

# Att bygga en nyttig nationell infostruktur

Utredning gjord av Metamatrix AB i maj 2001 för Statskontoret.

Pär Lannerö <[par.lannero@metamatrix.se](mailto:par.lannero@metamatrix.se)>

Johan Groth <[johan.groth@metamatrix.se](mailto:johan.groth@metamatrix.se)>

## Sammanfattning

Statskontoret har inför utredningen "Baskrav på infrastruktur för e-förvaltning" målat upp ett antal önskade tillämpningar och stödtjänster för "e-Sverige". Till exempel e-handläggning av ärendeprocesser. Det är uppenbart att här behövs en gemensam och öppen teknisk och informationsmässig bas, för att inte låsas till en enskild leverantörs system eller mängder av dyra konverteringstjänster. Man har identifierat Internet-teknik och XML som en möjlig sådan gemensam bas.

Detta dokument bekräftar den slutsatsen, presenterar XML-familjens medlemmar litet närmare och räknar sedan upp ytterligare ett antal strategiska komponenter som bör ingå i den gemensamma basen för att e-förvaltning ska bli verklighet:

- Tekniska standarder och specifikationer
- Scheman
- Ordlistor
- Verktyg
- Styrning och juridik
- Säkerhet
- Beständighet
- Referensdata
- Medvetenhet och användning

För var och en av dessa påbörjas en kartläggning: vad finns idag, vad behöver göras för att få resterande pusselbitar att falla på plats. Några konkreta förslag lämnas, men flera av områdena behöver en noggrannare genomlysning innan det går att avgöra vad som måste göras, och vem som bör göra det.

## Bakgrund

Under de senaste åren har en snabb utbyggnad skett av den tekniska, eller "hårda", infrastrukturen inom IT-området. På en i många avseenden kort tid har nästan alla företag, myndigheter, organisationer och medborgare i Sverige fått tillgång till Internet och andra digitala media. Därigenom har förutsättningarna för att hantera, sprida och lagra information och data på ett genomgripande sätt förändrats i samhället. Bara det faktum att information överhuvudtaget finns

tillgänglig i digital form har skapat nya tjänster och användningsområden. De satsningar som nu planeras och genomförs inom den "hårda" infrastrukturen (framför allt utbyggnaden av bredband) kommer att göra än mer information tillgänglig på nätet.

Samtidigt som mer och mer information blir tillgänglig ökar insikten om att blotta förekomsten av data på nätet inte utgör en tillräcklig grund för att utveckla framtidens digitala tjänster. Dessa tjänster, som växer fram ur verkliga behov hos användare och informationsproducenter, kräver att en "mjuk" infrastruktur - en informationsstruktur - skapas. Den "mjuka" infrastrukturen är för informationshantering vad Internet var för informationsspridning och kommunikation - en teknik som skapar nya förutsättningar vilket leder till nya tjänster.

Frågan blir alltså: "Hur bör den mjuka infrastrukturen - infostrukturen - utformas för att öka samhällets service gentemot medborgarna, spara pengar hos myndigheter, och ge industrin bästa tänkbara förutsättningar?". Nedan sammanställs ett antal frågeområden och ett antal förslag till insatser som kan föra Sverige in i "tredje vågens IT-användning".

## Situationen idag

Bredband utan infostruktur är som motorvägsnät när det bara finns häst och vagn. Ingen kan utnyttja nätet tillnärmelsevis så effektivt som dess kapacitet egentligen medger.

Följande uppräknig beskriver tyvärr dagsläget vad gäller informationen på Internet.

### **Begränsade sökmöjligheter**

Omöjligt sortera sökresultat respektive göra urval med avseende på viktigt informationsinnehåll i dokumenten (prisuppgifter, klockslag, postnummer, efternamn etc).

### **Informationen instängd i databaser**

Väldigt mycket information är instängd i databaser som i och för sig är öppna för mänsklig interaktion och sökning, men som inte kan indexeras av sökmotorer och därmed bara upptäcks av den som söker upp och manuellt loggar in på databasen.

### **Informationen låst till en form**

Det verkar förutsättas att målgruppen är en seende personer med 17-tumsskärm, PC med Windows, när det egentligen finns mängder av synskadade, folk med handdator, folk som vill skriva ut på papper, och framförallt elektroniska agenter (sökmotorer mm.)

## Ofullständighet

Många glömmar att ange viktig information, eftersom fokus ligger på formgivning.

## Låsning vid specialprogram

Ofta krävs en specifik visningsprogramvara (t.ex. Macromedia Flash) för att kunna tillägna sig informationen.

## Lagringsproblem

Dålig beständighet eftersom visningsmöjligheterna och därmed även dokumentstrukturen hela tiden förändras.

## Dålig maskinläsbarhet

Avsaknad av gemensamma maskinläsbara ordlistor gör det omöjligt att upprätta hantering av information utan dyr och långsam mänsklig bearbetning.

Allt ovanstående gäller i allra högsta grad myndigheternas information, även om det pågår lovvärda projekt som exempelvis sökmotorn på Sverige Direkt [[SD](#)].

## Önskvärd situation

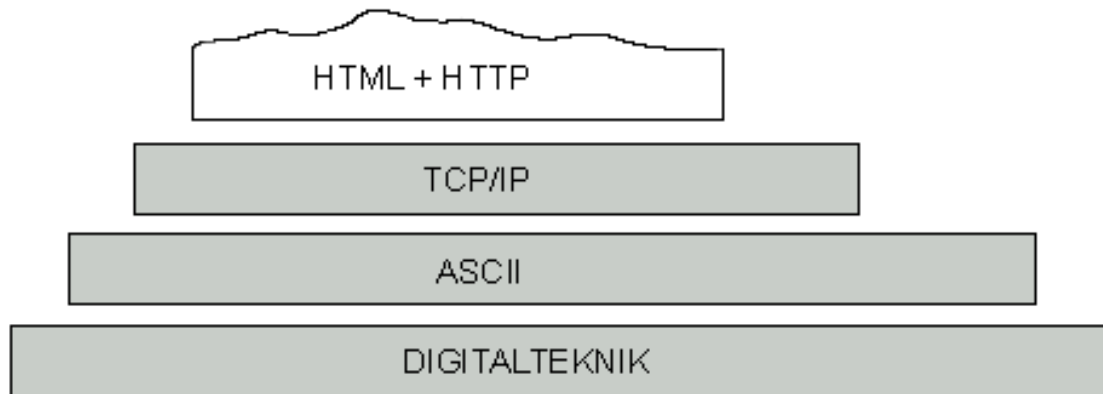
Det finns politiska målsättningar [[SamInfo](#), [Access](#)] att göra myndigheterna tillgängliga för alla medborgare i maximal utsträckning, exempelvis talar man om 24-timmarsmyndighet, gemensam förvaltningsportal etc. Ambitionen är att det skall vara enkelt att hitta informationen, eventuell nödvändig samverkan mellan myndigheter ska ske transparent för medborgaren, och naturligtvis ska man erbjuda en så bra service som möjligt för människor med speciella behov (t.ex. synskadade och språkliga minoriteter). Lika självklart är att säkerställa individens integritet.

Lösningen är infostruktur, dvs. överenskommelser om gemensamma scheman, protokoll, terminologier etc. Den mest realistiska konkreta tekniska plattformen för infostruktur heter XML. Att behovet av infostruktur är starkt understryks av flera statliga rapporter, till exempel IT-kommissionens skrivelse till Regeringen "Om behovet av insatser för utvecklingen av samhällets informationsinfrastruktur" [[ITKbrev](#)] och "SOU 2000:123 Hur blir en ny infostruktur motorn i e-Sverige" [[HEAR](#)].

## XML är en bra plattform

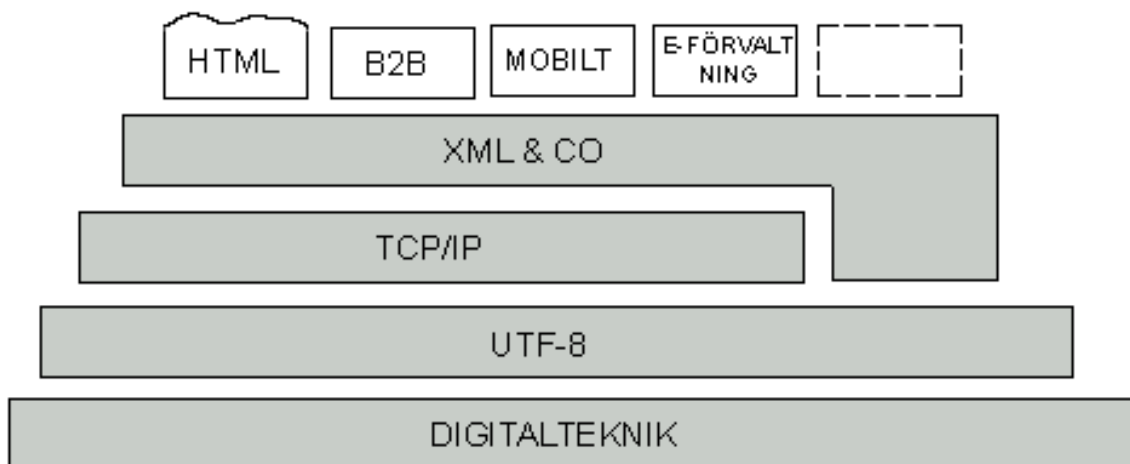
Databranschen går med raska steg mot XML (se t.ex. [[OIXML](#)]). Det är inte alltför djärvt vågat att påstå att XML är mer än bara ännu en förkortning i raden av "hypade" modeord. XML, och den familj av tekniska ideer och metoder som flockas kring denna relativt nya standard [[XML10](#)], är i själva verket ett mycket

välbehövt lyft av den gemensamma plattform som all kommunikation över Internet bygger på. Internet-revolutionen var ju inte riktigt planerad, utan den växte fram organiskt, baserad på några stabila grundläggande tekniska plattformar (TCP/IP, ASCII) och några enkla men också bristfälliga applikationsprotokoll (HTTP+HTML, SMTP m.fl.).



Bristerna i denna oerhört framgångsrika plattform medför bland annat de problem som nämnts ovan. XML är konstruerat för att lösa dem.

XML-familjen lyfter upp Internetplattformen till en nivå som länge varit självklar hos isolerade databssystem: ett standardiserat sätt att på ett strukturerat sätt lagra information.



## XML &co

Många säger redan idag att deras programvaror är "XML-anpassade", men alltför ofta menar de endast att det går att spara ner information på ett filformat med många "krokodilgap" ("<" och ">"). Risken är stor att de stirrat sig blinda på den (i och för sig. centrala) specifikation som heter XML 1.0 och missat att till

denna hör en hel familj av nya tekniska specifikationer, som *tillsammans* utgör den grundplatta vi har valt att kalla "XML & co":

För en närmare presentation av XML & co, se [bilaga 1](#).

## XMLism

XML bör alltså inte ses som en ersättning för HTML, utan som fundamentet i en ny teknikgeneration som kan komma att successivt ersätta HTML, SQL, RDBMS, många kommunikationsprotokoll och APIer, vissa filformat för lagring och dokumentation, osv. Kort och gott en ny era för informationstekniken, som kan realiseras tack vare Internet.

Möjligheter med XML & co:

- Gör att man kan söka med precision på Internet, och sortera resultatet automatiskt.
- Gör att man automatiskt kan publicera i många olika medier m.h.a. mallar.
- Gör informationen tillgänglig för syn- och hörselskadade och andra med speciella behov.
- Frigör från beroende av enskilda programvaruleverantörer.
- XML är ett framtidssäkert format lämpligt för långtidslagring.
- Stödjer strukturering av information och användning av metadata.
- Elimineras upprepade inmatning av samma information.
- Gör det billigt att bygga moduler till programvaror (t.ex. databaser) som gör att de kan kommunicera med andra programvaror oberoende av fabrikat och operativsystem.
- Möjliggör flöde av viktig information inom branscher, utan alltför stora kostnader (T.ex. rss, SKIcal. Jfr ebXML <-> EDI).

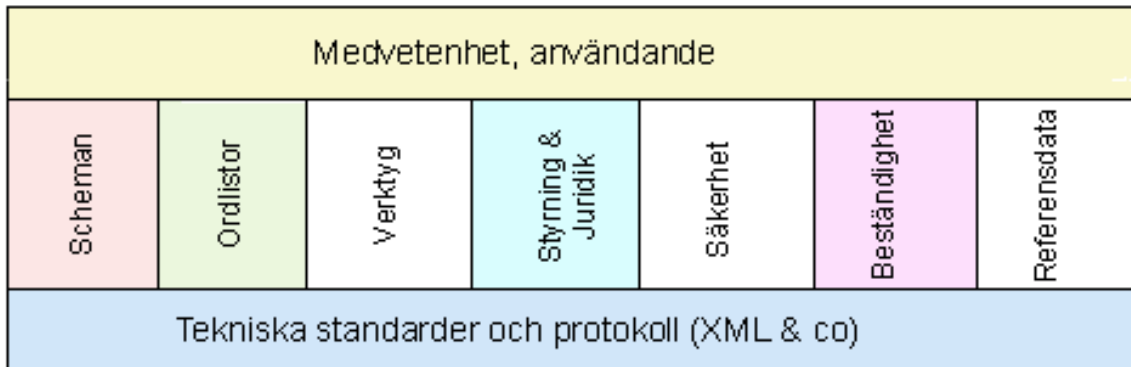
Svagheter och risker med XML & co:

- Osäkert vilka delar av XML-familjen som kommer att slå igenom.
- Det saknas gemensamma terminologier (kan åtgärdas, se nedan).
- Det saknas kunskap på många håll (kan åtgärdas, se nedan).
- Risk för att alla bygger sitt eget schema (kan åtgärdas, se nedan).

**Sammanfattningsvis: informationen blir starkare, och inte längre bunden till något operativsystem, någon applikation eller någon viss situation. Den står på egna ben. Denna filosofi kallas ibland för XMLism [\[GF\]](#).**

## XML räcker inte

XML ger goda förutsättningar för ett välstrukturerat, billigt och användbart flöde av information inom och mellan organisationer. Men det är många pusselbitar som måste komma på plats innan det hela blir verklighet!



Följande avsnitt beskriver de viktigaste pusselbitarna. Några av dem är redan idag någorlunda stabila (framförallt de tekniska specifikationerna i form av W3C-rekommendationer) medan andra helt återstår att ta tag i. Anledningen är att basen, XML & co är såpass ny att man helt enkelt inte hunnit med.

## Tekniska standarder och protokoll

XML-familjen och andra Internetstandarder (i praktiken rekommendationer från W3C [\[W3C\]](#) och RFC:er från IETF [\[IETF\]](#)) är som sagt en bra grundplattform. Ska man vara krass kan man konstatera att den svenska förvaltningen egentligen inte har så mycket att välja på om man vill bygga upp en välfungerande infostruktur. Man måste helt enkelt använda den teknik som övriga världen jobbar med. Men fördelarna är många, förutom de rent tekniska argumenten som summerats ovan:

- Specifikationerna finns öppet publicerade på Internet.
- De är oftast fria från patent- och licenskostnader.
- De har oftast mycket bred, global användning.
- En stor mängd standardprogramvaror följer redan specifikationerna.
- De är stabila tack vare att standardiseringen bygger på interoperabilitet och "running code" och inte på politiska kompromisser.

Något som kan vara svårt är att avgöra när man ska börja använda en ny specifikation. I och med att utvecklingen går så snabbt frestas många att hoppa på det senaste, även innan specifikationen kommit till en första stabil version.

I England har man inom "the e-GIF framework" [\[e-GIF\]](#) listat vilka Internetstandarder man vill att myndigheterna ska använda. Detta är ett bra

initiativ och det urval man gjort ser ganska vettigt ut. Ett problem är dock att det är svårt att hålla en sådan lista uppdaterad, och det kan ha en begränsande effekt om myndigheterna inte själva får utforska nytillkomna möjligheter.

**Förslag: Publicera riktlinjer för i vilka sammanhang det är lämpligt att börja använda Internetstandarder på olika nivåer.**

## Gemensamma scheman och begreppsmodeller

I XML-världen har man tidigare talat om DTD (Document Type Definitions). Nu talar man mest om XML Schemas som är ett rikare sätt att beskriva sina dokument- och informationsstrukturer. Ett schema beskriver vilken information som kan ingå i ett dokument (t.ex. avsändare, klockslag, rubrik och huvudtext i ett brev) och vilken typ av data varje informationselement kan ha (t.ex. är det skillnad på rubrik, som bör vara ett textvärde, och klockslag som har ett standardiserat format bestående av siffror och skiljetecken: "22:04:59"). Inom varje bransch behöver man ta fram scheman för den information man skickar mellan de olika aktörerna. Exempel på några tänkbara scheman är "recept", "faktura", "besiktningsrapport", "personakt" och kanske "lagfartsbevis".

### Vilka scheman finns idag?

De mest framsynta av branschorganisationer och regeringar har redan sett till att enas om datamodeller och scheman som man behöver för att kunna kommunicera effektivt. Här är några av de mest uppmärksammade:

- [ebXML](#) - Betydande och omfattande FN-projekt för elektronisk handel, kan komma att ta över efter EDI
- [NewsML](#) - Initiativ från AP för förmedling av nyhetsmaterial
- [HR-XML](#) - Initiativ för XML-information om arbetskraft (Human Resources)
- [XBRL](#) - eXtensible Business Reporting Language
- [OpenOffice.org](#) - Öppet XML-format för Office-dokument
- [Docbook](#) - Stort schema för framförallt teknisk dokumentation. Har funnits med ett tag.
- [Wf-XML](#) - XML för Workflow Automation
- [SKICal](#) - Svenskt initiativ för turism och evenemangsinformation.
- [LegalXML](#) - Juridiska dokument
- [Open Financial Exchange](#) - DTD för finansiell information
- [XMLEPR](#) - EU-projekt för vårdinformation

En förutsättning för att dessa ska komma i bruk (förutom att det måste finnas ett tydligt mervärde för branschen) är att de måste vara lätta att hitta och att många andra använder dem. Det finns ett antal initiativ för att publicera information om befintliga scheman, och driva fram de nya scheman som

behövs för olika ändamål. Inget av dem är i närheten av fullständighet. Närmast att nå sin målgrupp kanske engelska GovTalk är eftersom de har en viss möjlighet att påtvinga myndigheterna deltagande i sådana här register.

- [xml.gov](http://xml.gov) - USA:s CIO Councils samlingsplats ställer frågan: "Borde vi ha ett registry över scheman?"
- [GovTalk](http://GovTalk) - Brittiska regeringens webbplats för publicering av schemas.
- [BizTalk](http://BizTalk) - Microsoft försökte tidigt att samla alla scheman, man skulle publicera dem här var det tänkt.
- [schema.net](http://schema.net) - Välstrukturerat privat initiativ av en framstående XML-utvecklare.

## Ju fler standarder desto bättre?

I XML kan man tack vare Namespace-mekanismen [[NS](#)] kombinera element från flera scheman i ett och samma dokument. Det är därför inte nödvändigt att alla alltid använder ett standardiserat schema. Tvärtom öppnar XML för ett mycket mer dynamiskt förhållande till datastrukturer. Det viktiga är att verkligen återanvända märkord och strukturer när vedertagna sådana existerar!

Forskning pågår, bland annat inom Framkom [[Framkom](#)], om hur man på en högre begreppsmodellnivå ska kunna länka ihop och översätta mellan olika scheman.

Det är sannolikt att globala register över scheman kommer att dyka upp. Möjligen kan de komma att fungera som distribuerade databaser med väl definierade gränssnitt för att man ska kunna hitta scheman och översättningar mellan scheman på ett effektivt sätt.

**Förslag: Upprätta ett register över scheman, märkord och kända överensstämmelser/översättningar mellan scheman. Gärna i samarbete med EU-projektet Diffuse [[DiffStand](#)]. Redaktören för detta register bör följa med i utvecklingen internationellt för att kunna länka in sitt register i en större global infrastruktur av scheman.**

## Ordlistor

För effektiv kommunikation behövs inte bara ett medium. Båda parter behöver dessutom ha ett samförstånd om vad orden betyder. Av denna anledning finns mängder av ordlistor publicerade. Ett exempel är EUs ordbok "Europeiska unionens institutioner, politikområden och utvidgning" [[ORDBOK](#)]. Denna finns översatt till alla EU-språk och utgör en bra plattform för kommunikation inom unionens gemensamma intressesfärer. Ordlistor är viktiga även inom riket. I sitt lagförslag om "Samhällets grundläggande information" [[SamInfo](#)] skriver e-infogruppen "Det föreslås att en gemensam standard fastställs med sikte på en enhetlig terminologi."

När XML möjliggör god struktur på elektroniskt tillgänglig information banas väg för väldigt bra sökfunktioner, men även för mer avancerad hantering av informationen. Ett exempel på detta är automatisk översättning.

Om de som publicerar viktig information i en ökad omfattning börjar använda sig av termer som hämtats från gemensamma ordlistor, och även uppge från vilken ordlista man hämtat termer, så ökar informationsvärdet markant. Automatiska översättningar kan då ofta ge väldigt goda resultat. Kategorisering av information i en databas kan bevaras när informationen automatiskt flödar in i en annan databas, genom att man kan veta hur begrepp i den ena databasens terminologi motsvaras av den andra databasens.

För att det ska vara realistiskt att människor börjar använda termer från standardordlistor måste sådana göras elektroniskt tillgängliga och maskinläsbara, så att redigeringsprogrammen kan visa upp termerna som alternativ.

## Två typer av termer

För att det ska bli tydligt varför "maskinläsbara ordlistor" behövs görs här en jämförelse mellan märkord och elementvärde.

```
<kategori>barn</kategori>
```

Ordet "kategori" ovan är ett märkord, ordet "barn" är elementets värde. Märkorden hanteras bra i XML, de definieras entydigt i XML scheman och märkord från flera olika scheman kan kombineras utan risk för förvirring tack vare namespace-mekanismen [\[NS\]](#) i XML. Standarder finns för märkord. Till exempel Dublin Core [\[DC\]](#) med flera metadatainitiativ (Se t.ex. [\[SveMeta\]](#) och [\[NordMeta\]](#))

Men när det gäller *värdet* på elementen så lämnar de allra flesta scheman fältet öppet för vilka textsträngar som helst. Det finns ordlistor, men de behöver göras tillgängliga på ett standardiserat sätt.

Exempel på sådana ordlistor som skulle vara användbara är:

- Alla postorter och postnummer
- Sveriges kommuner
- Länen
- Landskapen
- Länder och landsbeteckningar
- Valutabeteckningar (ISO 4217)
- Betygsskalor
- Namn på skolor och universitet
- Yrkestitlar

- Namn på myndigheter
- Bolagsformer
- Medicinska ordlistor
- SAB - bibliotekens klassificeringar
- EUs ordlistor [[EUterm](#)]
- FNs ordlistor, t.ex. ISIC rev3

**Förslag: Utred hur man på bästa sätt kan publicera ordlistor på ett maskinläsbart sätt. Resultatet skulle till exempel kunna publiceras som ett standardförslag hos IETF, som redan hanterar ordlistor [[IANA](#)] (dock med rent tekniska termer). Se till att publicera alla viktiga ordlistor i det offentliga Sverige på detta format. Förslagsvis i samarbete med Terminologicentrum [[TNC](#)].**

## Referensdata

Staten äger mängder av värdefulla databaser, se t.ex. [[Eresund](#)] för en kartläggning. Dessa utgör en viktig del av nationens *immateriella naturtillgångar* och bör tas till vara precis som man tar till vara skogen och frekvensutrymmet i etern.

För att ytterligare öka värdet av den IT-infrastruktur som byggts upp kan det vara samhällsekonomiskt lönsamt att exponera vissa databaser på Internet.

Exempelvis Lantmäteriverkets geografiska data skulle kunna exploateras mycket om det gjordes till en allmän nytta. Svenska Kommunförbundet skriver "Ska Sverige vara bland de världsledande nationerna inom IT-området är det dags att investera i en grundläggande informationsinfrastruktur för åtminstone - och initialt - geografisk information." [[FriInfo](#)]

Tittar man på de större databaserna (fastighetsregister, SPAR osv.) är det mer komplicerat att exponera innehållet mot Internet. Tekniskt finns det lösningar (t.ex. [[GILS](#)] och [[Z3950](#)]) för att öppna dem för egna sökportaler, och det är inte omöjligt att även göra XML-export till webbservrar som kan indexeras av Internets sökmotorer, men här är det naturligtvis ofta integritetsfrågor som står i vägen. Däremot bör man mellan myndigheter kunna exponera ganska stora delar av databaserna, till exempel genom Spridnings- och hämtningsystem [[SHS](#)].

Det är dock inte bara de stora statliga databaserna som kan göras tillgängliga, utan minst lika värdefullt är att aktuellt basbelopp, skattesatser, datum för landets helgdagar, och varför inte formler, symboler och ljud (se MathML, SVG m.fl. ovan) publiceras på ett sådant sätt att kommersiella och offentliga aktörer kan bygga nyttiga webbtjänster ovanpå. Detta förutsätter att information kan hämtas från myndigheter dygnet runt på ett tekniskt standardiserat sätt.

**Förslag: Försök att uppskatta vilken information som skulle vara samhällsekonomiskt lönsam att exponera ut mot Internet på något klokt format. Sannolikt finns det väldigt många enkla listor och faktauppgifter som mycket enkelt kan publiceras, till stor nytta för många. Utred även hur ett sådant format bör utformas, vilket protokoll som bör användas för att kommunicera med dessa "register" och hur man på ett vettigt sätt kan hantera behörigheter i dessa sammanhang.**

## Beständighet

Allteftersom mer och mer viktig samhällsinformation görs tillgänglig för oss medborgare i elektronisk form som XML-filer eller som webbsidor vänjer vi oss vid att lita på informationen. Innehållet på en webbadress (URL) styr människors val och beteenden. Forskare inom vissa områden förväntar sig idag att det finns en URL som identifierar alla verk de citerar. Det blir alltså viktigare och viktigare att kunna referera till elektronisk information.

I och med att det blir så lätt att följa källhänvisningar så ökar också människors benägenhet att göra så, vilket har många positiva effekter. Inte minst demokratiska: medborgare kan enkelt kontrollera att myndigheter baserat sina beslut på rätt underlag, forskare frigörs helt från universitetsbibliotekets begränsningar, lagparagrafer blir omedelbart tillgängliga... Kraven på tillgänglig och pålitlig information ökar samtidigt. Det är därför inte acceptabelt att en myndighet som kanske byter webbpubliceringssystem helt plötsligt ändrar adresseringsschema för hela sitt informationsbestånd, och därmed tvingar alla andra myndigheter att länka om, svarar "404 document not found" på medborgares rättmätiga begäran att få ta del av ett viktigt dokument. Så är situationen idag.

Det finns tekniska lösningar på problemet med adressbeständighet på Internet. Det i dagsläget bästa sättet att hantera problemet är att använda så kallade Persistenta URLer [[Purl](#)]. Det innebär att URLen som man publicerar inte är knuten till en viss dator eller till en viss position i ett filträd/sökning i databas. Istället knyts URLen till en post i en tabell på en Purl-server. Purl-länkar är alltså indirekta via uppslagning, och förutsatt att Purl-servern aldrig flyttas och att de webmastrar som ansvarar för myndigheternas viktigare dokument ser till att anmäla ändringar av URLer till Purl-servern så kan man betrakta Purl-länkarna som eviga.

**Förslag: Metvetandegör problematiken bland myndigheternas webbansvariga. Utred om det finns ett behov av att starta en Purl-server för det offentliga Sverige, eller om det räcker att komma överens om adresseringsscheman och rutiner vid ändring av URLer.**

## Säkerhet

Precis som att man måste kunna referera till viktig information så måste man kunna lita på att den information man får är korrekt. Om Riksbanken publicerar

diskontot maskinläsbart på en offentlig URL så bör bankerna till 100% kunna vara säkra på att det stämmer. Vidare hanterar myndigheterna känsliga data, vilket gör att det ramverk som skapas måste ha ett genomgående högt skydd mot intrång och avlyssning.

Olika grader av säkerhet måste kunna finnas. Viss information bör göras tillgänglig för både medborgare och näringsliv, medan annan information bara får användas av vissa myndigheter i vissa situationer. Behörigheter bör vara en viktig fråga när det tas fram nya scheman.

Det naturliga sättet att lösa säkerhetskraven rent tekniskt är att använda krypteringsteknik med öppna och slutna nycklar, *Public Key Infrastructure* (PKI), av RSA-typ (RSA = Rivest, Shamir, Adelman) eller motsvarande. Detta är den matematiska grunden för all säker datorkommunikation. Sannolikt kan kommunikation mellan myndigheter ske på det sätt som etablerats inom ramen för Spridnings- och hämtningsystem [[SHS](#)].

En djupare analys av säkerhetsfrågor rymms inte inom ramen för denna studie, som bara konstaterar att säkerhetsaspekterna som alltid måste utredas ordentligt. På området finns redan stor kompetens hos Statskontoret.

## Verktyg

Precis som att man kan utbilda till användning av formatmallar i Word skulle man kunna utbilda till användning av semantiskt meningsfulla formatmallar. Det finns tillägsprogram till ordbehandlare som gör sådant, liksom separata XML-editorer, och Microsoft har någonting kallat "smarta etiketter" i nya Office XP, som sannolikt är ett steg i rätt riktning. Allteftersom sökmotorer och andra typer av elektroniska agenter börjar dra nytta av metadata och infostruktur kommer fler och fler att se värdet av att "tagga upp".

Statskontorets webbplats XML-akademien har en lista [[Verktyg](#)] på verktyg som finns idag. Sådana listor är alltid svåra att hålla uppdaterade, och slutgiltigt måste man överlåta till de kommersiella aktörerna att själva informera om sina verktyg. Statskontoret har redan upphandlat ramavtal med XML-konsulter, och bör i framtida ramavtal med programvaruleverantörer i allmänhet beakta hur programvarorna understödjer ett arbete med infostruktur.

**Förslag: Ställ krav på vettig hantering av metadata och informationsstruktur vid ramavtalsupphandling på programvaror i allmänhet.**

## Styrning och juridik

En så viktig informationsteknisk förändring som införande av e-förvaltning behöver gå hand i hand med en motsvarande juridisk utveckling.

Om till exempel information som publiceras i ett maskinläsbart "register" skall anses vara lagligen bindande så måste de juridiska aspekterna redas ut noggrant, eftersom människor (och agenter) kommer att fatta beslut baserat på informationen i registret. Likaså kan avtal komma att skrivas som t.ex. låter priset på en vara regleras av innehållet i ett register.

- Vem står risken för att information förvanskas?
- Hur bevisas vilket klockslag som information refererats? Måste man kräva att det förs en ändringslogg? Skall denna göras tillgänglig på samma sätt som övriga registret?
- Vad händer om registret vid något tillfälle är oåtkomligt? Kan detta tillåtas utgöra hinder för affärstransaktioner, myndighetsutövningar eller andra handlingar?
- Kan någon typ av försäkring användas för att skydda användare mot fel i ett register?
- Ska registret betraktas som auktoritativt original, eller ska den betraktas som en kopia?

De juridiska frågeställningarna kan naturligtvis vara ett hinder för utvecklingen, men man kan samtidigt tänka sig att via lagar och förordningar bana väg för utvecklingen. Om exempelvis alla myndigheter som genomför IT-satsningar förpliktigas att ta i beaktande relevanta infostrukturella överenskommelser kan man sannolikt förhindra att man investerar sig in i databas- och webblösningar som låser in informationen.

Observera att detta avsnitt ej har skrivits av en jurist. Det tar inte upp alla juridiska aspekter, utan bör ses som en uppmaning till att särskilt beakta dessa frågor vid det fortsatta arbetet. Därför ger vi inte heller några konkreta förslag på detta område.

## Medvetenhet och användning

Vi behöver få en ökad medvetenhet om infostrukturens betydelse på alla nivåer, alltifrån handläggare, utredare och forskare till informatörer, IT-personal och allmänhet. För de allra flesta är det naturligt att använda Internet när man behöver information, men få är medvetna om hur man själv kan bidra till den gemensamma informationsbasen.

När man eftersträvar bättre uppmärkning och struktur så ser vi två kategorier:

1. De som idag arbetar med traditionella databaser.
2. De som idag arbetar med ostrukturerade dokument.

Kategori 1 torde inte ha några problem att göra sin information tillgänglig. De kan enkelt förse sina databaser med XML-export som följer något relevant schema. Det är bara en ekonomisk fråga - När är det motiverat för dem att göra sin information tillgänglig utanför den egna slutna databasen?

Kategori 2 kan i vissa fall motiveras till att börja använda bättre struktur, om de får tillgång till bra verktyg. Ganska ofta orkar man dock inte med att explicit "tagga upp" den information man publicerar, och då är det enda hoppet att kontexten kan hjälpa programvaran att automatiskt sätta litet vettiga etiketter. Exempel: ICQ-meddelanden, e-post...

En avdelning inom AstraZeneca har infört ett dokumenthanteringssystem i vilket man kan lägga in metadata om alla dokument man vill lagra. De har märkt att ungefär fyra manuellt inknappade metadata (t.ex. nyckelord) är vad folk orkar med. Men sedan kan man utifrån sammanhanget automatgenerera metadata. Exempelvis hade de byggt sitt gränssnitt kring en virtuell hierarkisk mappstruktur, och beroende på var i denna struktur folk lagrar sitt dokument läggs motsvarande metadata in i dokumentet. Uppmärkningsproceduren transformeras sålunda till navigation. Klockslag, mallar och tillstånd, val av redigeringsprogramvara och mycket annat kan naturligtvis ge mängder av viktig implicit kunskap om dokument, som ett sofistikerat system registrerar och gör om till explicit uppmärkning i dokumentet.

En ökad medvetenhet om maskinläsbarhet, sökmotorer, agenter mm skulle troligen även få kulturella konsekvenser. Till exempel underlättas kontextuell uppmärkning om man håller sig till ett ämne per brev när man skickar e-post. Detta koncept, som skulle ge mailinglist-arkiv enormt mycket större informationsvärde om det infördes, kallas Semantic Mail [\[GF\]](#). Det kräver dock ett något ändrat beteende.

För att återknyta till bilden av motorvägar utan bilar som bredband utan infostruktur: Även om det finns bilar så måste man tillföra kunskap innan vägnätet kommer till sin rätt: allmänheten måste lära sig grundläggande trafikvett, och några måste lära sig att köra. Med infostruktur är det likadant: alla behöver ha översiktlig kunskap, och några bör lära sig hur det används i praktiken.

Det numera brett uppkopplade svenska folket i allmänhet, och offentliganställd personal i synnerhet behöver allmänbildas i infostrukturfrågor. Man behöver få övergripande kunskap om:

1. Varför bör man strukturera och inte bara formge sina dokument?
2. Hur gör man dokument värdefulla, hållbara och "hittbara" i en ocean av information?
3. Vad är XML och hur fungerar det?
4. Vad är metadata och hur kategoriserar man information?

Ett sådant kompetenslyft behöver inte kräva enorma insatser. Sannolikt kommer användandet och medvetenheten så snart nyttan blir uppenbar. Men eftersom detta är litet utav ett moment 22 - nyttan kommer ju först när användandet är igång - så är det viktigt att någon tar första steget.

**Förslag: Satsa på information om infostrukturtänkande. Samverka så brett som möjligt. XML-akademien kan vara en bra bas. Diffuse kan vara en strategisk partner, liksom Skolverket, Utbildningsradion m.fl. Intensifiera försöken att lyfta upp frågan i den offentliga debatten.**

## Sammanfattning av föreslagna åtgärder

De flesta av förslagen ovan handlar om att informera. Statskontoret har redan påbörjat detta i och med projektet XML-akademien, men vi ser behov av insatser som sträcker sig utöver det som XML-akademien hittills engagerat sig i. Bland annat att kartlägga förekomsten av scheman, ge vägledning till när man bör använda en viss Internetstandard, visa vägen i frågor som beständighet (starta purl-server?) och säkerhet. Men framförallt utöka informationsinsatserna om att infostrukturtänkande är viktigt, och varför.

Att publicera statliga databaser på XML-format som en allmän nytthet på Internet kan naturligtvis inte Statskontoret göra på egen hand. Däremot kan Statskontoret utreda hur detta skulle kunna gå till. Detta ligger helt i linje med e-infogruppens rekommendation i "Samhällets grundläggande information..." [[SamInfo](#)].

En satsning på att ta fram maskinläsbara ordlistor skulle vara oerhört nyttig, inte minst för myndigheterna själva, men där är det mer än information som behövs.

Framförallt är det viktigt att området utreds vidare. Man behöver inte ha någon enorm brådska - stora delar av XML-familjen är ännu inte färdiga standarder - men det är definitivt dags att intensifiera arbetet med denna nya nationella infrastruktur. Det tar lång tid att ändra beteenden, och det är just vad som måste till om man ska kunna dra nytta av de nya möjligheter som den tekniska utvecklingen medför.

Naturligtvis är det ett politiskt övervägande vilka initiativ man anser att en myndighet som Statskontoret bör ta. Ett sätt att se på saken är "om staten lämnar fältet fritt för näringslivet så slösas inga pengar bort" men i realiteten är staten en viktig aktör på marknaden, och om en viktig aktör förblir passiv så hindras utvecklingen. Inte minst i de sektorer inom vilka staten är en dominerande aktör. Dessutom har många av de frågor som berörs i detta dokument en karaktär av nationell infrastruktur, vilket ju traditionellt är ett område där staten varit mycket aktiv, till stor nytta för näringsliv och medborgare.

## Referenser

[Access] Tillgänglighet för handikappade  
[http://europe.eu.int/comm/information\\_society/eeurope/objectives/area07\\_sv.htm](http://europe.eu.int/comm/information_society/eeurope/objectives/area07_sv.htm)

[DanStat] Statistics for metadata in the Danish WEB  
Wed, 04 Aug 1999  
<http://nwi.dtv.dk/MD/>

[DC] Dublin Core Metadata Initiative  
<http://purl.org/DC>

[DiffClass] Diffuses guide till "Data Classification Standards"  
<http://www.diffuse.org/classify.html>

[DiffStand] Diffuses översikt över viktiga standarder och standardiseringsinitiativ  
<http://www.diffuse.org/standards.html>

[eEurope] eEurope  
[http://europe.eu.int/comm/information\\_society/eeurope/index\\_sv.htm](http://europe.eu.int/comm/information_society/eeurope/index_sv.htm)

[e-GIF] E-government interoperability framework  
Office of the e-Envoy, september 2000  
ISBN 0 7115 0401 6  
URL <http://www.citu.gov.uk/egif.htm>

[Eresund] Inventering av databaser över viktiga informationskällor  
(metadatabaser)  
En studie utförd åt Statskontoret av Bengt Eresund, Juli 1997  
<http://www.eresund.se/Statsk/Statsk.html>

[EUterm] Termlistor inom EU  
<http://europa.eu.int/comm/translation/bulletins/glosen.html>  
Sökbara på  
<http://eurodic.ip.lu/>

[Framkom] Keboni - Knowledge supported, message based interoperability on  
the web  
<http://www.framkom.se/print.asp?ArticleID=153920>  
Begreppsmodeller, Stig Berild

[FriInfo] Uttalande från Svenska Kommunförbundet om att viss offentlig  
information bör göras fritt tillgänglig  
<http://www.svekom.se/stat/statistik/remiss20001130.htm>

[GF] Greg FitzPatrick i W3Cs forum för RDF-diskussioner.  
<http://lists.w3.org/Archives/Public/www-rdf-interest/2000Nov/0076.html>

[GILS] GILS  
<http://info.er.usgs.gov/gils/>  
<http://www.dlib.org/dlib/december96/12christian.html>

[HBDB] Sammanställning över viktiga databaser  
<http://www.hb.se/bib/dat/data.htm>

[HEAR] Hur blir en ny infostruktur motorn i e-Sverige?  
IT-kommissionens rapport 3/2000  
ISBN 91-38-21374-5

[Hotmeta] XML-sökmotorer  
<http://purl.org/dstc/hotmeta>

[IANA] Internet Assigned Numbers Authority  
<http://www.iana.org/>

[IETF] Internet Engineering Task Force  
<http://www.ietf.org/>

[ITKbrev] Om behovet av insatser för utvecklingen av samhällets informationsinfrastruktur  
Skrivelse från IT-kommissionen till Regeringen, 2001-03-20  
ITK2001/18

[MedRDF] Metadata för medicinskt bruk  
John Parnefjord, Ulf Kronman och Elisabet Frigell  
Karolinska Institutets Bibliotek  
15 augusti 2000  
<http://kib.ki.se/kib/project/rdf/report/rdf-report-2000-08-15.html>

[NordMeta] Nordiskt initiativ för koordinering av Metadataaktiviteter  
<http://www.lib.helsinki.fi/meta/>

[NS] Namespaces in XML  
<http://www.w3.org/XML/Namespaces>

[OIIXML] Guide to XML and related standards  
<http://www.diffuse.org/oii/en/xmlguide.html>

[ORDBOK] ORDBOK: Europeiska unionens institutioner, politikområden och utvidgning  
<http://europa.eu.int/scadplus/leg/sv/cig/g4000.htm>

[Purl] PURL  
<http://purl.oclc.org/OCLC/PURL/SUMMARY>  
Dansk server: <http://www.purl.dk>

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

[SamInfo] Samhällets grundläggande information - Inventering, Analys, Förslag  
Promemoria från E-infogruppen  
Ds 2000:34  
[http://justitie.regeringen.se/propositionermm/ds/pdf/ds2000\\_34.pdf](http://justitie.regeringen.se/propositionermm/ds/pdf/ds2000_34.pdf)

[SD] Sverige Direkt  
<http://www.sverigedirekt.nu/>

[SemRoad] Semantic Web Roadmap  
Tim Berners-Lee, 1998  
<http://www.w3.org/DesignIssues/Semantic.html>

[SHS] Spridnings- och hämtningsystem  
<http://www.statskontoret.se/shs/>

[SveMeta] Nationellt initiativ för koordinering av Metadataaktiviteter  
<http://www.lub.lu.se/svenskmetadata/>

[TNC] Terminologicentrum TNC  
<http://www.tnc.se/>

[VadXML] Vad är XML?  
Statskontoret Rapport 1998:6. Rev uppl 1999.

[Verktyg] XML-akademins lista över XML-verktyg  
<http://www.xmlakademin.nu/verkstad/verktyg/verktygsglista.asp>

[W3C] WorldWide Web Consortium  
<http://www.w3.org/>

[XML10] XML 1.0  
Svensk översättning <http://www.xml.se/xml/REC-xml-19980210-sv.html>

[Z3950] Information and documentation -- Information retrieval (Z39.50) --  
Application service definition and protocol specification  
ISO 23950:1998  
<http://www.iso.ch/cate/d27446.html>

## Bilaga 1: XML-familjen

Följande är de idag mest centrala specifikationer som tillsammans utgör XML-familjen, eller "XML & co" om man så vill.

### [XML 1.0](#)

Detta är grunden för XML-familjen. Beskriver syntaxen för XML-dokument. Specifikationen rekommenderades av W3C i februari 1998.

### [XML Schema](#)

Standardiserat sätt att beskriva strukturen hos ett XML-dokument. Jämförbart med DTD, men modernare och mer fullständigt.

### [XML Namespaces](#)

Namespaces är ett kraftfullt koncept som gör att XML-dokument kan kombinera terminologi och struktur från flera olika scheman. I praktiken innebär detta att man kan återanvända ett standardschema, utan att behöva begränsa sig till *bara* detta.

### [XML Query](#)

Frågespråk som ska kunna användas för att ställa frågor mot samlingar av XML-filer, ungefär som SQL används för att ställa frågor mot samlingar av tabeller i en databas.

WebDAV, SOAP, XML-RPC, XMI, WDDX, Jabber...

Det finns ett antal olika protokoll för kommunikation med hjälp av XML-meddelanden. Vissa av dem har en bred uppbackning, andra är än så länge bara pappersprodukter. Protokollen är en förutsättning för att det som kallas *web services* ska fungera. Arbetsgruppen [XML Protocol](#) inom W3C försöker att hitta en minsta gemensam nämnare.

### [DOM](#) och [SAX](#)

Två olika specifikationer som beskriver hur information från XML-dokument kan hanteras på ett standardiserat sätt i datorns minne, för att programvaror enkelt ska kunna interagera med dokumentets olika delar. Exempelvis kan DOM lösa problemet för webbprogrammerare att behöva programmera olika för olika webbläsare. Istället kan de programmera för DOM.

### [RDF](#)

Resource Description Framework är ett sätt att lagra kunskap i form av maskinläsbara påståenden - inte dokument. I förlängningen kan detta leda till en "semantisk webb" [[SemRoad](#)] i vilken datorer själva kan härleda svar och leta upp relevant information. Utvecklingen drivs av webbens skapare Tim Berners-Lee.

### [CSS och XSL](#)

En viktig egenskap hos XML är att formgivning kan hanteras av separata stilmallar. Innehållet i dokumentet kan därmed utan manuell bearbetning presenteras på många olika format och med många olika formgivning, till exempel som formell pappersutskrift, som "flashig" webbsida, som webbsida med organisationsanpassad layout, som kompakt WAP-sida, som punktskrift, som ljud med hjälp av röstsyntes osv. CSS är stilmallar och XSL kan användas både för formgivning och mer avancerad bearbetning av innehållet.

### [XLink, XPointer och XML Base](#)

Länkning i XML har mycket större möjligheter än länkning i HTML. Bland annat kan man länka till ett enskilt element i ett dokument. Länkar behöver inte heller skapas av den som skapat dokumentet, utan ett separat dokument kan skapa länkar mellan element i flera befintliga dokument.

### [XHTML](#)

XHTML är ett specifikt XML-schema med den goda egenskapen att de dokument som följer schemat förutom att vara fullgoda XML-dokument även är HTML-dokument som kan visas i en vanlig webbläsare. Format med XHTML som bas är alltså väldigt lämpliga att använda idag om man vill börja fasa in metadata och struktur i sina webbsidor.

### [XForms](#)

XForms-projektet syftar till att ta fram ett modul till XHTML för generell formulärhantering. Det blir ett lyft för formulärhantering på webben i och med att man frigör presentationen från funktionen. Ett XForms-formulär kan alltså, precis som annan XML-information, presenteras på olika sätt i olika medier utan specialanpassning.

### [MathML](#)

Ett standardiserat sätt att representera matematiska uttryck på XML-format.

### [SMIL](#)

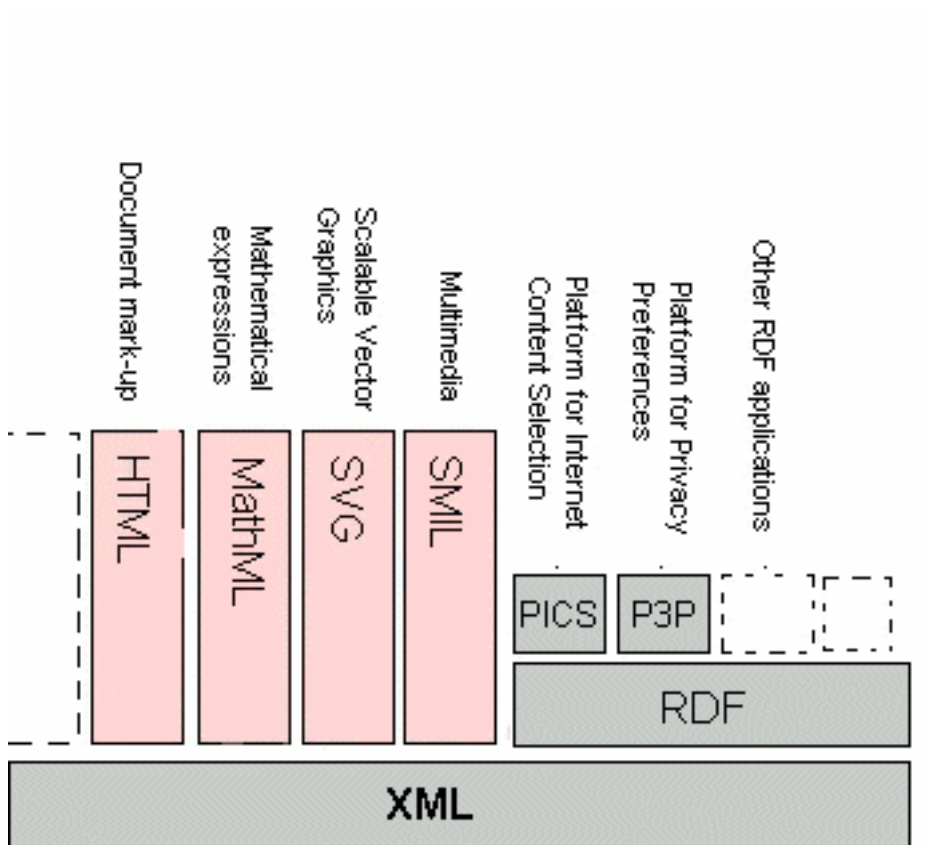
Ett standardiserat sätt att representera "streaming content" (rörlig bild, ljud, animationer mm.) på XML-format.

### [SVG](#)

Ett standardiserat sätt att representera grafik på XML-format.

### [XML Signature](#) och [XML Security](#)

Flera olika grupper arbetar med säkerhetsfrågor i XML-sammanhang. Här sker samarbete mellan bl.a. W3C och IETF.



XML & co enligt the W3C XML Activity Statement