

# Mycket 'meta' är det

*Stig Berild*  
(Myndigheten för skolutveckling)

## Inledning

Metadata är ”inne”. Det dyker upp inom snart när alla tillämpningsområden, inte minst de med bäring på Internet och webben. 'Metadata' utgör ett viktigt fundament för den mer avancerade användning av webben vi hoppas kunna realisera och till vissa delar redan ser växa fram. Allt detta är gott och väl. Men där finns problem bakom ytan, problem som måste tas om hand och hanteras. Dit hör att 'metadata' knappast är ett entydigt definierat begrepp. Inget ovanligt när det gäller nya trender och nya begrepp men icke desto mindre viktigt att belysa och om möjligt reda ut. Nästa avsnitt diskuterar några olika betydelser som tycks existera för närvarande. Därefter följer en kort diskussion kring en annan typisk effekt av en ny trend, nämligen den brist på strategi, samordning, standardisering som existerar och som indirekt får oönskade effekter.

## Olika betydelser

### Beskrivning av webbresurser

Vissa tolkar 'metadata' som beskrivningsinformation för olika typer av alster. Inom biblioteksvärlden har denna typ av information alltid varit vital i och för katalogisering och sökning av böcker, artiklar med flera typer av alster. Författare, titel, ämnesområde, ISBN-nummer med mera är kända beskrivningselement.

Med webbens intåg uppstod naturligt samma behov att beskriva de elektroniskt lagrade alstren. Egentligen blev behoven än större eftersom alstren nu även kunde vara av en typ som bibliotek kanske inte alltid hanterar. Bland möjliga typer finns textbaserade dokument av olika slag (böcker, artiklar, recensioner, ...), bilder, videofilmer, ljudinspelningar, .... Olika lagringsformat kan ha nyttjats. Ett alster kan dessutom vara en komposition av andra alster över en eller flera nivåer.

Webben omfattar en närmast outsinlig källa av alster att tillgå för tusen och en ändamål och med en närmast oändlig variation på innehåll. I metadata-sammanhang kallas dessa alster med en gemensam term för **resurser**.

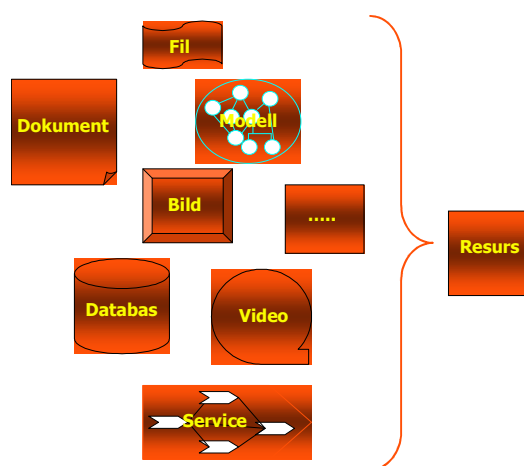
Med metadata-beskrivning öppnas nya möjligheter för en mer avancerad användning av webbens innehåll än vi hittills varit vana vid. Vi vill kunna upptäcka, hantera och utbyta dessa resurser oavsett var de finns eller i vilka format de är lagrade. Det kan gälla mer

formaliserad affärssamverkan eller varje annat sammanhang där informationen kan berika, liksom sammanhang där brist på information, feltolkning eller osäkerhet om dess innehåll kan få oönskade, kanske allvarliga, konsekvenser.

Men för att denna nya ambitionsnivå ska kunna realiserats krävs ett antal förutsättningar:

- att informationen kan beskrivas, sökas och återfinnas med exakthet och kvalitet
- att det går att förstå vilket innehåll den representerar i stort
- att det går att tolka/ta till sig övriga kännemärken av vikt
- att det går att konstatera vilka eventuella villkor som kan vara förknippade med tillgången till informationen (tekniska, ekonomiska, tidsstyrda, ....)
- med mycket mera.

För att göra bilden mer fullständig kan vi konstatera att de drivkrafter kring gränsöverskridande processsamverkan, webbtjänster, mm som på senare tid fått alltmer ökad uppmärksamhet likaledes kommer att involvera metadata för att realisera sina visioner. Processer eller tjänster som finns att tillgå på webben behöver ju också beskrivas av samma anledningar som information, ja i realiteten betydligt mer detaljerat. Allt för att de ska kunna lokaliseras, förstås, och tas i anspråk i enlighet med tillhandahållarens stipulerade villkor. I visionen ligger ju att nya tjänster ska kunna skapas, ja även exekveras genom en väldefinierad samverkan mellan ett antal separata, anropade undertjänster. Att kunna realisera sådan samverkan endast med ledning av tillgängliga metadata, d.v.s. utan mänsklig inblandning, vore ännu ett strå vassare.



Låter ju utmärkt. I alla fall på ytan. Tyvärr finns därunder ganska många frågetecken som behöver rätas ut. Här följer några:

- Resurser i elektronisk form kan/bör/behöver inte hanteras på exakt samma sätt som de som exempelvis finns i tryckt form. Inte heller är det säkert att de kan/bör/behöver beskrivas enligt samma principer. Alltså måste i varje enskilt fall klargöras vilken kategori resurser som är av intresse och vilken beskrivningsinformation som anses relevant i det givna fallet.
- Metadata kan "klistras" på respektive resurs alternativt finnas som en separat beskrivning - i det senare fallet ofta tillsammans med många andra resursers

beskrivningar i någon form av kunskapsbas. Med den princip man väljer följer olika konsekvenser att ta hänsyn till.

- Resurser av samma kategori behöver beskrivas på ett och samma sätt för att öppna sökningar ska kunna göras. Beskrivningen måste ske efter samma beskrivningsmall eller begreppsmodell. En viss begreppsmodell skapas och ges ett innehåll utifrån visst övergripande syfte, refererar entydigt till viss avgränsad population resurser av intresse samt visst givet sammanhang. Observera att samma resurs mycket väl kan vara refererad från många olika begreppsmodeller.
- Därutöver tillkommer det komplicerande faktum att olika nyttjargrupper bygger upp sin begreppsmodell efter sin uppfattning som kan variera även när övriga faktorer är desamma. Standardiserade, behovsrelaterade begreppsmodeller behövs och kommer säkert att tas fram i allt större utsträckning. Dock kommer de med största sannolikhet att bli svårare att etablera med "tyngd" ju mer avancerad användning de sätts att stödja. Komplicerade begreppsmodeller är komplicerade att standardisera.
- En absolut förutsättning för standardiserade begreppsmodeller är standardisering av det eller de språk som bör användas för att formulera begreppsmodeller, dvs begreppsmodelleringspråk. I vissa fall har man vid standardisering av begreppsmodeller inte ens noterat detta behov utan valt att rita och berätta efter "eget huvud" med hjälp av tabeller eller andra uppställningar, ibland grafiskt med hjälp av lite "bubblor och pilar". Visst kan detta vara tillräckligt för enklare behov men blir snabbt ohanterligt när det gäller precision i vad som uttrycks samt skapar utrymme för missuppfattningar.
- Representerar metadata endast en beskrivningsnivå? Rapporten " Metadata – Vad, När, Hur, Varför?" diskuterar ett scenario där en e-bok recenserar. Recensionen som läggs ut på webben utgör metadata. Någon väljer i sin tur att kommentera recensionen och lägga denna kommentar på webben. Är kommentaren att betrakta som 'metametadata'? Om någon väljer att ha synpunkter på kommentaren då? Osv. Verkar ju snurrt, men det gäller att motivera varför. Vilket också rapporten försöker göra.
- Är en webbresurs ur beskrivningssynvinkel i grunden så magiskt olik en artikel i ett lager eller en anställd i ett företag 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. Speciellt knepigt blir det om vi har behov av att beskriva både författare (personer) och de e-böcker (webbresurser) de medverkat i och göra det i ett sammanhang. Behövs 'meta' överhuvudtaget? Är månne gränsdragningarna artificiella?

## **Abstraktionsnivåer**

Vid konventionella databastillämpningar fångar man under konstruktionsarbetet upp informationsbehoven i form av begreppsmodeller. Begreppsmodellen uttrycker grovt vilka typer av uppgifter (begrepp) som är av intresse och hur de är relaterade till varandra. Den utgör en abstraktion av den avgränsade verklighet, det ämnesområde man är intresserad av att hantera information om. I en personaldatabas kan till exempel bland begreppen återfinnas 'anställd', 'utbildning', 'tidigare anställning', 'löneanspråk', 'befattning' och många fler. En sådan begreppsmodell ligger sedan till grund för hur databasen ska struktureras, dvs hur databasens schema ska definieras.

I en XML-miljö definierar man hur meddelanden ska vara uppbyggda med hjälp av dokumentmallar eller Document Type Definitions (DTDer). Även DTDn är en abstraktion men för utbyte snarare än hantering i en databas.

Ofta kallar man i dessa sammanhang abstraktionerna för 'metadata' just för att indikera att de inte är data i vanlig mening. Vanligt är också att kalla en formell definition av det begreppsmodelleringspråk som används för att uttrycka begreppsmodeller för 'metametadata'. Allt i det välmenande syftet att skilja på data och beskrivning av data och beskrivning av beskrivning av data. Mycket 'meta' blir det.

Nyttjande av 'meta' i detta sammanhang sker med en betydelse som helt avviker från det 'meta' som ovan använts för att indikera beskrivning av resurser.

Resursbeskrivningar har vi ju redan konstaterat är av samma häradsområde som person- eller artikeldata. Även i webb-sammanhang behöver man alltså kunna reglera vilka typer av uppgifter som är tillåtna. Dublin Core som används för att beskriva Webbresurser och standarden Learning Object Metadata (LOM) som används för att beskriva läroobjekt är två exempel. Om nu beskrivningselement för resurser anses vara 'metadata' borde väl i konsekvensens namn den abstrakta definitionen - begreppsmodellen - om vad som är tillåtet att formulera (exv Dublin Core och LOM) kallas 'metametadata'. I konsekvensens namn borde den formella definitionen av det språk som använts för att formulera dessa begreppsmodeller då kallas för 'metametametadata'. Eller? Puh! Rapporten " Metadata – Vad, När, Hur, Varför?" diskuterar ett förslag till lösning på meta-inflationen.

## **Data om dataelement**

För vissa ändamål behöver man kunna komplettera specifika dataelement med beskrivningsinformation. Det kan gälla uppgifter som behövs för att kunna tolka dataelementet och för att förstå dess relevans. Beskrivningselement kan vara uppgiftslämnare, tidpunkt för uppgiftslämnandet, trovärdighet, under vilka omständigheter uppgiften kom till, när uppgiften upphör att gälla, mm. Respektive dataelement kan finnas i en databas eller vara registrerat annorledes.

I grunden är detta ingen avgörande skillnad mot att beskriva resurser. Skillnaden ligger i att här beskrivs enskilda dataelement eller väldefinierade strukturer av dataelement

medan de beskrivna resurserna är mer sammansatta företeelser om vars interna innehåll vi ingenting vet, i alla händelser inte utifrån det aktuella perspektivet.

### ***Med mera uppfattningar***

Förmodligen finns fler uppfattningar. Dessa bör om möjligt fångas in, allt för att klargöra betydelse och i förlängningen för att om möjligt kunna applicera mer precisa begrepp för respektive betydelse.

### **Dagsläge**

Till de olika uppfattningarna om vad 'metadata' står för kan läggas bekymret av att endast befinna sig i början på en ny utvecklingstrend. De olika infallsvinklarna på metadata är visserligen inte alla nya men intresset har skenat iväg, sannolikt med webben som katalysator.

Framförallt när det gäller resursbeskrivning tycks alla rusa iväg för att definiera metadata inom sina respektive ämnesområden med den grundmurade uppfattningen att just deras ämnesområde är unikt och att det är just deras kall att definiera diverse metadata. Ofta resulterar det i att man tar tag i för mycket, bland annat sådant som inte specifikt har med ämnesområdet att göra. Om man stannade upp en stund, gjorde en mindre omvärldsanalys, skulle det i många av dessa fall säkert framgå att andra redan gjort jobbet, dessutom förmodligen bättre, i alla händelser om dessa metadata tagits fram inom ett område där den nödvändiga kompetensen funnits till hands.

Överlappningar, spagetti blir följd. Den indirekta följd blir förstås tolkningsproblem och därmed vattentäta skott mellan områden som annars kanske både kunde och borde samverka, samt i sämsta fall ett sådant misstroende mot de metadata som tagits fram att de aldrig kommer till användning.

Värre är att varje metadataprojekt med självaktning använder sig av sitt eget, ofta hemsnickrade sätt, att beskriva metadata, sitt eget beskrivningsspråk. Återigen, för att man underlåtit att söka och ta del av andras erfarenheter och kunskap. Att det finns flera decenniers erfarenhet från andra områden att bygga på upptäcks inte. Kanske inte att begära i alla situationer men vi kan konstatera att en sådan kunskap säkerligen skulle snabba upp det egna, specifikt ämnesinriktade arbetet alldeles betydligt. Dessutom med ett kvalitativt bättre resultat. Bland etablerade beskrivningsspråk finns Unified Modeling Language (UML) och enklare binära beskrivningsspråk. World Wide Web Consortium (W3C) har tagit fram ett beskrivningsspråk under beteckningen RDF (Resource Description Framework) med syfte att vara det generella beskrivningsspråket för resurser på webben. Listan kan göras lång. Ändå är det som sagt sällan dessa etablerade språk används i metadata-sammanhang. Tyvärr, eftersom en allmän acceptans av ett enda språk så mycket snabbare skulle föra utvecklingen framåt - utan att låsa in eller begränsa beskrivningsmöjligheterna. Respektive ämnesområde kunde koncentrera sig på sina specifika behov i vetenskapen om att man därmed öppnar för såväl intern som extern samverkan.

## **Sammanfattningsvis**

Ingen tvekan råder om att det så kallade metadataområdet måste skärskådas, struktureras, belysas. Kunskap måste genereras och spridas. Samordningsdörrar måste öppnas. Viktigt. Kanske bråttom. Det finns mycket att ta tag i.