

# Metadata

## – Vad, När, Hur, Varför?

Stig Berild  
(Myndigheten för skolutveckling)

”Metadata” upplevs av många som ett abstrakt och diffust begrepp. Följdriktigt förekommer också en palett av olika tolkningar. Föreliggande rapport syftar till att ge en personligt färgad introduktion till begreppet ”metadata” i kapitlen 2 och 3. Som avslutning redovisas i kapitlen 4 och 5 några lika personligt färgade reflexioner.

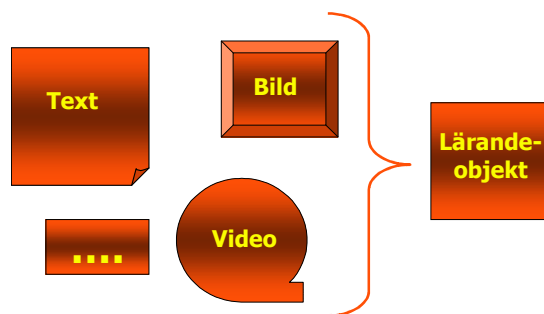
De exempel som redovisas är hämtade från ett projekt som Skolverket bedriver kring mjuk infrastruktur för informationshantering inom utbildningsområdet. Observera dock att rapporten försöker anlägga ett generellt perspektiv på metadata snarare än en specifik ansats för lärandeobjekt. Kapitel 1 förklarar varför.

## 1 (Lärandeobjekt – Lärkomponent) --> Resurs

Med lärandeobjekt menar vi någonting som

- har relevans i anslutning till kunskapsinhämtning
- representeras i någon multimedial form, t.ex. textdokument, fil, bild, video eller tal.

I många fall gäller dessutom att objektet ska finnas nåbart över Internet.

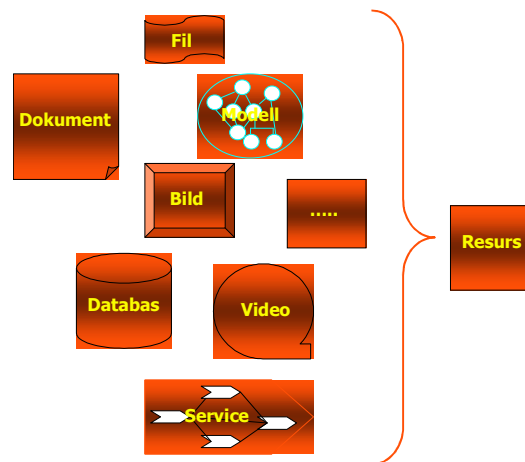


Figur 1

Webben omfattar en närmast outsinlig källa av objekt att tillgå för tusen och en ändamål och med en närmast oändlig variation på innehåll. Vedertaget i metadata-sammanhang är att kalla dessa objekt för **resurser**.

Resurserna finns där för olika syften och ofta representerande olika ämnesområden. Lärandeobjekt kan i sammanhanget sägas utgöra en viss kategori resurser. I det generella perspektivet representerar lärandeobjekt endast en liten del av den totala mängden resurser, men en väl så viktig del för dem som arbetar inom olika grenar av lärandeområdet.

I det fortsatta resonemanget görs ingen åtskillnad mellan olika typer av resurser. Den som så önskar får gärna sätta likhetstecken mellan resurs och lärandeobjekt eller vilken annan typ av objekt som helst – allt i syfte att placera diskussionen i känd miljö.

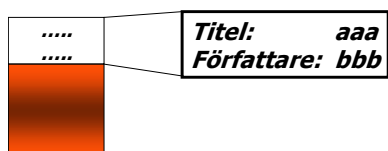


Figur 2

## 2 Metadata och metadatabeskrivningar

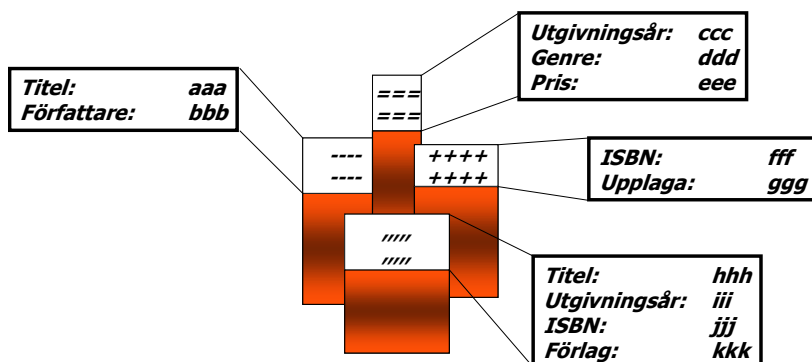
Oavsett typ av resurs kan av olika anledningar finnas behov att beskriva ett antal relevanta egenskaper som gäller för resursen. Varje anledning eller syfte ger upphov till sin beskrivning - sin uppsättning egenskaper. Resursbeskrivningar brukar kallas **metadatabeskrivningar** där varje beskrivningselement – uppgift eller data brukar kallas **metadata**.

I sin enklaste form kan beskrivningen bestå av en lista med metadata i form av ett antal **egenskaper** (eller **attribut** som man ibland använder som synonymt begrepp).



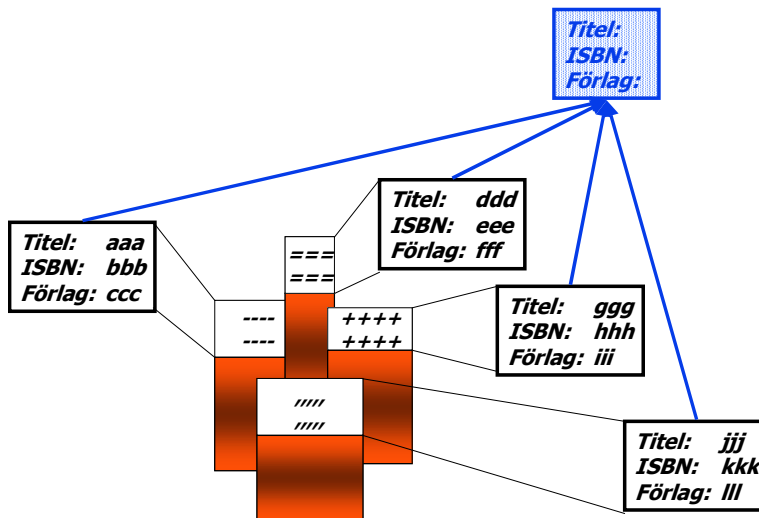
Figur 3

Inget hindrar i princip att varje unik resurs beskrivs med sin egen uppsättning metadata.



Figur 4

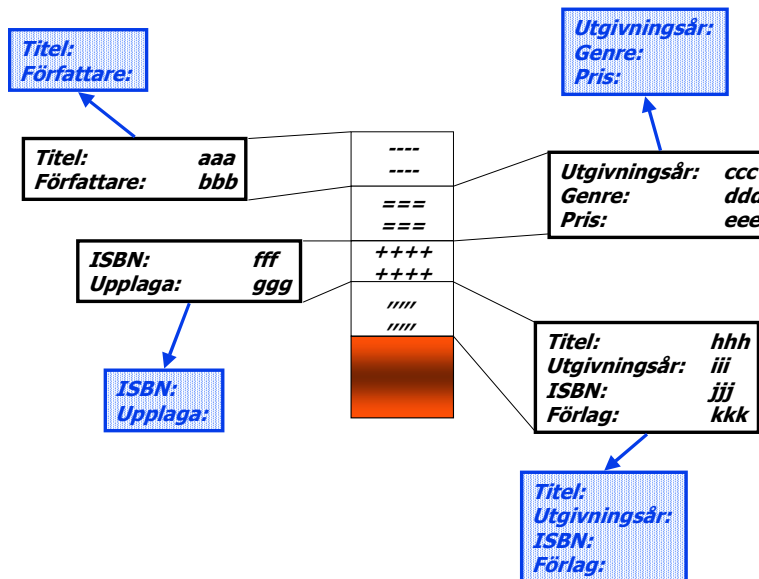
Vanligare är dock att det finns en överenskommelse om vilka typer av metadata (**mall**) som ska nyttjas för ett visst syfte, ett visst behov.



Figur 5

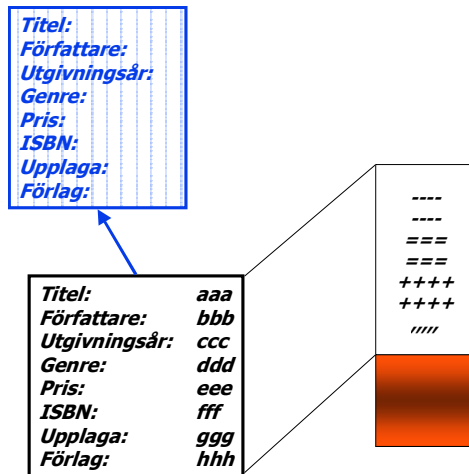
Ett av de mer välkända exemplen på en sådan överenskommelse är *Dublin Core* ([www.dublincore.org](http://www.dublincore.org)) som i sin ursprungsversion preciserade fjorton olika typer av egenskaper att användas för sådana resurser som primärt berör biblioteksområdet och dess behov.

En och samma resurs kan mycket väl bli föremål för beskrivning utifrån mer än ett perspektiv, för mer än ett syfte. Att "hänga på" ett antal olika uppsättningar metadatabeskrivningar (egenskapslistor) direkt i anslutning till en och samma resurs blir snabbt ohanterligt, speciellt som nya egenskapslistor kan förväntas tillkomma efterhand och andra tas bort.



Figur 6

Att föra ihop all beskrivningsinformation i en enda metadatabeskrivning kan vara en lösning, men ger upphov till andra konsekvenser.

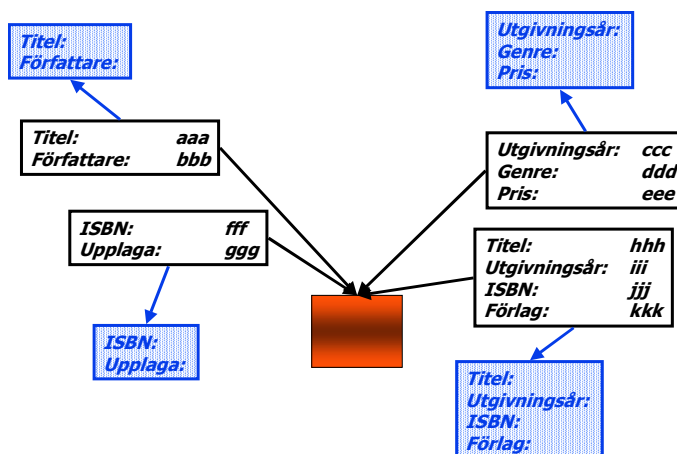


Figur 7

Dels krävs ett samordningsarbete och kompromissande parterna emellan om lämplig term för respektive metadata (attribut) – för att inte tala om den innebörd/semantik den avses representera. Behovet av en väldefinierad mall blir plötsligt mycket påtagligt.

Samma samordningsarbete måste sedan upprepas varje gång någon part önskar göra förändringar. Dessutom måste regler utformas för vem, vad, hur, när beskrivningsinformationen för en resurs ska formuleras och tillföras resursen. En part är ju i grunden bara intresserad av ”sina” egenskaper. Måste den ändå ta ansvar för övriga uppgifter om den står som ansvarig för resursen ifråga? Om inte, ska resursen och dess metadata skickas till de andra instanserna för kompletterande ifyllnad? Låter ohållbart tungt. Ska istället ofullständig uppsättning metadata accepteras? O.s.v.

Mycket talar för att varje syfte bör hantera sina egna metadata. En separering av resurs och metadata blir då den naturliga konsekvensen.



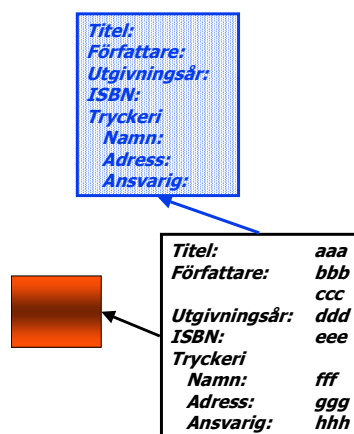
Figur 8

Samtidigt måste vi hålla i minnet att en separering ställer krav på viss disciplin i hantering av resursen. Metadata har formulerats utifrån de egenskaper resursen hade vid

ett visst tillfälle. Ändras därefter resursens utformning eller innehåll på så vis att det borde påverka metadatabeskrivningen gäller ju att ”någon” också kommer ihåg att uppdatera denna så att den korrekt svarar mot verkligheten. Innehållet i ett dokument kanske har disponerats om så att antalet sidor ökat från 23 till 27. Uppgiften *Sidantal* i metadatabeskrivningen borde följdriktigt sättas till det nya värdet. Man kanske väljer att ändra lagringen av dokumentet från Word- till PDF-format. *Lagringsformat* måste uppdateras. O.s.v.

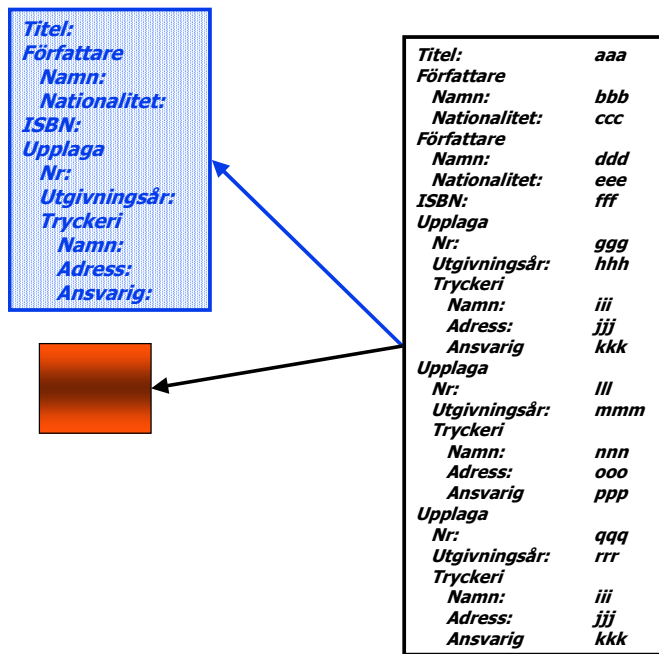
Låter självklart, men hur få den som ändrat resursen att tänka på att också ändra metadata, speciellt om metadata finns registrerat någon annanstans? Än mer problematiskt blir det om och när den ansvarige för metadata är någon annan än resursförändraren – en i realiteten sannolikt vanlig situation. Rutiner behöver etableras. Kanske även regler för versionshantering av resurser. Där flera är inblandade behövs principer för samspel mellan parterna.

Åter till metadatabeskrivningen, nu med koncentration på en viss uppsättning metadata, d.v.s. för viss unik resurs givet något relevant syfte. I vissa sammanhang är en mall i form av en enkel lista helt tillräckligt, exempelvis någon av de fyra alternativen i figur 8. I andra sammanhang ställs högre krav. Kanske ska mer än ett värde kunna ansättas för viss egenskap. Kanske bildar mallens innehåll en hierarki. Kanske .... Figur 9 visar ett exempel.



Figur 9

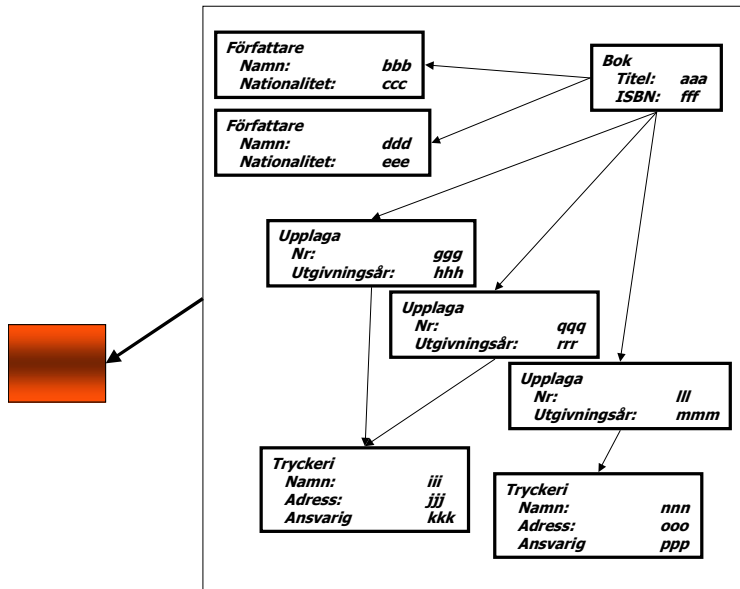
Visserligen skulle hierarkin här kunna undvikas genom att ta bort *Tryckeri* som ”rubrik” och istället mer explicit ange *Tryckerinamn*, *Tryckeriadress* och *Tryckeriansvarig*. Men i det generella fallet blir det knepigt. Anta att resursen är en bok som givits ut i ett antal upplagor och att olika tryckerier har varit engagerade i de olika upplagorna. Anta också att upplageuppdelad information måste kunna hanteras i metadatabeskrivningen. Anta därutöver att ytterligare upplagor kan komma att tryckas i framtiden.



Figur 10

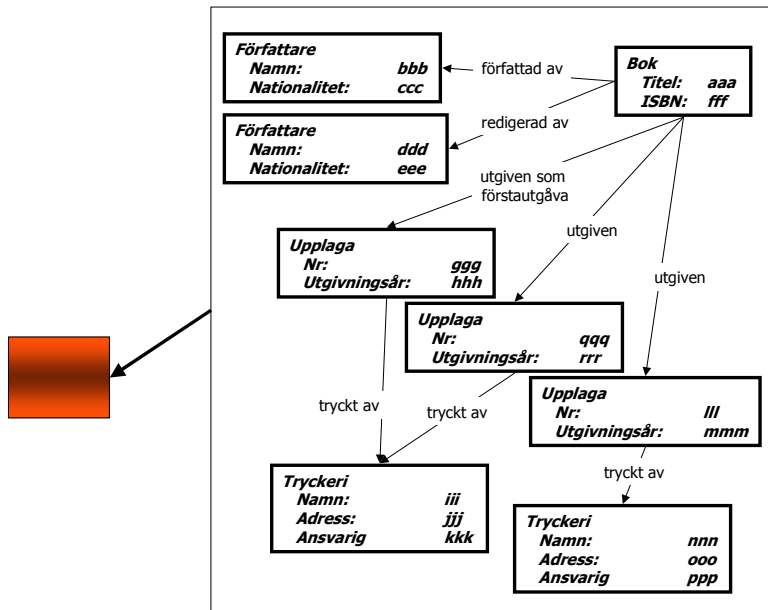
Den skarpögde noterar att Författare nu beskrivs med hjälp av såväl *Namn* som *Nationalitet*. Den skarpögde noterar därutöver att samma tryckeri (iii) varit med och tryckt både upplaga ett (ggg) och tre (qqq). Ganska klumpigt att behöva upprepa samma sak två gånger. Dessutom risk för bristande överensstämmelse. För att inte tala om risken att missa något ställe när en uppdatering behöver göras.

Båda upplagorna bör istället rimligtvis kunna referera till en och samma tryckeriinformation. Inte så lätt att uttrycka i en vanlig lista. Tryckeriinformationen bör alltså uttryckas i sin egen separata ”ruta” dit upplageinformationen kan referera. Varför inte tillämpa samma ansats när det gäller all ”grupperad” information, d.v.s. både *Författare*, *Upplaga* och *Tryckeri*? Figur 11 visar hur en sådan uppdelning kan se ut.



Figur 11

Än bättre vore det om vi även namnger semantiken i pilarna – den roll den utpekade beskrivningen spelar i det aktuella pil-sammanhanget - allt för att undvika alla missförstånd.

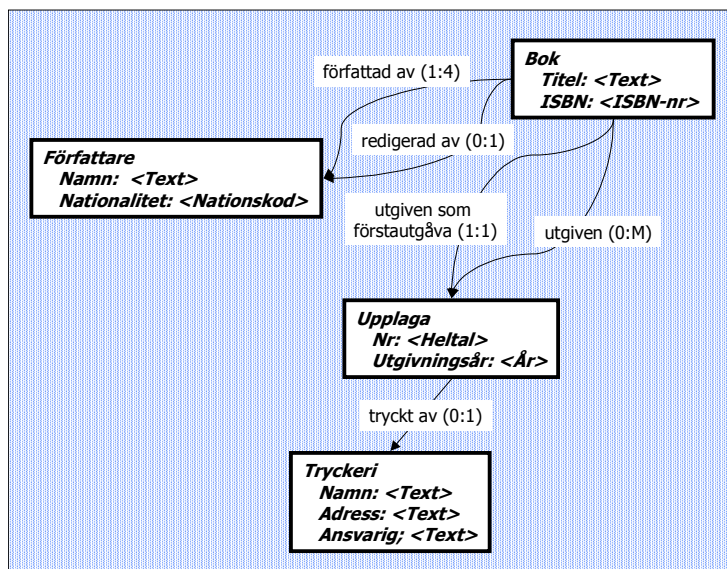


Figur 12

I ett mer komplicerat exempel skulle behovet av pilnamn bli än mer påtagligt.

Mallen behöver anpassas till den nya strukturen. Dessutom behöver mallen kompletteras med några uppgifter som bör finnas med för att ge en komplett vägledning om vad som är tillåtet.

Det saknas information om att *Upplaga* får repeteras valfritt antal gånger. Man har också kommit överens om att högst fyra *Författare* får anges, dock alltid minst en. Vissa böcker redigeras av en ansvarig redaktör. Denne ingår alltid i gruppen *Författare*. *Tryckeri* får utelämnas där det inte är känt. För varje typ av egenskap bör dessutom framgå hur respektive värde kan anges, d.v.s. vilken datatyp som gäller. På något sätt måste alla dessa uppgifter in i mallen. Figur 13 visar ett förslag. Observera att förslaget bara är ett av många olika möjliga sätt att uttrycka samma sak.



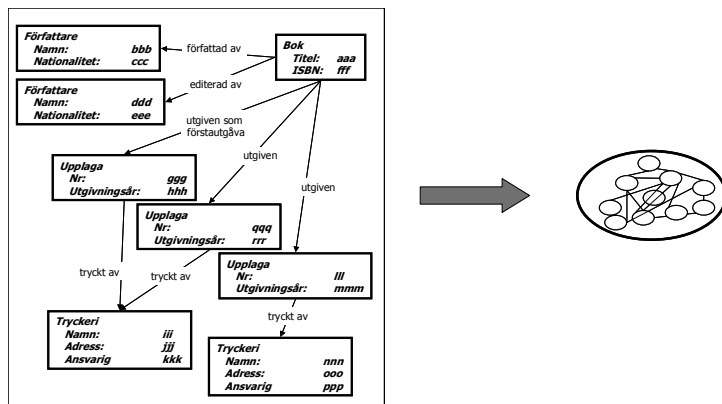
Figur 13

Inom hakparentes anges den datatyp som ska användas. Ibland räcker det med ett allmänt <Text> eller <Heltal>. I andra sammanhang gäller betydligt mer specifika datatyper som <ISBN-nr> eller <Nationskod>.

Parentesen efter ett pilnamn anger villkoret för hur många exemplar av den beskrivningsinformation pilen pekar på som får förekomma. När det gäller "författad av (1:4)" måste boken peka på minst en och max fyra uppsättningar författarinformation. Det får med andra ord – givet det aktuella syftet, vilket det nu är - inte finnas böcker som bara har en redaktör (d.v.s. ingen författare). M står för villkoret 'två eller flera'.

### 3 Metadatalager och begreppsmodeller

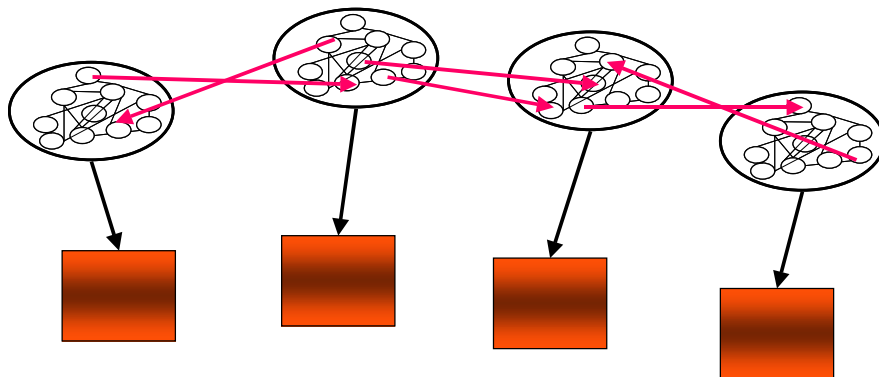
För det fortsatta resonemanget kommer vi för överskådlighetens skull att använda en oval innehållande lite bubblor och linjer som generell symbol för metadata.



Figur 14

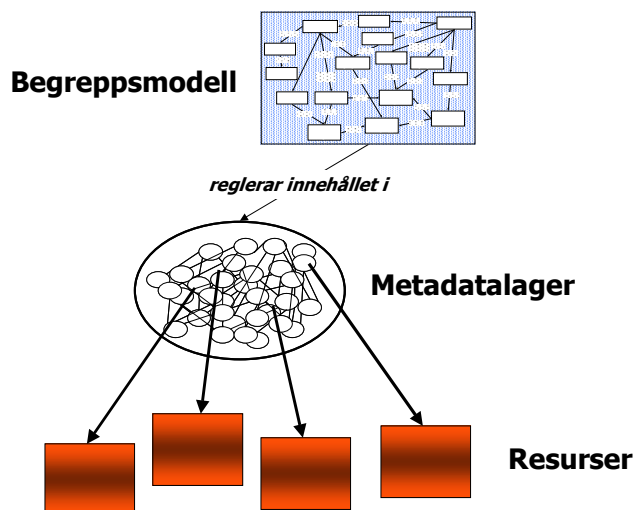
Än så länge har en metadata-beskrivning gällt en enda resurs. Men låt oss expandera resonemanget något. Ett syfte kan ju mycket väl vara att sammanföra metadata-beskrivningar för t.ex. ett antal böcker i en och samma databas. Anledningen är inte så mycket att "stuv in beskrivningen" på samma plats utan mer att göra det möjligt att vid behov skapa referenser mellan de olika resursbeskrivningarna. En annan anledning kan vara att frigöra en viss delbeskrivning från onaturlig direktkoppling till en viss resurs (bok).

Ta exempelvis Tryckeriet "iii". Förutom att ha tryckt två av upplagorna av vår ursprungliga bok har det sannolikt även tryckt ett antal andra böcker i olika upplagor. Naturligtvis vore det förträffligt om alla dessa upplagor hade möjlighet att referera till en och samma tryckeribeskrivning. Rationellt och i grunden helt naturligt. Dessutom effektivt eftersom exempelvis en adressändring bara behöver registreras på ett ställe – tryckeribeskrivningen – istället för på lika många ställen som tryckeriet nyttjats för tryckning.



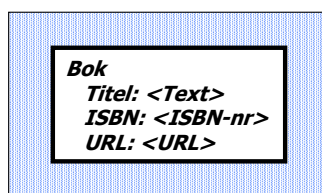
Figur 15

Informationen hänger alltmer ihop samtidigt som de olika delarna placeras där de logiskt hör hemma. Vill vi dessutom kunna uttrycka relationer mellan böckerna (resurserna), t.ex. för att beskriva lämplig nyttjandeordning ur ett inlärningsperspektiv eller för att notera vilka lärandeobjekt som tillsammans lämpar sig väl för ett visst tema, blir behovet av att samla informationen i ett gemensamt **metadatalager** för ett visst syfte än mer påtagligt. Mallen tenderar att bli en alltmer komplex sammansättning av begrepp och relateringar mellan begrepp. Detta faktum föranleder oss att nu ersätta termen ”mall” med det i andra sammanhang mer vedertagna **begreppsmodell**.



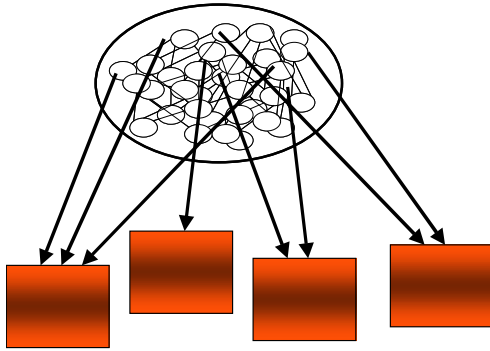
Figur 16

Dags också att explicit komplettera begreppsmodellen med den referens till resursen som hittills bara underförstått funnits med i form av en pil från metadata till resursen ifråga. Referensen är en vanlig URL (Uniform Resource Locator).



Figur 17

I många fall låter sig resursen representeras på en viss plats i metadatalagret (där bland annat URL-referensen till resursen sannolikt återfinns). I det generella fallet kan det dock mycket väl förekomma strukturer där en och samma resurs refereras från flera ställen i samma metadatalager. Där kan exempelvis finnas behov av att beskriva de olika roller den spelar ur olika perspektiv eller i olika sammanhang.



Figur 18

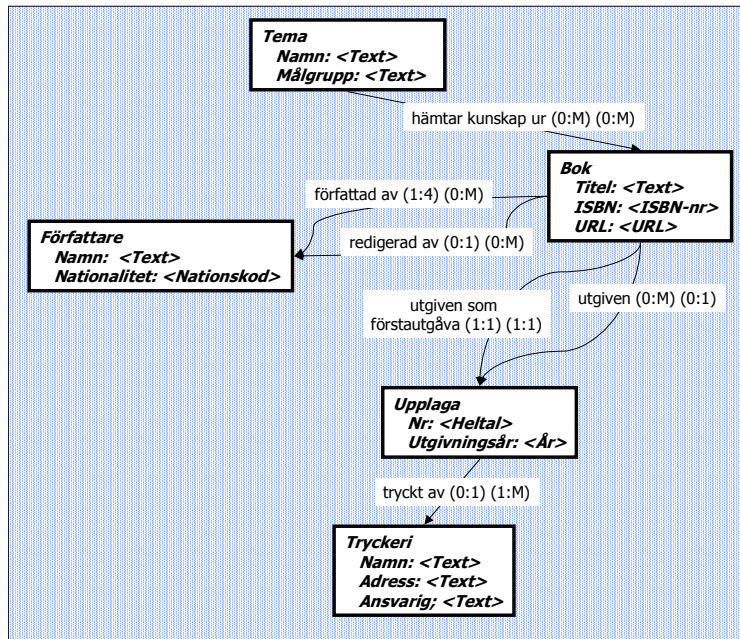
Med en alltmer inflätad och komplex begreppsmodell följer behov av att mer stringent tala om vilka regler och villkor som ska gälla för metadatalagrets innehåll (indikerat genom "reglerar innehållet i" i figur 16).

Vi vet redan utifrån viss bok att den författas av en till fyra författare. Men vad gäller för viss författare nu när ett antal böcker finns i metadatalagret? Får samma person vara författare till flera böcker? Ja, med största sannolikhet. Om nu inte det aktuella metadatalagret bara syftar till att redovisa den mest lästa boken av varje författare som fått Nobelpriset - förstås. För att undvika alla missförstånd införs den restriktion som ska gälla i begreppsmodellen. Pilen "författad av" kompletteras med ytterligare en parentes nämligen "(0:M)". Se figur 19. Därmed säger vi att utifrån en viss Författares synvinkel får denne vara relaterad till ingen (0) eller flera (M) böcker. "0" motiveras av att det kan finnas författare som bara figurerar i rollen som redaktör.

Observera att den använda notationen i begreppsmodellen bara är ett av många olika sätt att uttrycka samma sak. Varje modelleringsspråk har sin version. Det viktiga är egentligen inte hur villkoret noteras utan att det noteras på sätt som alla inblandade kan förstå.

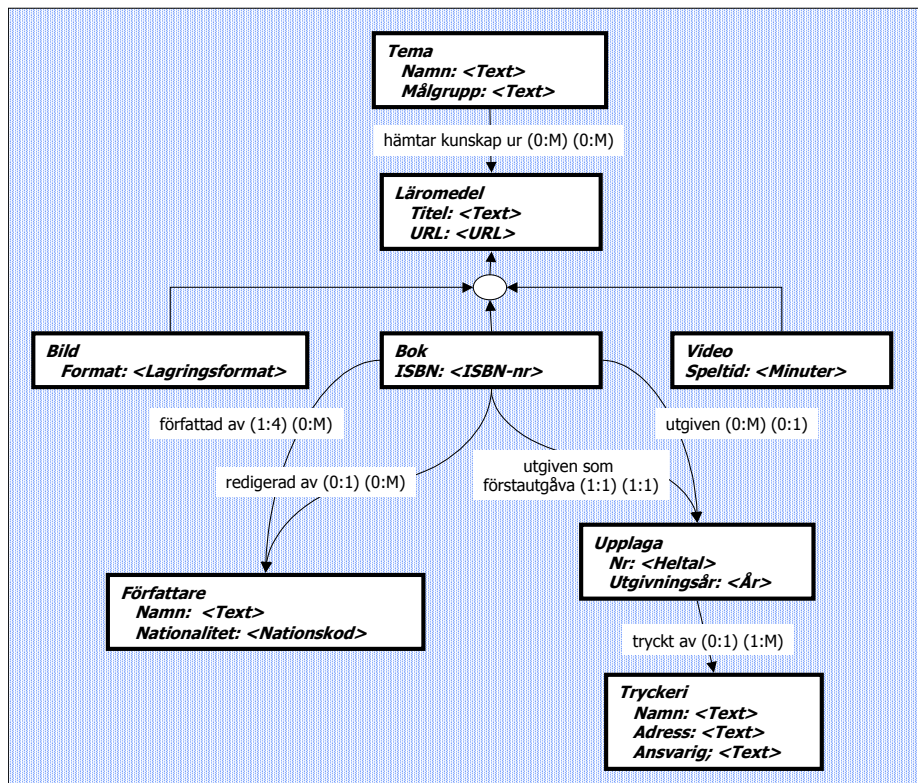
Från ett Tryckeris horisont gäller att det får refereras via "tryckt av" från flera Upplagor. Lämpligt eftersom detta är vad som gäller i verkligheten, nämligen att ett och samma Tryckeri kan engageras för att trycka vilka upplagor som helst av vilka böcker som helst. Det kan också vara värt att notera att det beslutats att bara Tryckerier som tryckt minst en Upplaga ska finnas med i metadatalagret. Dessa två villkor ur ett Tryckeris synvinkel deklarerar genom komplettering av "tryckt av"-pilen med (1:M). Övriga villkorskompletteringar framgår av figur 19.

Figuren har dessutom gjorts lite mer komplex genom kompletteringen av Tema. Varje Tema beskrivs med Namn och Målgrupp samt pekar ut till de böcker som förutsätts bli använda under instuderandet av Temat ("hämtar kunskap ur").



Figur 19

Nu visar det sig att temakunskap inte bara hämtas ur böcker. Även resurser i form av Bilder och Videofilmer är av intresse som Läromedel för Temat. Oavsett typ av Läromedel gäller att det har en Titel och finns tillgänglig via en URL. Varje specifik typ av Läromedel beskrivs därutöver med sin egen uppsättning metadata. För Bok är det ISBN-numret, för Bild det format det är lagrat i och för Videofilm hur lång speltid den har. Visst vore det bra om det fanns en enkel notation i grafen för att uttrycka att Bok, Bild och Videofilm alla är en slags Läromedel samt för att redovisa respektive metadata på sin naturliga plats. Figur 20 visar ett förslag på sådan notation.



Figur 20

Den lilla runda cirkeln står för en så kallad specialisering, d.v.s. att ett Läromedel mer specifikt kan vara endera en Bok, Bild eller Videofilm. Ur Bok-perspektivet gäller att varje Bok också är ett Läromedel. Av detta följer att förutom egna metadata (ISBN) så beskrivs Boken (som Läromedel) med hjälp av Titel och URL.

Syftet är inte att här och nu försöka göra läsaren till en fullfjädrad modellerare - alternativt röra till det - utan endast att visa några möjliga uttrycksvarianter för att öka precisionen på de villkor som ska gälla för de metadata som ska tillåtas lagras i metadatalagret.

Visst börjar metadatalagrets och begreppsmodellens likhet med hur en vanlig databas brukar struktureras att bli slående? Vilket inte är att undra på, eftersom metadata i grunden är vanliga data. Värt att notera är att de språk och den notation som i databassammanhang används för att formulera begreppsmodeller inte sällan innehåller betydligt fler möjligheter att uttrycka önskade villkor och restriktioner än de som visats ovan.

Förhoppningsvis har den förda diskussionen påvisat att även metadata behöver kunna struktureras med någon form av generellt **modelleringspråk** som inte bara kan definiera mallar i sin enkla form utan även kan definiera mer komplexa begreppsmodeller enligt ovan. W3C har specificerat ett enkelt språk för att lagra metadata i ett metadatalager, nämligen *RDF (Resource Description Framework)*. Ett annat språk som på senare tid fått ökad uppmärksamhet går under benämningen *Topic Maps*.

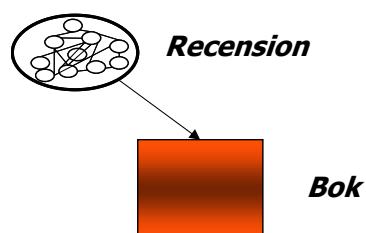
Notera att flera av de villkor som uttryckts i figur 19 och 20 ovan inte går att formulera med RDF men väl med andra modelleringspråk. Eftersom dessa typer av villkor med största sannolikhet kommer att efterfrågas i mer avancerade tillämpningar pågår arbete på olika håll att ta fram mer ”uttrycksfulla” modelleringspråk. Ett sådant är *DAML+OIL* (DAML=DARPA Agent Markup Language, OIL= Ontology Inference Layer), ett samarbete mellan två tidigare separat verkande organisationer.

Varför de som driver metadainriktningen inte istället, i alla fall till en början, väljer att direkt arbeta med något av de existerande, beprövade modelleringspråk som regelmässigt under flera decennier nyttjats i databassammanhang är en gåta.

## 4 Mycket 'meta' kan det bli ....

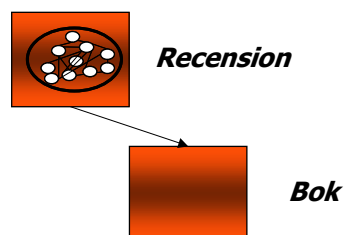
Till sist några exempel som förhoppningsvis sätter en slags knorr på detta med metadata. Kanske lite raljerande men förhoppningsvis med en seriös grundton. Syftet är att lägga en tankebas för de synpunkter som förs fram i kapitel 5.

En vanlig definition på metadata är ju "data om data" eller "information om information". Verkar vettigt med tanke på exemplen ovan. Men låt oss föra in en ny dimension. Anta att vi har en nyutgiven bok som finns att tillgå på webben. Den recenserar. Bland faktauppgifter i recensionen finns, förutom titel, författare, mm också referens till boken, d.v.s. dess URL. Recensionen är ett slags metadata.



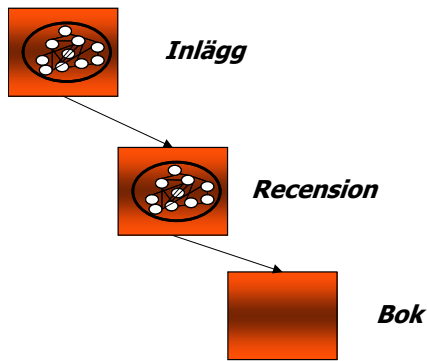
Figur 21

Recensenten lägger in Recensionen på den egna webbplatsen tillsammans med faktauppgifterna, allt snyggt strukturerat i enlighet med lämplig metadata mall (begreppsmodell). Recensionen får därmed en egen URL, blir en resurs i sig.



Figur 22

In på arenan kommer en bokälskare som läser recensionen, tycker den är helt "åt skogen" och väljer att skriva ett inlägg i en webbveckotidnings "Öppet forum" i enlighet med dess begreppsmodell. Inlägget är i praktiken en recension av recensionen. Liksom en ny webbresurs med egen URL.

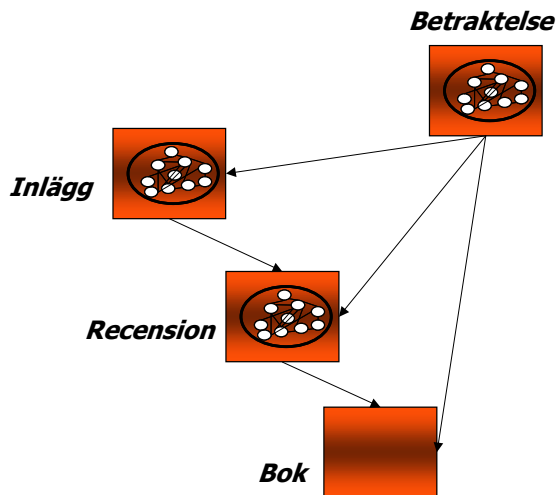


Figur 23

Är inlägget nu metametaddata? Mja, det är ju metadata av metadata. Men ....

Låt oss hålla ut ett tag till. En kulturskribent ser inlägget och söker fram recensionen. Eftersom hon läst boken väljer hon att skriva en humoristisk betraktelse över såväl bok som recension och inlägg. Beträktelsen läggs in på Kulturella klubbens webbplats.

Nu har vi plötsligt referenser till både data (boken), metadata (recensionen) och metametaddata (inlägget). Beträktelsen i sig placerar sig på metametametaddata-nivå. Eller?



Figur 24

Som ytterligare lök på laxen visar det sig att boken är en självbiografi som i mångt och mycket beskriver författarens egna excellenta förmågor i olika avseenden, d.v.s. den kapacitet eller resurs vederbörande besitter. Resursen ifråga är knappast möjlig att lägga in på webben eftersom den finns i vederbörandes kropp och själ. Otvetydigt är dock att boken är data om en resurs. Kanske är boken att betrakta lite som metadata ändå, även om inte just denna resurs finns på webben? Visst, väldigt långsökt.

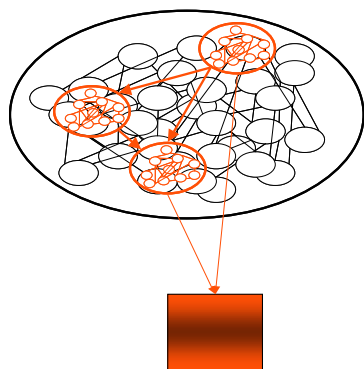
Lika långsökt är det att betrakta alla foton över författaren som finns i boken, som metadata. Rimligare då att åtminstone betrakta ett foto som data eller resurs. Värre blir det när vi konstaterar att det finns ett par foton i boken som faktiskt avbildar collage av

foton av författaren. Äntligen metadata i form av ett foto (data om data). Eller? I så fall får vi i konsekvensens namn lägga till en metanivå på alla övriga redan omnämnda metadataresurser som har bäring på boken.

Puh! Kan ju fara iväg och bli ganska många ”meta”-prefix. Något måste göras. Ett sätt är att rensa upp bland metanivåerna. Recensionen är ur bokälskarens perspektiv en artikel, d.v.s. en högst vanlig informationsresurs ur vilket följer att inlägget måste betraktas som metadata, inte metametadata. Boken, recensionen och inlägget är i kulturskribentens ögon alla data eller resurser av vad följer att betraktelsen blir metadata. O.s.v.

Aha, genom att vid varje tillfälle utgå från ett visst perspektiv går det att inskränka till endast en metadata nivå. Saker och ting blir snabbt lite enklare att hantera.

Alldeles påtagligt enklare skulle det bli om alla recensioner, inlägg, betraktelser, m.m. kunde beskrivas och hanteras inom ett och samma metadatalager. Nödvändiga referenser åstadkoms inom metadatalagrets ram. All data i lagret betraktas definitionsmässigt som metadata, alltså även våra recensioner, inlägg och betraktelser. Fixat och klart, punkt – slut.



Figur 25

Ingenting är tyvärr gratis. Nackdelen blir att de inte längre finns som ”fria” resurser i cyberrymden utan endast åtkomliga och relaterbara via metadatalagrets gränssnitt. Antagligen lika acceptabelt för vissa syften som oacceptabelt för andra.

En kompromiss skulle kunna vara att låta respektive recension, inlägg och betraktelse visserligen beskrivas i metadatalagret enligt ovan men dessutom samtidigt tillåtas existera som självständiga resurser med vissa vitala faktauppgifter på något sätt inbakade i den löpande texten. Helgardering, kan tyckas. Men nu återfinns plötsligt samma uppgifter på flera platser, med allt vad det kan föra med sig.

Det behövs ingen större uppfinningsrikedom för att generera ytterligare varianter på samma tema.

Är detta måne något den hägrande så kallade semantiska webben kommer att få dras med eller går det att ytterligare renodla resonemangen? I alla händelser hänger fortfarande kvar en del funderingar över vad ”meta” egentligen står för, varför det behövs. Om det nu alls behövs, förstås? Över till nästa kapitel.

## 5 Vaddå ”meta” egentligen?

Vi har diskuterat resurser (på webben) och behovet av att beskriva dem. I andra sammanhang finns behov av att beskriva artiklar i ett lagersystem, eller anställda i ett personalsystem, eller fakturor i ett ekonomisystem, eller ..... Världen är full av databastillämpningar för oräkneligt antal syften. Är en webbresurs ur beskrivningssynvinkel i grunden så magiskt olik artikel eller anställd när det gäller principer för hur de beskrivs?

Naturligtvis inte. De är alla föremål för beskrivningsbehov. Att den ena finns elektroniskt på webben i någon multimedial form, den andra ligger på en hylla i ett lager och den tredje sitter bakom ett skrivbord är ju knappast en avgörande skillnad i sammanhanget. (Möjligtvis i vissa fall som att den anställde rör på sig, har beteende. Men eftersom personalsystemet representerar tillståndsinformation snarare än dynamik kan även den nyansskillnaden här läggas åt sidan.)

I samtliga fall finns intresse av att beskriva tillståndsinformation av företeelser av intresse. För ändamålet har sedan flera decennier databashanteringssystem (dbms) stått till vårt förfogande. Varje tillämpnings behov av datahantering har i de flesta fall analyserats fram och beskrivits i form av en realiseringsoberoende datamodell (begreppsmodell) som sedan anpassats till de förutsättningar som aktuellt dbms arbetar under (exv enligt relationsmodellen –SQL). Personen sitter där den gör medan beskrivningen av denne ur något visst perspektiv ligger i en databas.

Observera att samma person vid skrivbordet mycket väl kan vara föremål för olika intressen, behöver beskrivas för olika syften. Varje syfte kan ge upphov till sin databas. Det kan gälla personalsystem (där utbildning, kvalifikationer mm finns beskrivna), lönesystem (där aktuella och tidigare löner finns dokumenterade), projektstyrningssystem (där aktuella arbetsuppgifter, ansvar, mm ajourhålls), .....

Den realiseringsoberoende begreppsmodellen uttrycks i ett modelleringspråk som förhoppningsvis valts med hänsyn till den uttryckskraft som önskas. Det kan vara alltifrån ett enkelt binärt modelleringspråk (ungefär så som modellen i figur 19 ovan uttryckts) till det betydligt mer komplexa UML (Unified Modeling Language).

RDF är för övrigt i huvudsak ingenting annat än ett enkelt binärt modelleringspråk.

Det viktiga att ha klart för sig är att varje begreppsmodell formuleras med avseende på

- en viss avgränsad ”verklighet” för vilken information bedöms vara intressant att hantera.
- ett visst syfte. (En konstruktionsritning beskriver sannolikt inte en bro på samma sätt som en turistbroschyr.)
- aktuella iakttagare, d.v.s. de erfarenheter, värderingar och personligheter som kännetecknar de tilltänkta informationsnyttjarna och deras förhållande till såväl ”verklighet” som syfte.

Med detta konstaterande, som vägledande princip för formulering av begrepps- och datamodeller för datahantering, behövs inte någon distinktion mellan data och metadata.
--

Det möjligtvis nya i metadavärlden är att man tillåter sig nyttja nya typer av dbms som klarar av att ta till sig den specificerade datamodellen och direkt utnyttja den som schema för lagring av informationen. I konventionella sammanhang används regelmässigt SQL (relationsdatabas). En implementering kräver då översättning av den realiseringsoberoende datamodellen till ett schema uttryckt i relationsmodellens språk. Tungt och inflexibelt, dessutom med risk för felöversättningar. Här banar metadataperspektivet vägen för nya attityder och realiseringsmönster. Detta kunde dock lika gärna åstadkommas utan att involvera "meta" i sammanhanget.

## 6 Till sist

Rapporten har introducerat ett antal infallsvinklar på metadata och metadatahantering. Alldeles tydligt är att datahantering och begreppsmodellering tycks nytt ur ett metadataperspektiv samtidigt som det är skäligen konventionellt ur ett dataperspektiv. Alla trender pekar i dagsläget på att beskrivning av de resurser som kan återfinnas och hanteras över Internet kommer att spela nyckelroller för framväxten av nästa generations webb, den som för närvarande går under beteckningen "den semantiska webben". För ändamålet kommer att behövas vedertagna och allmänt accepterade begreppsmodelleringsspråk med vars hjälp en förmodligen närmast oändlig mängd begreppsmodeller successivt kommer att skapas.

Modelleringspråk finns redan definierade. Begreppsmodeller med bäring på webbresurser börjar också genereras i allt stridare ström. Många är ännu av relativt enkel natur (attributlistor). Mer komplexa modeller kommer att tas fram allteftersom kunskap stabiliseras, erfarenheter samlas och kraven blir skarpare. Utvecklingen skulle ta betydligt snabbare och större kliv framåt om metadataföreträdare tillät sig ta del av den gedigna modelleringskunskap som finns inom konventionell data- och begreppsmodellering.

Vad som ännu tycks vara ett relativt öppet fält är hur metadata lämpligen hanteras med hänsyn till ett välformulerat nyttoperspektiv och med hänsyn till de olika användarroller som behöver kunna stödjas.

En försvårande omständighet som brukar påpekas kring webbresurser har att göra med incitament. Den som bäst kan formulera resursbeskrivningarna, ofta den som skapat resursen ifråga, har sällan själv någon nytta av att beskrivningsinformationen finns. Kanske har man idag alltför mycket arbetat med hur webbresurser kan och bör beskrivas utan att ställa beskrivningen mot ett nytto- och syftesperspektiv. Förmodligen gömmer sig bakom hörnet även betydligt svårare "nötter" att knäcka, inte minst när vi byter fokus från metadata som sådan till hur metadata kan komma att hanteras i de miljöer som nu håller på och tar form under trender som "Web Services" och "The Semantic Web".

Det är förmodligen inte alltför riskfyllt att påstå att det så kallade metadataområdet står inför spännande, kanske omtumlande, utvecklingstrender såväl i närtid som i ett längre perspektiv.