



rockad



The Stockholm Chamber
of e-commerce



INTERACTIVE
INSTITUTE

infobyte

0 | september 2000

Standarden *Namespaces in XML* – en björntjänst! [Stig Berild]

En artikel från Rockad

Stig Berild arbetar för närvarande vid Sveriges Tekniska Attachéers kontor i Silicon Valley med att huvudsakligen för industriforskningsinstitutens SITI och Framkom räkning spana efter, analysera och rapportera om ny teknik och intressanta trender inom ICT. Inte förvånande upptar XML och allt i dess närhet huvuddelen av intresset. Under en mer än 30-årig verksamhet inom IT har även många andra områden genom åren fångat intresset. Bland dem kan nämnas datamodeller, databaser, data warehouse, objekt/ komponent-modellering, repositorer, distribuerad samverkan, e-affärer. Under senhöst-ten återvänder Berild till Sverige och affärsområdet Ecom inom Framkom.

Kontakta redaktionen

Nicklas Lundblad,
Stockholms
handelskammare
nicklas@acm.org

Alex Jonsson,
Interactive Institute
alex.jonsson@interactiveinstitute.se

Per Melander,
Interactive Institute
per.melander@interactiveinstitute.se

PDF- & Webbproduktion

Pär Abrahamsson,
Svenska Infobyte
par@Infobyte.se

Standarden *Namespaces in XML* – en björntjänst!

Av: Stig Berild

Inledning

Denna artikel är indelad i fyra avsnitt. Det första avsnittet lägger grunden genom att på ett allmänt plan beskriva den pågående trenden inom informationsutbyte. Därefter följer en översiktlig introduktion till den av World Wide Web Consortium (W3C) framtagna standarden **Namespaces in XML**. Det egentliga syftet med artikeln är dock att göra en analys av standardens potential och brister, något som avsnitt tre ägnas åt. Det sista avsnittet innehåller några avrundande reflexioner.

Av naturliga skäl är synpunkterna subjektiva. Kanske är de färgade av en viss besvikelse. Förmodligen uppfattas de som lite elaka och med en kategorisk underton. Allt för att väcka intresse, kanske inspirera till eftertanke och vidare kunskapsinhämtning.

'Namespaces in XML' (i fortsättningen kort och gott **Namespaces**) är en av de tre XML-standarder (Recommendations) W3C hittills accepterat. De andra två är 'Extensible Markup Language (XML) 1.0' – förstas - och 'Associating Style Sheets with XML documents Version 1.0'. Namespaces antogs av W3C som Recommendation i januari 1999.

Till skillnad från många andra standarder är den kärnfullt kort, åtta sidor plus ett ickebindande appendix på tre sidor. Inte heller är det svårt att förstå den föreslagna syntaxen. Äntligen något rimligt gripbart, utbrister den luttrade läsaren av standarder.

Däri ligger fallgropen. I alla händelser för den som i standarden ser ett nödvändigt stöd för den dynamiska och globala informationsutbytesvärld vi förmodligen alla i Internets spår kan ana snart fångar oss och som vi behöver kunna hantera. Tyvärr tvingar oss Namespaces snabbt att skilja mellan förhoppning och fakta, mellan /trend och realitet. Att Namespaces är en kontroversiell standard visar den intensiva debatt som omger den.

Låt oss börja med en fri tolkning av den vision kring informationsutbyte som har bäring på det Namespaces försöker lösa och som föresvävar de flesta av oss.

Vision

I den nya framväxande cyberrymden kommer all världens information att utbytas mellan alla tänkbara kombinationer av sändare och mottagare. Viss in-

formation kommer att vara strukturerad, annan i fri form. Information kommer att vara paketerad i alla upptänkliga format och strukturer, med innehåll som kan förväntas vara sammansatt från många källor för många ändamål. Dokument och dokumenttyper kommer och går, förändras kontinuerligt efter nya förutsättningar. Utbytesparterna blir allt fler med hela tiden nya kombinationer partnerskap efter nya behov. Viss information kan hanteras fristående, annan information hänger ihop över mer eller mindre komplexa processer. Informationselementens semantik måste entydigt följa med deras rörelser i cyberrymden. Deras definitioner måste kunna hanteras och förstås på samma sätt i alla de sammanhang där det dyker upp.

Vi ser i XML-världen framväxten av en mängd olika vokabulärer för samma, olika eller delvis överlappande syften. Hypotesen är att dokument i ökad utsträckning kan komma att innehålla element och attribut som har sina definitioner i olika vokabulärer. Dessa dokumentet kan behöva hanteras av flera olika dataprogram där varje dataprogram bara är intresserat av de element och attribut det förstår och finner anledning operera på, d v s arbetar inom sin vokabulär. Då gäller det att kunna identifiera vilka komponenter i dokumentet som tillhör respektive vokabulär. Och gärna göra det på ett för dataprogrammet entydigt sätt. Om så, kan kanske dataprogrammet hantera ett antal olika dokumenttyper så länge som de intressanta delarna är enhetligt identifierbara och har en enhetlig struktur. Därtill helst en över tiden stabil identifiering.

Ett exempel kan vara två olika adressuppgifter i samma dokument. Kanske är de hämtade från olika källor, kanske representerar de olika semantik, kanske har de olika intern struktur. Förmodligen fyller de olika roller i dokumentet. För rätt hantering i de mottagande dataprogrammen måste de entydigt skiljas ut på något sätt. Att använda samma identifiering (märkord) fungerar inte. Är utbytet parvis mellan kända parter och inom en och samma organisation kan man säkert komma överens om att nyttja olika namn (levransadress, fakturaadress) eller på annat sätt indikera olikheter (x:adress, y:adress). Befinner man sig i en mer vidsträckt, kanske globalt orienterad utbytessituation blir det inte lika enkelt. Då måste en generellare lösning till.

En oändlig mängd informationsutbyten av alla de slag med de mest skiftande innehåll. Tala om komplexitet. Är kaos att förvänta? Inte enligt W3C. I alla fall inte om man identifierar informationen enligt principerna i Namespaces. Alternativa ansatser med högre ambitionsnivåer förs också fram från olika håll. Dit hör avancerade begreppsmodeller för att formulera den vokabulär XML-dokumentet hämtar sina märkord från.

Nå, varför alls prata visioner och trender i anslutning till en diskussion om Namespaces? Därför att denna inledning behövs som resonansbotten

att relatera den kommande diskussionen till. De flesta har heller inga svårigheter att relatera visionen till den aktuella utvecklingen runt Internet. Knappt ens en fråga om vision utan om krass nödvändighet för Internets fortsatta expansion som informationsutbyteskanal. Ytterligare ett argument är att just diskrepansen mellan vision och standardens realitet visat sig vara upprinnelse till både missförstånd och i vissa stycken infekterad debatt. Vad är då "realiteten"?

Realitet

Detta avsnitt innehåller en översiktlig beskrivning av Namespace-standarderna.

För korrekt hantering av ett inkommande dokument måste varje ingående element unikt kunna identifieras. Att på två platser finna märkordet *adress* men med olika innebörd är förstås oacceptabelt. Namespaces definierar en lösning som bygger på att åtminstone alla icke entydiga namn förses med ett prefix. Prefixet tillsammans med det lokala namnet ger entydighet, ger ett så kallat kvalificerat namn. Prefix och lokalt namn åtskiljs med kolon. XML-standarderna var förutseende nog, trots att den antogs redan 1998, att avsätta kolon som en karaktär att kunna användas för detta ändamål.

En lösning för dokumentet med två adresser skulle i princip kunna vara att kalla dem för *leverans:adress* respektive *bostad:adress*. Egenpåhittade, lokala lösningar löper dock risken att i något annat sammanhang inte längre ge unikheter. Det är inte otroligt att två parter helt fristående från varandra kommit fram till samma lokala prefix (*leverans*) för egna behov. Förmodligen har de något olika struktur eller är åtminstone baserade på något olika villkor. Kanske den ena skiljer på bostads- och leveransadress medan den andra inte gör det. Antag att dessa parter så småningom börjar samarbeta och önskar utbyta information med andra parter i form av dokument där bådas *leverans:adress* visar sig behövas av någon (outgrundlig) anledning. Återigen namnkonflikt men på en något allmänare nivå. Och så vidare. Lösningen ligger i att hitta ett prefix som garanterar universell unikheter.

Namespaces hittade Uniform Resource Identifier (URI). Den ger normalt universellt unik identifikation av Internetresurser. Något man bedömer ger tillräcklig garanti för att prefixet i kombination med det lokala namnet blir ett garanterat unikt kvalificerat namn.

Enligt standardens intentioner pekar en URI på ett namespace i vilken finns en uppsättning definierade namn ordnade i någon lämplig struktur. Om vi prefixerar *adress* med en viss URI (säg <http://www.detaljhandeln.se/leveranser>) menar vi den definition av *adress* som finns i det namespace URIn refererar till. Betydligt mer precist som

märkord än att endast använda *adress* isolerat. Nu kan vi tyvärr inte direkt kombinera ihop URI och lokalt namn med hjälp av kolon eftersom ett URI kan innehålla karaktärer som inte får förekomma i namn. Vårt '<http://www.detaljhandeln.se/leveranser>:adress' är inte entydigt tolkbart eftersom det innehåller två kolon. Alltså behövs en indirekt deklaration av ett lämpligt namn som kan användas istället för URIn som prefix inom dokumentet. När namnet väl är deklarerat kan det dessutom smidigt användas som prefix för alla andra lokala namn som ingår i eller har att hänföras till samma namespace. Ökar också läsbarheten under förutsättning att prefixnamnet väljs med omsorg. Det behöver ju bara vara unikt inom aktuellt dokument.

Deklarationen kan placeras inom valfritt element på valfri nivå i dokumentet och görs som ett attribut. Det deklarerade namnet har sedan sin giltighet på denna nivå och alla eventuella undernivåer. För vårt exempel skulle det kunna se ut som följer:

```
<order xmlns:leverans = 'http://www.detaljhandeln.se/leveranser'>
.....
</order>
```

Som synes har elementet *order* attributnamnet 'xmlns:leverans' med värdet '<http://www.detaljhandeln.se/leveranser>'. 'xmlns' står som förkortning för 'XML namespace'. Det är ett fixt namn i syntaxen med ett fixt syfte. Därefter kommer ett kolon och till sist det namn som fortsättningsvis inom sitt giltighetsområde ska användas som prefix istället för den fulla URIn. Vi har valt 'leverans' för att det i dokumentet ska framgå att referensen avser en leveransadress. Eftersom även fakturaadress förekommer i dokumentet väljer vi med hjälp av prefixnamnet 'faktura' att referera till det namespace där vi upplever att adress definieras i den rollen, nämligen i '<http://www.ekonomibranschen.se/betalning>'.

```
<order xmlns:leverans = 'http://www.detaljhandeln.se/leveranser'
xmlns:faktura = 'http://www.ekonomibranschen.se/betalning' >
.....
<leverans:adress> Storgatan 12 </leverans:adress>
....
<faktura:adress>
  <faktura:rad1> Box 123 </faktura:rad1>
  <faktura:rad2> 123 45 </faktura:rad2>
  <faktura:rad3> Storstad </faktura:rad3>
</faktura:adress>
.....
</order>
```

Figur 1

Som synes är de båda deklarationerna placerade under elementet *order*. På så vis kan *leverans* och *faktura* användas som prefix även för andra elementnamn hänförliga till respektive namespace som befinner sig på *order*-nivån eller därunder. Är båda namespaces bara aktuella för att unigöra respektive adressuppgift skulle vi lika gärna kunna ha deklarerat enligt figur 2. Observera att prefixet även gäller på den nivå där det deklarerats.

```
<order>
.....
<leverans:adress xmlns:leverans = 'http://www.detaljhandeln.se/leveranser'>
  Storgatan 12
</leverans:adress>
.....
<faktura:adress xmlns:faktura = 'http://www.ekonomibranschen.se/betalning'>
  <faktura:rad1> Box 123 </faktura:rad1>
  <faktura:rad2> 123 45 </faktura:rad2>
  <faktura:rad3> Storstad </faktura:rad3>
</faktura:adress>
.....
</order>
```

Figur 2

Sedan finns det varianter för att öka smidig användning. Anges bara 'xmlns' i namespace-deklarationen, d v s utan kolon och prefixnamn, ska det tolkas som att alla ickeprefixade namn på den nivån och därunder ska omfattas av angivet namespace. Man talar här om ett default namespace, d v s ett som är underförstått och utan explicit namn. Om flera defaultangivelser anges på olika nivåer hänförs ickeprefixade namn till närmast över deklarerat default namespace. Figur 3 nedan har exakt samma tolkning som figur 2 förutom att en landskod härstammande från namespace '<http://www.detaljhandeln.se/leveranser>' tillförts. Allt för att visa att blandning mycket väl kan förekomma.

```

.....
<order xmlns:leverans='http://www.detaljhandeln.se/leveranser'>
.....
  <leverans:adress> Storgatan 12 </leverans:adress>
.....
  <adress xmlns='http://www.ekonomibranschen.se/betalning'>
    <rad1> Box 123 </rad1>
    <rad2> 123 45 </rad2>
    <rad3> Storstad </rad3>
    <leverans:landkod> SE </leverans:landkod>
  </adress>
.....
</order>

```

Figur 3

Finns önskan om att på någon undernivå avbryta defaultmekanismen och övergå till lokala namn deklarerar en ny 'xmlns', denna gång med ett tomt värde (xmlns = ' ').

Märkordet 'order' indikerar att det har sin fulla definition inom dokumentet utan namespacekoppling, alternativt (under förutsättning att det inte representerar översta nivå) kan härstamma till ett namespace som är default-deklarerat på någon övre nivå. Skulle vi föredra att förknippa definitionen med namespace 'leverans' är det enkelt ordnat med hjälp av 'leverans:' framför inledande och avslutande märkord. En lustighet är att prefixet nyttjas redan innan det (i textordning) deklarerats. Helt i sin ordning eftersom giltighetsområdet är den nivå den deklarerats på och därunder.

```

.....
<leverans:order xmlns:leverans='http://www.detaljhandeln.se/leveranser'>
.....
</leverans:order>

```

Figur 4

I princip gäller samma regler för attribut. Figur 5 visar en ny variant på adressangivelsen med användning av attribut. Observera dock att default namespace bara fungerar för elementnamn, inte för ickeprefixerade attributnamn. För attribut gäller antingen lokal definition eller kvalificerat namn.

```
.....  
<order xmlns:bostad ='http://www.mantalsskrivning.se/data'  
.....  
  <adress          bostad:gata="Storgatan 12"  
                  bostad:postnummer="123 45"  
                  bostad:ort="Storstad"  
  />  
  .....  
</order>
```

Figur 5

Namespaces opererar helt inom XMLs Version 1.0 syntax. Två mindre begränsande tolkningar av syntaxen är egentligen det enda som behövs tillföras. Den första anpassningen är att kolon, som i XML 1.0 är vilket tecken som helst, i och med Namespaces blivit en avskiljare mellan prefix och lokalt namn. Den andra anpassningen är att 'xmlns:xxx' i och för sig syntaktiskt är ett attribut men nu får innebörden av en namespace-deklaration. Detta betyder att en parser som inte är anpassad till Namespaces (kanske är den av äldre datum) ändå mycket väl kan tolka ett dokument med namespace-deklarationer, dock utan att kunna förstå de namespace-relaterade konstruktionerna. Baksidan av detta är att en parser som kan tolka namespace-deklarationer feltolkar dokument inom vilka man, utan att känna till Namespaces-standarderna, har valt att blanda i kolon i sina lokala elementnamn.

Inga större konstigheter. Ser bra ut. Vari ligger då problemen?

Synpunkter

Den som främst läser denna artikel för att få en inblick i vad Namespaces är för något må sluta här, dock med tilläggsinformationen att det knappast finns någon Recommendation eller aktivitet inom W3C som rönt så högt debattläge som Namespaces. En sak är klar och det är att det mesta tycks vara oklart eller i varje fall föremål för påtagligt olika uppfattningar såväl hos allmänheten som inom W3C. Knappast en stabil grund att stå på för ett globalt nyttjande.

Resten av artikeln ägnas åt några av de debatterade aspekterna. Synpunkterna är mina egna men i vissa fall inspirerade av debattinlägg. Sakfel förekommer säkert. Åsikter kan vara baserade på missuppfattningar eller bristande kunskap inom vissa av XMLs tillämpningsområden. Kom gärna med rättelser, kommentarer, fakta och/eller synpunkter.

Allmän förvirring

Det tycks finnas två eller tre falanger inom Namespace-sfären. Den ena falangen betonar en krasst teknikfokuserad tolkning av syftet. Det ska vara enkelt för dataprogram att unikt identifiera element och attribut för att undvika namnkonflikter inom dokument och mellan dokument som delvis innehåller överlappande information. Dessutom bör om möjligt standarden syfta till sådan stabilitet vid namngivning att dataprogram inte behöver uppdateras eller anpassas i och för bearbetning av varje nytt dokument som genereras med informationen ifråga. Stabila namn underlättar också återanvändning av samma program inom flera av de tillämpningar informationen återkommer i. Standarden behöver en specifikation av en syntax som garanterar att inte element eller attribut riskerar förväxlas oavsett i vilket dokument det förekommer. Punkt slut.

Den andra falangen representerar en mer semantikorienterad tolkning av syftet. Varje element bör ha ett namn som vägleder till dess innebörd. Namespaces står i deras uppfattning för en uppsättning namn som svarar mot något avgränsat intresseområde och inom detta för ett visst aktuellt syfte. Man vill kunna ställa krav på namespaces såväl på dess ändamål som på dess innehåll. Man vill dessutom att namespaces ska kunna figurera som styrmedel, kravställare när det gäller namngivning, mm.

En tredje grupp ser i ett längre perspektiv namespaces utvidgade till en betydligt rikare beskrivning av de element och attribut som hanteras i dokument. De noterar behovet av en regelrätt begreppsmodell (schema). Förutom en betydligt rikare semantik skulle den helt omfatta den beskrivningsinformation som Namespaces nu lägger in i respektive dokument.

Det enklaste är att gå på syntax-linjen. Kräver inte lika omfattande analys av förutsättningar och behov. Inte heller ställer den några krav på hur namespaces ska kunna opereras. Kvar blir en syntax som garanterar (hmm?) unikheter. Vi vet redan hur standarden ser ut och att det är syntaxlinjen som genomsyrar standarden. Tyvärr med konsekvensen av en i högsta grad inskränkt nytta. Semantisyftet indikeras visserligen något försåtligt svävande i standardens inledande text, men utan att följas upp i den presenterade lösningen.

Sannolikt drevs standarden fram under en kombination av akut behov av någon snabb lösning samt ett dokument- snarare än begreppsmodellperspektiv hos arbetsgruppen. Ingen kunde heller i det läget (ca 1997-98) ana den explosionsartade expansion av formaliserat informationsutbyte vi senare kunnat konstatera, inte minst i anslutning till e-affärer.

Låt oss titta närmare på de tre perspektiven och om/hur deras respektive syften får stöd i standarden.

Syntaxperspektivet

En Namespaces-URI är en sträng uppbyggd enligt syntaxen för en konventionell URI. I Internetvärlden nyttjas URler för att identifiera olika typer av företeelser eller resurser. Där finns med andra ord en naturlig koppling mellan en referens och den resurs den refererar till. I Namespaces bryr man sig inte om denna koppling. En URI kan men behöver inte alls ha någon identifierande relevans i webbrymden. Det enda av vikt är att det är en syntaktiskt korrekt sträng. Och en förhoppning om att de som ska läsa URIn på något sätt kan läsa ut någon tolkning av vad den är tänkt att representera – om någonting.

Standarden deklarerar att användningen av URI som referens till ett namespace garanterar att det kvalificerade namnet är "universally unique" eftersom URI som prefix ger unik referens till ett namespace där det lokala namnet i sin tur bara förekommer en gång. Det är att ta i. Ta URI-referensen. En sträng är visserligen som sträng betraktad definitionsmässigt alltid unik. Självfallet finns det bara en unik sträng med teckenföljden 'http://www.w3.org/TR/1998/REC-xml-19980210' liksom en med 'http://tjo.hoppsan.nu/alltochinget' liksom en med 'tjohoppsan' eller alla andra upptänkliga strängar. Alla tre strängarna är globalt unika eftersom det inte finns två (än mindre fler) strängar med samma teckenföljd som är olika (givet samma inkodningsformat). Men det väsentliga borde vara vilken tolkning som kan göras utifrån den levererade strängen, vad strängen refererar till. Där ger Namespaces inget stöd.

Det enda som URI-kravet tillför i sammanhanget är en syntaktisk begränsning av hur strängen får vara uppbyggd. Där faller 'tjohoppsan' bort medan de övriga två accepteras. Att den andra inte är identifierar en existerande webbresurs spelar ingen roll. Namespaces stipulerar inga sådana krav. Att standarden ändå ställer krav på ett syntaktiskt giltigt URI måste bero på en förhoppning att dessa strängar ger ett bättre informationsvärde än andra mer fritt formulerade strängar och/eller minskar risken att samma sträng kommer till användning av flera för flera olika syften. Den första förhoppningen behöver inte alls vara sann. I de flesta fall skulle ett namn som har relevans och tolkbarhet inom ett intresseområde (de som utbyter information för något syfte) ge betydligt bättre vägledning. Mer om detta i anslutning till namespaceperspektivet. Den senare förhoppningen är däremot säkert korrekt under förutsättning att de som formulerar namespacesdeklarationerna är disciplinerade nog att bara nyttja URler som de har kontroll över och som ger en för dem själva och andra nyttjare tolkbar identifiering.

Obs, att "universally unique" inte behöver innebära globalt intresse, endast intresse utöver ett visst dokument. Oavsett spännvidd ska URI an-

vändas. Är det att gå över ån efter vatten? Kanske kan vi i positiv anda hoppas att arbetsgruppen såg användningen av webbidentifikator (i praktiken ofta URLer) som en bra bas inför ett nästa steg med större fokus på stöd till den skissade trenden.

Nå, är dessa lösliga förutsättningar en tillräcklig grund att etablera professionellt informationsutbyte på? Knappast, på sin höjd en närd förhoppning hos arbetsgruppen. Att låna URI-syntaxen och kalla de strängar syntaxen tillåter för URIer är att vilseleda eftersom så varken är fallet (med säkerhet) eller syftet, än mindre någon kunskap som kan nyttiggöras. Kanske refererar URIn till ett namespace i någon form, kanske något annat, kanske ingenting. I den mån URIn verkligen är en existerande referens har vi ingen aning om vad den refererar till, i alla händelser inte den som inte tillhört den gemenskap som formulerat URIn. Dock inget större bekymmer i syntaxperspektivet eftersom det är strängen som är det viktiga, inte vad den står för ☺.

En annan fundering: Behövs verkligen prefix med namespacesrelation för attribut? Inte om man ska tolka innebörden av attributbegreppet som i databassammanhang. Där är alltid attributet bundet till ett visst element. Är elementet unikt är alltså även attributet unikt med automatik. Om, å andra sidan, attribut i XML-dokument inte tänks ha någonting att göra med avbildning av ett intresseområde i en modell, utan beskriver syntaktiska, metadata-relaterade och andra omständigheter, finns ingen egentlig elementrelation annat än den syntaktiska närhet som attributets "verksamhetsområde" kräver. Det naturliga vore i så fall att element och attribut konsekvent kvalificeras med hjälp av helt separata namespaces. Bekymret har sin grund i den oklara betydelsen av attributbegreppet i XML.

Ytterligare en allmän fundering: Vilka behov tillgodoser och vilken service ger syntaxperspektivet XML-nyttjare? I de fall en tillämpning accepterar vilka dokument som helst för någon generell åtgärd, exempelvis för presentation av mångfacetterat innehåll ur många dokument enligt vissa bestämda generella regler, kan standarden mycket väl fylla en roll. Med hänvisning till tidigare exempel; oavsett i vilket dokument *adress* i betydelsen *leverans:adress* förekommer (eller mer strikt, oavsett i vilket dokument vi finner den URI som *leverans* representerar som prefix i anslutning till *adress*) ska viss font användas för presentationen. Dyker *faktura:adress* upp gäller en annan font och kursiverat, och så vidare.

Förutsättningen bygger på att typ av dokument är av underordnat intresse och att eventuellt länkad Document Type Definition (DTD) inte ger någon tilläggsinformation av värde. Förutsättningar som förmodligen har sin giltighet i HTML-orienterat dokumentutbyte för enhetliga presentationer i webbläsare eller för enhetlig redigering inför annan typ av exponering.

Mycket pekar dock på att XML i betydligt större utsträckning kommer att tillämpas för utbyte av formaliserade data mellan dataprogram. Det vill säga i miljöer där märkordets semantiska tolkning (tillsammans med dokumentets logiska struktur) är ett primärt krav snarare än bara indirekt som någon form av indikator för andra behov. Denna miljö bygger på förutsättningen att mottagare och avsändare önskar utbyta information för något avgränsat syfte. För ändamålet har de kommit överens om såväl dokumenttyper och deras strukturella uppbyggnad som innebörden av de märkord som ska sättas att representera de ingående elementtyperna. Inom givna syften etableras vokabulärer. Inom en vokabulär är vissa märkord lokala, andra är "lånade" från andra vokabulärer, ytterligare andra kan ha nått ett allmänt erkänt status som global de facto standard. En standard måste erbjuda anpassningsförmåga och nyanskraft.

Avslutningsvis, glöm garantier om unikheter. Här finns inga utpekade källor som beskriver vilka globalt eller regionalt eller på annat sätt "auktorisera" URler som finns tillgängliga (kanske dags för ett speciellt namespace för ändamålet ☺), än mindre regler för hur URler får överenskommas, definieras och användas. Vem skulle stå som garant för dem i så fall? Ingen garanti för att en URI pekar på en existerande webbsida, ingen garanti för att någon annan inte använder samma URI för något annat ändamål. Vi kan konstatera att namespacesdeklarationen endast med säkerhet kan nyttjas för att undanröja namnkonflikter inom dokument med hjälp av ett prefix refererande till en sträng uppbyggd enligt URI-syntax. Förhoppningsvis står strängen hos läsare (människa, program) för samma innebörd oavsett vilket dokument det förekommer i.

Namespaceperspektivet

Standarden är som en godispåse man lockar med men som rycks undan när någon försöker komma närmare. Redan titeln på standarden "Namespaces in XML" sätter hjärncellerna i rörelse mot att det är fråga om en utvidgning av XML med ett namespaceperspektiv. Titeln är godispåsen tillsammans med ett antal inledande fraser som talar om namespace som en "collection of names, identified by a URI reference", att det har en "internal structure", mm. Undanryckningen kommer när det snabbt visar sig att syntaxperspektivet ovan i realiteten är det enda som erbjuds, när det totalt saknas mekanismer för att tolka, nyttja och operera på ett namespace, än mindre garantera dess existens. Standarden säger A men menar B.

Icke förty, är det många som i standarden tycker sig se existensen av reellt refererade namespaces antingen explicit eller åtminstone som koncept. Dessa förhoppningsfulla är egentligen inte lurade in på detta tänkande.

De kanske inte ens tycker om godis. De drivs av en tro att namespaces är viktiga i sammanhanget, att de fyller en viktig roll för informationsutbyte i ett globalt perspektiv. Man vill på något sätt kunna deklarerat och få garantier för att det lokala namnet befinner sig inom det namespace prefixet indikerar. Eftersom ett namespace i detta perspektiv svarar mot ett syfte som motiverar dess avgränsning, tillför dess strängreferens dessutom en gnutta syftesrelaterad semantik.

Vad är då att säga om namespaceperspektivet? Över till några infallsvinklar. Antag att URIn refererar stabilt till visst namespace. Vad vet vi om dess innehåll? Ingenting! Är det en fil med en uppsättning unika namn listade eller är det en löpande text i vilken ett antal namn förklaras till sin innebörd? Är det månne en DTD eller ett XML Schema? Kanske bara en uppsättning länkar till en uppsättning andra webbresurser där respektive namn finns noterat och förklarat? Kanske en begreppsmodell uttryckt i UML? Eller? I den ickebindande delen av Namespaces anser man att det borde vara en "logical structure" av någon anledning. Månne en dokumentinfluerad syn på begreppsmodellering?

Även om vi på något sätt (kanske genom att fråga dokumentavsändaren) skulle kunna komma underfund med att en viss URI svarar mot någon specifik kategori enligt ovan, gäller total avsaknad av ett definierat gränssnitt för att operera på namespace ifråga (leta namn, leta definition, leta namnkontext, ...). Hur ska vi standardmässigt kunna validera lokala namn gentemot refererad namespace utan gränssnittsstandard? Här skulle antingen behövas ett standardiserat gränssnitt per typ av namespace alternativt helt rått ett krav att endast en viss typ är acceptabel i namespace-rolen. Vem kan tro att någon sådan standard skulle komma upp i XML-rabatten i närtid?

URler uttrycker någon tolkning. Den pekar inte rakt ut i cyberrymden. Men vilka formulerar URIn, ansätter tolkningen? Behöver vi inträde till hemliga klubben som inom sig hittat på URler de förstår och i bästa fall definierat? Kanske för vår räkning? Lite som XML-vokabulärer för olika intresseområden? Ser vi möjligen ett antal internationella överhögheter (läs standardiseringsorgan) växa fram som för den informationsutbytande globala gemenskapen självutnämnt deklarerar ett antal sanktionerade namespaces och deras URler? Vilka garantier kan man tro att de skulle lämna för stabilt innehåll och tolkning? Inga alls förmodligen. Hur får vi reda på eventuella uppdateringar? Här finns ingen koppling mellan dem som underhåller/ansvarar för ett namespace och dem som refererar till detta namespace. Vilket värde har en URI-referens till ett namespace vars innehåll vi inte fullt vågar lita på? En indikation på att intresseområdesanpassade namespaces som avtalas mellan utbytande parter är det enda stabilt hållbara. Låt oss därutöver hoppas

att namespaces växer fram i en kontinuerlig dynamik inom dessa olika intresseområden genom konsensus och efter behov och omständigheter. I så fall ligger återigen intresseområdesanpassade namn nära till hands istället för URler.

Ett annat bekymmer med URler; 'http://www.w3.org/TR/1998/REC-xml-19980210' och 'http://www.w3.org/TR/1998/REC-xml-19980210.ps' är två olika URler refererande till två olika namespaces. Vilket inte är sant eftersom de båda refererar till XML 1.0-standarden. Inte i samma textformat men väl vad avser innehåll - till de namn som där finns definierade. Är två olika inkodningsformat av samma uppsättning namn att betrakta som två olika namespaces? Ska vikt läggas på vad respektive namn i namespace representerar eller hur det uttrycks? En viktig frågeställning som tycks ha hamnat "mellan stolarna".

Om vi antar att innebörd är det viktiga, kan någon invända, att en och samma webblesurs mycket väl ha flera identifierare. Gott och väl, men har vi inte då tappat lite av vitsen med URIn eftersom den som letar efter element med namn ur visst namespace måste vara beredd på att reagera på ett antal olika URler som alla pekar på samma webbobjekt, d v s samma abstrakta namespace? I det generella fallet finns inget 1:1-samband mellan namespace och URI. Ett antal URler kan peka på ett och samma namespace. En viss URI kan, som en följd av deklaration inom olika intresseområden, genom okunskap eller missförstånd, med mera peka på olika namespaces. Suck!

På tal om intresseområdesanpassade namn kontra URler; Varför inte använda namn som ändå för något annat ändamål utnyttjas som unik identifikation på en resurs, till exempel 'REC-xml-19980210' (jmf ovan). Universellt unik eller ej, räcker den säkert som rimligt unik identifikation, dessutom med god sannolikhet tolkbar för dem som har att operera på dokument med detta namespace. Ett annat exempel är ISO-standarder. Varför inte referera till dem med hjälp av deras unika identifikation, exempelvis 'ISO/IEC 10027:1990'. Identifikationen är säkert välkänd för dem som har intresse av det tema den representerar, d v s för dem som kan tänkas bli mottagare av ett dokument där referensen förekommer. I alla händelser lika entydigt och välkänt som en URI. Varför inte använda ett ISBN-nummer för tryckta alster? Alla namespaces finns inte på Internet. Och så vidare.

Finessen med intresseområdesbaserade namn är att de står som en stabil abstrakt referens till ett objekt (exempelvis ett namespace) oavsett var det befinner sig eller vem som opererar på det. I de allra flesta exempel kring Namespaces tycks URler för övrigt uttryckas som URler(L=Locator), d v s med angivande av adress, vilket inte kan anses speciellt stabilt. Namespacefilen kan ju av olika skäl över tiden komma att flyttas till nya platser med

nya URLer till följd. Knappast en glädjande utsikt för alla de dataprogram som förlitat sig på URIns beständiga giltighet. (Ok, URLer är inget krav men ligger tydligen nära till hands.)

Observera att namespacedeklarationer inte kan göras i DTD för giltighet i de dokument som svarar mot denna DTD. Deklarationer förekommer endast i respektive dokument. Däremot måste DTD se till att xmlns-attribut finns deklarerade i DTDn för de element i dokumentet där de kommer att förekomma. DTDn måste också deklarerat de prefixade element- och attributnamn där de kommer att förekomma i dokumentet. Dock ser DTD ett prefixat namn som vilket namn som helst. Deklarationen av elementtypen leverans:adress kunde i DTDs ögon lika gärna ha varit leveransadress. Sträng som sträng. Att det i ett visst dokument innebär en avgörande skillnad ligger utanför DTDs intresseområde. På DTD-nivå finns alltså ingen kännedom om Namespaces. Ändå måste man nogsamt se till att deklARATIONERNA görs med tanke på den eventuella användningen av Namespaces-syntax i de individuella dokumenten som svarar mot DTDn. Inte speciellt upplyftande för utvecklare av XML-tillämpningar.

Sammanfattningsvis; att i standarden tolka in existensen av namespaces artikulerar avgränsningar och syften. Dock saknas helt valideringsmöjligheter. Inte heller bidrar bara namn, oavsett om de är lokala eller kvalificerade, till någon egentlig kunskap om vad namnet representerar. Därtill behövs betydligt utförligare beskrivningar. Användningen av URI som referensmekanism har sin uppsättning problem. Vi är lite bättre rustade än i syntaxperspektivet, men ack så lite. Vi får heller inte glömma att namespaceperspektivet är ett önsketänkande som inte har någon resonans i standarden.

Över till ännu vidlyftigare drömmar: Varför inte en gång för alla flytta definitionen av elementtyper och attributtyper dit där de hör hemma – till schemat, i alla händelser för alla de dokument som har referens till ett schema? Något som i de flesta fall är regel inom e-affärer och databasbase-
rade informationsutbyten.

Schemaperspektivet

Ett namespace är inget schema, det innehåller på sin höjd en unik uppsättning namn. Förhoppningsvis har det funnits någon idé bakom grupperingen av dem under samma namespace. Ingenting tvingar däremot vart och ett att referera till någon bakomliggande unik företeelse (objekt, entitet, aktivitet, händelse, ...). Två eller flera olika namn ur samma namespace kan mycket väl referera till samma företeelse. På samma sätt kan det tänkas, i alla fall i teorin, att två företeelser delar på samma namn. Ett schema erbju-

der rikare möjligheter att skilja på ett namn och vad det refererar till samt kan på andra sätt uttrycka semantik för ökad förståelse. Dit kan höra referens till externa namespaces, eller bättre, till andra scheman. De som från datamodellerings- eller databashållet kommit in i XML-världen tenderar att alldeles naturligt tänka i dessa banor.

Vi återkommer i en separat rapport om och hur den aktuella XML Schema-aktiviteten inom W3C företräder en sådan filosofi. Låt oss bara kort notera ett par saker i anslutning till schemaperspektivet.

I de flesta sammanhang där XML förväntas komma till användning har varje typ av dokument sitt givna syfte varför det också bearbetas av en specifikt avpassad rutin. I denna lokala värld arbetar man sannolikt också med bekräftade (valid) dokument, d v s sådana vars struktur och övriga villkor är definierade genom en DTD eller ett schema. Ett schema är i dessa sammanhang en naturlig ledsagare för den som vill förstå ett dokumentets innehåll.

Varje intresseområde har sin uppsättning begrepp, sin begreppsmodell ("vertikal" giltighet). Därtill kommer generellt överenskomna begrepp som bedöms ha en giltighet över ett antal intresseområden ("horisontell" giltighet). Erfarenhet har visat (se exempelvis motsvarande aktiviteter inom Object Management Group-OMG) att det är ett mycket tungt, ofta omöjligt arbete att ta fram den senare typen. Begrepp har sällan en bred intresseområdesneutral giltighet. När man inom OMG arbetade med att definiera *adress* (för övrigt ofta använt som exempel) – en vad man från början trodde vara ganska rättfram aktivitet - visade det sig att det fanns ett otal infallsvinklar på *adress*, både av generell karaktär och kopplat till olika syften. Med syften som ofrånkomliga styrinstrument för tänkandet föds snart mer precisa syftesrelaterade begrepp som *leveransadress*, *positionsreferens*, *postadress*, I de fall man lyckas standardisera några generella begrepp har det ofta föregåtts av ett kompromissande som resulterat i en urvattnad, profillös begreppsdefinition till ingens glädje, alternativt att någons definition genom intensivt lobbyarbete till slut vinner gehör. Att namespaces tycks fokusera på den horisontella kategorin pekar dess användning av universella URI-referenser på. Ett alternativt primärt fokus på intresseområde ökar markant möjligheten att såväl finna, få acceptans för och få ett brett utnyttjande av de för intresseområdet relevanta begreppen..

Helt kort en infallsvinkel på informationsutbyte i ett konventionellt begreppsmodellperspektiv: Refererar man till ett begrepp i en begreppsmodell pekar man även indirekt på dess omgivning, de krav som modellen fastställer om begreppets relateringar till sin omvärld. I datamodellsammanhang kallas detta för integrity constraints. Ta exempelvis *Person* i en begreppsmodell

där det bland villkoren (hämtade från det aktuella intresseområdet) framgår att *Person* måste ha *Adress* som i sin tur måste ha en *C/O*-angivelse som komplement till *gatuadress* samt ett *postnummer* som börjar på 19. Att vi vet det beror på att vi känner till intresseområdet. Dessutom är *Person* relaterad till *Företag* genom *anställd i*, där relateringen måste finnas till exakt två olika företag. I realistiska sammanhang vore ytterligare ett antal villkor knappast osannolika.

Hur hanteras denna semantik i anslutning till en namn i ett namespace? Inte alls! Är inte det viktig information, både för att formulera ett dokument korrekt och för att tyda det? Självfallet. Saknas denna information är det stor risk att avsändare och mottagare gör olika tolkning av ett dokument innehåll. De kanske ser intresseområdet med något olika ögon och har ingen chans att jämkna dessa i och för informationsutbyte. Är det överhuvudtaget rimligt att under sådana förutsättningar utbyta information? Bygger inte informationsutbyte på övertygelsen om samma uppfattning eller åtminstone samma tolkningskunskap om den information som utbyts? Är det inte rimligt att tänka sig att en informationsutbytesöverenskommelse (kontrakt) refererar till en sammanhållen semantisk deklaration - d v s en sammanhållen begreppsmodell vars syfte är att reglera just det avsedda informationsutbytet. Behövs inte alla kort på bordet? Absolut!

Har manne XML Schema tagit upp denna stafettpinne? Kommer det att behövas ett rikare modelleringspråk, kanske i grunden en ny filosofi? Vi återkommer!

Till sist

Namespaces är en halvmesyr. Antagligen har W3C haft bråttom. Möjligtvis har man utgått från en bristfällig analys av syftet med följd att standarden helt enkelt skjuter bredvid målet. I alla händelser tycks man ha underlåtit att först diskutera fram en gemensam idébas för informationsutbyte i cyberrymden. Vilka typiska utbytessituationer kan identifieras – idag och i ett längre perspektiv? Inom vilka av dessa kan Namespacesstandarderna komma till undsättning? Vilka kräver en rikare kapacitet? Vilka egenskaper bör denna kapacitet erbjuda? Med flera frågeställningar. Utan en sådan idébas blir resultatet ett hopkok som lämnar fältet öppet för tolkningar och alternativa ansatser.

För närvarande gäller kort och gott syntaxperspektivet. Tänk bort namespaces som något nödvändigtvis existerande. Glöm visionen om enhetlig semantisk definition av begrepp.

Den kontinuerligt pågående diskussionen – även inom W3C - om syfte med och tolkning av standarden är i sig en klar indikation om potentiella

problem kring dess utformning och tillämpning. Om en standard efter det att den antagits behöver kompletteras med ett antal klarlägganden och dessa därefter ger upphov till nya klarlägganden i form av motsatta ståndpunkter o s v i en mer eller mindre infekterad debatt, är det antingen ett tecken på en felaktig lösning, en ofullständig lösning eller i bästa fall på en undermålig förklaring av lösningen.

Här finns med andra ord utrymme för en genomgripande översyn utmynnande i en betydligt mer klagörande version 2 av standarden. Många skulle prisa ett initiativ som tar fram en pedagogisk översikt inklusive argumentation kring bakgrund, syfte inklusive ett antal väl valda exempel. En del har föreslagit att standarden lämpligen bakas in i XML Schema-standard, något som dock inte ser ut att bli fallet. Andra ser det som naturligt att Namespaces blir en del av nästa version av XML. Framtiden får utvisa.

Det är dags för ett större inflytande för dem som traditionsenligt arbetat inom databas- eller /begreppsmodellområdet. Förmodligen togs Namespaces-standard fram vid en tid då man ännu inte var genomsyrad av vikten av att operera med bekräftade (valid) dokument, d v s dokument vars definition återfinns i ett schema (DTD). Kanske såg man fortfarande framför sig endast fristående dokument flygande över världen? Finns man fortfarande ett HTML-tänkande i botten med webbläsare som måste vara kapabla att hantera allsköns inkommande dokument, med märkord som mer syftar till för tekniskt stöd än som tolkningsstöd mellan utbytande parter. Hur ofta kommer XML-dokument att existera på dessa fria grunder? Inte ofta, i alla händelser inte i den värld där XML just nu gör sin starkaste inbrytning, nämligen inom e-affärer.

Kanske fanns någonstans ett vidare perspektiv som i sin tur ledde fram till URI som prefix. Till ett schemaperspektiv nådde man inte. På gott och ont. Med den begränsade ambitionsnivån kunde en standard snabbt tas fram, dock med ett begränsat värde. En ambitiösare ansats skulle tagit betydligt längre tid, vara mer kontroversiell, förmodligen fortfarande befinna sig som Working Draft. Dock skulle en sådan standard när den väl tas lägga ett betydligt mer avancerat perspektiv på informationsutbyte och därmed etablera en utbytesplattform som skulle erbjuda en dignitet större nyttovärde för den informationsutbytande gemenskapen.

Ligger lösningen manne hos den med spänning emotsedda XML Schema-standard? Vi hoppas få återkomma!

