



Statskontorets Vägledning för IT-området

Nummer och utgåva: SV 1:3 XML-teknik och metadata

status: fastställd

datum: 2003-12-12

Vägledning: XML-teknik och metadata

Förord

Användning av XML-teknik och metadata är viktiga förutsättningar för att effektivisera informationsförsörjningen inom förvaltningen. Statskontoret har därför tagit fram denna vägledning.

Den ingår i Statskontorets serie Vägledningar för IT-området. Vägledningarna publiceras elektroniskt på www.statskontoret.se, i vissa fall även som pappersupplaga.

Syftet är att bidra till en effektiv och samordnad utveckling och användning av IT, i första hand inom den offentliga sektorn, genom att erbjuda aktuella råd och rekommendationer inom utvalda områden. Målgruppen är i första hand IT-strateger, IT-ansvariga, projektledare, konsulter och IT-leverantörer.

Nu föreligger version tre av denna vägledning. I förhållande till version två (februari 2003) har flera viktiga XML standarder uppdaterats och fått status som officiella W3C Recommendations. Vägledningen har även uppdaterats med avseende på XML-baserade ordbehandlingsformat och web services, samt fler hänvisningar till informativa XML-källor på Internet.

1. INTRODUKTION	5
2. MÅLSÄTTNING	7
3. MÅLGRUPP	7
4. AVGRÄNSNINGAR.....	7
5. BAKGRUND	8
5.1. ANVÄNDNINGSSOMRÅDEN FÖR XML.....	9
5.2. FAKTORER SOM PÅVERKAR STANDARDERS ACCEPTANS	10
6. ORGANISATIONER, TERMER OCH DEFINITIONER	10
6.1. WORLD WIDE WEB CONSORTIUM (W3C)	10
6.2. DUBLIN CORE METADATA INITIATIVE (DCMI).....	11
6.3. WEB SERVICES INTEROPERABILITY ORGANIZATION (WS-I).....	12
6.4. ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS (OASIS).....	13
6.5. INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO)	13
7. REKOMMENDATIONER	13
7.1. EXTENSIBLE MARKUP LANGUAGE (XML) 1.0 (SECOND EDITION).....	14
7.2. XML-SCHEMA VERSION 1.0	14
7.3. NAMESPACES IN XML VERSION 1.0	15
7.4. XML POINTER LANGUAGE (XPOINTER) VERSION 1.0	15
7.5. XML PATH LANGUAGE (XPath) VERSION 1.0.....	16
7.6. XML LINKING LANGUAGE (XLINK) VERSION 1.0	16
7.7. XML BASE.....	17
7.8. XHTML VERSION 1.0.....	17
7.9. EXTENSIBLE STYLESHEET LANGUAGE (XSL) VERSION 1.0	18
7.10. XSL TRANSFORMATIONS (XSLT) VERSION 1.0.....	19
7.11. ASSOCIATING STYLE SHEETS WITH XML DOCUMENTS VERSION 1.0.20	
7.12. XFORMS VERSION 1.0.....	20
7.13. SOAP VERSION 1.2.....	20
7.14. WS-I BASIC PROFILE VERSION 1.0 A	21
7.15. DUBLIN CORE METADATA ELEMENT SET (DCMES) VERSION 1.1....	22
8. VIKTIGA TILLÄMPNINGSOMRÅDEN	23
8.1. WEB SERVICES	23
8.2. WEB SERVICES DESCRIPTION LANGUAGE (WSDL) VERSION 2.0.....	24
8.3. UNIVERSAL DESCRIPTION, DISCOVERY, AND INTEGRATION (UDDI)	24
8.4. XML-BASERADE KONTORSdokument	25
9. RÅD OCH ANVISNINGAR	25
9.1. OM XML	26
9.2. OM METADATA	27



10.	BAKGRUNDSINFORMATION	28
10.1.	XML SOM KONCEPT	29
10.2.	SÖKFUNKTIONER OCH METADATA	31
11.	ORDLISTA	32
12.	AKRONYMER	33
13.	WEBB-BASERADE XML-KURSER	33
14.	REFERENSER, HÄNVISNINGAR, LÄNKAR.....	33
15.	KONTAKTPERSON	34

1. Introduktion

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

Anmärkning:

I detta dokument används termen "standard" som beteckning för det som brukar kallas "de facto standard", "öppen industristandard" eller "rekommendation". De "standarder" som nämnes i dokumentet är alltså inte alltid fastställda av något formellt standardiseringsorgan, utan produceras och förvaltas av industrigrupperingar med bred förankring i såväl användar- som leverantörsled.

Att en standard inte har tagits med skall inte tolkas som att Statskontoret avråder från att tillämpa den. Statskontoret har av olika skäl ännu inte tagit ställning.

Här påvisas också några andra framgångsfaktorer för att lyckas med införandet av strukturerad informationshantering, utöver valet av lämpliga tekniska lösningar, t.ex. följande:

- För att kunna fastställa strukturer och terminologier krävs noggranna förberedelser i form av informationsanalys, anpassning av verksamhetsprocesser, anskaffning av programvaror och utbildning av medarbetare.
- Informationsmodellering är också nödvändig för att kunna skapa gemensamma och kontrollerade vokabulärer, vilket ytterligare ökar möjligheterna till samverkan och återanvändning av information. I modernt språkbruk kallas detta för återanvändning av ontologier (begreppsmodeller).
- Man behöver ta hänsyn till att lösningar som bygger på manuell uppmärkning av stora informationsmängder blir personalintensiva och kompetenskrävande. Arbetet med informationsuppmärkning bör därför minimeras genom automatisering och stöd av bra programvaror och metoder.
- Man bör alltid först söka och om möjligt återanvända metoder och applikationer som redan är utvecklade och beprövade.

En viktig målsättning med allt informationsstruktureringsarbete är att framtidssäkra den investering som arbetet innebär. Med andra ord, sträva efter att hitta lösningar

vilka minimerar den mängd underhållsarbete som kan krävas i framtiden, då omvärlden, verksamheten och/eller processer ändras. Framtida kostnader beror dels på hur analys/strukturering/modellering genomförts och dels på vilka XML-applikationer (ev. branchspecifika) som man valt.

XML är en grupp standarder som har starkt stöd av företag och organisationer, i Sverige och internationellt. XML är inte en enda standard eller en färdig lösning utan en verktygslåda som innehåller en uppsättning basverktyg, men som även kontinuerligt tillförs nya verktyg. Statskontoret rekommenderar några grundläggande standarder som nu bedöms vara etablerade. I övrigt rekommenderas den princip (fritt översatt) som 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".

När det gäller frågor om språk för representation av verksamhetsinformation är grundrekommendationen att XML-baserade standarder skall användas. Speciellt för informationsutbytesgränssnitt mellan system och organisationer bör XML-lösningar övervägas.

Metadata (dvs. data om data) används för att underlätta informationssökning och för att hantera statusinformation. Den mest spridda modellen, Dublin Core, är numera en internationell 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.

En annan kategori av metadata är sådana data som dokumenterar datamodeller (informationsmodeller, informationsstrukturer). Den här typen av metadata anger krav som konkreta informationsobjekt måste uppfylla och beskriver hur olika informationsobjekt kan länkas samman. Sådan information är av stor vikt vid förändringsprocesser och informationen fungerar då som nyckelinformation om vilka effekter olika typer av förändringar kan få. Den teknologi som har bäring på detta område är t.ex. schema-information och DTD:er.

Denna utgåva av vägledningen är version tre. Vägledningen kan komma att utökas och bli mer detaljerad i kommande versioner. Det är också tänkbart att inriktningen kan komma att behöva justeras i takt med den snabba teknik- och marknadsutvecklingen på området.

Statskontoret rekommenderar därför alla berörda att själva kontinuerligt följa vad som sker, t.ex. på de webbplatser som anges i kapitlet "Referenser".

2. Målsättning

En effektiv och samverkande informationshantering är en grundförutsättning för att öka tillgänglighet och öppenhet hos myndigheterna, bl.a. för att förverkliga 24-timmarsmyndigheten.

Strukturerad informationshantering med hjälp av XML-teknik och metadata är viktig som möjliggörare för en effektiv informationsförsörjning inom många användningsområden.

Samtidigt måste verksamhetsprocesserna utvecklas, och den nya tekniken måste kunna samverka med befintliga lösningar för att nå önskat resultat.

Syftet med denna vägledning är att stärka beställarkompetensen inom offentlig sektor vid planeringen av informationsförsörjningslösningar. Den skall också förbättra förutsättningarna för beställarna att kommunicera med IT-leverantörer och XML-specialister.

3. 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 kapitlet "Referenser".

4. Avgränsningar

Version tre av denna vägledning begränsas till att endast ta upp generella, brett tillämpbara standarder.

För XML-området är grundläggande standarder fastställda av World Wide Web Consortium (W3C). Beträffande metadata hänvisas till rekommendationer framtagna av Dublin Core Metadata Initiative (DCMI).

Att en fastställd standard inte tas upp i denna vägledning skall inte tolkas som att Statskontoret avråder från att tillämpa den. Det innebär bara att Statskontoret av olika skäl inte har tagit ställning. Vissa ännu ej fastställda förslag till kommande "W3C Recommendations" har redan nått stor spridning och användning. Det kan

finnas skäl att redan nu använda dem, men man bör då vara medveten om risken att förslagen kan komma att förändras.

Det finns också ett stort antal standarder och initiativ från andra organisationer än W3C och DCMI som bygger på XML och metadata, t.ex. XML-applikationer för att stödja speciella behov i olika samhälls- och industrisektorer. Statskontoret är medvetet om att många av dessa är väl etablerade inom respektive område, men att bevaka och värdera speciella branschlösningar ligger utanför denna väglednings ambitioner.

Statskontoret rekommenderar därför myndigheterna att i sina projekt aktivt undersöka om gemensamma branschinitiativ finns inom respektive myndighets verksamhetssektor. I sådana fall bör myndigheten pröva möjligheten och lämpligheten av att ansluta sig till dessa initiativ.

Sökningångar till sådana samarbetsmöjligheter kan t.ex. finnas via några av de webbplatser som anges under referenser nedan.

Beträffande användning av metadata så bör man helst ha s.k. kontrollerade vokabulärer eller terminologier, kodlistor m.m. till sitt förfogande för att uppnå hög kvalitet och effektivitet. Arbetet med att skapa och fastställa sådana bör bedrivas sektorsvis, och det blir då ofta inom ramen för långsiktiga utvecklingsprojekt. Terminologifrågor är ett viktigt område som dock inte omfattas av denna vägledning. Däremot har Statskontoret utrett hur ett nationellt samverkansregister skulle kunna upprättas, där t.ex. återanvändbara strukturer, scheman, modeller, metoder och terminologier då kunde publiceras, kommenteras och vidareutvecklas. Något beslut om realisering har dock inte tagits.

5. 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 publicerades av W3C (World Wide Web Consortium) år 1998, men XML används också som ett samlingsbegrepp för hela XML-familjen, vilken omfattar ett stort antal relaterade standarder och initiativ, såväl inom som utom W3C. Dessa olika XML-"släktingar" har olika syften, omfattning och befinner sig i olika utvecklingsstadier. XML-konceptet har fått full acceptans inom IT-sektorn och en stor uppsättning leverantörer av basverktyg opererar på marknaden. Men utvecklingen står inte stilla, nya standarder kommer fram och arbete utförs för att utveckla programvaror och applikationer som följer kommande standarder. Vid utveckling av

standardsystem byggs XML-teknik numera ofta in i produkterna och på det sättet kommer vi alla med tiden förmodligen att automatiskt utnyttja XML utan att veta om det.

5.1. Användningsområden för XML

Tre grundläggande områden där XML-baserade standarder har naturlig användning är:

1. Lagring av information
2. Presentation av information (t.ex. via webbläsare)
3. Kommunikation av information (t.ex. mellan system i samma eller olika organisationer)

För att själv utveckla XML-baserade lösningar krävs planering och förberedelser. Informationsstrukturer och terminologier behöver fastställas och ändamålsenliga programvaror anskaffas eller tas fram.

Att "satsa på XML" behöver därför motiveras och förtydligas.

Det är värt att peka på möjligheten att i praktiken använda sig av en delmängd av en standard. Generella och uttrycksfulla standarder är i vissa fall omfattande, just för att de försöker vara kraftfulla på ett generellt sätt. Konkreta behov i en organisation kan ofta vara mer begränsade. Det finns då alltid möjligheten att man anammar en standard men bara använder en del av det som standarden tillåter. Detta kan vara ett sätt att få en större enhetlighet i standardanvändningen inom en organisation, men kräver naturligtvis att man klart identifierar den delmängd av standarden som man tillåter användas.

På motsvarande sätt behöver en satsning på metadataanvändning beskrivas närmare. Metadata, dvs. data om data, har använts länge i arkiv, bibliotek och i datasammanhang. Med växande informationsmängder har intresset för metadata ökat, framför allt för att de underlättar informationssökning på webben, men metadata har också betydelse för att hålla reda på statusinformation, vilket t.ex. behövs vid ärendehantering. Dessutom får metadata en växande betydelse som informationskälla vid förvaltning/underhåll av IT-system.

Metadata är i sig teknikneutrala och kan hanteras såväl med hjälp av kortlådor eller databaser som i HTML eller XML. Man behöver därför skilja på innehåll och form. Däremot finns en naturlig koppling till XML eftersom XML-formatet i princip bygger på strukturer av metadataelement.

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 samordna satsningarna inom XML och metadata.

5.2. Faktorer som påverkar standarders acceptans

Ett antal faktorer påverkar om, hur snabbt, och på vilket sätt en standard får stöd i kommersiell programvara.

Tidsfördröjningen mellan en standards publicering och kommersiell produkttillgänglighet varierar från område till område. Att leverantörer aktivt deltar i standardiseringsarbetet skapar förutsättning för att produkten kommer till marknaden. Produkter för större och mer kritiska marknadssegment kommer naturligtvis snabbare till marknaden.

Produktstabilitet stöds av (1) stabilitet i standarder, och (2) sätt att avgränsa standarder (dels allmänt, men även gränsdragning gentemot andra standarder). Ett omsorgsfullt och öppet standardutvecklingsarbete och ett arbete där grupper av standarder identifieras och specificeras som bitar i ett större pussel borgar för mer stabila, öppna standarder och därmed stabila produkter på marknaden. Betydelsen av öppen programvara (svensk översättning av den etablerade beteckningen OpenSource Software) skall inte underskattas. Tidigt implementeringsarbete i open-sourceorganisationer stöder i praktiken standardiseringsarbete genom identifiering av möjliga inkonsistenser och gap i föreslagna standarder. De identifierar även möjliga tekniska implementeringsansatser som förbereder för efterfrågan genom att tillhandahålla tidiga implementationer. Ett exempel på detta är Apache-gruppen (<http://www.apache.org>), där många XML-orienterade programprototyper av god kvalitet tagits fram (såsom webb-servern Apache och processorer för SOAP, XSLT och SVG).

6. Organisationer, termer och definitioner

6.1. World Wide Web Consortium (W3C)

Organisationen W3C bildades 1994 för att leda utvecklingen av gemensamma webbstandarder och därmed säkra framtida interoperabilitet över webben. W3C har idag cirka 450 medlemsorganisationer från hela världen, såväl myndigheter och användar- och forskningsorganisationer som produktproducerande företag. Arbetet i W3C drivs i arbetsgrupper, huvudsakligen bemannade med personal från W3C:s medlemsorganisationer. Processen är konsensusbaserad, dvs. resultat som produceras stöds av alla inblandade organisationer.

Statusbeteckningar för standardiseringsarbetet inom W3C:

Följande statusbeteckningar används på vägen fram till en fastställd "Recommendation"¹:

- **Working group note**
Endast information om kunskap framtagen i den specifika arbetsgruppen, inga officiellt erkända W3C dokument.
- **Working draft (WD)**
Arbetsversion av eventuell framtida standard
- **Candidate recommendation (CR)**
Specifikation som aspirerar på att bli en standard och väntar på kommentarer från andra arbetsgrupper inom W3C
- **Proposed recommendation (PR)**
Förslag till standard som nått en viss förankring och som nu är uppe för granskning
- **Recommendation (REC)**
Av W3C antagen standard. Detta innebär att W3C:s medlemsorganisationer granskat och godkänt förslaget.

Det finns förslag på att införa ytterligare en statusklassificering för dokument, Rescinded Recommendation (RR). Klassificeringen skall användas på standarder som W3C inte längre vill förorda, t.ex. därför att teknikutvecklingen tagit en annan riktning.

XML är namnet på den grundläggande XML-specifikation som publicerades som "Recommendation" av W3C år 1998, men XML används också som ett samlingsbegrepp för hela XML-familjen som omfattar ett stort antal relaterade standarder och initiativ.

6.2. Dublin Core Metadata Initiative (DCMI)

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) och fokuserade initialt på metadata som stöd för sökningar i biblioteksvärlden, dvs. sökningar i det

¹ <http://www.w3.org/2003/06/Process-20030618/tr.html>

man traditionellt kallar publikationer. I takt med att nya typer av objekt blir tillgängliga på webben, har DCMI:s resultat kommit till användning även utanför biblioteksvärlden för att uttrycka metadata om andra typer av informationsobjekt.

DCMI har för närvarande 18 aktiva arbets- och intressegrupper. Det finns runt 80 olika implementeringsprojekt av DCMI registrerade på DCMI-projektets hemsida, hemmahörande i alla olika världsdelar och inom många olika ämnesområden.

Statusbeteckningar för standardiseringsarbetet inom DCMI:

DCMI använder statusbeteckningarna Note, Working Draft, Proposed Recommendation och Recommendation på liknande sätt som W3C (se ovan).

6.3. Web Services Interoperability Organization (WS-I)

WS-I, grundat 2001, är en annan internationell sammanslutning av företag, med målsättning att arbeta för interoperabilitet mellan webbtjänster ("web services") baserade på accepterade standarder.

Det huvudsakliga resultatet av WS-I:s arbete är specificerade profiler, där en webbtjänst skall kunna deklarerat att den understödjer en viss profil. En sådan profil definieras i termer av ett antal specifika webbstandarder (t.ex. SOAP och WSDL), en nivå för varje standard (t.ex. SOAP 1.1) och hur vissa specifika element i denna nivå skall användas. Man kan uttrycka det som att en WS-I profil är en standard definierad som en konfiguration av explicita delmängder av existerande standarder. Arbetet bedrivs i arbetsgrupper, vilka skapas för att ta fram konkreta resultat, t.ex. testprogramvara, testapplikationer eller specifika standarder.

Man bör notera att WS-I:s insatser är ett komplement till andra organisationers arbete inom området webbstandarder, och WS-I samverkar med dessa andra organisationer. Denna samverkan underlättas av att WS-I:s medlemmar till största delen också är W3C-medlemmar.

Statusbeteckningar för standardiseringsarbetet inom WS-I:

Följande statusbeteckningar används av WS-I:

- **Draft specification**
Arbetsversion av en eventuell framtida standard.
- **Final specification**
Fastställd standard, granskad och godkänd av WS-I:s medlemmar.

6.4. Organization for the Advancement of Structured Information Standards (OASIS)

OASIS är ytterligare ett globalt konsortium, bildat 1993, som stöder utveckling, praktisk tillämpning och spridning av standarder. OASIS har över 600 medlemmar i 100 länder och arbetet är organiserat i ett stort antal tekniska kommittéer. Exempel på uppmärksammade projekt som drivs inom ramen för OASIS är ebXML, LegalXML och UDDI.

Två relativt nystartade tekniska kommittéer är av särskilt intresse eftersom de arbetar med Open Office XML Format respektive med standardiseringsfrågor kring e-Government.

6.5. International Organization for Standardization (ISO)

ISO är världens största standardiseringsorganisation. Sekretariatet är förlagt till Genève men ISO har medlemmar från 147 länder. Sedan 1947 har ISO publicerat fler än 13 700 internationella standarder inom de flesta områden men dess främsta uppgift är att ta fram tekniska standarder.

Standardiseringsaktiviteter förekommer även inom många andra standardiseringsorgan: Internet Engineering Task Force (IETF), European Committee for Standardization (CEN) och amerikanska National Information Standards Organization (NISO) är några exempel.

7. Rekommendationer

De XML- och metadata-standarder som tas upp i detta kapitel, har tagits fram inom ramen för W3C eller DCMI och har den högsta statusbeteckningen Recommendation.

Dublin Core, den metadatarekommendation som DCMI har tagit fram, har en speciell status eftersom den även har fastställts som en ISO-standard.

Nedan beskrivs de standarder som Statskontoret för närvarande har tagit ställning för. Att en fastställd standard inte tas upp i denna vägledning skall inte tolkas som att Statskontoret avråder från att tillämpa den. Det innebär bara att Statskontoret av olika skäl ännu inte tagit ställning. Vissa förslag till kommande W3C-standarder, dvs. kommande "W3C Recommendations", har redan nått stor spridning och användning. Det kan finnas skäl att redan nu använda dem, men man bör då vara medveten om risken att de kan komma att förändras.

Det finns även ett stort antal andra standarder och initiativ som bygger på XML och metadata, t.ex. applikationer för att stödja olika sektors speciella behov. Att

bevaka och värdera dessa branschlösningar ligger utanför denna väglednings ambitioner även om många av dessa är väl etablerade inom respektive område.

I de följande kapitlen ges ytterligare råd, anvisningar, information och källhänvisningar.

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

W3C Recommendation (februari 1998, 2nd edition: oktober 2000)

Grunden i hela XML-ramverket för strukturerad informationshantering är denna Recommendation. Den specificerar bland annat hur XML-dokument får se ut och skall byggas upp, dvs. syntaxen. Mot dessa krav kontrolleras att ett XML-dokument följer den gällande XML-syntaxen. XML 1.0 specificerar XML-dokumentets logiska struktur (element/attribut-struktur), deras fysiska struktur (lexikala/syntaktiska form), och begreppet validitet (gentemot en dokumenttyp).

XML 1.1 (Proposed Recommendation, november 2003) är en uppdatering av XML som i huvudsak orsakas av den utveckling som unicode-standarden genomgått sedan tidpunkten då XML 1.0 blev "Recommendation". Denna får effekt på vad som är legala namn i XML-dokument eftersom antalet definierade Unicode-tecken förändrats.

Läs mer på: <http://www.w3.org/XML/>
och <http://www.w3.org/TR/REC-xml> resp. <http://www.w3.org/TR/xml11/>

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

7.2. XML-Schema Version 1.0

W3C Recommendation (maj 2001)

XML Schema är ett modernare och mer fullständigt sätt att definiera strukturen för XML-dokument och kan användas istället för en DTD. 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. 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 som uttrycks i ett specialiserat språk. Förutom XML 1.0 bygger XML Schema även på XML Namespaces. Många XML-applikationer använder fortfarande DTD:er och många programvaror saknar fullt stöd för XML Schema, men användningen av XML Schema ökar, framför allt tack vare dess vidgade möjligheter att uttrycka krav på datatypning.

Läs mer på: <http://www.w3.org/XML/Schema>
<http://www.w3.org/TR/xmlschema-0/> ("XML Schema Part 0: Primer")

Teknisk specifikation:
<http://www.w3.org/TR/xmlschema-1/> ("XML Schema Part 1: Structures")
<http://www.w3.org/TR/xmlschema-2/> ("XML Schema Part 2: Datatypes")

7.3. Namespaces in XML Version 1.0

W3C Recommendation (januari 1999)

Det ökande behovet att sammanställa information från oberoende källor ställer krav på att olika uppmärkningspråk, använda i olika dokument, inte sammanblandas med varandra. Det måste vara möjligt att entydigt, men samtidigt effektivt och hanterbart, kunna särskilja en uppsättning XML-element/attribut från en annan uppsättning element/attribut. I XML-familjen åstadkommes detta genom användning av namnrymder ("name spaces").

"Namespaces in XML" är en W3C Recommendation som definierar hur namn på element och attribut får uttryckas, och hur två olika "namnförekomster" kan stå för samma namn. Användningen av Namespaces gör det möjligt att hitta det element eller attribut som specifikt används i ett XML dokument, även om det finns element med samma elementnamn, t.ex. "ID". Namespaces åstadkommer detta 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 applikation används och prefixet fungerar då som referens till rätt namnrymd.

I november 2003 blev Namespaces 1.1 en Proposed Recommendation. Skillnaden gentemot Namespaces 1.0 är, utöver rättelser av några fel, tillägg av en enda funktion: möjlighet att avdeklara ("undeclare") ett prefix. Namespaces 1.1 avses gälla för XML 1.1 och högre. Av dessa skäl kan användning av Namespaces 1.0 anses vara framtidssäkrad.

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

7.4. XML Pointer Language (XPointer) Version 1.0

W3C Recommendation (mars 2003)

XPointer, XML Pointer Language, är det språk som skall användas då man behöver uttrycka en länk till en bestämd del av ett XML-dokument. XPointer bygger vidare på XPath, och används för att identifiera delar av dokument. Identifieringen baseras på olika egenskaper i dokumentet, t.ex. element-typer, attributvärden, en position relativt ett annat element m.m.

Teknisk specifikation: <http://www.w3.org/TR/xptr-element/>,
<http://www.w3.org/TR/xptr-framework/> samt <http://www.w3.org/TR/xptr-xmlns/>

7.5. XML Path Language (XPath) Version 1.0

W3C Recommendation (november 1999)

XPath används för att referera till data inom ett XML-dokument. Dess primära användningsområde är att utgöra en gemensam grund för att specificera delar av XML-dokument, vilket är funktioner som behövs i t.ex. XSL Transformations (XSLT) och XPointer.

Data som kan refereras m.h.a. XPath är t.ex. element, attribut, text och andra komponenter. Notationen liknar sökvägar ("paths") i en katalogstruktur. XPath bygger förutom XML 1.0 även på XML Namespaces.

XPath 2.0 är publicerad i november 2003 som "Working Draft". Målsättningen med denna efterföljare till XPath 1.0 är huvudsakligen att förenkla XPath användning, samtidigt som bakåtkompatibilitet bevaras. Av detta skäl kan användning av XPath 1.0 anses vara framtidssäkrad.

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

7.6. XML Linking Language (XLink) Version 1.0

W3C Recommendation (juni 2001)

Ett av de områden där aktuella och framtida behov går utöver vad som traditionellt erbjuds av HTML är behovet av mer avancerade länkar mellan olika dokument (resurser). HTML erbjuder några språkkonstruktioner, varmed enkla riktade länkar kan specificeras. XLink erbjuder uttrycksmedel med vars hjälp man kan uttrycka sådant som att en länk kan koppla till mer än en resurs, vilken som visas och/eller väljs är sedan upp till webbläsaren. XLink gör det också möjligt att knyta metadata till en länk och att specificera länkar på platser som är skilda från de resurser de sammanlänkar.

För att förenkla hanteringen av HTML-liknande länkar (enkla länkar täcker en stor del av dagens behov av länkar), definierar XLink två kategorier av länkar: "extended links" och "simple links". Extended links erbjuder den fulla XLink-funktionen men kräver något mer omfattande specifikationer av länkar. Simple links har inte så rika uttrycksmöjligheter men möjliggör att länkar specificeras på ett sätt som påminner mycket om vanliga HTML-länkar.

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

7.7. XML Base

W3C Recommendation (juni 2001)

XML Base tillhandahåller uttrycksmedel för att ange en referenspunkt för relativa länkar i ett XML-dokument på ett sätt som motsvarar Base-elementets roll i ett HTML-dokument. Dess huvudsakliga användningsområde är att ge stöd åt XLink, men genom att definieras på ett modulärt sätt kan XML Base även användas i andra XML-sammanhang där relativa länkar används.

Denna W3C Recommendation, en av de textmässigt sett minsta i XML-floran, definierar ett enda attribut, `xml:base`, samt specificerar hur relativa länkar skall omtolkas för att bli absoluta länkar.

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

7.8. XHTML Version 1.0

W3C Recommendation (januari 2000)

Extensible HyperText Markup Language (XHTML) återskapar och utvidgar HTML, uttryckta i XML. XHTML 1.0 är den första XHTML Recommendation och den som främst används idag. Den omformulerar HTML 4.01 i XML.

Det finns tre varianter definierade i XHTML 1.0: Strict, Transitional och Frameset.

1. XHTML 1.0 Strict möjliggör en ren struktur-uppmärkning, helt fri från presentations-uppmärkning, så att presentationsegenskaper helt kan uttryckas i CSS.
2. XHTML 1.0 Transitional innehåller utvidgningar för att kunna uttrycka vissa presentationsegenskaper, t.ex. bakgrundsfärg. Avsikten med varianten "Transitional" är att under en övergångstid kunna stödja användning av äldre webbläsare som inte kan hantera CSS, och som därför behöver extra stöd för XHTML.
3. XHTML 1.0 Frameset möjliggör användandet av ramar vid presentation.

Andra XHTML- projekt

XHTML 1.1 eller som det också heter XHTML Modularization (W3C Recommendation, april 2001), specificerar en modularisering av XHTML-språket vilket gör det möjligt att olika delmängder av XHTML kan definieras och

kombineras på ett systematiskt sätt. Därigenom är det möjligt att specificera vilka moduler som en plattform stöder. Se <http://www.w3.org/TR/xhtml-modularization/conformance.html> för mer information om XHTML Modularization.

XHTML Basic (W3C Recommendation, december 2000), avsett som uppmärkningsspråk för mobiltelefoner, PDA:er, "set top boxes" (digitalTV-boxar) och liknande, är ett exempel på ett XHTML-språk specificerat m.h.a. XHTML Modularization.

XHTML 2.0 är under utveckling och baserat på XHTML 1.0 men specificerat med hjälp av XHTML Modularization.

XHTML och webbläsare

XHTML bör användas för webbsidor och stöds av de flesta moderna webbläsare. Senare versioner av Internet Explorer, Netscape och Mozilla har stöd för XHTML, se www.microsoft.se, www.netscape.com och www.mozilla.org för mer detaljerade uppgifter. Opera 7 stödjer XHTML Basic men har ett antal undantag i sitt stöd för XHTML 1.0 samt XHTML 1.1. Även tidigare versionen av Opera har ett visst stöd för XHTML, se www.opera.com för mer detaljerade uppgifter.

Teknisk specifikation XHTML 1.0: <http://www.w3.org/TR/xhtml1/>

Teknisk specifikation XHTML 1.1 (Modularization):
<http://www.w3.org/TR/xhtml11/>

Teknisk specifikation XHTML Basic:
<http://www.w3.org/TR/xhtml-basic/>

7.9. Extensible Stylesheet Language (XSL) Version 1.0

W3C Recommendation (oktober 2001)

XSL är ett är en sammansättning av tre olika XML komponenter som tillsammans används för att beskriva hur XML-dokument ska se ut när de ska presenteras, t.ex. i en webbläsare. En XSL-formatmall, ett "stylesheet", beskriver hur ett XML-dokument av viss typ skall formateras. I detta avseende har XSL samma målsättning som, och är i många avseenden kompatibel med, CSS2 (Cascading Style Sheets level 2), men XSL använder XML-syntax, vilket gör det mer uttrycksfullt än CSS2.

Vid användning av XSL för formatering så utnyttjas tre olika komponenter:

1. **XSLT**
Ett språk för att transformera XML-dokument
2. **XPath**
Ett språk för att referera till delar av dokument
3. **XSL Formatting Objects**
En notation för att specificera typografisk presentation

Dessa tre komponenter stöder två huvudsteg i presentationen:

- transformering (använder XSLT och XPath) och
- formatering (använder XSL Formatting Objects).

XSLT och XPath togs fram som komponenter för att stödja presentation av XML-dokument, men genom att på lämpligt sätt avgränsa och generalisera dessa två komponenter, så har de även kommit till användning vid allmän transformering och sökning av XML-dokument. XSL Formatting Objects är en del i XSL Version 1.0.

Cascading Stylesheet level 2, CSS2, framtaget för användning tillsammans med HTML, är föregångare till XSL och kan också användas för formatering av XML-dokument.

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

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

7.10. XSL Transformations (XSLT) Version 1.0

W3C Recommendation (november 1999)

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

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

XSLT 2.0 fick i november 2003 status "Working Draft, last call". Målsättningen med denna efterföljare till XSLT 1.0 är huvudsakligen att förenkla XSLT användning, samtidigt som bakåtkompatibilitet bevaras.

Teknisk specifikation: <http://www.w3.org/TR/xslt20/>

7.11. Associating Style Sheets with XML Documents Version 1.0

W3C Recommendation (June 1999)

Denna specifikation definierar hur man i ett XML-dokument kan uttrycka vilken formatmall som skall användas vid presentation av dokument.

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

7.12. XForms Version 1.0

W3C Recommendation (oktober 2003)

XForms är en XML-baserad specifikation för nästa generations webb-formulär. Den skall täcka det behov som för närvarande hanteras av formulär i HTML, men på ett mer modulärt och flexibelt sätt. XForms skall kunna användas i olika sammanhang, såsom XHTML och SVG, ge större möjligheter att deklarativt ange datatyper för inmatade data och skall ge stort apparatberoende.

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

XForms 1.0 Basic Profile är en version av XForms som skall hantera anpassning av formulär till mer resursbegränsade apparater och miljöer. I oktober 2003 presenterades XForms 1.0 Basic Profile som en W3 Candidate Recommendation.

Teknisk specifikation: <http://www.w3.org/TR/xforms-basic/>

7.13. SOAP Version 1.2

W3C Recommendation (juni 2003)

SOAP är en central komponent i W3C:s utvecklingsaktiviteter inom Web services, se även följande avsnitt samt <http://www.w3.org/2002/ws/Activity>.

SOAP är ett språk för att kommunicera funktionsanrop och strukturerad information i XML-format mellan tillämpningsprogram i en distribuerad och decentraliserad miljö.

SOAP 1.2 består av fyra delar. Del 0: Primer, Del ett: Messaging Framework, Del två: Adjuncts, och slutligen ett dokument som beskriver Specification Assertions and Test Collection.

Tekniska specifikationer:

<http://www.w3.org/TR/soap12-part0/>,

<http://www.w3.org/TR/soap12-part1/>,

<http://www.w3.org/TR/soap12-part2/> samt <http://www.w3.org/TR/soap12-testcollection/>.

7.14. WS-I Basic profile Version 1.0 a

WS-I Final Specification (augusti 2003)

Denna specifikation tas fram av WS-I (se avsnitt 6.3), till skillnad från de specifikationer som nämnts ovan, vilka tas fram av W3C. Basic Profile är den första webservice profilen som presenterats av WS-I, men kommer att följas av flera mer avancerade profiler. Genom att webservice skapas efter en vald WS-I profil hoppas man undvika de interoperabilitetsproblem som kan vara resultatet av att olika standarder används eller att de tolkas olika.

Basic Profile är en konfiguration definierad på standarderna XML Schema 1.0, SOAP 1.1, WSDL 1.1 och UDDI 2.0.

WS-I planerar att presentera fler profiler allteftersom det blir tydligare vilka behov som finns. Förmodligen kommer nya profiler att inkludera nya versioner av XML Recommendations, när dessa har fått en större spridning och acceptans.

Det föreligger således ett visst timing-problem, då W3C redan har publicerat uppdateringar av de specifikationer som WS-I har haft som grund för sitt arbete. Eftersom det finns ett antal viktiga skillnader mellan SOAP 1.1 och SOAP 1.2 är det troligt att SOAP 1.2 kommer att inkluderas i en eller fler framtida profiler. WSDL-komponenten i Basic Profile 1 är version WSDL 1.1 som får betraktas som en provisorisk version på vägen mot en fastlagd WSDL-rekommendation. Som beskrivs under avsnittet om "WSDL", nedan, så är WSDL 2.0 inte en ren utvidgning av WSDL 1.1, vilket kan medföra att rekommendationerna i Basic Profile inte blir direkt tillämpliga på WSDL 2.0, som troligen kommer att ligga närmare en kommande W3C-rekommendation. Det är rimligt att anta att det kommer att produceras en ny WS-I profil om hur WSDL 2.0 skall användas, men det är i dagens läge inte känt om och när detta kommer att ske.

Teknisk specifikation: <http://www.ws-i.org/Profiles/Basic/2003-08/BasicProfile-1.0a.html>

7.15. Dublin Core Metadata Element Set (DCMES) Version 1.1

DCMI Recommendation (2003)

DCMES har också fastställts som ISO-standard, ISO 15836-2003.

Denna standard är framtagen av DCMI , till skillnad från flertalet standarder som nämnts ovan, vilka tagits fram av W3C.

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 har specificerats för detta ändamål.

För att underlätta återsökning av dokument rekommenderas i första hand användning av relevanta delar av de 15 elementen i DCMES.

Teknisk specifikation: <http://dublincore.org/documents/2003/02/04/dces>

Läs mer på: <http://dublincore.org/>

8. Viktiga tillämpningsområden

8.1. Web services

Web services har under senare år varit ett av de hetaste IT-områdena och bygger på XML. Alla betydelsefulla teknologileverantörer har satsat på att leverera infrastruktur som skall göra det enkelt att såväl bygga upp som förvalta en organisations utbud av webb-tjänster. Många användarorganisationer ser också webb-tjänster som en infrastruktur för integration mellan egna tillämpningsprogram på ett intranät, vilket kan bidra till att minska beroendet av leverantörsspecifika integrationsplattformar.

Utvecklingstakten har bromsats upp något, bland annat beroende på att standarderna inte varit färdigutvecklade, på osäkerhet vid valet mellan olika initiativ från konsortier och leverantörer samt på frågetecken kring säkerheten. I takt med att standarder och produkter mognar fortsätter dock utvecklingen och spridningen av web services till nya områden.

Web services är inte ett helt entydigt definierat begrepp men de brukar allmänt anses bygga på XML, SOAP, WSDL och UDDI eller motsvarande.

Med en webbtjänst menas en automatisk tjänst som typiskt levereras över HTTP. För att kunna utnyttja en tjänst måste man veta vad tjänsten kan utföra och hur man kan anropa den. WSDL, "Web Service Description Language", är ett språk i vilket man kan uttrycka sådana beskrivningar. Med andra ord, WSDL stöder upptäckt ("discovery") och beskrivning ("description") av tillgängliga webbtjänster. Sådana tjänstebeskrivningar publiceras av tjänsteleverantören, i någon publikt sökbar databas, så att klienter kan hitta lämpliga tjänster (t.ex. via UDDI, "Universal Description, Discovery, and Integration"). När väl klienten har fått WSDL-beskrivningen av tjänsten så kan klienten anropa denna genom att sända ett meddelande till tjänsten (t.ex. till en URL som angivits i tjänstens WSDL-beskrivning). Såväl det sända meddelandet som det svar klienten får tillbaka uttrycks i SOAP, som är ett språk varmed man kan överföra strukturerad information mellan tillämpningsprogram.

Flexibiliteten i de standarder som webb-tjänster bygger på gör det möjligt att definiera såväl mycket komplexa som mycket enkla tjänster. För att underlätta ett snabbt införande av dessa standarder har några enklare kombinationer av grundstandarderna fastlagts i form av "webb-tjänst-profiler". Genom att en klient direkt kan identifiera vilken profil tjänsten utnyttjar så kan klienten enkelt bygga de meddelanden som sändes till tjänsten och enkelt avkoda det svar som erhålles i

retur. Sådana profiler definieras t.ex. av en till W3C komplementär gruppering, WS-I (se föregående avsnitt).

SOAP 1.2 har nu blivit en W3C Recommendation och UDDI V2 kan betraktas som en öppen standard. Den återstående centrala komponenten för web services som ännu inte har uppnått en fastställd status är beskrivningsspråket WSDL.

8.2. Web Services Description Language (WSDL) Version 2.0

W3C Working Draft (november 2003)

WSDL är ett XML-baserat språk varmed man kan beskriva webb-tjänster, "web services". Beskrivningsmodellen baseras på meddelandeformat och protokoll, operationer och tjänster ("services"). WSDL definierar också bindningar till ett antal protokoll, såsom SOAP, HTTP och MIME.

WSDL 1.1 är en tidigare arbetsversion av WSDL-språket, ett steg på vägen mot WSDL 2.0 (föret kallad 1.2), med målet att bli fastlagd W3C-standard. Det nya i version 2.0 (gentemot version 1.1) är att den är bättre integrerad i grundläggande specifikationer (såsom XML Schema), "service descriptions" kan uttryckas enklare, vissa potentiella interoperabilitetsproblem har eliminerats, och att bindningarna mot HTTP har förbättrats. Detta betyder att version 2.0 bör betraktas som mer än en ren utvidgning av version 1.1.

Läs mer på <http://www.w3.org/2002/ws/desc/>

WSDL 2.0 är för närvarande en Working Draft.

Teknisk specifikation: <http://www.w3.org/TR/wsdl20/>

8.3. Universal Description, Discovery, and Integration (UDDI)

OASIS Open Standard

Universal Description, Discovery and Integration är en specifikation för distribuerade, webbaserade informationsregister över tillgängliga webbtjänster. UDDI finns också som publikt tillgänglig implementering av specifikationen, som gör det möjligt att registrera information om de webbtjänster man erbjuder så att andra kan finna dem.

UDDI-specifikationen tas fram inom ramen för organisationen OASIS öppna standardisering. För närvarande gäller UDDI version 2 som standard, men UDDI version 3 publicerades i augusti 2003.

Läs mer på: <http://www.uddi.org/>

Tekniska specifikationer:

<http://www.oasis-open.org/committees/uddi-spec/doc/tcspecs.htm#uddiv2>

8.4. XML-baserade kontorsdokument

Ett annat viktigt tillämpningsområde för XML är kontorsdokument. De senaste versionerna av OpenOffice och Sun StarOffice, dvs. kontorsapplikationspaket som bygger på öppen källkod, använder ett XML-baserat format som standard. Eftersom Microsofts Office-format, t.ex. .doc idag kan anses vara de dominerande formaten, så inkluderar OpenOffice ett stöd för att hantera de formaten, även om konverteringen av dokument mellan de två programmen i .doc-formatet inte är helt säker².

Microsofts nya office-version, Office 2003, inkluderar även den hantering av XML-format. Dock ersätter Microsoft inte sina egna format med XML, utan utökar med en möjlighet att spara dokument i XML-format, vilka finns publicerade och kan användas även av andra programvaror. I de professionella utgåvorna av Office 2003 finns också ett utvidgat XML-stöd, med möjlighet att använda sig av andra XML-scheman än Microsofts egna. XML-formaten i kontorsapplikationspaketen från Sun och Microsoft är dock baserade på helt olika filstrukturer baserade på DTD:er resp. XML scheman i de versioner som nu finns. Detta betyder att de i praktiken inte är kompatibla, trots att de båda bygger på XML. Det är ändå ett utvecklings-steg i rätt riktning, som bland annat förbättrar kontorsdokumentens arkivbeständighet. Ett förbättrat och utvidgat programvarustöd för olika XML-format kan också förväntas, vilket på sikt minskar beroendet av att använda specifika kontorsapplikationer.

9. Råd och anvisningar

Ovanstående standarder 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.

² Se *Interoperability Test and XML Evaluation of StarOffice Writer 6.0 and Office Word 2003 Beta 2*, Statskontoret (2003:122)

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.

Valet av plattformar och standarder beror också på med vilken tidshorisont man arbetar. Särskilt i tidiga projektsteg, när man har långt kvar till den konkreta tekniska implementeringen, kan det vara en fördel att göra antaganden om den troliga framtida utvecklingen, snarare än att bara utgå från det faktiska dagsläget.

Eftersom XML-familjen utgör en uppsättning beskrivningar och specifikationer som i sin tur ligger till grund för marknadens utveckling av strukturer och programvaror så måste man räkna med en viss marknadsfördröjning innan nya standarder kan realiseras i praktiken.

Tillgången på användarvänliga och stabila programprodukter är därför en väsentlig komponent som bör tillmätas stor vikt. Se t.ex. Statskontorets ramavtal för informationsförsörjning 2002 som inkluderar XML-produkter <http://it-upphandling.statskontoret.se/Uhw/> eller www.xmlsoftware.com, en privat översikt över programvaror som stöder XML-standarder.

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. Statskontoret kompletterar ovanstående med ytterligare några allmänna, praktiska råd:

9.1. Om XML

- Vid utveckling av webbplatser, använd i första hand XHTML.
- I projekt där nya informationsstrukturer behöver skapas, satsa i första hand på lösningar baserade på XML Schema.
- Vid utbyggnad av befintliga informationsstrukturer kan dock befintliga DTD:er användas tills vidare, om fungerande sådana redan existerar.

För att få en robust informationsstruktur som tål störningar och förändringar så bör man sträva efter att skapa informationsmängder som i sig själva är så ”intelligenta” som möjligt, dvs. information som är självbeskrivande och som bibehåller sin struktur och betydelse när den flyttas runt. Detta blir t.ex. väsentligt för en hållbar framtida infrastruktur för 24-timmarsmyndigheten. Särskilt viktigt är det att anpassa gränssnitten för att importera och exportera information mellan olika

system, t.ex. med hjälp av XSLT och XSL. Detta är ett praktiskt tillämpningsområde där XML är på väg att få stor betydelse som en enklare metod för att skapa enhetliga överföringsformat mellan system som har olika interna informationsformat.

Delar av XML-standarderna och deras praktiska användning kan ha mindre relevans eller vara mer avlägsna i tiden vad gäller realiserbarhet. Det kan t.ex. dröja innan marknader erbjuder anpassade standardprogramvaror. XML bör alltså inte ses som en generell metod utan som en teknisk lösning som man väljer när målen är klart formulerade och förutsättningarna är tekniskt, resursmässigt och organisatoriskt klarlagda. Att i egen regi utveckla och införa helt nya XML-lösningar är krävande, så möjligheten att använda standardprodukter och att återanvända befintliga XML-applikationer bör alltid undersökas.

Statskontoret kan också rekommendera att följa den princip som (fritt översatt) återfinns på W3C:s (World Wide Web Consortium, dvs. XML-grundarnas) egen webbplats:

"XML är inte alltid den bästa lösningen men den är alltid värd att överväga"

9.2. Om Metadata

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

Resurs- och kompetenskraven som det innebär att manuellt definiera metadata av hög kvalitet medför ibland att man av praktiska och ekonomiska skäl vill söka alternativa metoder. De kan baseras på sådana metadata och sökvägar som är möjliga att fånga med automatik. Det kan vara metadata som kan härledas ur sammanhanget (dvs. informationens plats och ursprung i befintliga informationsstrukturer och -hierarkier, såsom filsystem eller webbplats). Man kan också komplettera med fulltextindexering, genom utnyttjande av avancerade sökmotorer. För exempel på sökverktyg, se Statskontorets ramavtal för informationsförsörjning 2002
<http://it-upphandling.statskontoret.se/Uhw/>.

Den tekniska utvecklingen av indexerings- och söksystem går snabbt och det finns exempel på programprodukter och informationstjänster som effektivt kombinerar metadata- och fritextsökning.

För att avgränsa ett projekt eller dela in det i steg som anpassar metod och ambitionsnivå efter tillgängliga resurser, så kan man exempelvis bestämma sig för

att indexera stora volymer av befintlig information på ett enklare sätt än den information som skall skapas i framtiden.

Man kan då successivt bygga upp informationsstrukturen:

- För befintlig information så kan arbetsinsatsen begränsas genom automatiserade indexeringsmetoder och sökmotorer enligt ovan. Det gör informationen begränsat sökbar utan att kräva omfattande manuell bearbetning av källinformationen.
- För nyskapad information kan strukturerade och/eller automatiserade metoder och programvaror införas som redan vid källan fångar de metadata som skall underlätta framtida återsökning.

Metadata är ett viktigt hjälpmedel för att återsöka information. För att kunna utnyttja det mervärde som väldefinierade metadata ger vid sökning, måste dock användare ha en viss förståelse av hur metadata termer bör användas. Vilka metadata som är relevanta för att ge bra sökresultat beror på sammanhanget. Det har också påvisats att olika individer kan göra helt olika val av vilka metadata som skall anges. Metadata skall därför ses som ett av flera sökhjälpmedel. Bra sökresultat kan ofta uppnås om man kombinerar metadata dels med fritextsökning, dels med information om den logiska platsen inom en informationshierarki (geografiskt, organisatoriskt, kronologiskt osv.).

10. Bakgrundsinformation

Detta kapitel ger ytterligare, kompletterande bakgrundsinformation kring strukturerad informationshantering.

Informationsmodellering ligger i tiden. Det ständigt växande informationsöverskottet gör det uppenbart att man behöver nya hjälpmedel för att kunna navigera rätt och hitta fram till den information man behöver i en viss situation.

Informationsmodellering måste dock sättas in i sitt sammanhang. Begreppet kan tolkas på många sätt och behoven bör lösas med olika metoder beroende på utgångspunkt. All information kan inte enkelt passas in i strukturerade mallar. I själva verket är majoriteten av den information som de flesta arbetar med idag inte välstrukturerad (ur dokumentteknisk synpunkt).

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)

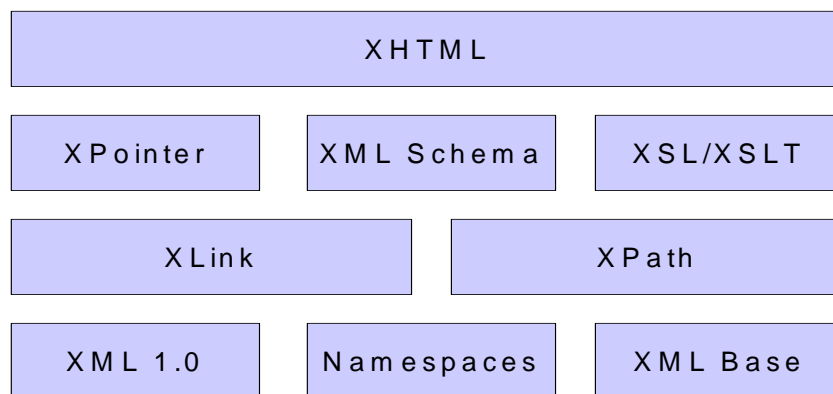
- Ingår informationen i en styrd ärendeprocess eller är den mer oförutsägbar avseende när, var och hur den uppstår (ur mottagarens/användarens synpunkt)?
- Handlar det om redan befintlig och/eller nyskapad information?
- Om informationen är befintlig, i vilken form existerar den nu? (Som papper i ett 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?)

10.1. 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-standarden fastställdes så sent som 1998 och även om målen och visionerna är formulerade och en uppsättning generella standarder är fastlagda så återstår det en hel del arbete för att skapa kompletta, branschspecifika ramverk. Först skall W3C samordna och besluta om alla pågående förslag till standarder. Sedan skall marknads aktörer dels definiera och enas om olika branschapplikationer, dels utveckla nödvändiga strukturer och programvaror som bygger på de överenskomna standarderna.

Statskontoret har i detta dokument identifierat några viktiga XML-baserade standarder och kort beskrivit vad de erbjuder. Eftersom dessa standarder bygger på varandra kan en figur illustrera hur standarderna är uppbyggda på ett komponentorienterat sätt.



I denna figur gäller att standarder i ett visst lager utnyttjar standarder i underliggande lager. Branschspecifika ramverk definieras sedan i lager ovanpå dessa lager av generella standarder.

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, till exempel i sektorer som handel, finans och resor.

Med de många fördelar som XML-konceptet utlovar så är sannolikheten stor för att XML-baserade informationsstrukturer 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. Frågan om att satsa på XML blir därför till slut en fråga som den globala IT-branschen avgör genom sina val av tekniska infrastrukturplattformar. Då nästan alla ledande leverantörer för närvarande visar stort intresse för XML, Web Services m.m., så är det en god strategi att även i användarledet ligga långt framme. Genom att bygga upp egna kunskaper och skaffa egna erfarenheter så skapas förutsättningar att såväl kunna bedöma och påverka marknadsutvecklingen som att optimalt utnyttja de tekniska lösningar som marknader erbjuder.

Det som tar längst tid och kräver mest resurser är inte den tekniska systemutvecklingen. Det är den verksamhetsinriktade informationsanalysen och det

därpå följande arbetet med att strukturera, standardisera och anpassa informationshanteringen inom och mellan berörda organisationer.

10.2. 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 ett arbetsområde inom W3C centrerat kring standarder vilka definierar en grund för hur metadata bör struktureras och användas. "The Semantic Web" (läs mera på <http://www.w3.org/2001/sw/>), ett annat arbetsområde i W3C, är en framtidsvision för intelligent informationshantering, bl.a. informationssökning. En hel del utveckling återstår dock innan det kan bli praktiskt verklighet.

Att ange metadata, dvs. "data om data", är en etablerad metod för att underlätta återsökning. Den vanligaste metamodellen, Dublin Core, är att betrakta som en de facto-standard inom sitt område. Dublin Core (DC) bör användas när begrepp som titel, författare, datum, ämne m.m. är relevanta. DC-element som inte känns aktuella utelämnas då lämpligen.

Kungliga bibliotekets enhet för bibliografisk utveckling och samordning, BUS, har sidor om metadata (<http://www.kb.se/Bus/Metadata/>) som 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 har t.ex. initiativ tagits för att skapa en enhetlig "Government"-variant av Dublin Core, under arbetsnamnet "EU Government Metadata Framework".

Ibland krävs andra uppsättningar metadata för att de skall bli meningsfulla. Exempel på sådana branschtillämpningar finns inom turism och utbildning, och fler vertikala metadata modeller är under konstruktion.

För statusorienterade metadata, t.ex. inom dokumenthantering, ärendehantering och arbetsflöden, ställs åter andra krav och här finns olika initiativ med idéer till lösningar.

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. På sikt kan informationskvaliteten kraftigt förbättras genom att införa mer standardiserade terminologier. Att fastställa teknik, format och strukturer är viktigt, men svårast och mest avgörande för att bygga upp en framgångsrik informationshantering är att fylla strukturerna med rätt innehåll!

11. Ordlista

Följande lista beskriver hur vissa engelska termer har översatts till svenska. Avsikten med denna lista är att minska riskerna för missuppfattningar, beroende på att svensk teknisk vokabulär inte har stabiliserats.

- apparatoberoende = device independence. Förekommer i sammanhang som "apparatoberoende uppmärkning", vilket betyder att information med denna uppmärkning kan presenteras på många olika typer av apparater, som webb-läsare, mobiltelefon, PDA, etc.
- attribut [hos ett element]= attribute [of an element]. Egenskapsvärde hos ett element. T.ex. "postnummer-attribut" i ett "adress-element".
- datatypning = data typing. Specifikation av vilka typer av data som t.ex. ett attribut får ha. T.ex. att attributet svenskt postnummer måste vara ett fem-siffrigt heltal.
- element [i ett dokument] = element [in a document]. Ett objekt som i sig är ett dokument eller är en sammanhängande del av ett dokument. T.ex. adress-element, beskrivande adress-information med attribut som postnummer, m.m.
- formatmall = style sheet. Se XSL, XSLT och CSS
- presentationsuppmärkning = presentation markup. Den uppmärkning som ett dokument har då det skall presenteras på någon apparat, såsom en webb-läsare. T.ex. att postnummer skall visas i fetstil.
- strukturuppmärkning = structural markup. Den uppmärkning som beskriver den innehållsliga strukturen av information. T.ex. att en faktura har ett fakturanummer, adressinformation, datum, m.m.
- uppmärkning = mark up. Förekommer i sammanhang som "uppmärkningsspråk" = "markup languages" och "strukturuppmärkning".
- XML-applikation = XML application. Förekommer i sammanhang som "MathML är en XML-applikation", vilket betyder att MathML är ett XML-baserat språk. Skall ej sammanblandas med "applikationsprogram".

12. Akronymer

Följande lista beskriver några tekniska akronymer som refererats i detta dokument, men som inte definierats ovan.

- CSS: Cascading Style Sheets - beskrivning av hur dokument skall presenteras på datorskärm, hur de skall skrivas ut, skall översättas till tal etc. De mest kända standarderna är *CSS1* (W3C Recommendation, december 1996) och *CSS2* (W3C Recommendation, maj 1998). Ett subset av *CSS2*, *CSS Mobile Profile*, avsett för apparater som mobiltelefoner, har nått status W3C Candidate Recommendation i oktober 2001. CSS kan användas för såväl HTML-dokument som för XML-dokument. Mer information: <http://www.w3.org/Style/>. CSS har en speciell syntax, inte XML-baserad. Ett XML-baserat alternativ är XSL (se XSL ovan).
- DTD: Document Type Definition - beskrivning av hur en viss typ av dokument får representeras, i termer av element, attribut och attributvärden. Den typ av DTD som XML understöder skiljer sig i vissa avseenden från DTD:er i SGML. I samband med att XML Schema standardiserats (se ovan) har behovet av DTD:er i XML minskat.
- RDF: Resource Description Framework - ett ramverk för att representera metadata. Det erbjuder interoperabilitet mellan applikationsprogram som utbyter information i XML-format. RDF kan användas i många olika sammanhang, såsom resurs-upptäckt ("resource discovery") vid intelligent sökning, för att beskriva innehåll på enskilda informationsobjekt och relationer mellan objekt samt för att beskriva egenskaper hos samlingar av informationsobjekt. Mer information: <http://www.w3.org/RDF/>.

13. Webb-baserade XML-kurser

<http://www.w3schools.org> - Innehåller ett antal olika kurser inom olika XML-standarder.

<http://www.xfront.com/#schema> - Omfattande introduktion till XML Schema och XSLT.

14. Referenser, hänvisningar, länkar

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

<http://www.w3c.org/> - W3C: s officiella webbplats.

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

<http://w3c.sics.se> - Svenska W3C-kontorets webbplats. Här återfinns bl.a. svenska översättningar av officiella W3C-dokument.

<http://www.dublincore.org/> - Dublin Core Metadata Initiative's officiella webbplats.

<http://www.ws-i.org/> - Web Services Interoperability officiella webbplats.

<http://www.oasis-open.org/> - OASIS officiella webbplats.

<http://www.xml.org/> - Register sponsrat av OASIS, internationellt konsortium som stöder användning av olika standarder. Innehåller både teknikorierad information och branschorienterad information.

<http://xml.coverpages.org/> - Register sponsrat av OASIS, med information om såväl grundläggande XML-standarder som branschorienterade XML-tillämpningar. Här finns en relativt omfattande samling länkar.

http://dmoz.org/Computers/Data_Formats/Markup_Languages/XML/Applications/ - dmoz Open Directory Project, listar ett urval XML-tillämpningar.

<http://www.xmlsoftware.com/> - Översikt över XML-programvaror, privat.

<http://www.kb.se/> - Kungliga Biblioteket, har Dublin Core översatt till svenska och ger bra praktiska hänvisningar.

<http://www.xmlakademin.nu/> - basinformation om XML och ett informellt nätverk för informationsstrateger inom svensk offentlig sektor.

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

<http://it-upphandling.statskontoret.se/Uhw/> - se Statskontorets ramavtal för informationsförsörjning 2002 som bland annat erbjuder sökmotorer och XML-produkter och tjänster.

15. Kontaktperson

Karl Wessbrandt, Statskontoret, tel. 08-454 47 12
e-post: karl.wessbrandt@statskontoret.se