initiativ för att förbättra informationsåtervinningen av nätverksbaserade resurser, med särskilt fokus på webben. Strävan var att skapa samförstånd kring ett generellt metadatasystem präglat av enkelhet. Systemet kom sedermera att få namnet *Dublin Core*, ofta förkortat DC [DC].

Dublin Core använder 15 standardiserade metadataelement för att beskriva de centrala aspekterna hos resurser på webben. Elementens betydelse är avsiktligt brett definierade för att kunna tillgodose olika intressegruppers behov.

Tabellen nedan visar elementen indelade i kategorier utifrån vilka aspekter de avser att beskriva.

De tre kategorierna är innehåll, upphovsrätt samt manifestation.

Innehåll	Upphovsrätt	Manifestation
Title	Creator	Date
Subject	Publisher	Format
Description	Contributor	Identifier
Source	Rights	Type
Language		
Relation		
Coverage		

När man beskriver en webbsida väljer man fritt bland de olika elementen. Det är också tillåtet att upprepa samma element flera gånger i beskrivningen.

Användningen av Dublin Core i HTML-sidor bygger på en utvidgning av HTMLs META-standard.

Genom att använda attributen NAME och CONTENT till HTMLs META-taggar specificerar man olika element.

En beskrivning av ovan nämnda webbsida skulle enligt Dublin Core kunna se ut på följande vis:

```
<HEAD>
<TITLE>Metadata för medicinskt bruk</TITLE>
<META NAME="DC.Title" CONTENT="Metadata för medicinskt bruk">
<META NAME="DC.Subject" CONTENT="Metadata,Resource Description Framework, RDF">
<META NAME="DC.Description" CONTENT="Metadata för medicinskt bruk">
<META NAME="DC.Creator" CONTENT="Frigell, Elisabet">
```

```
<META NAME="DC.Creator" CONTENT="Parnefjord, John">  
<META NAME="DC.Creator" CONTENT="Kronman, Ulf">  
</HEAD>
```

För att möjliggöra fördjupade beskrivningar erbjuder Dublin Core även påbyggnad av underordnade element, så kallade *kvalificerare* (*qualifiers*). Kvalificerare syftar till att ytterligare precisera eller specificera innehållet i ett DC-element. Vid angivande av upphovsman kan det till exempel vara intressant att skilja mellan författare, illustratör och fotograf.

Användning av kvalificerare i Dublin Core kan exempelvis se ut på följande sätt:

```
<META NAME="DC.Creator. Author"CONTENT="Frigell, Elisa-  
bet">
```

Dublin Core löser en del av problemen med HTML META, men skapar också ett antal nya problem. Det största problemet är Dublin Cores brist på utbyggbarhet i kombination med dess tolkningsfrihet, som har lett till att olika organisationer har använt det på helt olika sätt.

Dublin Core framställs som ett väl etablerat metadatasystem, representerat på 21 språk, men Natures undersökning visar att endast 0,3 % av sidorna på 2500 undersökta webbservrar använde Dublin Core [Nature].

De stora söktjänsterna på nätet stödjer inte heller metadata enligt Dublin Core, varför den som vill få sina sidor väl synliga måste lägga till metadata enligt HTML META. I vissa fall verkar det vara så illa att användningen av Dublin Core tillsammans med HTML META av vissa söktjänster uppfattas som så kallad "keyword spamming" – dubblering av sökord för att hamna högt i träfflistorna. Detta leder istället till att sidorna rensas ut ur sökindexen.

Problemen med Dublin Core kan sammanfattas med:

- * För generellt för många användningsområden

- * Ingen standardiserad utbyggbarhet

- * Liten spridning

- * Brist på stöd i söktjänster

RDF – Resource Description Framework

Vi ser alltså att det finns problem med de standarder för metadata som idag används på Internet: HTML META är relativt spritt men har stora begränsningar i sin beskrivning av dokumenten. Dublin Core är mer detaljerat, men dåligt spritt och har svårt att vinna acceptans från de som har specialiserade behov med sin metadatahantering. För att råda bot på detta initierade webbkonsortiet W3C ett arbete för att ta fram en ny standard för metadata – *Resource Description Framework*.

Till grund för sitt arbete med RDF använde W3C webbindustrins tidigare metadatasystem Platform for Internet Content Selection [PICS] och Meta Content Framework [MCF]. Andra modeller som legat till grund för arbetet är Extensible Markup Language [XML] och Dublin Core [DC].

RDF är, precis som namnet anger, inte en metadatastandard i sig, utan ett utbyggbart *ramverk* för metadatastandarder. Inom detta ramverk kan man använda olika standarder för metadata som i RDF kallas för *scheman*. En finess med RDF är att man också kan använda flera olika scheman i samma beskrivning, något som gör att RDF blir utbyggbart och därigenom anpassningsbart för olika behov.

En av grundtankarna med RDF är att möjliggöra automatiserad behandling av metadata. Med tanke på webbens omfattning och webbsidors föränderliga natur är det orimligt att insamling och indexering ska ske manuellt. För att RDF-beskrivningar ska kunna tolkas maskinellt krävs mycket tydliga standarder. RDF definierar standarder på tre nivåer:

- * **Struktur** – RDF *datamodell*
- * **Syntax** – RDF *syntax*, uttryckt i Extensible Markup Language (XML)
- * **Semantik** – RDF *schema*, till exempel Dublin Core

RDF:s datamodell anger hur resurser struktureras och hur relationer mellan dessa kan beskrivas. För själva kodningen av metadata används RDF-syntax, som skrivs med hjälp av Extensible Markup Language (XML). Slutligen krävs en semantisk enhetlighet, vilket åstadkoms med hjälp av så kallade scheman. Dublin Core är ett bra exempel på ett sådant schema [DC]. De tre följande avsnitten kommer vi att ägna åt en grundläggande genomgång av RDF:s tre byggstenar: datamodellen, syntaxen och schemat.”.....

Sidan 19 i rapporten: RDF i framtiden

Vad kan vi då vänta oss av RDF i framtiden? Vi både hoppas och tror att RDF kommer att bli mycket viktigt för resurshanteringen på Internet sett i ett längre perspektiv. Tyvärr tror vi dock också att det kommer att dröja ett bra tag innan detta inträffar.

Förutsättningarna för att RDF ska bli användbart kan sammanfattas i följande punkter:

- * Det behövs stöd vid indexering och sökning av resurser genom att webbläsarna och söktjänsterna utnyttjar RDF vid insamling och presentation.
- * Det behövs stöd vid produktion av dokument, till exempel i form av editorer, konverterare och (halv-)automatiska indexerare.

* En ökad användning av XML eller övergångsstandarden XHTML kommer troligen också att underlätta införandet av RDF [XHTML].

Om vi försöker skåda in i framtiden kan vi troligen se en rad milstolpar och utvecklingstrender framför oss inför en allt mer utbredd användning av RDF på Internet:

Fastlagda standarder. Det första vi väntar på är att specifikationen för RDF schema ska fastställas av W3C. Den har idag (augusti 2000) status "Release Candidate 1", vilket betyder att den sannolikt kommer att bli en sanktionerad W3C-standard inom ett par månader.

Stöd i webbläsare. Här ligger Netscape långt framme med sitt projekt för webbläsaren Mozilla [NS-MOZ] som använder sig av RDF. Koden från Mozilla ligger till grund för den kommande webbläsaren Netscape 6, varför vi kan vänta oss ett visst stöd för RDF i den. Microsoft har inte visat något uttalat intresse för RDF, men eftersom RDF är en tillämpning av XML och Microsoft har annonserat att de kommer att basera sin framtida programutveckling på XML kan man nog räkna med att de relativt snabbt kommer att stödja RDF i sin webbläsare Internet Explorer.

Stöd i specialiserade söktjänster. De som producerar söktjänster där de har kontroll över både själva sidorna och sökfunktionen borde kunna börja arbeta med RDF ganska snart utan större hinder, eftersom man kontrollerar både indata och indexeringsprocess. I Sverige har vi till exempel goda möjligheter att förbereda oss för den kommande RDF-standard genom att centrala söktjänster som SveSök och SAFARI som i dagens läge arbetar med Dublin Core i HTML-form redan nu börjar använda sig av Dublin Core i RDF-form för sina metadatabeskrivningar.

Stöd i globala söktjänster och webbkataloger. Även här ligger Netscape långt framme med projektet för webbkatalogen Open Directory [NS-ODP]. Redan nu finns det möjlighet att skapa RDF-beskrivningar för nätverksresurser och anmäla dem till katalogen. För de kommersiella söktjänsterna ser läget lite mörkare ut – de är ju tvungna att vänta ut en bred användning av RDF innan de kan börja stödja sökning och strukturering med hjälp av RDF.

Stöd i redigeringsredskap och servrar. Efter att webbläsare och söktjänster har börjat stödja RDF finns det incitament för att även stödja RDF i de redskap som skapar webbsidor. Vi kommer troligen att se hur metadatahanteringen byggs in i vanliga redskap för produktion av webbsidor. Här kan vi kanske vänta oss en samverkan mellan tekniker för så kallad text-mining och mänskliga bedömningar på så vis att programmen automatiskt genererar ett förslag till metadata som sedan slutredigeras av webbredaktören. I system som hanterar kommersiella transaktioner med hjälp av XML kommer grundläggande metadata troligen att skapas automatiskt avservrarna.

Ökad användning av XML underlättar. Vartefter sidorna på webben i ökande omfattning blir kodade med hjälp av XML kommer det successivt att bli lättare att skapa RDF-data ur dem. Detta beror på att XML-dokument är så logiskt strukturerade och taggade att de i princip kan översättas direkt till metadata.

MARC och MeSH blir scheman i RDF? Det är inte osannolikt att våra gamla beprövade standarder för indexering och metadata som MARC (Machine Readable Cataloguing) och MeSH (Medical Subject Headings) med flera i en framtid kommer att anpassas för resurser på Internet genom att bli RDF-scheman.

Slutsatser

Resource Description Framework är ett utbyggbart ramverk där man samtidigt kan använda flera olika system för metadata. RDF är så konstruerat att dessa metadata-system kan utvecklas och användas utan central kontroll och koordinering. RDF har därför alla förutsättningar att kunna bli den framtida enhetliga standarden för metadatahantering på Internet.

Metadata-systemet Dublin Core är ett av de centrala schemata för RDF. Utbyggbarheten hos RDF kommer sannolikt att säkra framtiden för Dublin Core, eftersom RDF tillåter att varje organisation lägger till sina egna system för specialiserade metadata utan att förändra något av de 15 grundelementen i Dublin Core. Dublin Core kommer alltså att användas för allmänna beskrivningar av resurser.....”

Referenser

[DC] Dublin Core Metadata Initiative.

[<http://purl.org/dc/>]

[DCDM] Dublin Core Metadata Initiative: Data Model Working Group.

[<http://purl.org/dc/groups/datamodel.htm>]

[DCES] Dublin Core Metadata Element Set, Version 1.1: Reference Description.

[<http://purl.org/dc/elements/1.1/>]

[DC-dot] DC-dot: Dublin Core generator.

[<http://www.ukoln.ac.uk/metadata/dcdot/>]

[DC-HTML] Encoding Dublin Core Metadata in HTML. – Red. J. Kunze. - RFC 2731.

[<http://www.ietf.org/rfc/rfc2731.txt>]

[DC-RDF] Guidance on expressing the Dublin Core within the Resource Description Framework

(RDF). – Red. Eric Miller, Paul Miller och Dan Brickley.

[<http://www.ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/>]

[DT-1] Tångring, Jan. - Med metadata är Internet en distribuerad databas // Datateknik 3.0. –Nr. 5, 1999. - S. 29.

[DT-2] Tångring, Jan.-Biblioteksfolket städar upp på nätet // Datateknik 3.0.– Nr 5,1999.-S.30– 32.

[HTML] HyperText Markup Language: Home Page.

[<http://www.w3.org/MarkUp/>]

[HTML-META] Meta data // HTML 4.01 Specification: W3C Recommendation 24 December 1999.–

Red. Dave Raggett, Arnaud Le Hors och Ian Jacobs.

[<http://www.w3.org/TR/REC-html40/struct/global.html#h-7.4.4>]

[IANNELLA] Iannella, Renato. - An Idiot's Guide to the Resource Description Framework.

[<http://archive.dstc.edu.au/RDU/reports/RDF-Idiot/>]

[MCF] Guha, R.V. - Meta Content Framework.

[<http://www.xspace.net/hotsauce/mcf.html>]

[med-CERTAIN] med-CERTAIN. Rating system enabling patients and consumers to filter harmful health information on the Internet.

[<http://medpics.org/medcertain/1.htm>]

[Nature] Lawrence, Steve och Giles, C. Lee. - Accessibility of information on the web // Nature. –

Vol. 400, nr 107, 1999. – S. 107-109.

[<http://www.nature.com/server-java/Propub/nature/400107A0.frameset?context=search>]

[<http://www.nature.com/server-java/Propub/nature/400107A0.pdf>]

[NS-MOZ] RDF in Mozilla.

[<http://rdf.netscape.com/rdf/doc/>]

[<http://rdf.netscape.com/rdf/doc/api.html>]

[NS-ODP] Open Directory Project.

[<http://dmoz.org/>]

[NS-Z3950] Mozilla RDF / Z39.50 Integration Project

[<http://www.mozilla.org/rdf/doc/z3950.html>]

[PICS] Platform for Internet Content Selection (PICS).

[<http://www.w3.org/PICS/>]

[RDF] Resource Description Framework (RDF).

[<http://www.w3.org/RDF/>]

[RDF-ABBREV] Basic Abbreviated Syntax // Resource Description Framework (RDF) Model and

Syntax: W3C Recommendation 22 February 1999. – Red Ora Lassila och Ralph R. Swick.

[<http://www.w3.org/TR/REC-rdf-syntax/#abbreviatedSyntax>]

[RDF-SCHEMA] Resource Description Framework (RDF) Schema Specification: W3C Proposed

Recommendation 03 March 1999. – Red Dan Brickley och R.V. Guha.

[<http://www.w3.org/TR/PR-rdf-schema/>]

[RDF-SPEC] Resource Description Framework (RDF) Model and Syntax Specification: W3C

Recommendation 22 February 1999. – Red Ora Lassila och Ralph R. Swick.

[<http://www.w3.org/TR/REC-rdf-syntax/>]

[Reggie] Reggie: The Metadata Editor.

[<http://metadata.net/dstc/>]

[RFC2396] Uniform Resource Identifiers (URI): Generic Syntax.– Red.T. Berners-Lee, R. Fielding och L. Masinter. - RFC2396.

[<http://info.internet.isi.edu/in-notes/rfc/files/rfc2396.txt>]

[RFC1766] IETF RFC 1766: Tags for the Identification of Languages. – Red. H Alvestrand

[<http://www.ietf.org/rfc/rfc1766.txt?number=1766>]
[SiRPAC] SiRPAC - Simple RDF Parser & Compiler.
[<http://www.w3.org/RDF/Implementations/SiRPAC/>]
[vCard] Representing vCard v3.0 in RDF.
[<http://www.dstc.edu.au/RDU/RDF/draft-iannella-vcard-rdf-00.txt>]
[W3C] W3C - World Wide Web Consortium.
[<http://www.w3.org/>]
[XHTML] XHTML 1.0: The Extensible HyperText Markup Language. A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000.
[<http://www.w3.org/TR/xhtml1/>]
[XML] Extensible Markup Language (XML).
[<http://www.w3.org/XML/>]
[XML-NAME] Namespaces in XML: World Wide Web Consortium 14-January-1999. – Red. Tim Bray, Dave Hollander och Andrew Layman.
[<http://www.w3.org/TR/REC-xml-names/>]
[XML-SPEC] Extensible Markup Language (XML) 1.0:W3C Recommendation 10-February-1998.–
Red. Tim Bray, Jean Paoli och C. M. Sperberg-McQueen.
[<http://www.w3.org/TR/REC-xml>]

XML löser alla dina problem!

<http://www.xml-forum.com/fakta/enator.shtml> Artikel av Daniel Björkman, en av författarna bakom Statskontorets rapport 1998:6 "Vad är XML?"

“eXtensible Markup Language (XML) kommer att medföra en revolution på Internet! Det är en plattform för handel över nätet, ett koncept för dokumenthantering, ett format för databaskommunikation, storskalig webbpublicering, intranät, extranät och dataöverföring i allmänhet. Det utgör en bas i strategiska applikationer såsom Office 2000, SAP R3 och Oracle 8. Det kommer att ersätta HTML på Internet. Strategier på ledande IT-företag som Microsoft, Netscape, Oracle, Sun är alla lyriska över de nya möjligheter som erbjuds. Alla verkar vara överens om att XML kommer att frälsa världen.

Detta är den bild som förmedlas om det nya utökade märkordsspråket XML på Internet, mässor och i fackpress. Att säga att XML är hett är därför närmast ett 'understatement'. Självfallet är det många som blir intresserade av vad som egentligen ligger bakom den här uppståndelsen.

Vad är XML?

XML är en specifikation som beskriver en syntax för hur man kan skapa märkordsuppsättningar som i sin tur kan användas för att beskriva och strukturera information eller data. Det mest kända exemplet på en märkordsuppsättning är HTML (Hypertext Markup Language).

Med XML kan man skapa sitt eget HTML, specialanpassat för den information man vill beskriva. Organisationen bakom XML-specifikationen heter World Wide Web Consortium (W3C), en mycket tung aktör med ca 250 medlemsorganisationer. Man brukar också säga att XML är en standard. Officiellt sett är det en rekommendation från W3C vilket inte skall förväxlas med t.ex. ISO- eller ANSI-standarder eftersom W3C inte är någon standardiseringsorganisation. XML är

däremot nära besläktad och fullständigt kompatibel med ISO-standard Standard Generalized Markup Language (SGML).

Varför behöver vi XML?

Grunden för att ta fram det nya märkordsspråket XML var i första hand att användningen av Internet har utvecklats och att HTML inte räcker till för att hantera mer komplex information. När man studerar Internets i och för sig kortvariga historia kan man se en ganska tydlig trend hur användningen av detta media förändrats. När World Wide Web (WWW) fick sitt kommersiella genombrott i mitten av 90-talet använde vi det mest för att publicera information. Alla skulle ha en egen hemsida, såväl företag, organisationer som enskilda personer. Inte många företag eller organisationer vågade så småningom stå utanför detta potentiella marknadsfönster. Man skapade hemsidor med många roliga funktioner som gjorde allt annat än bidrog till företagets resultat. Så småningom började man använda Internettekniken även internt, man byggde intranät. Fortfarande handlade det mycket om publicering men man började också laborera med kopplingar till databaser. Efter en tids laborerande vågade man sig på att bygga Internet-tillämpningar där man släpper in kunderna direkt mot de interna systemen genom väldefinierade gränssnitt. De första Internetbankerna såg dagens ljus. Man blev också alltmer våghalsig i att använda den här tekniken internt för allt från bokning av möteslokaler till tidrapportering och kvalitetssystem. Man började upptäcka att HTML hade ganska stora begränsningar men med lite trixande gick det ofta att komma runt problemen. Den största utmaningen var att hålla systemen uppdaterade i takt med utvecklingen av nya webbläsare, utökning av HTML mm. Idag har vi tagit ytterligare ett steg uppför komplexitetstrappan i Internetutvecklingen och det finns gott om exempel på affärskritiska tillämpningar på Internet. Vi har företag som Bokus, Boxman, Amazon, Leknet, nischbanker som Skandiabanken och Ikanobanken m.fl. som har Internet som sin enda markandsförings- och försäljningskanal. Självfallet är det helt avgörande att tekniken bakom de här lösningarna är pålitlig, att det är lätt att uppdatera informationen, att säkerheten är

tillförlitlig och att systemen är skalbara när marknaden växer. En timmes avbrott får direkt ekonomiska konsekvenser och medför också stora förluster i goodwill för dessa företag. Det är här XML kommer in i bilden, som ett mer avancerat format för att hantera information. HTML klarar helt enkelt inte att svara upp mot alla dessa krav eftersom det är utvecklat för att presentera relativt enkel information, inte för att bygga affärskritiska tillämpningar på.

Varför standarder?

Det är t.ex. ganska frustrerande att inte kunna öppna den rykande färska försäljningsrapporten, som man erhållit via e-post, för att man har fel version av kalkylprogram. Samma frustration känner sannolikt webbmastern som precis har lagt sista handen vid den nya hemsidan bara för att upptäcka att den inte fungerar med Netscape. Det är då man plockar fram ikonerna med "denna sida visas bäst i Internet Explorer". Standarder gör livet enklare och avsaknaden av dem ger ofta upphov till stora problem. Dokumenthantering är ett i detta avseende mycket eftersatt område. Vi lägger ner mycket tid, pengar och adrenalin på att konvertera dokument från t.ex. FrameMaker till WordPerfect, till MS Word95, till MS Word97 o.s.v.

Syftet med XML är att det skall vara ett globalt och universellt format för att beskriva data och information på ett sätt som är oberoende av enskilda applikationer, operativsystem och leverantörer.

Hur funkar det?

Den viktigaste skillnaden mellan XML och HTML är att man i XML själv har möjlighet att definiera sina egna märkord för att beskriva sin information. X:et i XML står för eXtensible, d.v.s. utökningsbart. I HTML är man låst till att använda den fasta uppsättning märkord som finns definierad i t.ex. HTML 4.0. De märkord som finns i HTML är till för att tala om hur information skall presenteras. Exempelvis använder man märkordet <p> för att tala om att informationen inom detta

märkord skall presenteras som vanlig stycketext, <h1> skall presenteras som en rubrik på nivå ett, <table> används för tabellpresentation o.s.v. HTML är ett mycket fattigt språk för att beskriva innehåll d.v.s. vad informationen handlar om.

För att underlätta sökning och för att göra det möjligt att bearbeta informationen med hjälp av datorer vill man märka upp den på ett otvetydigt sätt. Om man lösryckt ur sitt sammanhang skriver <p>75</p> är det omöjligt för dig att avgöra om 75 syftar på min vikt, ett artikelnummer, mitt födelseår eller antal enheter som finns i lager. Precis så lösryckt ur sitt sammanhang måste vi ofta kunna bearbeta information eftersom datorer inte är speciellt bra på att förstå sammanhang. Om man i stället skriver <pris>75</pris> vet du att det gäller ett pris. Om man dessutom förser priset med ett attribut som talar om vilken valuta det är blir det ännu tydligare

<pris valuta="SEK">75</pris>. Med hjälp av syntaxen i XML kan vi också knyta ett visst pris till en viss produkt, t.ex. <produkt>Lättmjölk<pris valuta="SEK">6</pris></produkt>.

Eftersom vi med XML kan beskriva informationen mer precist kan vi också söka efter information med större träffsäkerhet. Avsaknaden av effektiva sökmetoder är ett stort problem på Internet idag. Om man till exempel vill söka efter artiklar skrivna av Göran Persson kan man i XML ta hjälp av märkorden i sökfrågan och söka efter t.ex. <författare>Göran Persson</författare>. På så sätt slipper man få träff på alla artiklar där Göran Persson omnämns, eller ännu värre, där Göran **eller** Persson omnämns.

I takt med att alltmer information finns tillgänglig i elektronisk form ställer vi också högre krav på att kunna bearbeta den på ett effektivt sätt. För att detta skall vara möjligt måste den märkas upp på ett intelligent sätt med en stringent syntax som kan läsas av datorer. Antag att man vill göra en enkel adresslista och publicera den på Internet eller ett intranät. Kodad enligt HTML skulle en post i listan kunna se ut på följande sätt:

<p>Sven Svensson</p>
<p>Stora vägen 2</p>

<p>123 67 Malmö</p>

Problemet med att märka upp informationen på detta sätt är att endast en människa kan avgöra om Sven är namnet på en person eller en gata. Märkorden talar bara om att informationen skall presenteras som vanlig stycketext. Om man i stället kodar den i XML skulle man kunna använda mer passande märkord och koda den på följande sätt:

<adresspost>
<namn><förnamn>Sven</förnamn><efternamn>Svensson</efternamn></namn>
<gatuadress>Stora vägen 2</gatuadress>
<postadress><postnummer>123
67</postnummer> <ort>Malmö</ort></postadress></adresspost>

På detta sätt kan adresslistan enkelt sorteras på förnamn, efternamn, gatuadress, postadress eller ort. Om man bara är intresserad av namn och ort kan man enkelt dölja de andrafälten vid presentation. Man kan också med enkla medel importera listan till en databas.

Om man skall lägga upp adresslistan på Internet eller ett intranät räcker det inte att märka upp informationen. Man måste också tala om hur man vill att innehållet i märkorden skall presenteras. Detta gör man genom att skapa en formatmall som beskriver hur varje märkord skall presenteras, t.ex. att namn skall presenteras med teckensnittet Arial i 14 punkters storlek och fetstilt. Presentationsmallen kodas i XML enligt en speciell syntax som W3C definierat i specifikationen XSL (eXtensible Stylesheet Language). Det är en mycket stor fördel att på detta sätt skilja mellan själva informationen och dess presentation eftersom man då kan välja att presentera informationen på olika sätt utan att för den skulle behöva ändra själva innehållet. Man kan t.ex. ha olika formatmallar för olika medier eller för olika målgrupper. I HTML behöver man inte någon formatmall eftersom webbläsarna vet hur märkorden som är definierade i HTML skall presenteras. Man kan säga att presentationsmallen i detta fall är inbyggd i webbläsaren.

Ofta vill man förhöja värdet på presentationen genom att skapa så kallade hypertextlänkar till relaterad information när man publicerar information på Internet eller intranät. Den typ av länkar som finns i

HTML lämnar mycket övrigt att önska, de är svåra att underhålla. W3C har därför tagit ett samlat grepp när det gäller länkhantering och genom två specifikationer skapat förutsättningar för nya intressanta möjligheter. Man kan till exempel skapa länkar som går till flera mål, själv bestämma hur länken skall presenteras och vad som skall hända när användaren klickar på den.

Informationshantering baserad på XML

XML i sig löser inte alla problem. Det måste du som vanligt göra själv. Standarden XML är ett redskap som du kan använda för att få ordning och reda på din information.

Eftersom XML är en generell standard kan den också användas för handel över nätet, dokumenthantering, databaskommunikation, storskalig webbpublicering, intranät, extranät och dataöverföring i allmänhet. Det används i Office 2000, SAP R3 och Oracle 8. Det lär dock dröja innan XML helt ersätter HTML eftersom HTML är ett utmärkt format för att på ett enkelt sätt presentera enkel information.

Vi på Enator förespråkar att alla företag och organisationer som hanterar för verksamheten strategiskt viktig information bör ta fram en informationsstrategi. I denna strategi definierar man målgrupper och deras karaktäristika, vilken information som skall hanteras och på vilket sätt man skall hantera den.

Vi menar att du bör fundera på vad XML erbjuder för möjligheter när du utformar din informationsstrategi eftersom XML kan bidra till att du får bättre ordning och reda på din information och därmed säkerställer värdet av gjorda investeringar.

Välj att hantera din information i ett neutralt och standardiserat format i stället för att välja ett format som är knutet till en enskild leverantörs applikationer.

Ofta vet man inte på förhand på vilket sätt och i vilka sammanhang en viss informationsmängd kommer att användas i framtiden. Exem-

pel på information som ofta används på en mängd olika sätt och i olika sammanhang är artikelregister, personalregister, försäljningsstatistik, anvisningar, företagspresentationer, tekniska data för produkter m.m.

Grundtanken med XML är att du en gång för alla skall analysera, strukturera och märka upp din information så att du kan använda den på en mängd olika sätt i framtiden.

Vi på Enator är övertygade om att det är det bästa sättet att bringa ordning i ditt informationskaos och säkerställa att dina investeringar i informationshanteringssystem inte är bortkastade pengar.

Som sagt, XML löser inte alla dina problem med automatik. Innan du kan beskriva din information med XML måste du vara helt klar över vad den innehåller och hur du vill att den skall se ut. Du måste känna till informationens innehåll och struktur. Processen för att ta reda på det kallas ***informationsanalys***.”

XML-Forum <http://www.xml-forum.com/>

XML Forum startar om. XML-forum har varit vilande medan användningen av XML gått framåt med stormsteg. När XML-forum bildades för några år sedan var XML en ny teknik och kunskapen om XML väldigt låg. På några år har XML utvecklats till en accepterad teknik och den utvecklas åt många olika håll i olika applikationsområden.

Bland annat:

Lär dig mer om XML Här publicerar vi en rad artiklar om XML, författade av medlemmar i XML-Forum. <http://www.xml-forum.com/fakta/>

XML: Öppna informationssystem med strukturerade standardiserade dokument
<http://www.xml-forum.com/fakta/xml.shtm>

XML - en verktygslåda för språkbyggare!
<http://www.xml-forum.com/pekare/>

Information om XML

- * [XML-Forum Fakta: Artiklar om XML](http://www.xml-forum.com/fakta/index.shtml) <http://www.xml-forum.com/fakta/index.shtml>
- * [Encyclopedia Intranetica: Om XML](http://www.webway.se/intranetica/5steg/xml.shtm)
<http://www.webway.se/intranetica/5steg/xml.shtm>
- * [W3C XML](http://www.w3.org/XML) - hemsidan för XML på World Wide Web Consortium
<http://www.w3.org/XML>
- * [XML.COM](http://www.xml.com/) - en webbserver om XML <http://www.xml.com/>
- * [XML FAQ](http://www.ucc.ie/xml/), underhållen av Peter Flynn <http://www.ucc.ie/xml/>
- * [Introduksjon til XML](http://www.stud.ifi.uio.no/~larsga/download/xml/xml.html) Norsk introduktion av Lars Marius Garshol.
<http://www.stud.ifi.uio.no/~larsga/download/xml/xml.html>
- * [XML Page](http://www.sil.org/sgml/xml.html), skapad av the Summer Institute of Linguistics

<http://www.sil.org/sgml/xml.html>

- * [WebReview om XML](http://webreview.com/xml/) <http://webreview.com/xml/>
- * [The XML Files](http://www.gca.org/memonly/xmlfiles/) <http://www.gca.org/memonly/xmlfiles/>

* [XML for managers\(Arbortext\)](http://www.arbortext.com/Think_Tank/XML__Resources/XML_for_Managers/xml_for_managers.html)

http://www.arbortext.com/Think_Tank/XML__Resources/XML_for_Managers/xml_for_managers.html

* [Yahoo-sökning på XML](http://www.yahoo.com/Computers_and_Internet/Information_and_Documentation/Data_Formats/XML/)

http://www.yahoo.com/Computers_and_Internet/Information_and_Documentation/Data_Formats/XML/

XML och Java

* [The Java-XML home page](http://ala.vsms.nottingham.ac.uk/vsms/java/)

<http://ala.vsms.nottingham.ac.uk/vsms/java/>

* [Suns Java-division om Java och XML](http://java.sun.com/xml/) <http://java.sun.com/xml/>

* [Cafe con leche on XML](http://metalab.unc.edu/xml/) <http://metalab.unc.edu/xml/>

Information om XSL

* [W3C XSL](http://www.w3.org/Style/XSL/) - "hemsidan" för XSL <http://www.w3.org/Style/XSL/>

* [XSL Page](http://www.sil.org/sgml/xsl.html), skapad av the Summer Institute of Linguistics.

<http://www.sil.org/sgml/xsl.html>

* [Arbortext: Introduction to XSL](http://www.arbortext.com/Think_Tank/XML__Resources/XSL_-_a_proposed_stylesheet_fo/xsl_-_a_proposed_stylesheet_fo.html)

http://www.arbortext.com/Think_Tank/XML__Resources/XSL_-_a_proposed_stylesheet_fo/xsl_-_a_proposed_stylesheet_fo.html

* [Microsoft XSL Home Page](http://www.microsoft.com/xml/xsl/xslintro.htm)

<http://www.microsoft.com/xml/xsl/xslintro.htm>

* [Microsoft XSL Tutorial](http://www.microsoft.com/xml/xsl/tutorial/contents.htm)

<http://www.microsoft.com/xml/xsl/tutorial/contents.htm>

* [XSL Jumpstart](http://www.jeremie.com/JS/XSL/all.html) - mängder av pekare till information om XSL.
<http://www.jeremie.com/JS/XSL/all.html>

XML - Extensible Markup Language - Språket som vidgar webben

<http://www.xml-forum.com/fakta/xml-intro-cnet.shtml>

Statskontoret.Vägledning XML-teknik och metadata.19 sidor 020315.Från Internet 020515

<http://www.statskontoret.se/pdf/200208.pdf>

Innehåll

VÄGLEDNING: XML-TEKNIK OCH METADATA

Förord

1. Sammanfattning
2. Målsättning
3. Målgrupp
4. Avgränsningar
5. Bakgrund
6. Termer och definitioner
- 7 Rekommendationer

7.1 Extensible Markup Language (XML) 1.0 (Second Edition)

7.2 XML Schema

7.3 Namespaces in XML

7.4 XML Path Language (XPath) Version 1.0

7.5 XML Linking Language (XLink) Version 1.0

7.6 XML Base

7.7 XHTML Version 1.0

7.8 XSL Transformations (XSLT) Version 1.0

7.9 Extensible Stylesheet Language (XSL) Version 1.0

7.10 Dublin Core Metadata Element Set, Version 1.1

8. Råd och anvisningar

8.1 om XML

8.2 om Metadata

9 Information

9.1 XML som koncept

9.2 Sökfunktioner och metadata

10.Referenser, hänvisningar, länkar

11.Kontaktperson

Citat :

“Denna vägledning behandlar XML-teknik och metadata, två samverkande teknikområden som kan bidra till att bygga en effektiv informationsstruktur. Ambitionen är att initialt ange några få, generella rekommendationer som bedöms vara stabila och praktiskt användbara.”

“**XML** är en grupp rekommendationer som har starkt stöd av företag och organisationer, i Sverige och internationellt. XML är inte en enda standard eller en färdig lösning utan ett koncept under utveckling. Vi föreslår här några grundläggande rekommendationer som nu bedöms vara etablerade. I övrigt rekommenderas den princip som (fritt översatt) återfinns på XML-grundarnas webbplats (World Wide Web Consortium, www.w3.org): ‘XML är inte alltid den bästa lösningen men den är alltid värd att överväga’.”

“**Metadata** (dvs. data om data) används för att underlätta informationssökning och för att hantera statusinformation. Den mest spridda modellen, Dublin Core, kan betraktas som en de facto-standard. Dublin Core täcker dock inte alla behov. Vidareutveckling och standardisering pågår på olika håll och inom vissa sektorer etableras andra, mer branschanpassade metadatamodeller. Dessa behandlas dock ej närmare i denna utgåva. Denna vägledning är den första versionen. Den skall kunna utökas och bli mer detaljerad i kommande versioner.”

“**Målgrupp.** Målgruppen för denna vägledning är IT- och informationsstrateger, projektledare, systemarkitekter och systemkonstruktörer med ansvar för utformningen av informationshanteringslösningar inom offentlig sektor. Läsaren förutsätts ha grundläggande kunskaper om informationshantering, XML och metadata.(För litteraturhänvisningar och länkar, se avsnittet Referenser nedan.)”

“**Avgränsningar** Version 1 av denna vägledning begränsas till att endast ta upp generella, brett tillämpbara rekommendationer. För XML-området är de fastställda av World Wide Web Consortium (W3C). Beträffande metadata hänvisar vi till rekommendationer från Dublin Core Metadata Initiative (DCMI).”

“Beträffande användning av metadata så bör man helst ha s.k. kontrollerade vokabulärer eller terminologier, kodlistor mm. till sitt förfogande för att uppnå hög kvalitet och effektivitet. Att skapa och fastställa sådana måste bedrivas sektorsvis och det blir då ofta frågan om långsiktiga utvecklingsprojekt. Terminologifrågor är ett viktigt område som dock inte omfattas av denna vägledning. Däremot utreder Statskontoret hur ett nationellt samverkansregister skulle kunna upprättas, där t.ex. återanvändbara strukturer, scheman, modeller, metoder och terminologier då kunde publiceras, kommenteras och vidareutvecklas.”

“Bakgrund. XML (eXtensible Markup Language) har under de senaste åren fått stor uppmärksamhet som framtidens teknik för informationshantering. Med XML-teknik öppnas nya möjligheter att strukturera, söka, länka, utbyta och presentera information. XML är namnet på den grundläggande XML-specifikationen som rekommenderades av W3C (World Wide Web Consortium) år 1998, men XML används också som ett samlingsbegrepp för hela XML-familjen som omfattar ett stort antal relaterade rekommendationer och initiativ, inom och utom W3C. Dessa olika XML-"släktingar" har olika syften och omfattning och de befinner sig i olika utvecklingsstadier. Ännu befinner vi oss i början på XML-utvecklingen och mycket arbete återstår, bl.a. med att utveckla programvaror och tillämpningar som följer de fastställda rekommendationerna.”

“Idag pågår ett stort antal projekt som bygger på användning av XML och metadata, såväl inom offentlig som privat sektor. I takt med utbyggnaden av 24-timmarsmyndigheten ökar behovet av enhetliga metoder och standarder för informationsutbyte. Denna vägledning skall bidra till att prioritera och samordna satsningarna inom XML och metadata.”

“W3C (World Wide Web Consortium) bildades 1994 för att leda utvecklingen av gemensamma webb-standarder och därmed säkra den framtida interoperabiliteten över webben. W3C har idag över 500

medlemsorganisationer från hela världen. **XML** är namnet på den grundläggande XML-specifikationen som rekommenderades av W3C år 1998, men XML används också som ett samlingsbegrepp för hela XML-familjen som omfattar ett stort antal relaterade rekommendationer och initiativ.”

“Statusbeteckningar för standardiseringsarbetet inom W3C: Följande statusbeteckningar används på vägen fram till en fastställd ‘Recommendation’:

- **Note** En idé eller kommentar som publicerats.
- **Working draft** Arbetsversion av eventuell framtida standard.
- **Candidate recommendation** Specifikation som aspirerar på att bli en standard och väntar på kommentarer från andra arbetsgrupper inom W3C.
- **Proposed recommendation** Förslag till standard som nått en viss förankring och som nu är uppe för granskning.
- **Recommendation** Av W3C antagen standard.”

“Dublin Core Metadata Initiative (DCMI) är en internationell, ämnesöverskridande gruppering som syftar till att utveckla mekanismer för att beskriva resurser i elektronisk miljö. Arbetet med Dublin Core påbörjades 1995 vid en workshop i Dublin (Ohio), därav namnet. DCMI har f.n. 17 aktiva arbetsgrupper med över 75 aktiva deltagare.

Statusbeteckningar för standardiseringsarbetet inom DCMI: DCMI använder använder statusbeteckningarna Note, Working Draft, Proposed Recommendation och Recommendation på liknande sätt som W3C (se ovan).”

“De XML- och metadata-specifikationer som tas upp i denna vägledning har utvecklats inom ramen för W3C eller DCMI och har den högsta statusbeteckningen **Recommendation**. Nedan listas de rekommendationer som vi f.n. har tagit ställning för.”

“7.1 Extensible Markup Language (XML) 1.0 (Second Edition) W3C Recommendation (oktober 2000)

Grunden i hela XML-ramverket för strukturerad informationshantering är denna rekommendation. Den specificerar bland annat hur XML-dokument får se ut och skall byggas upp (syntaxen).

Läs mer på: <http://www.w3.org/XML/>

Teknisk specifikation: <http://www.w3.org/TR/REC-xml>

7.2 XML Schema

W3C Recommendation (maj 2001)

XML Schema är ett modernare och mer fullständigt sätt att bestämma strukturen för XML-dokument och kan användas istället för en DTD. Ett XML Schema ser ut som ett XML-dokument, vilket beror på att det i sig är en XML-applikation, till skillnad från en DTD. Förutom XML 1.0 så bygger XML Schema även på XML Namespaces. XML Schema erbjuder mer stöd för modularitet och datatypning än vad man kan specificera i en DTD, dock blir schemat i regel längre och mindre överskådligt än motsvarande DTD. De flesta XML-applikationer använder fortfarande DTD:er och många programvaror saknar fullt stöd för XML Schema men det blir mer och mer populärt, framför allt tack vare de ökade datatypningsmöjligheterna. (DTD se nedan)

Läs mer på: <http://www.w3.org/XML/Schema>

Teknisk specifikation: <http://www.w3.org/TR/xmlschema-0/>

7.3 Namespaces in XML

W3C Recommendation (1999)

Om begrepp från olika XML-applikationer används i samma XML-dokument, dvs. element- och attributnamn med olika innebörd, så ökar risken för missförstånd. Namespaces eliminerar denna tvetydighet genom att man för varje element anger vilken definition av elementet som gäller, dvs. vilken namnrymd det tillhör. Ett prefix sätts in varje gång ett begrepp från en viss applikationen används och prefixet fungerar då som referens till rätt namnrymd.

Teknisk specifikation: <http://www.w3.org/TR/REC-xml-names/>

7.4 XML Path Language (XPath) Version 1.0

W3C Recommendation (1999).

XPath används för att adressera data inom ett XML-dokument. Data är till exempel element, attribut, text eller andra komponenter. Notationen liknar sökvägar i en katalogstruktur. XPath bygger förutom XML 1.0 även på XML Namespaces.

Teknisk specifikation: <http://www.w3.org/TR/xpath>

7.5 XML Linking Language (XLink) Version 1.0

W3C Recommendation (juni 2001)

XLink är den rekommendation som tillhandahåller länkningsmöjligheter till XML. Den erbjuder även komplexa länkar. Dessa används för att beskriva kopplingar från flera källor till flera destinationer. Med detta kan man bygga upp nät av relationer mellan resurser. Länkarna kan antingen finnas inuti en källa eller så kan destination ligga utanför, en s.k. extern länk.

Teknisk specifikation: <http://www.w3.org/TR/xlink/>

7.6 XML Base. W3C Recommendation (juni 2001)

XML Base tillhandahåller länkningskonstruktioner till XLink och definierar hur länkar ska skrivas i XML 1.0. Ett av kraven är att stödja motsvarande länkar som finns i HTML, där det är möjligt att sätta upp en bassökväg för att inte behöva skriva ut hela sökvägen. Med relativa sökvägar får man ett förkortat skrivsätt.

Teknisk specifikation: <http://www.w3.org/TR/xmlbase/>

7.7 XHTML Version 1.0 W3C Recommendation (januari 2000)

Extensible HyperText Markup Language (XHTML) är en XML-applikation som direkt motsvarar HyperText Markup Language (HTML 4). HTML är ursprungligen en SGML-applikation och efter att XML introducerats arbetade W3C fram XHTML som beskriver HTML i XML. XHTML bör användas för webbsidor och stöds av de flesta moderna webbläsare.

Teknisk specifikation: <http://www.w3.org/TR/xhtml1/>

7.8 XSL Transformations (XSLT) Version 1.0

W3C Recommendation (1999)

Ursprungligen en del av XSL (se nedan), men sedan 1999 en egen specifikation. XSLT används för att beskriva hur XML-dokument kan översättas till andra XML-dokument. Här avses översättning av själva märkningen och inte av innehållet i dokumentet. En sådan översättning kallas för en transformation och översätter märkningen från en XML-applikation till en annan.

Teknisk specifikation: <http://www.w3.org/TR/xslt>

7.9 Extensible Stylesheet Language (XSL) Version 1.0

W3C Recommendation (oktober 2001)

XSL är en rekommendation för att beskriva hur XML-dokument ska se ut när de visas, dvs. hur de ska formateras. Ett XSL-dokument kallas för stylesheet, en formatmall, och reglerna uttrycks i s.k. Formatting Objects (XSL FO). XSL använder sig av XML Namespaces och av XPath för att adressera strukturer inom XML-dokument. (XSL innehöll från början även en avdelning om transformering av ett dokument till ett annat. Denna del har sedan 1999 varit en egen specifikation, se XSL Transformations XSLT). (Cascading Stylesheet (CSS), ett formateringsspråk för HTML, är föregångare till XSL och kan också användas i vissa fall.)

Teknisk specifikation XSL: <http://www.w3.org/TR/xsl/>

Läs mer om CSS på: <http://www.w3.org/Style/CSS/>

7.10 Dublin Core Metadata Element Set, Version 1.1

DCMI Recommendation (1999)

Dublin Core Metadata Initiative är ett internationellt, ämnesöverskridande arbete som syftar till att utveckla mekanismer för att beskriva resurser i elektronisk miljö. Ett antal metadata-element finns specificerade för detta ändamål. Arbetet med Dublin Core påbörjades 1995 vid en workshop i Dublin (Ohio), därav namnet. Standardiseringsaktiviteter förekommer inom flera standardiseringsorgan: Internet Engi-

neering Task Force (IETF), European Committee for Standardization (CEN) och amerikanska National Information Standards Organization (NISO). För att underlätta återsökning av dokument rekommenderas i första hand användning av relevanta delar av de 15 elementen i Dublin Core.

Teknisk specifikation:

<http://dublincore.org/documents/1999/07/02/dces/>

Läs mer på: <http://dublincore.org/>

“Ovanstående rekommendationer är grundläggande specifikationer som nu bedöms vara stabila och accepterade av marknaden. De täcker dock inte alla behov. Det finns många fler standardförslag och ibland behöver man bygga lösningar med hjälp av förslag som kanske stöds av en grupp leverantörer men som ännu är under bearbetning och därför kan komma att förändras. Statskontoret följer utvecklingen i stort men varje projekt behöver självständigt ta ställning till vad som krävs och vilka tekniska utvecklingsrisker som kan vara acceptabla. Man behöver inte heller alltid börja med ett stort projekt. Ofta är det bra att börja i begränsad skala och skaffa egna erfarenheter. Ett första steg kan t.ex. vara att anpassa sina webbsidor till XHTML.”

“Tillgången på användarvänliga och stabila programprodukter är därför en väsentlig komponent som bör tillmätas stor vikt. Se t.ex. Statskontoret ramavtal för XML-produkter <http://it-upphandling.statskontoret.se/Uhw/> eller www.xmlsoftware.com, en privat översikt över XML-programvaror. Såväl i Sverige som internationellt finns ett stort antal projekt där man har samlat värdefulla erfarenheter med XML och metadata. Att sammanställa dessa faller f.n. utanför denna väglednings ambitioner men Statskontoret utreder olika möjligheter för att sprida erfarenheter och uppmuntra till återanvändning. Tills vidare rekommenderas att utnyttja befintliga nätverk och sökvägar för att skapa samband med liknande satsningar.”

“om XML

- Vid utveckling av webbplatser, använd i första hand XHTML.
- I projekt där nya informationsstrukturer behöver skapas, välj i första hand XML Schema.
- Vid utbyggnad av befintliga informationsstrukturer, utnyttja tills vidare befintliga DTD:er om användbara sådana redan existerar.”

“om Metadata

Dublin Core rekommenderas i de fall dess uppsättning metadata passar för tillämpningen, dvs. när begrepp som titel, författare, datum, ämne mm. är relevanta. I annat fall bör i första hand andra existerande initiativ sökas och utvärderas, t.ex. olika förekommande förslag till Dublin Core Extensions.”

“Information

Detta kapitel ger ytterligare, kompletterande bakgrundsinformation kring strukturerad informationshantering:.....

Att definiera och införa strukturer är därför ett omfattande åtagande.

För varje aktuell tillämpning bör följande frågor ställas och besvaras:

- I vilken grad har informationen redan en enhetlig logisk struktur avseende innehåll och form? (Jämför t.ex. skillnaden mellan blanketter och fritt formulerade brev.)
- Ingår informationen i en styrd ärendeprocess eller är den mer oförutsägbar avseende när, var och hur den uppstår (ur mottagarens/ användarens synpunkt)?
- Handlar det om redan befintlig och/eller nyskapad information?
- Om informationen är befintlig, i vilken form existerar den nu? (Som papper i ett arkiv, som dokumentfiler i ett filsystem, som poster i en relationsdatabas, som webbsidor på Internet?)
- Om informationen skall nyskapas, i vilken mån kan man då styra (organisatoriskt, tekniskt) hur den skall framställas och hanteras? (Av vem, med vilken programvara, enligt vilka arbetsrutiner?)”

“XML som koncept

XML är inte en färdig lösning utan ett koncept med en verktygslåda, vilket framgår av namnet: XML = eXtensible Markup Language: Alltså ett *språk*, avsett för att *märka upp* information, konstruerat för att *utvidgas*. Den grundläggande XML-rekommendationen fastställdes så sent som 1998 och även om målen och visionerna är formulerade och ramverket existerar så återstår mycket detaljarbete. Först skall W3C samordna och besluta om alla pågående rekommendationsförslag. Sedan skall marknadens aktörer dels definiera och enas om olika branschtillämpningar, dels utveckla nödvändiga strukturer och programvaror som bygger på de överenskomna rekommendationerna. XML som idé har etablerat sig snabbt och stöds av åtskilliga offentliga projekt och organisationer, både i Sverige och internationellt, liksom av de flesta tongivande aktörer inom IT-industrin. Även inom det privata näringslivet utanför IT-branschen pågår många intressanta XML-initiativ, t ex för handel, finans och resor.

Med de många fördelar som XML-konceptet utlovar så är sannolikheten stor för att XML-baserade informationsstrukturer så småningom kommer att bli dominerande. XML skall inte ses som en ersättare för HTML på webben utan snarare som fundamentet för en helt ny teknikgeneration som successivt kan komma att ersätta exempelvis HTML, SQL, Edifact och många andra dokument- och filformat, kommunikationsprotokoll, programmeringsgränssnitt osv. På vissa områden går det fort men tidsperspektivet för att nå fullt genomslag är långsiktigt, det kan handla om decennier snarare än om år!”

“Sökfunktioner och metadata

Effektiva sökfunktioner har avgörande betydelse för att med rimliga insatser kunna hitta rätt i det växande informationsöverflödet. Att strukturera informationen är ett sätt att underlätta återsökningen. RDF (Resource Description Framework) är t.ex en rekommendation inom W3C som fokuserar på dessa frågor. The Semantic Web är en framtidsvision för informationsspridning men en hel del grundläggande utveckling återstår innan den kan bli praktisk verklighet.

Att ange metadata, dvs. ”data om data”, är en etablerad metod för att underlätta återsökning. Den vanligaste metadatamodellen, simple Dublin Core, är att betrakta som en de facto-standard inom sitt område. Dublin Core bör därför användas när begrepp som titel, författare, datum, ämne mm. är relevanta. DC-element som inte känns aktuella utelämnas då lämpligen. Kungliga biblioteket (www.kb.se) ger praktiska råd med länkar till program som t.ex. skapar HTML-kod för Dublin Core metadata. I andra situationer behöver Dublin Core kompletteras med andra sökbegrepp. Inom EU finns t.ex. initiativ för att skapa en enhetlig ”Government”-variant av Dublin Core.

Ibland krävs andra uppsättningar metadata för att de skall bli meningsfulla. Exempel på sådana branschtillämpningar finns inom turism och utbildning och fler vertikala metadatamodeller är under konstruktion. För statusorienterade metadata, t.ex. inom dokumenthantering, ärendehantering och arbetsflöden, ställs åter andra krav och det finns olika idéer till lösningar, t.ex. ett svenskt initiativ som återfinns på www.wfmd.org.

Att införa metadata av god kvalitet är resurs- och kompetenskrävande. Först behövs noggranna analyser av det aktuella materialet. Sedan krävs att terminologier fastställs, sprids och används, att rollfördelning, arbetsrutiner och programvaror för inmatning av metadata anpassas och att regelverket därefter följs av alla berörda med god disciplin. Det krävs utbildning, träning och motivation, både för att lägga in och för att framgångsrikt kunna söka med hjälp av metadata”

“Referenser, hänvisningar, länkar

Följande webbplatser kan rekommenderas för att söka aktuell och fördjupad information:

www.w3c.org - W3C:s officiella webbplats.

<http://www.w3.org/TR/#Recommendations> - Aktuell statuslista för W3C:s rekommendationer.

www.xml.org - Register sponsrat av OASIS, internationellt konsortium som stöder användning av olika standarder.

www.oasis-open.org/cover/ - Tidigare privat register, nu sponsrat av OASIS.

www.schema.net - Informativt, privat register.

www.xmlsoftware.com - Översikt över XML-programvaror, privat.

www.dublincore.org - Dublin Core Metadata Initiative's officiella webbplats.

www.kb.se - Kungliga Biblioteket, har Dublin Core översatt till svenska och ger bra praktiska hänvisningar.

www.xmlakademin.nu - basinformation om XML, informellt nätverk för informationsstrategier inom svensk offentlig sektor.

www.statskontoret.se - Statskontoret rapportserie: Vad är XML-familjen? (2000:30-36).

<http://it-upphandling.statskontoret.se/Uhw/> - Statskontorets ramavtal XML Produkter och tjänster.

Inför publiceringen av denna vägledning har bl.a. medlemmarna i följande av Statskontorets nätverk getts möjlighet att granska och kommentera innehållet:

- E-forum
- XML-akademin
- Ramavtalsleverantörerna för XML

Kontaktperson

Karl Wessbrandt, Statskontoret, tel. 08-454 4712,

e-post: karl.wessbrandt@statskontoret.se”

(Till 7.2: DTD, Document Type Declaration. Inleder HTML-dokument och talar om för webbläsaren vad det är för typ av dokument.)

Exempel där användning av XML och metadata kan förekomma: IT- plattformar för informationsförsörjning

Presentation av en förstudie med inriktning på 24- timmarsmyndighets behov, 10 april 2002 Karl Wessbrandt, Statskontoret. 30 sidor

<http://www.statskontoret.se/seminarier/semdok/itplattwess.pdf>
