

2006-06-27

Dokumentation av LISA-mötet den 1 juni 2006 om
Metadatas roll inom rättslig informationsförsörjning
Metadata: letadata eller vetadata?

1. Bakgrund

Den 1 juni 2006 samlades ett trettiotal personer hos SICS (Swedish Institute of Computer Science) i Kista för att diskutera ”metadatas roll inom rättslig informationsförsörjning”. Mötesinbjudan presenterade temat på följande vis:

Metadata, dvs uppmärkning och beskrivningar av dokument på webben (eller i databaser), syftar bl.a. till att öka möjligheterna att finna relevant information med god precision vid användning av olika slags söksystem. Detta ställer givetvis krav på beskrivningarna. En särskild fråga är vad som gäller för metadata knutna till rättsliga dokument. Vem – om någon över huvud taget – kan ställas till svars för ett icke återfunnet dokument? En komplikation i sammanhanget är att beskrivningen av ett dokument i sig utgör en utsaga om dokumentets innehåll som möjligen kan ha juridiska implikationer. Seminariet avser att belysa dessa frågeställningar, sedda i ljuset av den teknikutveckling på sökmotorområdet som gör att vissa betraktare anser att metadata blir obsolet i framtiden.

Den dokumentation som nu finns publicerad på LISA:s hemsida består dels av bilder från de inledande inläggen, dels minnesanteckningar från den efterföljande diskussionen sammanställda av Lars Klasén. Först här några definitioner av ”metadata” som Lars läste upp inledningsvis:

- ”Metadata definieras som strukturerade, standardiserade beskrivningar av resurser, digitala eller fysiska, som stöder upptäckt (discovery), sökning och användning av dessa resurser” (National Science Digital Library/National Science Foundation)
- ”Med metadata menas data om data, dvs en beskrivning eller en utsaga om något slags data. (...) Idag används termen ofta inom IT-området. Där betyder metadata en beskrivning av en webbsida. Beskrivningen kan ligga på själva sidan – som om man fäste ett katalogkort direkt på det elektroniska dokumentet – eller separerat från dokumentet, i t ex en databas.” (Kungl. biblioteket)
- ”Metadata är en datorrelaterad term som kan betyda data/information om data, eller de egenskaper ett objekt har.” (Wikipedia)

2. Presentationerna

Mötet inleddes med tre presentationer. Här bilder från dessa:

- Metadata ”lever”: Incitament för produktion och förvaltning av metadata. *Bengt Eriksson, Riksdagen*
- Metadata som hävstång: Om användning av metadata som medel för att förbättra kvalitet och effektivitet. *Olle Olsson, Swedish Institute of Computer Science (SICS)*

- Metadata och juridisk metod: Hur metadata utmanar rättsutvecklingen. *Erik Helling*, Institutet för rättsinformatik (IRI)

3. Anteckningar från diskussionen

Cecilia Magnusson Sjöberg, Institutet för rättsinformatik, inledde med att förmedla en synpunkt från *Gunilla Lilie Bauer*, Stockholms universitetsbibliotek. Gunilla, som varit tvungen att avvika efter de inledande inläggen, saknade biblioteksperspektivet och hoppades att ett kommande seminarium kunde ta upp hur man inom bibliotekssektorn arbetar med metadata med hjälp av moderna verktyg.

Lars Klasén, InfoTorg AB, upplyste om att Strukturgruppen (se mer nedan) under förberedelserna för seminariet hade dryftat frågan om biblioteksmedverkan men beslutat att avstå inför risken att de juridiska aspekterna skulle hamna i skuggan av de diskussioner om metadata som pågår inom biblioteksvärlden.

På tal om biblioteken berättade *Peter Krantz*, Valtech AB och knuten till Verva, att Kungl. Biblioteket (KB), som ju har i uppdrag att hantera pliktexemplar av svenskt tryck, även samlar in juridisk information. Denna avser man nu att märka upp med metadata. Statligt producerad information bör vara relativt enkel att hantera. Detta planeras nu också inom ramen för ett uppdrag benämnt ”Gemensam söktjänst för statliga rapporter och skrifter”, bakom vilket bl a Verva och KB står.

David Eriksson, White & Case Advokat AB, framhävde här den roll som webbens sökmotorer kan och bör ha vad gäller att tillgängliggöra statliga informationskällor. Statens viktigaste uppgift bör vara att ”öppna” alla statliga informationskällor för åtkomst av dessa sökmotorer. Det är inte möjligt att ”tagga” all information som det offentliga producerar. Viss information kanske inte heller är lämplig att nyttja för metadata, t ex personuppgifter. Visserligen kan sökmotorer inte ”tolka” och därmed indexera allt korrekt. Vad står t ex en datumangivelse för? Vad i ett dokument är en referens? Vad är rubrik? Etc. Men det är bättre att informationen görs potentiellt tillgänglig än att den inte är tillgänglig alls! Ju mer desto bättre. För övrigt blir sökmotorerna med tiden allt bättre på att tolka och göra sådan information som inte är taggad sökbar. Och ju mer information som görs tillgänglig desto större blir möjligheten att ”kombinera” dokument eller delar av dokument – vilket möjliggör nya och ytterligare uttolkningar av information. David pekade särskilt på juridikens föränderlighet. Därför är myndigheternas bästa strategi fullständig öppenhet – vilket ger bäst förutsättningar för tolkningar.

Olle Olsson, SICS, knöt an med att påpeka att den information som kan nås via den ”öppna” webben (och indexeras av sökmotorerna) bara utgör ”fragment” av all information. Exempelvis kan en del bara nås via exakt adress. Men det allra mesta är begravt i databaser – och bara att göra sådant tillgängligt är en välgärning.

David fortsatte sitt tidigare resonemang med att säga att det i synnerhet inom juridiken är av stor betydelse med en så bred tillgång som möjligt till den information som staten producerar. Det skulle kunna rätta till det missförhållande som idag råder, där domstolar och domare kan lyfta fram sådan information som advokatsidan inte har tillgång till eller ens känner till.

Sökmotorer och metadata i all ära; *Cecilia* uppmärksammade oss på behovet av annotering av dokumentets rättsliga status, t.ex. hur "allmänna råd" tillgängliga via en myndighets hemsida skiljer sig från annan information om rättsläge och praxis som också ligger ute på webben. Det är många gånger också viktigt att klargöra skillnaden mellan ett offentligt producerat/bearbetat dokument och ett kommersiellt producerat/bearbetat.

Ordet gick nu till *Harry Hartikainen*, STEMA Form. Han arbetar med blanketter och inmatningsformulär och redogjorde för behovet av "metadata om metadata"; ett medel för att kunna spåra källan till de regler som manifesteras i blanketterna/formulären. Ett kryss i en ruta på en blankett – vad betyder det? Regeln kan ju ha ändrats. Vilka andra blanketter berörs av en regelförändring?

Olle påpekade i sammanhanget att den föränderliga värld vi lever i också kan påverka metadata. Hur hanterar vi den förändringsprocessen? Och hur gör vi med konfigurationshantering, versionshantering etc?

Joakim Andersson, Sjöfartsverket, instämde i att det finns behov av definitioner. Därefter återförde han diskussionen till den tidigare tråden kring sökmotorer i det han påpekade att det i vissa fall är nödvändigt att känna till en viss samling av regler – och menade att det finns en definitiv fara med system som kanske gör att man som användare inte hittar alla regler. Inte är väl dagens sökmotorer så "duktiga" att de hittar allt?

Även *Niklas Lindström*, Valtech AB, demonstrerade en avvaktande inställning till sökmotorernas förmåga. De kan ju inte upptäcka sådant som inte är entydigt uttryckt. Här tjänar metadata till att precisera informationen: "Det går ju inte att lita på maskinella system som arbetar på data som i sig inte är entydiga!". Till exempel måste den vara försedd med någon form av identitetsbegrepp, något som i och för sig inte vara särskilt besvärligt att åstadkomma. Ofta räcker det att i inledningen ange något som identifierar det hela.

Självklart är det så att strukturerad information, gärna uppmärkt i XML, ger större chans för sökmotorer att extrahera informationen – i den mån de är försedda med sådana möjligheter. Men *Joakim* påpekade att detta i sig inte är tillräckligt för bra återvinning – det gäller också att användarna inte bara söker i fritext (och får alltför många svar) utan också förmår att nyttja avgränsa sina sökningar (till t ex fält, som författare etc). Och för att få dem dithän krävs vägledning.

Visst kan användning av metadata öka "precision" och "recall". Men i samband med hantering rättslig information finns det anledning att vara försiktig, påpekade *Cecilia*, eftersom bearbetningen många gånger omfattar subjektiva inslag. Faktum är att konstruktionen av en DTD/ett schema (vanligen) förutsätter juridiska

ställningstaganden, liksom användningen av olika format, annoteringar, strukturell uppmärkning (av vem?), anpassning till sökmotor etc.

Frågan är här, menade *Joakim*, om man bara applicerar metadata i ett tydliggörande syfte; något som är inriktat på användning inom myndigheten ifråga, eller om den nyttjas i ett vidare perspektiv; rättsfilosofiskt. Vidare refererade han till Eriks påpekade om att det finns behov av en objektiv ansats.

Ett önsketänkande, menade *David*, i synnerhet inom juridiken. Man kommer *alltid* att missa något. Juridiken förändras ständigt och det gäller att inte sätta ribban för högt. Ett system som kräver hundra procentig uppmärkning kräver också ett ständigt hundra procentigt underhåll. Och något sådant är inte realistiskt. Och ändå kan användaren inte vara säker på att han/hon söker ”rätt” (t ex i ”rätt” fält eller kategori). Webbens sökmotorer är visserligen inte ”exakta” men om du säkert vet att alla statliga dokument indexerats där så vet du i alla fall att du har *möjlighet* att få fram det – och du är bekant med förutsättningarna, dvs att man inte kan lita på svaren.

Niklas höll med om att det gäller att inte sätta ribban för högt – men ansåg ändå metadatamärkning viktig inte minst för att det kan underlätta arbetet på lokal nivå, med t ex egna sökmotorer. Därtill ger det ledning för/underlättar ”tänkandet” för författarna.

David höll med om att metadata såklart berikar systemet men vidhöll ändå sin åsikt om att inte ställa krav på t ex en förutbestämd standard eftersom det gör arbetet svårt och långsamt och riskerar att göra att det inte alls blir sökbart.

Anders Jonasson, Norstedts Juridik (NJAB), menade att juridik i sig är en ”metadatavetenskap”. NJAB, som är ett gammalt förlag, arbetar med den ”metadatastruktur” för Sveriges lagar som skapades redan på 1700-talet. Den nyttjas av de författare som NJAB samarbetar med, den utgör stöd och hjälp för tolkningar och tillämpningar – den utgör sammanfattningsvis en ”kunskapsmassa” som genererats i samspel med omgivningen och som kan sägas utgöra en manifestation av gemensamma överenskommelser på marknaden.

Med anledning av tidigare inlägg pekade *Erik Helling*, Institutet för rättsinformatik, på de risker som användningen av olika typer av verktyg innebär. Det finns alltid en risk att förlora något. Samtidigt invaggas användaren psykologiskt i en (falsk) säkerhet vid användningen. Det rör inte bara en övertro på fullständighet utan också att verktygen styr användning och uttolkning och därmed minskar incitamentet för egna tankar, självkritik etc.

Peter anknöt nu till orden ”letadata och vetadata” vilka båda diskuterats här: letadata i samband med sökmotorer och vetadata, som kräver en mer avancerad ansats, av Erik. Visserligen kompletterar de varandra men Peter menade att det vore intressant att följa upp var gränsen mellan dessa går och framför allt hur kraven och behoven *egentligen* ser ut.

Joakim påtalade här att systemen och applikationerna bör anpassas till behoven. Räcker det att finna 80 % av relevant information eller är det mödan värt att satsa

på manuell uppmärkning för att också hitta resterande 20 %? Vilka konsekvenser får det om en redare inte finner alla regler som gäller vid fartygsbygge och som därför inte får det certifierat? Man måste veta konsekvenserna när man rekommenderar någon att använda ett visst verktyg. Och ibland kan det vara bättre att inte ge ut någon information alls om innehållet och hur man hittar det.

David var inne på samma linje och menade att det i alla sammanhang är frågan om avvägningar. Vad gäller t ex juridiken kan det vara bättre att få ett någorlunda rätt svar snabbt än ett helt korrekt om ett år.

Olle konstaterade att det uppenbart finns ett intresse för metadata inom rättsinformatikområdet men att han i diskussionen saknat en sak, nämligen det som händer på webben i form av användarmedverkan (det som går under begreppet "Web 2.0"). Se t ex på det fria uppslagsverket Wikipedia, som helt byggs av användarna och som – mot förmodan – faktiskt har en hygglig kvalitet. Kanske kan ansvaret för metadata i många fall föras över till användarna, "communities", istället för speciellt tillsatta kommittéer? Därmed uppstår en ständigt tillväxande "metadatamassa", direkt driven av användarnas behov.

Seminariet avslutades med att *Lars* efterlyste teman för kommande seminarier samtidigt som han konstaterade att det redan fanns två uppslag. Ett givet tema är en uppföljning vad gäller metadatadiskussionen. Ett annat är ett seminarium kring det arbete kring rättsinformationssystemet som *Verva* nyligen igångsatt med *Peter* som projektledare.

Därmed tackade **LISA** och **Strukturgruppen** för sig och för deltagarnas intresse och synpunkter.

Kort om Strukturgruppen (<http://www.lisan.org/sg/>) som förberett just detta LISA-möte:

Strukturgruppen arbetar för att information, i synnerhet samhälls- och rättsinformation, ska produceras, utväxlas, sökas, presenteras och återbrukas på ett standardiserat sätt. Ett reflekterande förhållningssätt kring informationens strukturering och användning av metadata står i förgrunden.