

# METADATA FÖR DET DIGITALA BIBLIOTEKET

---

OBJEKTBEKRIVNING AV ELEKTRONISKA RESURSER

---

**Miriam Björkhem**  
**Jessica Lindholm**

Examensarbete (20 poäng) för magisterexamen i Biblioteks- och informationsvetenskap vid Lunds universitet  
Handledare Mats Lindquist och Colm Doyle

BIVILs skriftserie 2000:7

ISSN 1401-2375

© Lunds universitet. Biblioteks- och informationsvetenskap 2000

## **Metadata for the digital library. Cataloguing electronic resources**

### **Abstract**

Cataloguing electronic documents makes new demands on libraries.

This thesis discusses the metadata formats Dublin Core and MARC which are, individually or in combination, commonly used for object description in the digital library.

The formats are studied with particular emphasis on characteristics such as underlying reasoning (ambitions during the development of the formats), flexibility (possibilities of expansion and adaption to new demands) and interoperability (co-operation between different metadata formats).

The formats are dealt with in their context, i.e. we discuss what types of documents the formats are meant to describe, what environment they were originally created for, what functions they are meant to fill (to find, identify, select and acquire/obtain entities), and also how these factors are related. The Swedish projects SAFARI (which uses Dublin Core) and NetSök (which uses MARC) are presented briefly.

The thesis also discusses tools for co-ordination between formats; Z39.50 and RDF. A short survey of OCLC's CORC-project exemplifies co-ordination put to practice.

Cornerstones in the construction of functional metadata are simplicity, flexibility, interoperability and standardization. Simplicity and flexibility guarantee that the format is easy to use, and fulfils universal description. At the same time it can be developed to satisfy the specific description needs of individual projects or institutions, such as libraries.

The standardization, in turn, guarantees that there is no need to centralize the cataloguing work, and it is also an important condition for interoperability with other collections. The functional metadata format makes it possible for the library to share and copy catalogues. After elaborating on these issues we focus more specifically on a library perspective in the second part of the thesis.

We discuss the digital library and the issues that arise when libraries build up repositories for electronic resources. Setting out from the strategy for cataloguing electronic materials that is currently being developed for Lund University Library we finally recommend a combined model: where cross-searchable catalogues using MARC and Dublin Core coexist. We also recommend different description levels for printed and more transitory electronic resources.

## *Uppsatsen tillägnas Björn Tell*

Och hela jorden hade enahanda tungomål och talade på enahanda sätt. Men när de bröto upp och drogo österut, funno de en lågslätt i Sinears land och bosatte sig där. Och de sade till varandra: ”Kom, låt oss slå tegel och bränna det.” Och teglet begagnade de såsom sten, och såsom murbruk begagnade de jordbeck. Och de sade: ”Kom, låt oss bygga en stad åt oss och ett torn vars spets räcker upp i himmelen, och så göra oss ett namn; vi kunde eljest bliva kringsspridda över hela jorden.” Då steg HERREN ned för att se staden och tornet som människobarnen byggde. Och HERREN sade: ”Se, de äro ett enda folk och hava alla enahanda tungomål och detta är deras första tilltag; härefter skall intet bliva dem omöjligt, vad de än besluta att göra. Välan, låt oss stiga ditned och förbistra deras tungomål, så att den ene inte förstår den andres tungomål.”

(Första Moseboken 11:1-7)

# INNEHÅLLSFÖRTECKNING

---

Inledning .....	5
1.1 Syfte .....	6
1.2 Frågeställningar .....	7
1.3 Metod, upplägg och avgränsningar .....	7
Bakgrund.....	10
2.1 Bibliografisk kontroll .....	10
2.2 Metadata.....	12
2.3 De nya dokumenttyperna .....	14
2.4 Katalogisering av de elektroniska dokumenten.....	15
Format för objektbeskrivning.....	19
3.1 MARC.....	24
3.1.1 <i>Formatets struktur och användning</i> .....	26
3.1.2 <i>Resonemang</i> .....	28
3.1.3 <i>Formatets flexibilitet</i> .....	30
3.1.4 <i>Interoperabilitet</i> .....	32
3.1.5 <i>MARC i praktiken: exemplet NetSök</i> .....	34
3.2 Dublin Core .....	37
3.2.1 <i>Formatets struktur och användning</i> .....	38
3.2.2 <i>Resonemang</i> .....	41
3.2.3 <i>Formatets flexibilitet</i> .....	42
3.2.4 <i>Interoperabilitet</i> .....	43
3.2.5 <i>Dublin Core i praktiken: exemplet SAFARI</i> .....	45
3.3 Samordning mellan metadataformaten .....	48
3.3.1 <i>Z39.50</i> .....	48
3.3.2 <i>RDF &amp; XML</i> .....	53
3.3.3 <i>Samordning i praktiken: exemplet CORC</i> .....	54
Metadata i biblioteket.....	56
4.1 Det elektroniska dilemma .....	56
4.2 Virtuella samlingar och digitala bibliotek .....	57
4.3 Digitala strategier .....	60
4.4 MARC eller Dublin Core? .....	62
4.5 Objektbeskrivning i heterogena samlingar .....	64
Sammanfattning och slutsatser.....	67
5.1 Formatets betydelse.....	67
5.2 Det digitala biblioteket: Babels torn <i>revisited</i> .....	69
Källförteckning.....	73
Förkortningar .....	80

## INLEDNING

---

”Noas efterkommande” skriver Alf Henriksson,

samlades på en lågslätt i Sinears land och tog itu med att bygga en stad med ett torn som skulle räcka upp till himlen. Av någon anledning misshagade detta Gud, som sade sig att efter detta första tilltag kunde människorna förväntas klara vad som helst; han steg därför ned och gjorde slut på sämjan och endräkten genom att förbistra deras språk så att allas tal lät som ett babbel för de andra. Han spridde också ut dem över jorden och förhindrade så allt internationellt samarbete.<sup>1</sup>

Låt oss spola fram filmen några tusen år. Det talas just nu mycket om ”globalt tänkande” och ”informationskultur”, och återigen inbillar vi oss att vi står i en spännande brytningstid, där såväl språkförbistring som nationella gränser är på väg att slutligt överbryggas. Det nya är att den här gången kan det tänkas att vi till viss del har rätt. Snicksnacket om att informationsteknologin i allmänhet och Internet i synnerhet skall förändra hela världen har börjat tystna en aning. Helt klart är emellertid att biblioteken och bibliotekariens roll i grunden har förändrats. Vi bibliotekarier har stigit i graderna - degraderats säger en del - från att vara samlingsvårdare till att bli informationspedagoger eller ”informationsnavigatörer”.

Information vill, och skall, få vara fri. En av bibliotekens mer ärofulla uppgifter är att bidra till att alla människor får fri tillgång till den information och den kunskap de vill ha. Uppdragets ädla natur är ett mer tacksamt diskussionsämne än de praktiska implikationer denna obegränsade tillgång medför. För som alla vet är tillgång till information i sig inget värt. I alla fall inte om informationen består i några hundra miljoner hemsidor som ligger utslängda huller om buller på vad som kallats den elektroniska motorvägen. På ett fåtal år har miljontals resurser okontrollerat publicerats på Internet, och man har aldrig stannat upp för att inventera dem. Idag är Internet gigantiskt, svårhanterligt och oöverskådligt. Det påminner mer om resultatet av en informationsexplosion än av en dito revolution. En bibliotekarie skulle, om han/hon fick arbetsuppgiften att bringa ordning i ett bibliotek med miljontals resurser kunna börja i en av två ändar. Antingen försöka att med olika medel bringa röran i något slags ordning, eller också koncentrera sig på att utveckla alltmer avancerade och ”intelligenta” sökstrategier. Tillsammans bildar dessa bägge ansträngningar underlag för den bibliografiska kontroll som gör biblioteket till något mer än bara en samling böcker. Information måste få vara fri, för detta är en grundval för vår demokrati. Men för att informationsmängden skall kunna användas, måste den också vara strukturerad.

I dagens mediasamhälle blir frågor om urval och struktur viktigare än någonsin. Vad kan några stackars bibliotekarier göra åt det här, kan man fråga sig? Förslagsvis kan de ju sluta fråga sig vad Internet kan göra för biblioteken, och istället fråga vad biblioteken kan göra för Internet. Vi bibliotekarier har en tradition vilken är nästan lika gammal som ovannämnda Gudsingripande att samla och ordna information. Bibliotekarierna blev också mycket riktigt bland de första att försöka ställa upp Internet i bokstavsordning. Men nya tider kräver nya seder, eller i alla fall nya termer.

---

<sup>1</sup> Henriksson, Alf, *Bra Böckers Hexikon. En sagolik uppslagsbok*, 1988, uppslagsord ”Babels torn”

”Objektbeskrivning av elektroniska resurser” låter nu också så mycket tjugigare än ”katalogisering av hemsidor”. Och vem vill hålla på med ”katalogiseringsregler” då man istället kan använda sig av ”metadataformat”?

Den här uppsatsen handlar om metadata för det digitala biblioteket. Vi kommer att besvara frågorna Vad är metadata? Varför är metadata och arbetet för att effektivisera metadataformaten viktigt? Vilka problem ställs metadata kreatörerna inför idag? och Hur förhåller sig biblioteken till det elektroniska materialet?

I bibliotekslagens 2 § står det att ”Folkbiblioteken skall verka för att databaserad information görs tillgänglig för alla medborgare”.<sup>2</sup> Ännu en diffus formulering, med andra ord. Det finns olika organisationer som försöker skapa ”biblioteksordning” för de elektroniska resurserna inom ett visst ämne eller med en viss målgrupp i tankarna. Frågan är om det finns någon lösning som verkar långsiktigt och passar alla elektroniska resurser, samtidigt som det är flexibelt nog att kunna användas av det enskilda biblioteket? För även om likheterna mellan bibliotekariens gamla arbetsuppgifter och de man nu står inför i det digitala biblioteket är många, finns det också stora skillnader. De elektroniska resurserna är nämligen inte bara billiga, snabba, lätta att kopiera och förvara, utan också flyktiga, svårbeskrivna och omöjliga att ha i möblerade bibliotek.

## 1.1 Syfte

*Katalogisering av elektroniska dokument ställer nya krav på biblioteken. I föreliggande uppsats avser vi att dra ett antal slutsatser om lämpliga strategier för uppbyggnaden av ett digitalt bibliotek. Som underlag för dessa kommer vi att undersöka två metadataformat som, var för sig eller kombinerade, används för katalogisering av elektroniska resurser i biblioteksmiljö, samt de samordningsmöjligheter som är aktuella i sammanhanget.*

I vår undersökning av formaten kommer vi att lägga särskild tyngdpunkt på egenskaper som bakomliggande resonemang (ambitioner vid utarbetande av formaten), flexibilitet (möjlighet till utbyggnad och anpassning till nya krav) och interoperabilitet (samverkan mellan olika metadataformat). Vi vill beskriva formaten i sitt sammanhang - det vill säga att vi kommer att diskutera vilka dokumenttyper formaten är tänkta att beskriva, vilka funktioner de är tänkta att fylla, hur de förhåller sig till varandra, samt vilken miljö de ursprungligen är skapade för. Utifrån denna analys kommer vi att diskutera olika modeller för bibliotekens hantering av det elektroniska materialet. De katalogiseringsprojekt och samordningsansträngningar vi studerat skall tjäna att exemplifiera möjliga förhållningssätt visavi de elektroniska resurserna. Vi avser vidare att utifrån vårt formatperspektiv visa på några av de problem och möjligheter som uppstår vid uppbyggnaden av ett digitalt bibliotek.

---

<sup>2</sup> Bibliotekslagen, SFS 1996:1596

## 1.2 Frågeställningar

### Formaten i sitt sammanhang

*Vilka metadataformat kan vara lämpliga att använda för bibliotekskatalogisering av elektroniska resurser?*

Hur fungerar dessa beskrivningsformat?

I vilka sammanhang är formaten tänkta att verka? Vad skall de beskriva, vilka funktioner är de tänkta att fylla?

Vilken funktion fyller beskrivningsformatet för den bibliografiska kontrollen?

### Formatens struktur och användning

*Hur är de studerade formaten uppbyggda?*

Krävs en bibliotekarie/katalogisatörs kunskaper för att använda och katalogisera i formaten?

Hur fungerar de i mötet med andra beskrivningsformat (interoperabilitet) och med nya dokumenttyper (flexibilitet, utbyggbarhet)?

Problematiken med olika format och formatdialekter kontra det världsomspännande informationsnätet.

Är det möjligt att utveckla ett standardformat som passar alla? Finns den ambitionen?

### Samordningsförsök

*På vilka sätt kan formaten tänkas interagera och samverka?*

Vad krävs för att interoperabilitet och samordning mellan olika format för objektbeskrivning skall vara möjlig?

Vilka är de samordningsverktyg som är aktuella för de studerade formaten?

Gör den senare tidens utveckling av samordningsverktyg frågan om vilket metadataformat som används onödig?

### Metadata i biblioteket

*Vilken roll spelar beskrivningsformatet i uppbyggnaden av digitala bibliotek?*

Hur kan biblioteket förhålla sig till det elektroniska materialet?

Hur skall det digitala biblioteket kunna möta användarnas behov utan att för den skull förlora den bibliografiska kontrollen över sina stora och heterogena samlingar?

## 1.3 Metod, upplägg och avgränsningar

Vår uppsats bygger i huvudsak på litteraturstudier. Vi har inte velat göra någon empirisk studie, då en sådan - utförd inom våra tidsramar - endast skulle kunna ge upplysningar av marginellt intresse för vårt uppsatsämne. Emellertid hänvisar vi till en del olika projekt, av vilka de flesta närmast kan betraktas som explorativa.

Uppsatsen inleds med en kort bakgrundsteckning av området och vår begreppsapparat. Efter denna följer en analys av två olika format för objektbeskrivning: *MACHine Readable Cataloguing* (MARC) och *Dublin Core Metadata Element Set* (Dublin Core). Analysen av formaten koncentrerar sig kring frågorna om resonemang, interoperabilitet och möjlighet att bygga ut formatet (flexibilitet). Efter denna analys fördjupar vi oss i två verktyg för samordning mellan metadataformat: *Z39.50* och *Resource Description Framework* (RDF), för att bättre visa formatens styrkor och svagheter. Vi avser även att klargöra samband och

beroenden mellan olika format och regelsystem. Användningen av formaten exemplifieras med två svenska projekt: SAFARI i vilket Dublin Core används för resursbeskrivningen, samt NetSök, där man katalogiserar sina resurser med MARC. Projekten syftar till att upprätta "bibliotekskataloger" för elektroniska resurser. Att valet föll på just dessa projekt beror delvis på att vi under höstterminen 1999 haft tillfälle att studera dessa närmare, då vi arbetat extra med att skapa Dublin Core-poster till SAFARI och korrekturläsa NetSök-poster.

I avsnittet "Metadata i biblioteket" kommer vi att vidare diskutera dessa projekts för- och nackdelar ur biblioteksperspektiv. Vi tar även upp förhållningssätt biblioteken kan ha gentemot de elektroniska resurserna. Vi diskuterar möjligheten att använda sig av de virtuella samlingar som redan finns, och vilka konsekvenser detta innebär för biblioteket. Vi utreder vidare vad som skiljer den virtuella samlingen från det digitala biblioteket, och vad man måste tänka på i uppbyggnaden av ett sådant.

Efter en sammanfattande slutdiskussion följer så en förteckning över de källor vi använt oss av. Vi har valt att här endast redovisa de dokument vi direkt citerar eller hänvisar till i notapparaten.

Såväl biblioteks- som datavärlden vimlar av svårtydda akronymer. För att inte onödigt tynga texten med förklaringar av dessa redovisar vi dem vi använder och deras uttydning i en separat lista, "Förkortningar". Givetvis skrivs termen ut första gången den omnämns, men förhoppningsvis skall listan bespara våra läsare en del blädbrände.

Biblioteksperspektivet avgör vilka miljöer vi huvudsakligen kommer att röra oss i. Vi har funnit det nödvändigt att även avgränsa oss vad gäller vilka elektroniska resurser vi skall behandla. Ett av problemen med de nya dokumenttyperna, vilket vi kommer att gå närmare in på i avsnittet "Katalogisering av de elektroniska dokumenten", är den bredd materialet har. Det kan röra sig om ljudfiler, bilder, animationer, textdokument, multimediaprodukter eller mjukvara. I värsta fall är resursen en kombination av allt detta. I huvudsak ansluter vi oss till den definition man enades om i det inledande arbetet med vad som skulle komma att bli Dublin Core. Dokumentliknande objekt eller DLO (*Document-Like Objects*) kallade man de resurser där det intellektuella innehållet presenteras huvudsakligen som text och vars bibliografiska beskrivning mycket liknar den det tryckta materialet kräver.<sup>3</sup> Senare kom Dublin Core-samhället att ytterligare vidga definitionen av ett DLO, begreppet skulle bland annat täcka bilder och videoklipp. Vi kommer att beröra arbetet med att få Dublin Core att följa med i denna utveckling, men vårt fokus förblir hos de resurser som mer liknar det "traditionella" tryckta materialet, där det intellektuella innehållet kan representeras i en pappersutskrift (till skillnad från exempelvis en ljudfil). Som vi kommer att se blir svårigheterna att korrekt katalogisera dessa resurstyper stora nog ändå. Litteraturens definition av ett DLO är emellertid så vag och vacklande att vi - inspirerade av en annan typ av svårbeskrivna objekt, om än flygande och mer tefats- än dokumentliknande - under uppsatsarbetet valt att sinsemellan kalla dessa dokumentliknande objekt för UDLO:s (*Unidentified Document-Like Objects*).

---

<sup>3</sup> Weibel, Stuart et al., *OCLC/NCSA metadata workshop report*, 1995



I uppsatsen behandlas metadataformaten MARC och Dublin Core, men dessa är på intet vis de enda möjliga formaten för objektbeskrivning av elektroniska resurser. Andra möjliga format för vår undersökning skulle vara exempelvis MCF (*Meta Content Framework*), SHOE (*Simple HTML Ontology Extensions*) eller TEI (*Text Encoding Initiative*). I de miljöer vi valt att studera, www (*World Wide Web*) och bibliotekskataloger är det dock sedan några år tillbaka MARC och Dublin Core som gäller och till vilka biblioteksvärldens förhoppningar fästs. Det finns format för beskrivning av andra dokumenttyper eller i andra miljöer än de vi vill undersöka, till exempel EAD (*Encoding Archival Description*) för arkiv; IAFA/Whois++ (*Internet Anonymous ftp Archives*) som huvudsakligen används för ftp-arkiv snarare än för www; GILS (*Government Information Locator Service*) som används för beskrivning av statligt producerade dokument och PICS (*Platform for Internet Content Selection*) som snarare används för att filtrera bort resurser (exempelvis de som är olämpliga för barn) än för att beskriva själva resursen.<sup>4</sup> Vissa av ovannämnda format kommer att beröras kortfattat, men då endast i förhållande till ”våra” format, Dublin Core och MARC.

---

<sup>4</sup> Information om andra metadataformat finner man till exempel Dempsey, Lorcan och Heery, Rachel, *Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats*, 1997 samt i IFLA:s *Digital Libraries: Metadata Resources*

## BAKGRUND

---

Michael Day vid *UK Office for Library and Information Networking* (UKOLN) skriver ”Internet has many virtues, but it – and in particular the World Wide Web – was not designed specifically for information retrieval.”<sup>5</sup> I detta avsnitt kommer vi att gå genom begreppet bibliografisk kontroll och kort beskriva Internets explosiva utveckling och den utmaning detta bångstyriga medium i allmänhet och de nya dokumenttyperna i synnerhet har inneburit och fortfarande innebär för bibliotekarierna. Vidare kommer vi att introducera metadata som en av bibliotekskatalogens grundstenar.

### 2.1 Bibliografisk kontroll

Bibliografisk kontroll har definierats som konsten att organisera information så att man kan återfinna den.<sup>6</sup> Det handlar alltså om att på något vis inordna dokument i en katalog. Man talar emellertid inte längre så mycket om katalogisering som om bibliografisk kontroll eller *bibliographic management*. För att uppnå denna behöver man en informationsstruktur vilken förutom att erbjuda återvinningsmöjligheter även bereder plats åt framtida dokument (skalbarhet). För att det skall vara möjligt att återvinna informationen måste den beskrivas. Beskrivningsmallen skall helst vara utformad med avseende på förutsägbarhet, logik och standardisering. Detta sista gäller även det innehåll man fyller mallen med, för att bibliografisk kontroll skall vara möjlig är det viktigt att man använder sig av något slags kontrollerad vokabulär. Hänsyn måste vidare tas till användarnas behov, det vill säga vad man skall använda informationsstrukturen till, vilka funktioner den skall fylla.

Bibliotekskatalogen bygger på en sådan formaliserad beskrivning av dokumenten. Beskrivningsmallen gör att alla dokument representeras på ett likartat, jämförbart och förutsägbart sätt. Modern deskriptiv katalogiseringsteori och -praktik har under de senaste 150 åren utvecklats för att organisera information i bibliotek så att den kan återvinnas på bästa möjliga sätt. Som den första moderna regelsamlingen för bibliotekskatalogisering brukar *British Museums* katalogregler, eller Panizzis kod räknas. Koden fastställdes 1839, och kom ut i reviderad upplaga så sent som 1936. 1876 utkom *Rules for a printed dictionary catalogue*, vilken innehöll Cutters berömda regler. I dessa definierar han ändamålen med en katalog och de medel genom vilka dessa kan uppnås. I enlighet med Cutters regler har bibliotekarier de senaste hundra åren arbetat med huvuduppslag vilka gör det möjligt för informationssökaren att finna ett dokument där antingen verkets författare, titel eller ämne är känt. Datortekniken har sedermera gjort det möjligt att veta mindre än så och ändå få träffar vid sökning.

De regler bibliotekskatalogen följer har utformats med avseende på bibliotekets heterogena samlingar. Förutom det traditionella tryckta material man kanske i första hand tänker på, innehåller dessa även tidskrifter, kartografiskt material, musiknoter, småtryck, handskrifter med mera. Möjligheterna att bygga ut katalogen är också viktiga, och man försöker utveckla katalogiseringsreglerna för att klara av nya medier

---

<sup>5</sup> Day, Michael, ”Extending metadata for digital preservation”, *Ariadne*, 1997

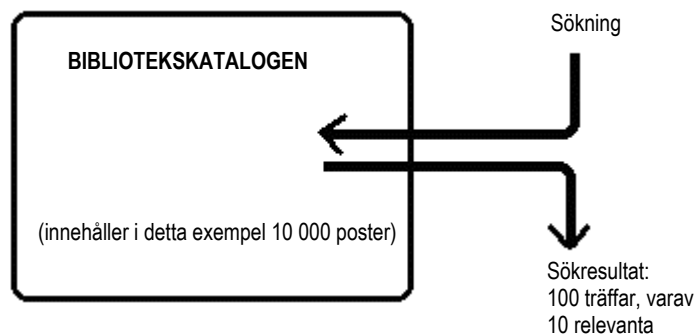
<sup>6</sup> Benito, Miguel, *Bibliografisk kontroll*, 1995, s 3

när de blir aktuella för bibliotekens samlingar. Som sådana nya medier kan vi nämna microfiche, musikinspelningar, videofilmer, filmer och datafiler som inte alltid varit självklara delar av bibliotekets samlingar. En riktigare term än dokumentbeskrivning skulle i ljus av detta kanske vara ”objektbeskrivning”.

För att uppskatta hur god bibliografisk kontroll en informationsstruktur eller katalog tillåter brukar man använda termerna *precision* och *recall*. *Precision* mäter antalet relevanta återfunna dokument i förhållande till hela antalet återfunna dokument och anger hur mycket ”brus” sökresultatet innehåller. Resultatet uttrycks enligt formeln

$$p = \frac{\text{relevanta återfunna dokument}}{\text{återfunna dokument totalt}} \times 100\%$$

I exemplet nedan ser vi hur en sökning resulterar i 90% brus.

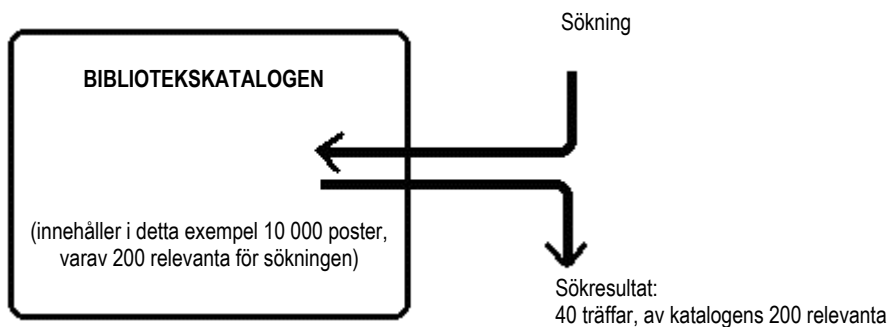


**Figur 1 Exempel på dålig precision (10%)**

*Recall*, eller återvinningsgrad mäter antalet relevanta återfunna dokument i förhållande till det totala antalet relevanta dokument i samlingen, och anger hur uttömmande sökningen är. Detta uttrycks enligt

$$r = \frac{\text{relevanta återfunna dokument}}{\text{relevanta dokument i samlingen}} \times 100\%$$

Sökningen nedan ger 40 av 200 relevanta dokument, vilket innebär att man missar 80% av de dokument som skulle varit intressanta.



**Figur 2 Exempel på katastrofalt dålig återvinningsgrad (recall 20%)**

Detta är klassiska mått på en informationsstrukturs effektivitet. I enlighet med dessa värden kan man också justera katalogiseringsrutinerna. Kanske är dokumenten för allmänt beskrivna, vilket ger många träffar men dålig precision? Kanske gör ofullständig objektbeskrivning att viktiga dokument inte dyker upp på träfflistorna, och sänker katalogens återvinningsgrad? För www med dess enorma mängder dokument kan värdena för precision endast med svårighet mätas, och för *recall* inte alls mätas, då ”relevanta dokument i samlingen” inte kan beräknas.<sup>7</sup>

Katalogisering och klassifikation har alltmer kommit att erkännas som en grundläggande och viktig del av biblioteksarbetet. Fred Ayres menar i sin artikel ”What is the future for catalogues and cataloguers?” att bibliotekarieyrkets framtid hänger på hur man väljer att utveckla tekniker för bibliografisk kontroll. Han förutspår att frågor kring detta kommer att fokuseras än mer under de kommande decennierna.<sup>8</sup> Detta är ett inte alltför vågat antagande, då det i alla tider varit bibliotekariens uppgift att finna, strukturera och värdera information. Teknikens framsteg har fört med sig ett informationsöverflöd som gör denna sällning och strukturering, denna bibliografiska kontroll, viktigare än någonsin. Och bibliografisk kontroll uppnår man med hjälp av metadata.

## 2.2 Metadata

Metadata kan förstås som data om data, eller hellre information om information. Ordet metadata har kommit att alltmer knytas till datateknik i allmänhet, och till beskrivning av internetresurser i synnerhet. Vi menar att all information som beskriver och därmed representerar ett objekt är en form av metadata. I sin uppsats *Er lykken gjort?* skriver Rasmus Bruun

Indførslerne i en bibliografi eller en katalog er således metadata, idet de beskriver de egentlige dokumenter. Men metadata er meget andet. Når man læser anprisningerne på bagsiden af en roman, læser man metadata.<sup>9</sup>

Ett utmärkt exempel på metadata är det gamla klassiska katalogkortet, utan vilket generationer av biblioteksbesökare för alltid förlorat sig bland bibliotekets hyllor.

Då man skall ordna och strukturera stora mängder information på ett bibliotek använder man sig alltså av metadata. Angående det kaotiska Internet versus den ordnade bibliotekskatalogen skriver Lorcan Dempsey och Rachel Heery vid UKOLN att ”it is recognised that in an indefinitely large resource space, effective management of networked information will increasingly rely on effective management of metadata.”<sup>10</sup> Metadata är den information om objekten som ”relieves their potential users of having to have full advance knowledge of their existence or characteristics.”<sup>11</sup> Med andra ord den tilläggsinformation som, ofta på ett strukturerat och standardiserat sätt, beskriver objektets egenskaper på ett sätt som gör det möjligt att inordna och återfinna det i en samling. Som resonemangen kring precision-*recall* låtit antyda ger

---

<sup>7</sup> Formlerna för *precision – recall* hämtade från Benito, 1995, s 106f

<sup>8</sup> Ayres, Fred, ”What is the future for catalogues and cataloguers?”, *Catalogue & Index*, 1996, s 2

<sup>9</sup> Bruun, Rasmus, *Er lykken gjort? – om metadata som indexeringsgrundlag ved bibliografisk kontrol af webdokumenter*, 1997

<sup>10</sup> Dempsey, Lorcan och Heery, Rachel, *Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats*, 1997, s 2

<sup>11</sup> Dempsey och Heery, 1997, s 5

en rikare och mer strukturerad objektbeskrivning möjlighet till mer preciserade frågeställningar.

I litteraturen finner vi många olika definitioner av begreppet metadata. Att det handlar om "information om information" är de flesta helt ense om, men då man kommer in på vilket slags information det handlar om går åsikterna isär. Priscilla Caplan ansluter sig till vad vi konstaterade i uppsatsens inledning, "nya tider kräver nya termer"

Metadata really is nothing more than data about data; - - - We could call it cataloging, but for some people that term carries extra baggage, like *Anglo-American Cataloging Rules* and USMARC. So to some extent this is a "you call it corn, we call it maize" situation, but metadata is a good neutral term that covers all the bases.<sup>12</sup>

Grundstenar i konstruktionen av funktionella metadata är enkelhet, flexibilitet, interoperabilitet och standardisering. Enkelheten och flexibiliteten garanterar att formatet är lätt att katalogisera i, och samtidigt som det uppfyller allmänna beskrivningskrav även kan utvecklas så att det tillgodoser den enskilda instansens (exempelvis ett biblioteks) speciella beskrivningsbehov. Standardiseringen garanterar i sin tur att katalogiseringsarbetet inte behöver centraliseras, då alla som katalogiserar följer standarden. Standardiseringen är vidare en viktig förutsättning för att interoperabilitet med andra samlingar skall vara möjlig. Det funktionella metadataformatet tillåter olika instanser att dela och kopiera kataloger.

I biblioteksvärlden är fördelarna med metadata sedan länge kända. Hittills har vi huvudsakligen talat om de fördelar användning av metadata medför vid sökning av resurser. Information om information kan dock fylla även andra funktioner. Den kan vara behjälplig för att presentera resursen, för att identifiera och lokalisera ett specifikt objekt, och i samband med de elektroniska resurserna även tjäna att länka katalogen direkt till de beskrivna objekten. Beroende på vilken av dessa funktioner man prioriterat har olika format för metadata utvecklats, format där enkelheten ibland fått stryka på foten för krav på noggrannhet och detaljrikedom, eller där specifika förutsättningar gjort standardisering omöjlig. I biblioteksvärlden är standardisering och interoperabilitetsmöjligheterna kanske de viktigaste kriterierna.

Jörgen Eriksson, LUB NetLab, konstaterar att

There are two main approaches to metadata creation. One is to use elaborated and specialized schemes like the MARC-record. This requires time, money and qualified staff and will probably only be used on a small, choice selection of valuable and stable resources. The other approach is to create a simple record that would be possible for the author of a document to create.<sup>13</sup>

Utvecklingen av datatekniken har inneburit ett stort genombrott för möjligheterna att till fullo utnyttja metadata. Information kan snabbt och enkelt kopieras, skickas och användas i nya sammanhang. För slutanvändaren har utvecklingen av dataprogram som hanterar metadata erbjudit snabbare och bättre möjligheter att ställa precisa frågor och erhålla bättre svar. Den lokala bibliotekskatalogen är inte längre självklart

---

<sup>12</sup> Caplan, Priscilla, "You call it corn, we call it syntax-independent metadata for document-like objects", *The public-access computer systems review*, 1995

<sup>13</sup> Eriksson, Jörgen, *Metadata – the description of electronic objects, documents and services*, 1997

knuten till en fysisk samling, och de framväxande fulltextdatabaserna är det närmaste vi kan komma den uråldriga drömmen om ett världsbibliotek. Redan på den gamla goda tiden var det emellertid svårt för katalogisatorerna att hinna indexera den mängd material biblioteken tillhandahöll. Den informationsexplosion Internet inneburit ifrågasätter deras noggranna arbete, och mer eller mindre tvingar fram alternativa lösningar.

## 2.3 De nya dokumenttyperna

Internets historia börjar 1969 med *Advanced Research Projects Agency Network* (Arpanet). Som ett försvarsvapen i det kalla kriget kopplades fyra amerikanska datorer samman i ett militärt experiment. Sammankopplingen skulle parera risken för informationsförlust genom att datorerna kommunicerade på ett nytt sätt; informationen skickades ”paketvis” och inte via en fast linje, utan på vilken väg som helst som för tillfället var framkomlig. Tekniken kom sedan att användas för att länka samman de stora universitetens datorer med varandra, och blev en viktig kanal för forskarkontakt. I 1970-talets USA får Internet sitt namn genom skapandet av *Transmission Control Protocol/Internet Protocol* (TCP/IP), den standard för kommunikation mellan datorer som används på Internet. De första e-postmeddelandena skickades 1971 och andra protokoll lanserades, till exempel ftp (*File Transfer Protocol*) för överföring av filer.

Under 1980-talet ökar antalet anslutna länder. 1988 anslöts det svenska universitetsnätverket SUNET till Internet via nordiska samarbetet NORDUNET. Möjligheterna till ett grafiskt gränssnitt kom först 1989 med markeringsspråket HTML (*Hyper Text Markup Language*), protokollet http (*Hyper Text Transport Protocol*) som används när två datorer skickar hypertext mellan varandra och de webläsare som behövs för att ta del av hypertextmiljön. I och med detta skapades Internets grafiska del www (*World Wide Web*), som blev Internets genombrott för de breda massorna. Frammot mitten av 1990-talet har www vuxit sig ofantligt stor. I februari 1999 uppskattades antalet indexerbara sidor på www att uppgå till 800 miljoner, vilket motsvarar en tillväxt med nästan 500 miljoner sidor sedan januari 1997.<sup>14</sup> Det är idag populärt att publicera information på www, och det är mer regel än undantag att såväl företag, organisationer som enskilda personer, har egna hemsidor där man ”marknadsför” sig och informerar. Med tiden har interaktionsmomentet ökat, och nu finns det exempelvis företag som sköter all försäljning och marknadsföring över Internet.

Internet betraktas av många som den största informationsrevolutionen sedan Gutenberg började trycka sin bibel. På Internet kan vem som helst lägga ut i princip vad som helst, utan den flaskhals pressläggningstider, ogina förlag och brist på allmänt medhåll innebär. Givetvis är detta en sanning med modifikation. Många har varken medel eller möjlighet att vara med om IT-revolutionen, och att då tala om Internet som ett uttryck för sann demokrati är måhända en aning storvulet. Att det publicistiska klimatet drastiskt har förändrats är emellertid inte att säga för mycket. Med uppsatsens tidigare avsnitt i ryggen inser läsaren dock omedelbart att informationsrevolutionen inte är enbart en välsignelse. Den bibliografiska kontroll vi tidigare talat om är som sagt mycket svår att uppnå i det anarkistiska Internet-klimatet. Resurserna visar här inga större likheter med det statiska, ”traditionella”

---

<sup>14</sup> Lawrence, Steve och Giles, C. Lee, ”Accessibility of information on the web”, *Nature*, 1999, s 107

biblioteksmaterialet. Det finns inget centralt beslutande organ, vilket medför publicistisk frihet, men också enorma svårigheter att få någon överblick över det samlade materialet.

Den dynamiska www-resursen kan också flyttas, förändras, förekomma i många olika format, förekomma i många olika (odaterade) versioner, bestå av många olika sammanlänkade sidor, ha många olika författare et cetera.

Problemen med www-resursernas flyttbarhet och flyktighet, vilket medför risken att länkar plötsligt blir ogiltiga, adresseras av bland annat identifikationssystem som DOI (*Digital Object Identifier*). Dessa system genererar permanenta och globalt unika koder till ett objekt. Till skillnad från URL:en (*Uniform Resource Locator*) har dessa koder ingen koppling till objektets tillfälliga adress. Så länge resursen över huvud taget finns utlagd skall det alltså vara möjligt att nå det via dess DOI och identifiera den med hjälp av dess URN.<sup>15</sup> De kan sägas motsvara de tryckta publikationernas ISBN (*International Standard Book Number*).

För det elektroniska materialet talar man vidare om dess konvergens, det vill säga att olika medieformer växer samman. Ett exempel på detta är hur de elektroniska versionerna av dagstidningar alltmer använder filmade nyhetsklipp och ljudinspelningar som inte tillhör det traditionella tidningsmediet. Det elektroniska materialet kännetecknas även av alltmer tilltagande individualisering. Med detta avser man de möjligheter som finns för konsumenten att ”skraddarsy” den information som kommer honom/henne till del. Även här får de elektroniska nyhetstidningarna tjäna som exempel. Här kan man exempelvis välja att få vissa geografiska områden specialbevakade, eller att få se en fotbollsmatch ur en speciell kameravinkel et cetera.<sup>16</sup>

Detta är några av de fenomen som bereder speciella problem för den stackare som skall beskriva objektet. Katalogisatorer världen över sliter sitt hår. Och då de slutat med det biter de samman tänderna och försöker lösa problemen.

## 2.4 Katalogisering av de elektroniska dokumenten

Fenomenet Internet har bidragit till den rollförändring bibliotekarier världen över upplevt de senaste åren. I en tid som redan före IT-revolutionen präglades av ett informationsöverflöd har bibliotekarierna från att kanske främst betraktas som kulturens och samlingarnas väktare alltmer kommit att ses som ett slags informationspedagoger. Svårigheten är inte längre att hitta information - vi står faktiskt redan upp till knäna i den - svårigheten är att lära sig navigera i den. Internet har även tillfört helt annorlunda metoder för bibliotekens anskaffning, lagring, bearbetning och förmedling av information. Mycket har blivit tekniskt enklare, allt går snabbare, men samtidigt innebär utvecklingen ett ständigt ifrågasättande av beprövade metoder. Aspekter som kvalitet och sökningars *precision - recall* har blivit närmast omöjliga att uppskatta. Man har alltså ingen möjlighet att kontrollera om man har fått ett relevant svar på sin sökning i sökmotorer och databaser. Mängden av dokument har gjort att kontrollen sedan länge glidit användarna ur händerna.

---

<sup>15</sup> Se exempelvis Lindquist, Mats G., ”Not your father’s references: Citations in the digital space”, *The journal of electronic publishing*, 1999

<sup>16</sup> SOU 1998:111, *E-plikt. Att säkra det elektroniska kulturarvet*, 1998, s 59

Det är dock helt klart att biblioteken kan dra nytta av Internet, att de faktiskt redan nu inte kan vara utan den mängd information som ligger där (och ofta endast där). Vi ställde inledningsvis den retoriska frågan vad biblioteken kan göra för Internet. Som vi ser det är det ganska mycket.

Vad skiljer det jäsande Internet från det välordnade bibliotek vi känner och älskar? Det som först faller en bibliotekarie i ögonen är väl den totala bristen på bibliografisk kontroll och effektiv användning av metadata. Katalogisatören Amanda Xu vid Massachusetts Institute of Technology sammanfattar fördelarna med att applicera metadata på det elektroniska materialet i fyra punkter

- Metadata erbjuder ett enkelt sätt att hantera stora mängder information. Katalogerna behöver endast behandla dokumentrepresentationen, och inte det fullständiga dokumentet.
- Metadata möjliggör avancerade sök- och urvalsprocesser, då dokumentrepresentationen eller dokumentsurrogatet är utformat särskilt för detta.
- Standardiserade metadataformat tjänar att integrera även mycket heterogent material, också sådant från olika databaser. Detta är ovärderligt, då det gör det möjligt för de olika baserna att jämföra, dela, integrera och återanvända både dokument och katalogposter.
- Metadata kan inte bara användas för att välja, utan också för att välja bort, och sålunda fylla en viktig funktion som *gatekeeper*. Med hjälp av metadata kan man filtrera bort oönskat material, informera om dokumenträttigheter et cetera.<sup>17</sup>

Metadata kan vara antingen *embedded* (inklistrad eller inbäddad) i dokumentet, eller fristående. Inbäddad metadata används bland annat för att göra automatgenererad indexering bättre. Vid katalogisering av www-resurser är det vanliga att man placerar – ”klistrar in” – metainformationen i HTML-dokumentets ”huvud” (mellan <HEAD> och </HEAD>). Här kan dokumentproducenten själv lägga sin beskrivning, som då den är en fysisk del av resursen kan den hämtas och läsas av många olika sökmotorer. Metadata kan även användas fristående. Dokumentbeskrivningen är i detta fall separerad från dokumentet, och fungerar som dokumentsurrogat i exempelvis en katalog eller databas. Ett annat exempel på fristående metadata är de traditionella katalogkorten.

Redan 1996 skrev Fred Ayres

Just as the information held in the library must be subject to bibliographic management so the information held on the Internet must eventually come under some sort of control if its full potential is to be realised.<sup>18</sup>

Här har biblioteken redan hård konkurrens från andra informationsförsörjande tjänster, vilka använder modifierad bibliotekskatalogisering för att strukturera sina samlingar. Även den simplaste sökmotor uppställer ju ett slags katalog eller index. ”Katalogposterna” kan i detta fall vara dokumenten mer eller mindre i sin helhet. Denna metods svagheter torde vara uppenbara, och vi kommer att vidare diskutera dem i nästkommande avsnitt. Det gör det också möjligt för den enskilde, inte alltför seriöse hemsideskaparen att få fler träffar på sina föreläsningssanteckningar i

---

<sup>17</sup> Xu, Amanda, *Metadata conversion and the library OPAC*, 1997

<sup>18</sup> Ayres, 1996, s 5



molekylärbiologi genom att helt enkelt upprepa frasen ”*kinky sex*” ett hundratal gånger i början av dokumentet. Denna vilseledande information kan också läggas i sidans kod, vilket gör att besökaren oftast inte har en chans att förstå vad som är så ”*kinky*” med molekylärbiologi. Vissa sökmotorer försöker vara lite intelligentare, och räknar orden, hur nära de ligger varandra, hur tidigt i dokumentet de förekommer och så vidare. En liknande metod har även använts för att klassificera Internetresurser. Arbetet med automatiserad klassifikation är emellertid ännu på experimentstadiet, och verkar passa bäst för ämnesbaserade samlingar.<sup>19</sup> Automatiskt skapad metadata kan kanske minska mängden brus något, men innan forskningen om artificiell intelligens kommit längre får vi acceptera att maskinerna måhända är snabba, men att de inte besitter någon som helst bedömningsförmåga.

Vanligt är numera att de virtuella samlingarna strävar att bygga upp en funktionell informationsstruktur. Resurserna placeras här i ett slags ämnesbaserad hierarki, ofta med en särskild målgrupp i tankarna. Yahoo! är ett exempel på en virtuell samling. Vi kommer att vidare diskutera termen ”virtuella samlingar” ställd mot ”digitala bibliotek” i avsnittet ”Metadata i biblioteket”.

Frågan om granularitet, det vill säga hur långt ned i siterstrukturerna man skall bry sig om att ge varje sida en fullständig beskrivning, är något som kan variera från fall till fall. En vettig jämförelse från vedertagen katalogiseringspraxis är här boken och den vetenskapliga tidskriften. I en bok får inte varje kapitel sin egen katalogpost. Är boken en essäsamling kanske man kostar på en indexering av kapitelrubrikerna och författarna. På ungefär samma sätt gör man med den vetenskapliga tidskriften. Då dessa alltmer började utgöra ryggraden i forskningen - åtminstone den naturvetenskapliga - blev biblioteken tvungna att ställa sig frågan huruvida man hade de ekonomiska resurser som krävdes för att kosta på alla dessa artiklar sina egna katalogposter. Det hade man oftast inte, utan fick nöja sig med att lista artikelrubrikerna (mindre lustigt för de forskare som valt en skämtsam eller mångtydig titel på sitt arbete). Fullständiga poster för de vetenskapliga publikationernas artiklar kan man emellertid finna i ämnesspecifika artikelförteckningar eller databaser.

Priscilla Caplan, verksam vid universitetsbiblioteket i Chicago, skrev redan 1993

How do we accommodate networked electronic information when our cataloging rules are designed to describe physical items owned by and residing in libraries? How do we provide access to that information? - - - Questions like these must be addressed before we can move into the future and provide our patrons with information the way they are coming to expect it.<sup>20</sup>

Ursprungligen var det vanligt att man försökte beskriva material publicerat över de nätverk som biblioteken var kopplade till - närmast Internet - med samma regler som biblioteksbundet material. Den mängd resurser att beskriva www-explosionen resulterade i fick dock många att tänka om.<sup>21</sup>

---

<sup>19</sup> Koch, Traugott och Ardö, Anders, ”Automatic classification of full-text HTML-documents from one specific subject area”, 2000

<sup>20</sup> Caplan, Priscilla, ”Cataloging Internet resources”, *The public-access computer systems review*, 1993, s 61

<sup>21</sup> Chapman, Ann, Day, Michael och Hiom, Debra, ”Metadata: Cataloguing practice and Internet subject-based information gateways”, *Ariadne*, 1998

I denna bakgrundsteckning har vi tagit upp begreppen bibliografisk kontroll och metadata, samt diskuterat de speciella svårigheter det elektroniska materialet ställer katalogisatören inför. Som avslutande ord väljer vi att med Clifford Lynch, CNI, konstatera

In short, the Net is not a digital library. But if it is to continue to grow and thrive as a new means of communication, something very much like traditional library services will be needed to organize, access and preserve networked information.<sup>22</sup>

---

<sup>22</sup> Lynch, Clifford, "Searching the Internet", *Scientific American*, 1997

## FORMAT FÖR OBJEKTBESKRIVNING

---

The need for metadata services is already clear in the current Internet environment.  
- - - Metadata is not only key to discovery, it will also be fundamental to effective use of found resources ... and to interoperability across protocol domains.<sup>23</sup>

Som vi tidigare noterat har uttrycket ”metadata” tenderat att knytas till objektbeskrivning av elektroniska resurser.<sup>24</sup> I sin artikel med den retoriska titeln ”Cataloguing vs. metadata: Old wine in new bottles?” menar Stefan Gradmann, PICA, att termen ”metadata” uteslutande skall användas för objektbeskrivning av elektroniskt material. Vidare menar han att denna och mer traditionell bibliotekskatalogisering har så lite gemensamt att bibliotekariens roll i försöken att skapa ordning bland www-resurserna närmast kan beskrivas som konsultens. Genom att redovisa var det gått snett i utvecklingen av bibliotekens omhuldade MARC-format skall bibliotekarierna kunna förhindra att samma fel upprepas i arbetet med metadataformaten.<sup>25</sup> Vi är medvetna om att olikheterna resurserna emellan är stora och bereder särskilda problem för den som skall beskriva dem, och vi kommer att närmare diskutera detta i de två följande avsnitten. De specifika krav det elektroniska materialet ställer på sitt beskrivningsformat är en central fråga i föreliggande uppsats. Vi tror oss dock inte ha något att vinna på att trycka på olikheterna formaten emellan, utan avser snarare att betona likheterna. Först då kan vi få den helhetssyn på beskrivningsformaten som gör jämförelser fruktbara. Vi håller oss därför till vår något bredare definition av begreppet metadata.

I bibliotekskatalogen fungerar varje post som ett dokumentsurrogat. Posten består av en beskrivning av dokumentet, en beskrivning som ofta är högt strukturerad och innehåller olika beskrivningselement, exempelvis ”titel”, ”författare” och ”ämnesord”. Den deskriptiva representationen utgör erbjudet oftast många olika sökmöjligheter. Som motpol kan vi ställa den fulltextindexering vi tidigare berört. Redan 1997 skriver Jörgen Eriksson

When the possibilities of fulltext-searching of documents arose, there was a trend that this would be enough to get precise information retrieval. Today, with search engines indexing 50 milj. [sic!] URLs, and the number of relevant documents only too often being a fraction of the total number retrieved from a search, there is a growing trend towards improvement by using old library tools, such as cataloging and indexing.<sup>26</sup>

I teorin verkar fulltextindexering vara det optimala, eftersom hela dokumentet är sökbart, och informationssökaren alltså har möjlighet att finna alla de dokument som nämner det begrepp han eller hon är intresserad av. I praktiken blir nackdelarna snart uppenbara. Det är inte bara det att den mängd dokument sökmotorerna indexerat gör träfflistorna fullständigt ohanterliga, det är också så att utan hjälp av ett strukturerat metadataformat är det omöjligt för en sökmotor att avgöra om ”Gandhi” är ämnet för,

---

<sup>23</sup> Dempsey och Heery, 1997, s 5

<sup>24</sup> Se exempelvis Eriksson, 1997 eller Heery, Rachel, ”Review of metadata formats”, *Program*, 1996

<sup>25</sup> Gradmann, Stefan, ”Cataloguing vs. metadata: Old wine in new bottles?”, 64<sup>th</sup> IFLA general conference, 1998

<sup>26</sup> Eriksson, 1997

titeln på eller författaren till en text. Alla som någonsin använt sig av Internet är väl förtrogna med detta fenomen och de kniviga frågor som uppstår då sökresultaten levereras. Är en privat hemsida om författarens motvilja mot Microsofts produkter lika intressant som företagens årsredovisning? Varför kommer man till Volvos hemsida då man söker på SAAB? Att ett ord har hög frekvens i en text är dessutom i sig ingen garanti för att texten i vanlig mening ”handlar” om detta. Och tvärtom, ”Metadata cannot only improve precision, it can increase recall - - - a document will be retrieved from properly applied metadata even if it never uses the controlled term in its text.”<sup>27</sup> Vi känner igen tankarna från det föregående avsnittet. Den strukturerade dokumentbeskrivningen, metadataformatet, är alltså ett sätt att förbättra sökningens *precision - recall*, och minska sökresultatets delmängd brus.

Inom DESIRE-projektet (*Development of a European Service for Information on Research and Education*) utfördes 1997 en grundlig undersökning av olika metadataformat för beskrivning av elektroniska resurser.<sup>28</sup> DESIRE-rapportens författare, Lorcan Dempsey och Rachel Heery, har valt att gruppera formaten i tre olika ”band” (eller grupper) enligt följande

	<b>BAND 1</b>	<b>BAND 2</b>	<b>BAND 3</b>
<i>Kännetecken för formaten -</i>	Enkla format  Anpassade efter sammanhanget  Fulltextindexering	Strukturerade format  Framväxande standarder  Grundläggande fältstruktur	”Rika”, fullödiga format  Internationellt standardiserade  Strängt reglerade fält
<i>Format - (exempel)</i>	AltaVista  Yahoo!	Dublin Core  IAFA/Whois++	MARC  TEI

**Figur 3. Typologi över metadataformat (efter DESIRE)<sup>29</sup>**

Grundstenarna för uppbyggnad av ett effektivt metadataformat måste vara flexibilitet och om möjligt även enkelhet. Formatet skall dessutom gärna vara standardiserat och tillåta interoperabilitet. Givet detta finns det som synes ändå mycket utrymme för variationer. Som vi berört kan metadata ha många olika funktioner. De viktigaste är sökning, identifikation, urval av specifika objekt samt åtkomst (lokalisering och hämtning). Det är sammanhanget som avgör vilka funktioner som prioriteras.

Band 1 innehåller i detta fall relativt enkla format. Det är här oftast inte möjligt att söka i olika fält, och fulltextindexeringen gör att hela dokumentet (eller i vissa fall de

<sup>27</sup> Milstead, Jessica och Feldman, Susan, ”Metadata: Cataloging by Any Other Name”, *ONLINE*, 1999a

<sup>28</sup> Dempsey och Heery, 1997 Anm. Denna utvärdering användes även av Nordiska metadataprojektet, då man ansåg den så fullödig att det var onödigt att göra en egen. (Hakala, Juha et al., *The Nordic metadata project. Final report*, NORDINFO-rapport, 1998:1, NORDINFO, Helsingfors 1998, s 10). Ännu så sent som 1999 hänvisar Hakala till den i sin artikel ”Internet metadata and library cataloging” (s 22) så trots sin i Internetsammanhang höga ålder äger undersökningen ännu relevans.

<sup>29</sup> Dempsey och Heery, 1997, s 8

inledande meningarna) får fungera som sin egen katalog- eller metadatapost. Här finner vi AltaVista, vars automatiska indexering är ett kraftfullt, men trubbigt verktyg. Yahoo! har haft stor framgång med de enkla hierarkier man byggt upp. De metadata denna virtuella samling använder sig av är inte maskingenererade, utan skapade av människor. Yahoo! kritiserar ofta för sin mycket allmänna ämnesstruktur - ”Hälsa och sjukvård”, ”Konst och kultur”, ”Utbildning” - men som Tim Bray, W3C, skriver ”Yahoo! is pitiful; but its popularity is clear evidence of the power of (even limited) metadata.”<sup>30</sup>

Band 2 innehåller de så kallade ”strukturerade” formaten. Juha Hakala vid universitetsbiblioteket i Helsingfors kallar dem ”an interesting compromise between simple and rich formats.”<sup>31</sup> De strukturerade formaten medger sökning i olika fält (exempelvis sökning endast i titelfält). Detta skall i princip göra det möjligt för informationssökaren att redan utifrån metadataposten kunna avgöra huruvida resursen är intressant eller inte. Givetvis beror detta också på hurpass specifik information man söker efter. Formaten i band 2 skapas för det mesta manuellt. Om posterna skapats maskinellt krävs det oftast mänsklig inblandning för att förbättra dem. Det för gruppen typiska formatet kräver något mer än vad maskiner kan åstadkomma. Däremot behövs det inga specialistkunskaper för att skapa poster.

I band 3 finner vi de format som kan användas inte enbart för identifikation och åtkomst av olika resurser, utan även kan ses som ”dokumentsurrogat” i ordets verkliga bemärkelse. De rikaste formaten beskriver snart sagt varje egenskap dokumentet och dess innehåll har. Produktion av denna sorts metadata kräver ingående kunskaper om formatet och de regler som styr fältinnehållet.

Om bibliotekariekåren fritt fick styra och ställa är det av allt att döma ingen tvekan om vilken formatgrupp den skulle välja då hela världen sorterades in i bokstavsordning. Olyckligtvis är de rikare formaten inte bara en tidskrävande specialistangelägenhet, utan också en ganska dyrbar historia. Även om man lyckas drilla dokumentproducenterna att själv göra en noggrann och uttömmande beskrivning av de resurser de publicerar, eller om man lyckas skapa program som utvinner metadata ur dokumenten på automatisk väg – och ingetdera alternativet är egentligen intressant för biblioteket – ”the creation of full, structured metadata will remain expensive, wherever along the use chain that cost falls.”<sup>32</sup>

Som vi nämnt skall metadataformatens strukturerade katalogposter användas till mycket annat än att förbättra sökresultaten. Den bibliografiska posten används också vid inköp, katalogisering, inventering, cirkulation, fjärrlån, bevarande, som referens och för informationsåtervinning. Samma bibliografiska post skall också helst kunna återanvändas i exempelvis bibliografiska databaser och förlagskataloger, förutom att den tillgodoser bibliotekskatalogens krav. Inom katalogiseringssteori och -praktik är *International Federation of Library Associations and Institutions* (IFLA:s) studie *Functional Requirements for Bibliographic Records* från 1997 en viktig utgångspunkt för gemensam förståelse av hur den bibliografiska postens olika beståndsdelar svarar mot de krav som ställs på den. Även om IFLA:s studie inte fyllde någon annan funktion, skulle den vara oundgänglig för sitt klara sätt att spalta upp de begrepp som

---

<sup>30</sup> Tim, *RDF and Metadata*, 1998

<sup>31</sup> Hakala, ”Internet metadata and library cataloguing”, 1999, s 22

<sup>32</sup> Dempsey och Heery, 1997, s 7

utgör grunden för praktisk bibliografisk kontroll. Undersökningen är utförd mot bakgrund av bibliotekens verklighet: nödvändigheten av att ekonomisera katalogiseringen samtidigt som den fortsätter uppfylla användarbehoven, och detta i en värld där dokumenttyperna att beskriva blir fler och fler. I studien rekommenderas följaktligen en basnivå (*minimal core level*) av funktionalitet för poster i nationalbibliografier och bibliotekskataloger. Man utgår från användarbehoven sökning, identifikation, urval och åkomst. För att förtydliga innebörden av dessa hänvisar vi till IFLA:s egna definitioner

---

**Sökning** (*find*)

using the data to find material that corresponds to the user's stated search criteria (e.g., in the context of a search for all documents on a given subject, or a search for a recording issued under a particular title);

**Identifikation** (*identify*)

using the data retrieved to identify an entity (e.g., to confirm that the document described in a record corresponds to the document sought by the user, or to distinguish between two texts or recordings that have the same title);

**Urval** (*select*)

using the data to select an entity that is appropriate to the user's needs (e.g., to select a text in a language the user understands, or to choose a version of a computer program that is compatible with the hardware and operating system available to the user);

**Åtkomst** (*acquire/obtain*)

using the data in order to acquire or obtain access to the entity described (e.g., to place a purchase order for a publication, to submit a request for the loan of a copy of a book in a library's collection, or to access online an electronic document stored on a remote computer).<sup>33</sup>

---

De olika typer av information en bibliografisk post innehåller relateras i studien till användarbehoven, och listas efter hurpass viktiga de är för de olika uppgifterna. IFLA skiljer i sammanhanget noga mellan att exempelvis söka ett verk (*work*), ett uttryck (*expression*), en manifestation (*manifestation*) eller ett exemplar (*item*). Vi hänvisar återigen till studiens definitioner

---

**Verk** (*work*): a distinct intellectual or artistic creation.

**Uttryck** (*expression*): the intellectual or artistic realization of a *work* in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms.

**Manifestation** (*manifestation*): the physical embodiment of an *expression* of a *work*. - - - As an entity, *manifestation* represents all the physical objects that bear the same characteristics, in respect to both intellectual content and physical form.

---

<sup>33</sup> IFLA, *Functional Requirements for Bibliographic Records. Final Report*, 1997, s 8-9

**Exemplar** (*item*): a single exemplar of a *manifestation*. The entity defined as *item* is a concrete entity.<sup>34</sup>

---

Som brukligt är exemplifierar vi med föreliggande uppsats. *Metadata för det digitala biblioteket* är alltså ett *verk*. Filmatiseringen Spielberg planerar av detsamma är ett *uttryck*, en expression av *verket*. Ett annat *uttryck* är den text du just har framför dig. Som olika *manifestationer* av texten kan vi nämna pappersexemplaret, Word-dokumentet, HTML-versionen samt pdf-filen. Det specifika pappersex på vilket våra handledare skrivit sina uppmuntrande tillrop är ett *exemplar* av pappers*manifestationen*. Verk, uttryck, manifestation och exemplar är bra begrepp att ha i bakhuvudet då man undersöker de olika formaten för objektbeskrivning.

Den basfunktionalitet eller *minimal core level* IFLA menar att nationalbibliografiska poster skall ha är att de skall tillåta användaren att

- 
- Find all *manifestations* embodying:
    - The *works* for which a given *person* or *corporate body* is responsible
    - The various *expressions* of a given *work*
    - *Works* on a given subject
    - *Works* in a given series
  - Find a particular *manifestation*:
    - When the name(s) of the *person(s)* and/or *corporate body(ies)* responsible for the *work* is (are) known
    - When the title of the *manifestation* is known
    - When the *manifestation* identifier is known
  - Identify a *work*
  - Identify an *expression* of a work
  - Identify a *manifestation*
  - Select a *work*
  - Select an *expression*
  - Select a *manifestation*
  - Obtain a *manifestation*<sup>35</sup>
- 

De data posten måste innehålla för att tillåta ovanstående kan delas upp i två grupper, deskriptiva dataelement och dataelement i normativ form (exempelvis auktoriserade namnformer). Olika dokumenttyper kräver olika kompletteringar av elementen för att posten skall uppfylla *the functional requirements*. De deskriptiva element som tillkommer för elektroniska resurser är anmärkningar om systemkrav, åtkomstsätt (telnet, http eller annat) och åtkomstadress (exempelvis URL).

---

<sup>34</sup> Ibid s 16-23, citaten tagna från s 16, 18, 20 och 23

<sup>35</sup> Ibid s 97

Verk och manifestation skall tjäna som auktoritetsposter, och de funktionella kraven för posten studeras i olika hänseenden, huvudprincipen är att informationen är strukturerad på ett sådant sätt att de olika delarna kan återanvändas oberoende av varandra, alltifrån från vid inköp till bevarande.

I klassifikationssammanhang talar man om ett systems fullständighetsgrad (*exhaustivity*), det vill säga till vilken grad de ämnen och begrepp som behandlas i en publikation kan återfinnas med hjälp av de indexerade termerna.<sup>36</sup> På samma sätt kan vi tala om olika nivåer av *exhaustivity* för olika metadataformat. Vi har tidigare diskuterat olika metadata definitioner. I DESIRE-rapporten inleder man med att fastställa att metadata "is data which describes attributes of a resource."<sup>37</sup> För att återvända till DESIRE:s tre "band" kan man mycket förenklat säga att de skiljer sig åt genom hurpass många och väldefinierade attribut formaten innehåller. De enklare formaten inbegriper alltså i princip endast titel och (delar av) innehållet. Detta räcker för att användaren skall kunna göra enkla sökningar, och exempelvis få reda på att *x* antal dokument innehåller ordet "Strindberg". Nästa formatgrupp stöder även sökning i fält, vilket gör att användaren kan begränsa sin sökning till exempelvis de dokument som har "Strindberg" som ämnesord. Detta gör att inte bara *sökning* och *identifikation*, utan också ett visst *urval* är möjligt. I den tredje och mest avancerade gruppen är katalogposterna såpass fullständiga att termen dokumentsurrogat börjar verka rimlig. Tillämpade i sin fullständiga nivå är det endast formatgrupp två och tre som kan sägas uppfylla de krav IFLA ställt upp i sin *Functional Requirements* för den bibliografiska posten.

I det följande kommer vi att detaljstudera två av de format som är aktuella för bibliotekskatalogisering av elektroniska resurser, först det gamla hederliga MARC-formatet och därefter uppstickaren Dublin Core.

### 3.1 MARC

MARC-formatet (*MAchine Readable Cataloguing*) kan definieras som ett standardiserat format för utbyte och lagring av bibliografisk information i maskinläsbar form. Och det låter ju inte så upphetsande. Poängen med MARC är emellertid de hisnande möjligheter till utbyte och återanvändning av bibliografisk information datatekniken ställde biblioteken inför. MARC började lite blygsamt som ett system för att centralisera produktionen av katalogkort. Library of Congress använde det maskinläsbara MARC för att massproducera katalogkort som sedan användes i de lokala bibliotekssystemen över hela Amerika.<sup>38</sup> Jämsides med detta initiativ utvecklades teknologin snabbt och gjorde det plötsligt möjligt att ställa oerhört mycket högre krav på katalogerna. Biblioteken förstod att om deras katalogiseringssystem skulle ha något att betyda i framtiden var det viktigt att man höll sig på den teknologiska utvecklingens barrikader.

1965 påbörjade Library of Congress alltså sitt arbete med att utforma ett maskinläsbart format. Med hjälp av olika pilotbibliotek skapade och utvärderade man det som kom att kallas MARC I. Förutom att konsultera några av de andra bibliotek

---

<sup>36</sup> Benito, 1995, s 106

<sup>37</sup> Dempsey och Heery, 1997, s 5

<sup>38</sup> Furrie, Betty, "Understanding MARC Bibliographic: Machine-Readable Cataloging", avsnitt "What is a MARC record, and why is it important?"



som skulle använda formatet hade man dessutom ambitionen att arbeta nära systemutvecklarna. 1967 hade man kommit så långt att man kunde inleda arbetets andra fas.<sup>39</sup> Den viktigaste erfarenheten man gjort i pilotstudien var att MARC-formatet måste vara mycket flexibelt för att kunna fungera som standardformat. Spektrat av olika bibliografiska uppgifter var nämligen mycket brett.<sup>40</sup> Här hade man nu MARC, ett tungt och rikt metadataformat som var komplext nog att användas för att göra mycket noggranna beskrivningar, och trots en viss tungroddhet ändå flexibelt nog att kunna anpassas till olika biblioteks speciella krav.

Flexibiliteten var kanske det som lockade de lokala biblioteken mest. Olika länder införlivade sina respektive katalogiseringstraditioner, och trettio år efter det att formatet skapades finns det ett tjugotal olika nationellt anpassade MARC-varianter. Bland de mest betydelsefulla märks det amerikanska USMARC och det brittiska UKMARC. Man har under senare år arbetat med att försöka ”harmonisera” USMARC med det kanadensiska CANMARC och engelsmännens UKMARC. Resultatet, det nya USMARC, kallar man *MARC 21 Format for Bibliographic Data*. Det har funnits tillgängligt sedan sommaren 1999.<sup>41</sup> I Sverige står Kungliga Biblioteket just i begrepp att byta MARC-format, från LIBRISMARC till SweMARC. SweMARC ”ansluter nära till det internationellt ledande formatet ... [MARC 21, vår anm.] samtidigt som det tillgodoser väsentliga nationella behov och nationell katalogiseringspraxis.”<sup>42</sup> Detta innebär bland annat att man behåller vissa av LIBRIS-formatets grundläggande drag, exempelvis bruket att vid katalogisering av flerbandsverk upprätta dels en huvudpost, dels ett antal delposter.<sup>43</sup>

Rachel Heery talar om de olika dialekterna som ett tecken på den inre konflikt MARC-formatet alltid lidit av, konflikten mellan kravet på allmängiltighet och nödvändigheten av anpassning till lokala behov.<sup>44</sup> Man är i högsta grad medveten om att de olika formatdialekterna är en av MARC-familjens största svagheter. Eller som systembibliotekarien Alan Hopkinson vid Middlesex University, London uttrycker det ”We are talking about data of which many users demand a very high level of uniformity. - - - Conformity and uniformity are required rather than agreement.”<sup>45</sup> I sammanhanget ifrågasätter han huruvida MARC-familjen, vilken fungerar som ett slags biblioteksstandard, egentligen kan anses vara standardiserad.

IFLA har sedan många år kämpat för att genomdriva UNIMARC, vilket är avsett att fungera som ett standardformat för internationellt utbyte av elektroniskt lagrade bibliografiska poster. Formuleringen är snarlik de mål man ursprungligen hade för USMARC. Sedan dess har emellertid USMARC:s olika dialekter blivit så sinsemellan olika att man förlorat möjligheterna till internationellt utbyte. Tanken med UNIMARC är att de olika nationella MARC-formaten endast skall behöva

---

<sup>39</sup> Byrne, Deborah J, *MARC manual. Understanding and using MARC records*, 1998, s 6ff

<sup>40</sup> *Ibid* s 8

<sup>41</sup> Library of Congress Network Development and MARC Standards Office, ”Guidelines for Coding Electronic Resources in Leader/06”. Mer om MARC 21 kan man läsa på deras ”MARC 21 concise format for bibliographic data”

<sup>42</sup> Kungliga bibliotekets hemsida, ”SweMARC formatöversikter”

<sup>43</sup> Mer information om detta på Kungliga bibliotekets hemsida, ”Libris: Frågor och svar”

<sup>44</sup> Heery, 1996, s 345-373

<sup>45</sup> Hopkinson, Alan, ”Traditional communication formats: MARC is far from dead”, *International cataloguing and bibliographic control*, 1999, s 19

översättas till och från ett enda annat format, ”mellanledet” UNIMARC.<sup>46</sup> Utbytesformatet har dock svårt att hävda sig i konkurrensen, och mest använt är utan tvekan det amerikanska USMARC.<sup>47</sup> De olika biblioteken ute i världen är mer intresserade av att förhålla sig till det större USMARC än att lägga ned arbete på anpassning till UNIMARC. Försöket kan jämföras med ”världsspråket” esperanto, ett konstlat sätt att återskapa det ursprungliga samförstånd som rådde före språkförbistringen. Även i detta fall har det visat sig svårt att få nationerna att acceptera ett nytt språk då engelska redan är internationellt gångbart inom de flesta vetenskapliga områden.

Vi kommer att återkomma till de skillnader som föreligger mellan de olika tillämpningarna av MARC, men då vi i det följande utan att närmare specificera oss talar om ”MARC” och ”formatet” berör vi enbart de drag som är gemensamma för medlemmarna i MARC-familjen.

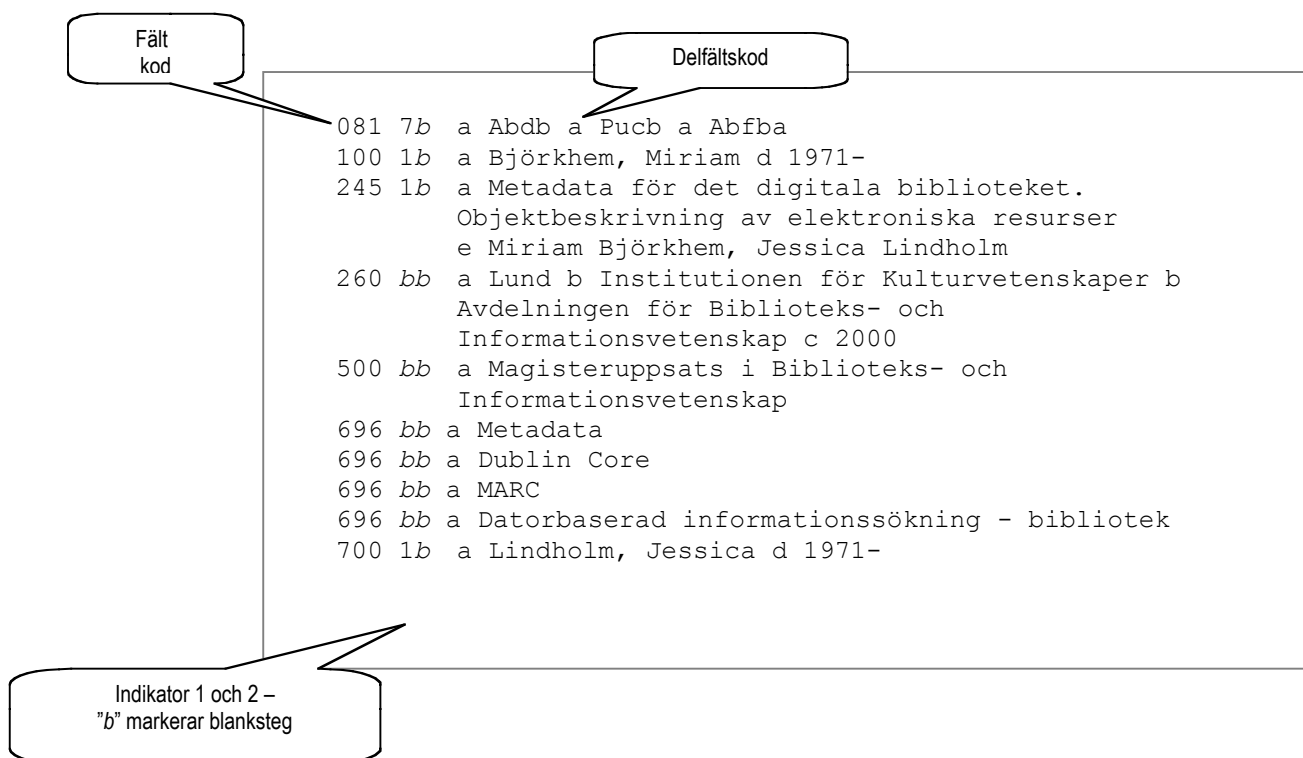
### **3.1.1 Formatets struktur och användning**

En MARC-post består av ett otal fält, numrerade med en tresiffrig etikett eller fältkod (000 – 999). I början av varje katalogpost finns ett särskilt utrymme, fält 000 - 008, för teknisk och statistisk information. Uppgifterna avser bland annat utgivningsland, publikationstyp, objekttyp och utgivningstid. I de övrigafälten (009 – 999) läggs den bibliografiska informationen. Fälten med bibliografisk information varierar i längd och antal. De kan innehålla ett eller flera delfält, markerade med delfältskoder. I vissa fall kan även så kallade indikatorer förekomma. Några av fälten och ett antal delfält är upprepningsbara.

---

<sup>46</sup> Ibidem

<sup>47</sup> I presentationen av SweMARC kallas MARC21 ”det internationellt ledande formatet”, Kungliga bibliotekets hemsida, ”SweMARC formatöversikter”



**Figur 4. MARC-posten i sin prydno**

De numrerade fälten är grupperade i olika block. Så innehåller exempelvis 0XX-fälten som vi nämnt ovan kontrollinformation och proveniens, medan de övriga kan få huvudrubriker enligt följande:

FÄLTKOD	RUBRIK OCH FÖRKLARING AV FÄLTINNEHÅLL
01X-08X	Kontrollinformation och proveniens
1XX	Huvuduppslag (personnamn, institutionsnamn, konferensnamn eller uniform titel)
2XX	Titel, upphov och upplageinformation (även varianttitel och adress)
3XX	Fysisk beskrivning (även restriktioner för spridning och "tillgänglighetstid")
4XX	Seriebiuppslag (personnamn, institutionsnamn eller konferensnamn, även uppgifter om serie)
5XX	Anmärkningar (exempelvis "utgiven tillsammans med", även restriktioner för tillgång och utlån)
6XX	Ämnesordsbiuppslag
7XX	Biuppslag
8XX	Seriebiuppslag
9XX	Fält reserverade för lokal information

**Figur 5. MARC-formatets huvudgrupper<sup>48</sup>**

<sup>48</sup> Ibid, under avsnitt "Excelfiler med fullständiga formatöversikter"

MARC-formatets fält kan vara antingen generella eller lokala. Generella fält (i princip alla mellan 000 och 899) innehåller bibliografiska uppgifter om verket som sådant, exempelvis beskrivningen och generella anmärkningar. De lokalafälten används för uppgifter om det aktuella bibliotekets bestånd, klassifikation och hyllsignum.

MARC-posten har ett slags inre logik som den hjälpligt följer. Så kan vi se att personnamn ofta placeras i X00-fält, institutionsnamn i X10-fält och geografiska namn i X51-fält. Bibliotekarien och katalogisatören Deborah Byrne talar i sin *MARC manual* också om att delfältsmarkörerna ger en antydning om delfältets innehåll.<sup>49</sup> Detta är emellertid en sanning med modifikation, som möjligtvis kan tjäna som minnesknep för vana katalogisatörer, men lämnar alla oss andra hjälplöst förvirrade.

### 3.1.2 Resonemang

Library of Congress har alltid haft det huvudsakliga ansvaret och tjänat som den ultimata auktoriteten för MARC-formaten. Library of Congress fungerar också som utgivare av dokumentationen om USMARC. För de olika nationella varianterna svarar särskilda instanser. I Sverige har Kungliga biblioteket huvudansvaret för detta. Förutom Library of Congress arbetar även ALA-kommittén MARBI (MACHINE-Readable Bibliographic Information) och USMARC Advisory Group med MARC-formatens utveckling.<sup>50</sup> MARBI består av representanter för olika American Library Association-avdelningar. Dessa representerar katalogisatörerna som skapar MARC-poster, systemutvecklarna som koordinerar bruket av MARC-poster i systemen, och de primära användarna av MARC-baserade bibliotekssystem. MARBI träffas två gånger om året, vid ALA-konferenserna. Mötena fungerar som ett diskussionsforum, där även andra organisationer kan delta i utvecklingen av MARC-formatet.<sup>51</sup>

I USMARC Advisory Group ingår, förutom medlemmar från MARBI, representanter från Library of Congress och andra nationalbibliotek i USA, Kanada och Australien, representanter från olika "bibliographic utilities" som exempelvis *Online Computer Library Center office of research* (OCLC) och representanter från olika specialbibliotek. Deras uppgift är att fungera som rådgivare åt Library of Congress angående utvecklingen av MARC-formaten. På Library of Congress hemsida står det att "the common characteristic of these groups is that they are large scale users of MARC, thus have serious concerns for the development and stability of the format."<sup>52</sup> Ambitionen är att alla användare av formatet skall kunna bidra med förslag på hur det skall kunna förbättras.

Arbetsgången är sådan att Library of Congress publicerar de slutsatser och förslag som framkommit. Några månader innan USMARC Advisory Group möts och diskuterar (i likhet med MARBI gör man detta två gånger om året) skall förslagen dessutom ligga ute på Library of Congress hemsidor. På mötena tar rådgivningsgruppen ställning till förslagen. Gruppens MARBI-medlemmar röstar om huruvida förslaget skall bifallas. Sista ordet har emellertid Library of Congress. Efter

---

<sup>49</sup> Byrne, 1998, s 55f. Hela detta avsnitt, betitlat "Patterns in the MARC Format" (s 47-56), behandlar dessa svårtydda mönster i USMARC.

<sup>50</sup> Ibid, s 10

<sup>51</sup> Ibid, s 11

<sup>52</sup> Library of Congress Network Development and MARC Standards Office, "MARC Advisory Committee"

det att ett förslag eller en modifikation på detta sätt godkänts publiceras det återigen som en uppdatering av MARC-formatet.<sup>53</sup>

För att diskutera de regler och standarder som styr formatet måste vi återvända till dess uppbyggnad. I formatsammanhang talar man om ett systems syntax och semantik. Semantiken är ”meningen”, den information formatet eller det enskilda fältet representerar, och syntaxen är det sätt på vilket man uttrycker denna information. Eric Miller skriver

Individual resource description communities define the semantics, or meaning, of metadata that address their particular needs. Syntax, the systematic arrangement of data elements for machine-processing, facilitates the exchange and use of metadata among multiple applications.<sup>54</sup>

Vi urskiljer tre olika aspekter av MARC-formatet. Dessa är struktur, fältbeteckning och fältinnehåll.<sup>55</sup> Här kan struktur och fältbeteckning sägas motsvara MARC-formatets syntax, medan fältbeteckningarnas bakomliggande innebörd motsvarar formatets semantik. För klarhetens skull kommer vi att presentera de tre aspekterna i något omkastad ordning.

MARC-formatens struktur följer ISO-standarden 2709, vilken för USMARC motsvaras av NISO Z39.2.<sup>56</sup> ISO 2709 bestämmer alltså hur posterna skall se ut, vilka koder som skall användas och så vidare.<sup>57</sup>

Fältinnehållet i MARC-formaten styrs av olika regler i olika nationella varianter av MARC.<sup>58</sup> USMARC följer exempelvis AACR2 (*Anglo-American Cataloguing Rules, 2<sup>nd</sup> edition*) och LIBRISMARC följer KRS (*Katalogiseringsregler för svenska bibliotek*). KRS är emellertid närmast att betrakta som en direktöversättning av AACR2, med vissa revideringar för att passa svensk katalogiseringspraxis bättre. På samma sätt är det med de regler som styr andra varianter av MARC-formatets fältinnehåll. Det är i detta sammanhang man talar om MARC som en ”tillämpning” av AACR2. MARC-formaten växer jämsides med de amerikanska katalogiseringsreglerna, ja Fred Ayres påstår till och med att MARC följer dem med ”pathological rigidity”<sup>59</sup>.

Nu återstår endast regler för fältbeteckningen, eller vilken bibliografisk information som skall ligga i vilket (etiketterat) fält. Detta är vad som gör det möjligt för en dator att förstå var i posten den skall söka efter viss information. Beteckningen ger även med hjälp av fältetiketterna maskinen möjlighet att ”förstå” vilket slags information det är den funnit, och hur den skall användas.<sup>60</sup> Fältbeteckningen bestämmer att delfält

---

<sup>53</sup> Ibidem

<sup>54</sup> Miller, Eric, ”An Introduction to the Resource Description Framework”, *D-Lib Magazine*, 1998

<sup>55</sup> *Structure, content designation* och *content*. Se Byrne, a.a. s 17ff

<sup>56</sup> NISO Z39.2 (National Information Standards Organization) kallades tidigare ANSI Z39.2 (American National Standards Institute), se Hopkinson, 1999, s 19. Dempsey och Heery, 1997, s 57

<sup>57</sup> Den senaste ISO-standard (International Standard Organization) för överföring av bibliografiska data heter ”ISO 2709:1996 Information and documentation - Format for Information Exchange”.

<sup>58</sup> Dempsey och Heery, 1997, s 56

<sup>59</sup> Ayres, 1996, s 2

<sup>60</sup> ”ANSI [numera NISO, vår anm.] Z39.2 stipulates a structure that is followed for MARC records; however, it is not the same thing as MARC format. USMARC and other MARC formats actually supply specific meaning to various parts of the ”shell” provided by Z39.2 [eller ISO 2709, vår anm.]”

245 a är ett titelfält, oavsett om det är ifyllt eller inte. I och med fältbeteckningen kopplas fältkod och fältinnehåll samman. Hur denna sammankoppling skall ske styrs av MARC-formatet självt, och uttrycks i de olika nationella MARC-varianternas katalogiseringsmanualer. Manualerna skapas och utges oftast av ländernas respektive nationalbibliotek, och fungerar som *de facto*-standarder.<sup>61</sup> Man strävar inte efter att standardisera dessa regelverk ”på riktigt”. Tendensen är, och har nästan alltid varit, att Library of Congress introducerar nya regler som sedan allteftersom plockas upp av de andra.

MARC har aldrig varit avsett att användas av den breda massan. Det krävs specialistkunskap för att kunna katalogisera i MARC, och denna indexering utförs ofta centralt, vid ett huvudbibliotek eller motsvarande. Andra bibliotek nöjer sig med att kopiera posterna och lägga till sina lokala noteringar. För den händelse att man inte är en dator kan det vara en skrämmande upplevelse till och med att endast läsa en fullständig MARC-post. Då man funnit den post man sökt i LIBRIS erbjuds man att beskåda den fullständiga MARC-posten. Erbjudandet beledsagas dock med den avskräckande texten ”Bibliotekens interna format”.<sup>62</sup> MARC-formatet är i högsta grad en specialisternas och maskinernas angelägenhet.

### 3.1.3 Formatets flexibilitet

Vi talade om MARC-formatets flexibilitet som något som ursprungligen lockade biblioteken att adoptera formatet. Denna flexibilitet gäller dock som vi sett främst möjligheten att använda anpassade regelsystem för fältinnehållet, samt att använda 9XX-fälten för att addera lokal information. Detta är en av de bidragande orsakerna till att MARC har blivit så stort. Stort och trögt, kan man kanske tycka då man betänker den komplicerade beslutsgången innan fält för nya aspekter av objektbeskrivning kan definieras och inlemmas i formatet. Under 1990-talet har MARC-samhället försökt utveckla formatet för att svara mot de nya och alltmer komplexa krav de elektroniska resurserna ställer katalogiseringen inför.<sup>63</sup> I arbetet med att inlemma de resurser vi intresserar oss för, de nätburna, följde man upp de ansträngningar man redan gjort för att kunna katalogisera datafiler i allmänhet. Om dessa första stapplande steg för att omforma USMARC skriver Priscilla Caplan

This may sound simple, but for historical reasons the computer files format is surprisingly limited in its ability to describe computer files. It was originally designed with only a single type of file in mind - statistical data sets - - - Later it was expanded to handle the microcomputer software that libraries had begun to collect. As such, it's like a house with only two rooms: a kitchen and a bedroom. OK until you want to take a bath.<sup>64</sup>

Är i detta sammanhang bandet mellan MARC och AACR2 (eller motsvarande regelsystem) en black om foten? Fred Ayres konstaterar

---

For example, Z39.2 establishes the structure of tags, indicators and subfield codes, but the MARC formats actually specify which particular tag (and so forth) will be used and what their meaning will be (i.e., content designation).” Byrne, 1998, s 135

<sup>61</sup> Dempsey och Heery, 1997, s 56

<sup>62</sup> Se exempelvis Kungliga bibliotekets hemsida, ”LIBRIS: Enkel sökning”

<sup>63</sup> Dempsey och Heery, 1997, s 57f

<sup>64</sup> Caplan, 1993, s 62

In fact, library automation is innovative and has produced solutions to complex problems which have led the way in some aspects of automation. - - - However, there is an Achilles heel. That is the reluctance to accept that the standards which were used in the production of the manual card catalogue do not work equally well in an automated situation and are quite often unnecessary. - - - We still treat AACR2 as sacrosanct and spend valuable resources in trying to maintain it.<sup>65</sup>

Rachel Heery har kallat AACR2 ”a code of the age of the catalog card and printed bibliography”, vilket väl får sägas mer än antyda att systemet är något av en dinosaurie.<sup>66</sup> Problemen med att på detta sätt behålla ett system som utvecklats med helt annorlunda förutsättningar än dagens är legio. Som ett exempel kan vi nämna att AACR2, och därmed MARC, ännu tar stor hänsyn till huvud- och biuppslag, det vill säga de aspekter som ansågs så viktiga att ett katalogkort fick tillverkas för att sorteras in även under dessa uppslagsord. Möjligheten att söka på hela postens innehåll, eller endast i en lång lista ämnesord, ifrågasätter detta förfarande.

Men tillbaka till de elektroniska resurserna. Man enades slutligen om att använda fält 856 till de elektroniska resursernas fromma. En finess med MARC-formatet är att långt ifrån alla fält ännu är definierade, och det är alltså i stort sett bara att enas om ett fälts nödvändighet och sedan börja använda det. URL-fältet kom i detta fall att hamna under ”Seriebiuppslag”. Fält 856 skall alltså innehålla den elektroniska resursens adress samt uppgifter om hur man kommer åt den.<sup>67</sup> Olika indikatorer anger huruvida åtkomstvägen är via e-post, ftp, telnet eller http. Men det finns även andra uppgifter som skall klämmas in i fältet för att det skall kunna fungera som en någorlunda tillfredsställande beskrivning av en elektronisk resurs. Uppgifter om DOI/URN, rättigheter, lagringsformat (exempelvis HTML eller ASCII)... Hela alfabetet, med undantag för e och y, går åt till att beteckna delfälten i fält 856.<sup>68</sup>

Fält 856 är ett upprepningsbart fält. Enligt Library of Congress finns det flera skäl att upprepa fältet. De flesta av dessa känner vi igen från den diskussion av problematiken med de nya dokumenttyperna vi haft i tidigare avsnitt. Den mest flagranta anledningen att upprepa fält 856 är emellertid att detta enda fält skall rymma otroligt mycket information.

---

<sup>65</sup> Ayres, 1996, s 1f

<sup>66</sup> Heery, 1996, s 345-373

<sup>67</sup> ”Field 856 ... is used for Electronic Location and Access information to an electronic resource and contains information related to the resource. - - - The field may be used ... when that resource or a subset of it is available electronically. In addition, it may be used to locate and access an electronic version of a non-electronic resource described in the bibliographic record”, Library of Congress Network Development and MARC Standards Office, ”Guidelines for the Use of Field 856: Revised August 1999”, 1999

<sup>68</sup> Dempsey och Heery, 1997, s 59f

```

081 7b a Abdb a Pucb a Abfba
100 1b a Björkhem, Miriam d 1971-
245 1b a Metadata för det digitala biblioteket. b
      Objektbeskrivning av elektroniska resurser m
      Elektronisk resurs e Miriam Björkhem, Jessica
      Lindholm
256 bb a Text
260 bb a Lund b Institutionen för Kulturvetenskaper b
      Avdelningen för Biblioteks- och
      Informationsvetenskap c 2000
500 bb a Magisteruppsats i Biblioteks- och
      Informationsvetenskap
696 bb a Metadata
696 bb a Dublin Core
696 bb a MARC
696 bb a Datorbaserad informationssökning - bibliotek
700 1b a Lindholm, Jessica d 1971-
856 40 u http://www.lub.lu.se/UB2proj/LIS_collection/
      MiriamJessica/toc.html q text/html

```

**Figur 6. Beskrivning av elektronisk resurs enligt MARC**

Sammanfattningsvis kan vi kanske säga att trots att MARC och hela apparaten kring formatet arbetar långsamt finns det goda möjligheter att anpassa formatet till att kunna beskriva även de elektroniska resurserna. Ett stort antal odefinierade fält står tomma och väntar på att definieras och tas i bruk av framtida, ännu otänkta, dokumenttyper. Frågan är snarare om AACR2 kan och skall sätta upp regler för alla dessa, och om formatet inte borde begagnas främst till de dokumenttyper det utvecklats för, de statiska, ”traditionella” biblioteksresurserna?

### 3.1.4 Interoperabilitet

Interoperabilitet, som så mycket annat, *begins at home*. Som vi sett har man fått lägga ned mycket arbete bara på att få MARC-formatet att vara kompatibelt inom familjen. Samordningsfrågorna är alltså inget nytt för MARC-samhället. En fördel i detta sammanhang är att MARC-familjen som helhet följer erkända regelsystem och standarder. Nackdelen är att olika formatvarianter eller –dialekter i viss mån följer olika regler och standarder. Dempsey och Heery uppmärksammar en MARC-formatets svaghet som blir uppenbar först då det blir tal om samordning

There is no attempt in the standard to identify whether records searched or delivered are in the US or UK or other MARC formats. This can cause problems for interoperability ... because the ”flavour” of MARC format is not identified.<sup>69</sup>

Hur skulle man i en MARC-post på detta sätt kunna ange vilken MARC-variant det är fråga om? Här är vi inne på meta-metadata, ”metadata in relation to the record itself” som Heery tidigare kallat dessa uppgifter. Som sådan ”meta-metadata” skulle man även kunna beskriva andra postens attribut: exempelvis datum för postens skapande, katalogisatörens signatur eller en postnumrering.<sup>70</sup>

<sup>69</sup> Ibidem

<sup>70</sup> Heery, 1996



Då MARC på sitt sätt är högt standardiserat innebär samordning och ”översättning” inte några omedelbara problem. Problemet ligger snarare i den MARC:s grundidé vi tidigare berört då vi konstaterade att formatet inte är något för den breda massan. AACR2 är ett komplicerat och ytterst precist regelsystem. Ett enklare format än MARC kan få svårigheter att uttrycka denna bestämda kod, med oundviklig informationsförlust som följd. Dessutom skapas MARC-poster med nödvändighet av specialister, katalogisatörer som har lång träning i att noggrant beskriva objekt.

Varför, kan man fråga sig, skall biblioteken över huvud taget sträva efter interoperabilitet i sin objektbeskrivning av elektroniska resurser? Med MARC och AACR2 kan man beskriva nästan vad som helst mycket grundligt. Behovet att fungera tillsammans med format utanför familjen är inte alldeles självklart. AACR2 har dock alltid haft det statiska, mer ”traditionella” biblioteksmaterialet för ögonen. Endast med svårighet kan katalogiseringsreglerna tillämpas på dynamiska resurser, och beskrivningen av dessa kan inte heller ses som lika uttömmande som den av de statiska resurserna. MARC-katalogisering är också som vi tidigare noterat en dyrbar historia. Det finns helt klart fördelar med att försöka uppnå samförstånd med de nyare format som skapats särskilt för beskrivning av det elektroniska materialet, och även att engagera sig i arbetet med att vidareutveckla sådana format. Stefan Gradmann skriver

There are, after all, a few good reasons – some explicit, others implicit – why the metadata community did **not** start off proposing MARC amendments but created a completely new frame of attributes.<sup>71</sup>

Detta är väl något av en sanning med modifikation. Som vi visat arbetar man jämsides med samordningsförsöken med att vidareutveckla AACR2 och MARC för att tillfredsställande kunna beskriva även de elektroniska resurserna.

De format som är aktuella för samordning är de som redan används för att beskriva elektroniska resurser. Så har man gjort mappningar mellan olika format för att få MARC att fungera tillsammans med exempelvis Dublin Core och GILS.<sup>72</sup> I arbetet med dessa konverteringsprogram har det lustigt nog visat sig att trots att det vanligtvis är lättare att konvertera från ett komplext till ett enklare format än tvärtom, ”converting from MARC to anything is harder than converting from anything to MARC.”<sup>73</sup> Nordiska metadataprojektet har dock färdigställt ett testprogram för konvertering av Dublin Core-poster till MARC.<sup>74</sup> De framtida systemen måste kunna söka metadata uttryckt i olika syntaxer. Det är alltså viktigt att dessa olika syntaxer inte är alltför inbördes olika. Library of Congress arbetar på att skapa vad man kallar en DTD (*Document-Type Definition*) för MARC. Målet är att underlätta konvertering mellan SGML-familjen (*Standard Generalised Markup Language*) och MARC.<sup>75</sup>

---

<sup>71</sup> Gradmann, 1998

<sup>72</sup> Library of Congress Network Development and MARC Standards Office, *Dublin Core/MARC/GILS Crosswalk*

<sup>73</sup> Hakala et al., 1998, s 27. Hakala tar åter upp detta i artikeln ”Internet metadata and library cataloguing”, 1999, s 23, och menar där att det beror på MARC-formatets ”peculiar structure”.

<sup>74</sup> Nordiska Metadataprojektet, ”d2m: Dublin Core to MARC converter”

<sup>75</sup> Library of Congress Network Development and MARC Standards Office, *Dublin Core/MARC/GILS Crosswalk*

Att översätta de färdiga posterna är ett sätt att nå samförstånd mellan olika format. Ett annat är att använda sig av protokoll som exempelvis Z39.50, vilket vi kommer att behandla närmare i ett kommande avsnitt. Varde tills vidare nog sagt att det i arbetet med att utveckla formaten är viktigt att förhålla sig till dessa. Vad just Z39.50 beträffar är detta protokoll "which enables search and retrieval of bibliographic information over the Internet" särskilt utformat för att kunna hantera MARC-poster.<sup>76</sup>

Den stora mängd MARC-poster som redan finns är ett mäktigt argument för att fortsätta utveckla formatet och sätta det i stånd att göra saker det ursprungligen aldrig var avsett för. I en artikel, vilken har den talande underrubriken "MARC is far from dead" skriver Alan Hopkinson att "it is clear that MARC is needed because it is used."<sup>77</sup> Han avslutar med att konstatera att "MARC has led to some fossilization of bibliographic data processing which has provided a stable environment for the development of systems" och fortsätter

Any standard has a built-in anomaly: standards stifle innovation but they also enable all users to reach the same platform or plateau from which they can all jump together to a higher level. We have not yet jumped but we are on the same level!<sup>78</sup>

### **3.1.5 MARC i praktiken: exemplet NetSök**

Bibliotekstjänsts NetSök är ett kvalitetsurval av www-resurser, katalogiserade med korta MARC-poster. Målgruppen är folkbiblioteken, men man riktar sig även till skolbibliotek, olika företag och privatpersoner. De resurser basen innehåller väljs ut av Bibliotekstjänsts ämnesspecialiserade redaktörer, eller importeras från OCLC:s NetFirst, som även den är en samling av uteslutande elektroniskt material. Posterna från NetFirst måste översättas, SAB-klassificeras och förses med svenska ämnesord och en svensk beskrivning innan de blir en del av NetSök. I NetSök är kvalitetskriteriet det avgörande, och basen innehåller både specifika resurser och olika länklister av varierande omfattning. Tyngdpunkten ligger på nordiskt och svenskt material. NetSök beräknas att under våren år 2000 innehålla cirka 5500 poster, med en årlig tillväxt på ett par tusen adresser.<sup>79</sup>

---

<sup>76</sup> Dempsey och Heery, 1997, s 57

<sup>77</sup> Hopkinson, 1999, s 18

<sup>78</sup> Ibid s 20

<sup>79</sup> Bibliotekstjänsts (BTJ) hemsida, <http://www.btj.com>

```

041 bb A swe
081 bb A Abdb S DR A Pucb S DR A Abfba S DR
100 Eb A Björkhem, Miriam
245 bb A Metadata för det digitala biblioteket M Elektronisk
resurs B Objektbeskrivning av elektroniska resurser
257 bb A Text
260 bb A Sverige B M. Björkhem C 1999-
513 bb A Magisteruppsats i ämnet Biblioteks- och
informationsvetenskap som behandlar först och främst
metadataformaten MARC och Dublin Core, men även olika verktyg för
samordning mellan formaten (RDF och Z39.50) samt diskuterar hur
bibliotek kan förhålla sig till det elektroniska materialet.
553 bb A Tidigare med titeln: Objektbeskrivning av elektroniska
resurser. En studie av metadataformaten MARC och Dublin Core ur
biblioteksperspektiv
553 bb A Tidigare adress: http://opera.kult.lu.se/~biv98mb/format
553 bb A Även utgiven i tryckt form
564 bb A Titel från titelskärbild
700 Eb A Lindholm, Jessica
856 Zb 2 http u
http://www.lub.lu.se/UB2proj/LIS_collection/MiriamJessica/toc.html
945 bb m [i de fall resursen även ingår i BURK innehåller fältet
BURK-numret]

```

**Figur 7. MARC-post från NetSök. Observera att exemplet (ännu) inte ingår i NetSöks samlingar**

NetSök använder sig av en egen tillämpning av Bibliotekstjänsts interna MARC-variant. Denna post är dock snarlik den MARC-post för elektroniska resurser vi tidigare visat. En finess med posten är att man i fält 553 kan lägga information om eventuell tidigare adress. Detta kan underlätta sökning av sidan om den nuvarande adressen plötsligt blir ogiltig. I ett annat anmärkningsfält noteras att titeln hämtats från titelskärbilden. Ett annat tänkbart ställe att hämta titeln från är HTML-kodens <TITLE></TITLE>-tag. I katalogiseringssammanhang talar man om primära och föreskrivna källor. För böcker är den primära källan exempelvis titelsidan. Hämtas uppgifter från något annat ställe måste detta noteras. I den dynamiska www-miljön är det inte säkert att titelskärbilden ser likadan ut eller innehåller samma information nästa gång sidan visas.

I de fastafälten (vilka inte visas i figuren ovan) ingår information om när och av vem posten skapats och ändrats, ”meta-metadata”. Att posten inte är daterad i presentationslayouten kan göra det svårt för användaren att veta när ett dokument indexerats, vilket i sin tur gör det svårt att veta hur pass aktuell beskrivningen av den dynamiska resursen är. Att detta inte meddelas är emellertid inget unikt för NetSök. I fullposten visas nio fält enligt följande:

FÄLT	KOMMENTAR
TITEL	www-resursen titel, hämtad från HTML-kodens <TITLE></TITLE>-fält eller titelskärbild
UTGIVARE	Ansvarig utgivare eller författare
INNEHÅLL	Kort beskrivning (50 – 80 ord) av resursens innehåll. Skall vara stringent men informationsrik för att stödja fritextsökning
SPRÅK	Det språk (ett eller flera) resursen är avfattad på
WWW-ADRESS	Adress; direktlänkad
KLASSIFIKATION	Enligt SAB-systemet
ÄMNESORD	Automatgenererade, enligt SAB-systemet
LAND	Resursens "utgivningsland"
NUMMER	NetSök-nummer, samt eventuellt BURK-nummer

**Figur 8 NetSök – kommenterad fullpost<sup>80</sup>**

Den resursbeskrivning som författats av NetSöks redaktörer och som ingår i varje post (se fält 513 i Figur 7) skall underlätta fritextsökning och göra det möjligt att utvärdera resursens relevans utan att lämna basen.

Bibliotekstjänsts idé med NetSök är den att de enskilda biblioteken skall avlastas. Man är medveten om att det tunga arbetet inte ligger i att finna informationen, utan i att sälla den. Med basens tydliga inriktning på svenska folkbibliotek erbjuder man en helhetslösning som är kompatibel, kanske inte direkt med bibliotekets samlingar, men med svenska bibliotekariers traditionella kunskaper. På NetSöks avancerade sökskärm har man exempelvis möjlighet att använda sig av booleska söktermer, trunkering och maskering. Att klassningarna görs med det allmänna, men på svenska bibliotek förhärskande SAB-systemet, är ett försäljningsargument i sig.

<sup>80</sup> Tillgång till NetSök kräver användaridentitet och lösenord.

Orsaken till att man i NetSök valt att använda sig av MARC-formatet är att posterna skall kunna inlemmas i Bibliotekstjänsts stora bas över folkbibliotekens resurser; BURK. Att man använder sig av MARC underlättar också importen av poster från NetFirst. Ambitionen att man redan utifrån den funna posten skall kunna göra ett urval gör också att det verkar vettigt att använda sig av ett format från band 1.

### 3.2 Dublin Core

Dublin Core (*Dublin Core Metadata Element Set*) är ett metadataformat som skapats mot bakgrund av behovet av ett vettigt sätt att finna och identifiera www-resurser. Grunderna till formatet arbetades fram under en konferens våren 1995, *the OCLC/NCSA Metadata Workshop*. Under konferensen sammanstrålade femtiotvå personer från olika professioner, "librarians, archivists, humanities scholars and geographers, as well as standards makers in the Internet, Z39.50 and Standard Generalized Markup Language (SGML) communities" som med gemensamma krafter skulle kunna diskutera och enas kring vilka aspekter av en resurs som behövde beskrivas för att den skall kunna återfinnas och värderas.<sup>81</sup>

Arbetsplanen var enkel, Priscilla Caplan beskriver den i sin rapport från konferensen som att helt enkelt "find a group of interested people from a variety of constituencies, make them come to Dublin, Ohio, and don't let them leave until they've agreed on a core data element set."<sup>82</sup> Det låg alltså i högsta grad i konferensdeltagarnas intresse att nå en tillfredsställande lösning. Tre dagar senare hade deltagarna lokaliserat ett antal "kärnelement" (*core elements*) som behövdes för återvinning och värdering av en elektronisk resurs. De tretton element de enats om påminde till stor del om den traditionella bibliografiska beskrivningen, med författar-, titелеlement och så vidare.

Man kallade de resurser formatet skulle klara av att beskriva för "dokumentliknande objekt" (DLO). Begreppet DLO definierades inte närmare, betydelsen klargjordes istället utifrån exempel: ett dokumentliknande objekt kunde vara "an electronic version of a newspaper article or a dictionary" medan; "an unannotated collection of slides is not."<sup>83</sup> Med DLO avsåg deltagarna på *the OCLC/NCSA Metadata Workshop* resurser där det intellektuella innehållet representerades huvudsakligen som text och att "the metadata required for describing DLOs ... [skulle, vår anm.] bear a strong resemblance to the metadata that describes traditional printed texts."<sup>84</sup> Vidare påpekar Priscilla Caplan i sin konferensrapport att oavsett om deltagarna inte ägnade mycket tid åt att definiera vad en DLO var, var det en viktig aspekt att "the core data element set did not have to handle every type of resource that could theoretically be available on or through the network. It only had to handle things like, well... document-like objects."<sup>85</sup> Under den tredje Dublin Core-konferensen i slutet av 1996 kom begreppet DLO att även få innefatta beskrivning av elektroniskt publicerade bilder, fotografier och videoklipp. Kärnelementen behövde inte förändras alltför mycket för att även

---

<sup>81</sup> Weibel et al., 1995

<sup>82</sup> Caplan, 1995

<sup>83</sup> Weibel et al., 1995

<sup>84</sup> Ibidem

<sup>85</sup> Caplan, 1995

kunna hantera dessa objekt. Några element fick emellertid byta namn för att inte vara så tydligt textinriktade och två nya element fick definieras, *Description* och *Rights*.<sup>86</sup>

Utvecklingen av Dublin Core sker i informella workshops. Det har dock hela tiden funnits en kärntrupp människor kring formatet, *Dublin Core Metadata Initiative* (DCMI) eller Dublin Core-initiativet, med formatets ”fader” Stuart Weibel i spetsen.<sup>87</sup> Mellan konferenserna behandlar arbetsgrupper varsitt problemområde och presenterar förslag inför nästa konferens. Så bildas arbetscykeln konferens – enskilt arbete i grupperna – förslag – konferens. Mellan konferenserna hålls diskussionen vid liv i ett antal *mailing lists* eller sändlistor.<sup>88</sup> Allteftersom formatet har blivit mer uppmärksammat har arbetsprocessen blivit mer formaliserad, Stuart Weibel poängterar att ”the goal is to achieve a stable procedural foundation for the Dublin Core that retains the interdisciplinary, international consensus-building culture that has grown up around the Dublin Core initiative.”<sup>89</sup> Ett steg i denna process har varit att OCLC Office of Research under 1998 tog ansvaret för stöd och finansiering samt att två fasta arbetsgrupper har inrättats.

### 3.2.1 Formatets struktur och användning

Dublin Core består av femton element som tillsammans utgör kärnan för vad som behöver berättas om resurs för att den skall kunna återfinnas och hämtas. Tillsammans beskriver elementen objektets intellektuella innehåll, objektets fysiska format och vilka personer som ansvarar för dokumentets skapande och upprätthållande.<sup>90</sup>

---

<sup>86</sup> Förändringen av DLO-begreppet behandlas exempelvis av Stuart Weibel och Eric Miller i ”Image Description on the Internet. A Summary of the CNI/OCLC Image Metadata Workshop September 24-25, 1996. Dublin, Ohio”, *D-Lib Magazine*, 1997

<sup>87</sup> I Nordiska Metadataprojektets slutrapport kallas Weibel ”the father of the Dublin Core”, Hakala et al., 1998, s 12

<sup>88</sup> Sändlistorna är tillgängliga via <http://www.mailbase.ac.uk>.

<sup>89</sup> Weibel, Stuart, ”The state of the Dublin Core Metadata Initiative April 1999”, *D-Lib Magazine*, 1999

<sup>90</sup> Motsvarande engelska begrepp är *content*, *instantiation* och *intellectual property*, Hakala, Juha, ”Internet metadata and library cataloguing”, *International cataloguing and bibliographic control*, 1999, s 22

BESKRIVNINGSDELAR	KÄRNELEMENT	FÖRKLARING
Innehållsrelaterade element ( <i>Content</i> )	<b>TITEL</b> <i>DC.TITLE</i>	Utgivaren, skaparen eller författarens namn på objektet.
	<b>ÄMNESORD OCH KLASSIFIKATION</b> <i>DC.SUBJECT</i>	Ämnesord eller klassifikation för att beskriva dokumentets intellektuella innehåll. Man bör välja kontrollerade ämnesord och formaliserade klassifikationsscheman, specificerade med en <i>qualifier</i> .
	<b>BESKRIVNING</b> <i>DC.DESCRPTION</i>	Kortfattad beskrivning av resursens innehåll; exempelvis innehållsförteckning eller <i>abstract</i> .
	<b>KÄLLA</b> <i>DC.SOURCE</i>	Referens till objektets källa, om sådan finns. Bör anges som unik identifikationssträng såsom DOI, ISBN eller <i>Uniform Resource Identifier</i> (URI).
	<b>SPRÅK</b> <i>DC.LANGUAGE</i>	Det språk resursens intellektuella innehåll uttrycks i. Här är det rekommendabelt att använda en standardiserad språkkod. Vilken standard man valt specificeras med en <i>qualifier</i> .
	<b>RELATIONER</b> <i>DC.RELATION</i>	Relation (till andra resurser), exempelvis om den är en del av ett större verk
	<b>TÄCKNINGSGRAD</b> <i>DC.COVERAGE</i>	Resursens geografiska eller tidsmässiga täckning eller täckningsgrad. Bör uttryckas med hjälp av kontrollerad vokabulär, specificerad med en <i>qualifier</i> .
Upphovsrelaterade element ( <i>Intellectual Property</i> )	<b>FÖRFATTARE/ TEXTANSVARIG</b> <i>DC.CREATOR</i>	Person(-er) eller organisation som ansvarar för skapandet av objektets intellektuella innehåll. För exempelvis fotografier avses här fotografen.
	<b>ANSVARIG FÖR UTGIVNING</b> <i>DC.PUBLISHER</i>	Dokumentets ansvarige utgivare eller förläggare.
	<b>ÖVRIGA UPPHOVSMÄN</b> <i>DC.CONTRIBUTOR</i>	Medförfattare, bidragsgivare och sekundära upphovsmän.
	<b>UPPHOVS RÄTT</b> <i>DC.RIGHTS</i>	Rättigheter: upphovsrätt, copyright.
Versionsrelaterade element ( <i>Instantiation</i> )	<b>DATUM</b> <i>DC.DATE</i>	Tidsangivelse för dokumentet. Datum för skapande av dokumentet, publicering, "bäst föredatum" eller liknande. Man bör använda ett standardiserat datumformat, specificerat med en <i>qualifier</i> .
	<b>TYP AV RESURS</b> <i>DC.TYPE</i>	Kategori för resursen intellektuella innehåll, exempelvis dikt, avhandling, artikel et cetera.
	<b>LAGRINGSFORMAT</b> <i>DC.FORMAT</i>	Lagringsformat (exempelvis HTML, ASCII eller Word97) – vilket/vilka format objektet lagrats och kan läsas i. Man kan även ange objektets storlek, varaktighet (om man bara kan använda resursen en viss tid) och/eller vilka programvaror som behövs för att läsa objektet. Här rekommenderas användning av kontrollerade benämningar på dataformat, <i>Internet Media Types</i> (IMT).
	<b>WWW-ADRESS/ UNIK IDENTIFIERING</b> <i>DC.IDENTIFIER</i>	Unik sträng för identifiering av objektet, exempelvis ISBN, DOI eller URI/URL.

Figur 9. Dublin Cores femton element<sup>91</sup>

Kärnelementen har dessutom sällskap av *qualifiers*, vilka används för att förtydliga betydelsen av själva elementet eller av dess innehåll så att en exakt tolkning skall kunna göras. *Qualifiers* finns i två former så kallade *encoding schemes* och *element refinements*:

<sup>91</sup> Enligt Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set, Version 1.1: Reference Description*, OCLC, 990702

### Element refinements

Dessa *qualifiers* läggs till elementet för att förtydliga och precisera elementets betydelse. För *Date*-elementet rekommenderas *Date.Created*, *Date.Issued*, *Date.Modified* och *Date.Valid*.<sup>92</sup> Utan detta förtydligande blir datumangivelsen tämligen svårt att tolka.

### Encoding schemes

För att kunna förstå beskrivningen används två typer av *encoding schemes*. Gemensamt är att de anger enligt vilken standard eller vokabulär beskrivningen är skapad. För titel, beskrivning och ämnesord är det möjligt att ange vilket språk beskrivningen är skriven på (se `LANG="sv"` i Figur 10).

*Scheme*-tillägget är den andra typen av *encoding scheme* och används för att berätta att elementinnehållet följer någon standardiserad form, exempelvis att ämnet för resursen uttrycks enligt SAB:s klassifikationssystem (se `SCHEME="SAB"` i Figur 10) eller att ett datum i *Date*-elementet uttrycks enligt en viss standardiserad form (se användning av `SCHEME="ISO8601"` i Figur 10).

*Qualifiers* är i likhet med elementen upprepningsbara och frivilliga, men kan vara avgörande för både människa och maskins förståelse av beskrivningen.

Är det någon del som saknas för att formatet skall fungera för ett visst material är det fullt möjligt att skapa egna element och *qualifiers*. Stuart Weibel påpekar att "few applications have found that the 15 elements satisfy all their needs."<sup>93</sup> Ett projekt kan alltså välja att använda exempelvis tio av de femton elementen, samt definiera ett antal egna element. På så sätt kommer posternas utseende att variera mellan olika projekt som använder Dublin Core.

---

<sup>92</sup> DCMI, *Dublin Core Metadata Element Set Qualifiers 1.0*, Working Draft, 1999

<sup>93</sup> Weibel, Stuart, "The state of the Dublin Core Metadata Initiative April 1999", *D-Lib Magazine*, 1999



En Dublin Core-post uttryckt i HTML kan se ut så här:

```
<HEAD>
<META NAME="DC.Title" CONTENT="Metadata för det digitala biblioteket.
Objektbeskrivning av elektroniska resurser">
<META NAME="DC.Subject" CONTENT="Abdb">
<META NAME="DC.Subject" SCHEME="SAB" CONTENT="Pucb">
<META NAME="DC.Subject" SCHEME="SAB" CONTENT="Abfba">
<META NAME="DC.Subject" SCHEME="SABAO" CONTENT="Metadata">
<META NAME="DC.Subject" SCHEME="SABAO" CONTENT="Dublin Core">
<META NAME="DC.Subject" SCHEME="SABAO" CONTENT="MARC">
<META NAME="DC.Subject" SCHEME="SABAO" CONTENT="Datorbaserad
informationssökning - bibliotek">
<META NAME="DC.Description" CONTENT="Magisteruppsats i ämnet
Biblioteks- och informationsvetenskap som behandlar först och främst
metadataformaten MARC och Dublin Core, men även olika verktyg för
samordning mellan formaten (RDF & Z39.50) samt diskussion kring hur
bibliotek kan förhålla sig till det elektroniska materialet vid
uppbyggnad av ett digitalt bibliotek.">
<META NAME="DC.Language" SCHEME="ISO639-1" CONTENT="swe">
<META NAME="DC.Creator.PersonalName" CONTENT="Björkhem, Miriam">
<META NAME="DC.Creator.PersonalName" CONTENT="Lindholm, Jessica">
<META NAME="DC.Publisher.CorporateName" CONTENT="Lunds universitet:
Institutionen för Kulturvetenskaper">
<META NAME="DC.Date.Created" SCHEME="ISO8601" CONTENT="2000-01-19">
<META NAME="DC.Type" CONTENT="Text.ThesisMaster">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="text/html">
<META NAME="DC.Identifier" SCHEME="URL" CONTENT="
http://www.lub.lu.se/UB2proj/LIS_collection/MiriamJessica/toc.html">
<META NAME="DC.Identifier" SCHEME="URN" CONTENT="URN:NBN:se-d200011">
</HEAD>
```

**Figur 10. Dublin Core-post**

Dublin Core-posten uttrycks vanligen i HTML, som idag är det gängse filformatet för www-resurser. Dublin Core-posten läggs då mellan <HEAD> och </HEAD>-taggarna i HTML-dokumentet, det vill säga den del av HTML-dokumentet som innehåller information om själva dokumentet och inte är synligt i webläsaren.

### 3.2.2 Resonemang

Ett av de viktigaste resonemangen kring formatet handlar om att formatet skall vara enkelt att använda (*simplicity*). Det skall vara lätt både att skapa och underhålla Dublin Core-poster. Enkelheten speglas på alla nivåer, formatet är ursprungligen avsett att beskriva endast det man absolut måste veta för att söka och hämta en resurs. Avsikten är också att indexering i Dublin Core skall kunna utföras även av personer helt utan erfarenhet av katalogisering. I Nordiska Metadataprojektets rapport ställer man detta mot biblioteksvärldens hävdvunna metoder, "as libraries do not have resources for describing even the most relevant Web documents, it is necessary to build tools with which anyone can create metadata for his or her document."<sup>94</sup> Om Nordiska Metadataprojektets "anyone" som skapar "...metadata for his or her document" kan antas vara den som skapar HTML-dokumentet är det med all sannolikhet en person med vissa kunskaper om HTML-redigering och om publicering

<sup>94</sup> Hakala et al., 1998, s 13

på www.<sup>95</sup> En viss kunskap om klassifikation och/eller specifik ämneskunskap underlättar också en god beskrivning. För att indexera med Dublin Core krävs således vissa kunskaper, om än inte en professionell utbildning. För att underlätta Dublin Core-indexering erbjuds olika, oftast projektanknutna, inmatningsformulär i vilka man skriver in dokumentbeskrivningen, och sedan erhåller en färdig HTML-kod att infoga i dokumentet. Om man händelsevis inte är expert på ISO-standarder underlättas skapandet av *encoding scheme-qualifiers* rejält av inmatningsformulären, eftersom man erbjuds förvalda alternativ för vilka *qualifiers* som finns att välja bland. Ett exempel på ett sådant inmatningsformulär visas i avsnittet ”Dublin Core i praktiken - SAFARI” (Figur 12).

MARC/KRS kan sägas beskriva exemplaret (*item*) mer ingående än vad Dublin Cores beskrivningar av elektroniska resurser möjliggör. En anledning till detta är att formatet ursprungligen skapats för att beskriva de resurser som fanns i bibliotekets samlingar. Å andra sidan kan en utförlig Dublin Core-beskrivning ge mer information om resursens innehåll, exempelvis är det närmast en *de facto*-standard när man beskriver elektroniska resurser att ha ett kort sammandrag av vad resursen handlar om. I mappning med MARC får ett flertal av Dublin Core-elementen läggas i MARC-formatets anmärkningsfält för fritextsökning, bland annat den beskrivning av resursen som bland Dublin Core-fälten räknas som ett av de viktigaste.<sup>96</sup>

Formatet beskriver resursen, men sätter inte upp regler för hur själva posten skall se ut. Dublin Core-initiativet skapar rekommendationer för hur element och *qualifiers* skall utformas, samtidigt är alla element och *qualifiers* upprepnings- och uteslutningsbara. Vad gäller postens utformning så är fältordningen oviktig. Dublin Core strävar efter ett språkligt (semantiskt) oberoende, att underlätta interaktion med andra format. De få och formbara elementen skall lätt kunna förstås av andra format.<sup>97</sup> Men är det enbart en fördel att Dublin Core inte sätter upp några strikta regler för hur formatet skall användas? Olika beskrivningsmodeller innebär problem för utbyte och samsökning. Hakala skriver om semantiska metadata och deras ”commonly understood semantics”<sup>98</sup> På Internet samsas resurser inom olika ämnesområden och det används olika beskrivningsmodeller - en beskrivningsmodell som kan förstå de andra ökar möjligheterna för utbyte och kan tjäna som en sammanförande faktor mellan olika discipliner.

### 3.2.3 Formatets flexibilitet

Beroende på vilken beskrivningsnivå man vill uppnå och utifrån vilken typ av objekt man beskriver kan man med Dublin Core skapa allt från den enklaste beskrivning helt utan *qualifiers* till poster vilka fullständigt uttrycker AACR2:s regelverk. I Nordiska Metadataprojektets slutrapport påpekas att ”although the Dublin Core is simple, this simplicity is deceptive in the sense that with qualifiers one may build a format, which surpasses MARC in complexity.”<sup>99</sup> Formatet skall vara flexibelt, men avser inte att kunna uppställa nationalbibliografiska poster eller att ersätta andra mer komplexa metadataformat. Michael Day påpekar mycket riktigt att ”Dublin Core is intended to

---

<sup>95</sup> Ibidem

<sup>96</sup> För presentation används oftast DC.Title, DC.Description och DC.Identifier, se exempel på detta i avsnittet ”Dublin Core i praktiken: exemplet SAFARI”, Figur 11

<sup>97</sup> DCMI:s hemsida, *Dublin Core Metadata Initiative*, styckerubrik ”Semantic Interoperability”

<sup>98</sup> Hakala et al., 1998, s 11

<sup>99</sup> Ibid, s 13

be a simple structured record which can, if required, be enhanced or mapped to more complex records.”<sup>100</sup> Formatet skall alltså tjäna som ett komplement mellan olika format, och erbjuda interoperabilitet med andra mer komplexa metadataformat.<sup>101</sup>

Aspekter som prioriteras i det framtida utvecklingsarbetet är förhoppningen om att de stora söktjänsterna skall förstå även kontrollerad metadata, att HTML-editorer skall stödja skapandet av Dublin Core-poster och att inmatningsformulär som gör formatet enkelt att använda skall underlätta spridningen av formatet.

Bland de olika discipliner som använder sig av metadata finns olika behov, och det är viktigt att det finns stöd för användning av metadata på olika sätt. Det är viktigt att utveckla modeller som kan främja metadataanvändningen. Ett exempel är W3C:s *Resource Description Framework*, RDF, (som vi behandlar i avsnittet RDF & XML) där DCMI bevakar och talar för de digitala bibliotekens intressen under RDF:s utvecklingsarbete.

### 3.2.4 Interoperabilitet

Interoperabilitet med andra format och miljöer är grundläggande för Dublin Core-filosofin. I jämförelse med MARC-formatet som är skapat för en sorts domän, biblioteken, är Dublin Core inte specialiserat och inriktat på att passa bäst i en viss miljö. Domänspecifikt blir Dublin Core först när ett projekt anpassat formatet till sin egen verksamhet.

Dessutom är avsikten med Dublin Core inte att kunna katalogisera allt, som MARC/KRS teoretiskt sett kan göra. Dublin Core har en annan ambition - minst lika vacker – utan att kunna beskriva allt, skall Dublin Core kunna samverka med andra metadataformat och användas i olika miljöer. I mötet med andra format är översättning av elementen inga problem, eftersom de i grunden är både enkla och vedertagna för objektbeskrivning. Vi har tidigare beskrivit hur formatet kan byggas ut på två olika sätt, antingen med hjälp av *qualifiers* eller med hjälp av egna element. Den stora utmaningen för interoperabiliteten ligger i användandet av *qualifiers*. Stuart Weibel uppmärksammar att ”the range of possible qualifiers is limitless. If applications are to interoperate, it is desirable to constrain these possibilities.”<sup>102</sup> Weibel fortsätter ”standardization of the semantics and methods for qualification of the basic elements is necessary if such qualification is to be widely interoperable.”<sup>103</sup> Att rekommendera exempelvis ett visst internationellt datumformat kanske kan öka standardiseringen och underlätta vid samordning. För nationella system som exempelvis SAB-klassningen är möjligheterna för samordning sämre. Men som Nordiska Metadataprojektet framhåller ”all well built Dublin Core applications will behave gracefully when meeting a qualifier they cannot process.”<sup>104</sup>

Dessutom finns möjligheten att med XML:s så kallade *namespaces* hänvisa till olika klassifikationsscheman i maskinläsbar form. Med *namespaces* i XML kan man även ange olika formatstandarder, exempelvis Dublin Core 1.0 eller USMARC.

---

<sup>100</sup> Day, 1997

<sup>101</sup> ”The Dublin Core is not intended to supplant other resource descriptions, but rather to complement them.”, Weibel et al., 1995 samt; ”Dublin Core may have a role as a minimum metadata set in order to allow interoperability between other, more complex, metadata formats”, Day, 1997

<sup>102</sup> Weibel, 1999

<sup>103</sup> Ibidem

<sup>104</sup> Hakala et al., 1998, s 14

Utbyte mellan olika format kan göras genom mappning. Mappningen styrs av en specifikation över vilka element i de olika formaten som motsvarar varandra semantiskt. I *Dublin Core/MARC/GILS Crosswalk* ställs fältsyntaxen för exempelvis URL:er ”856 \$u” (MARC) och ”DC.Identifier” (Dublin Core) upp som semantiska motsvarigheter.<sup>105</sup> Mappning mellan olika format minskar tidsåtgången för att skapa och underhålla metadata. För Dublin Core är det viktigt att kunna mappas med ett stort format som MARC då det finns mängder av information/metadata världen över lagrade i MARC-formatet. På motsvarande sätt kan biblioteken dra nytta av redan indexerade elektroniska resurser när de vill tillföra elektroniska resurser till biblioteket, eller som Juha Hakala skriver ”use these resource descriptions as a basis of MARC cataloguing, given that conversion to MARC is possible.”<sup>106</sup> För Dublin Core-kollektivet innebär hårdare regler för användningen av formatet ökade möjligheter för interoperabilitet. Men samtidigt är det en fördel att formatet inte är alltför starkt reglerat om man vill göra en anpassning mot ett mer avvikande format för att skapa interoperabilitet dem emellan. Detta är en avvägning att begrunda i arbetet med formatet.

I debatten om Dublin Cores framtida implementering och utveckling urskiljer sig två grupper; minimalisterna och strukturalisterna. Jessica Milstead och Susan Feldman skriver om en ”growing tension between the original design for simplicity and the desire of specialized groups to increase the level of detail to meet their needs.”<sup>107</sup> Minimalisterna menar att enkelheten är det viktigaste, och att man med alltför mycket *qualifiers* riskerar att förlora interoperabilitetsmöjligheterna. Strukturalisterna å andra sidan menar att visst finns den risken, men att *qualifiers* kan ge specifika grupper större utbyte av formatet.<sup>108</sup>

Enkel Dublin Core ger inte samma möjligheter till bibliografisk kontroll som den data biblioteken är vana vid och den metadata som produceras av dokumentproducenten kan inte ersätta bibliotekskatalogisering. *Task Force on Metadata and the Cataloging Rules* undersökning av olika metadataformat, utförd med utgångspunkt från de användarbehov som utvecklats i *Functional Requirements*, ger vid hand att enkel Dublin Core fungerar bäst för att söka (*find*) och för att erhålla de resurser som gav träffar vid sökningen (*acquire*).<sup>109</sup>

Dublin Core används i ett flertal www-projekt, och har över tid kommit att fungera som en rekommenderad standard för metadata-beskrivningar av elektroniska resurser för bland annat Kungliga biblioteket i Stockholm. Någon exakt siffra på hur många projekt världen över som använder Dublin Core har vi inte funnit, men i mars år 2000 listade DCMI 57 Dublin Core-projekt på sin hemsida, även om det finns skäl att anta att långt fler projekt i själva verket använder sig av formatet.<sup>110</sup> Ett annat sätt att skaffa sig en bild av användningen av Dublin Core är genom Steve Lawrence och C. Lee Giles artikel i *Nature*. Artikelförfattarna har i februari 1999 gått igenom 2.500 webbservers hemsidor för att bland annat se hur många av dem som använder kontrollerade eller okontrollerade metadata-beskrivningar. Av dokumenten var 34,2%

---

<sup>105</sup> Mappning mellan GILS, MARC21 och Dublin Core 1.0 enligt Library of Congress Network Development and MARC Standards Office, *Dublin Core/MARC/GILS Crosswalk*, 1999

<sup>106</sup> Hakala, 1999, s 22

<sup>107</sup> Milstead, Jessica och Feldman, Susan, ”Metadata Projects and Standards”, *ONLINE*, January 1999b

<sup>108</sup> Weibel, Stuart, et al., ”The 4<sup>th</sup> Dublin Core Metadata Workshop Report”, *D-Lib Magazine*, juni 1997

<sup>109</sup> *Task Force on Metadata and the Cataloging Rules, Final Report*, August 21, 1998

<sup>110</sup> DCMI:s websida *Projects using Dublin Core metadata*, URL: <http://purl.org/dc/projects/>

vara metadatamärkta (det vill säga att <META></META>-taggar återfinns i HTML-dokumentet) och av dessa använde 0,3% Dublin Core, vilket sannerligen inte kan anses vara någon stor andel.<sup>111</sup>

För att öka tillämpningen är det viktigt att formatet når utanför de begränsade projekten. Charles Thomas och Linda Griffin menar i en artikel i *First Monday* att ”after metadata schemes are established for the Internet, the next major obstacle will be to persuade information creators to use these schemes.”<sup>112</sup> Michael Day menar att om Dublin Core-metadata ”is attached to or embedded in a large number of Web documents, a new generation of Web robots or harvesters could collect this metadata and make it available through improved search services.”<sup>113</sup> Dublin Core-initiativet angriper företeelsen från bägge dessa håll, å ena sidan genom att verka för att få etablerade sökmotorer att förstå och använda sig av Dublin Core-beskrivningen, å andra sidan arbetar man för att HTML-editorerna skall stödja skapande av Dublin Core-poster. Det är lite av Moment 22 över det hela, för samtidigt som det är svårt att motivera hemsidesproducenterna att metadatamärka sina resurser om de stora sökmotorerna ändå inte drar nytta av informationen, vill dessa söktjänster gärna se att det finns ett stort underlag av Dublin Core-märkta resurser innan de ändrar sina indexeringsrutiner.

En aspekt DCMI borde ha i åtanke är vikten av att standardisera användningen av formatet. Detta för att inte riskera att formatets idé förlorar sig i en stor mängd av osökbara poster. Denna risk kommer att stå klar för alla, om inte förr så den dag både stöd för formatet och stora mängder Dublin Core-märkta poster finns på www. DCMI marknadsför Dublin Core som en ”prominent candidate for description of electronic resources” på grundval av dess enkelhet att underhålla, skapa poster i, dess möjligheter till interoperabilitet och mappning mellan andra format. DCMI pekar dessutom på det stöd man ger till andra näraliggande utvecklingsgrenar och styrkan i det internationella samarbetet med formatet.<sup>114</sup>

### 3.2.5 Dublin Core i praktiken: exemplet SAFARI

SAFARI är en svensk söktjänst, skapad på regeringsuppdrag för att tillgängliggöra www-baserad forskningsinformation för allmänheten. Högskoleverket ansvarar för tjänsten, som rymmer kortfattade beskrivningar om vad institutioner gör, presentationer av forskare, pressmeddelanden om ny forskning och avhandlingsresuméer.<sup>115</sup>

Marita Fagerlind och Gunilla Gisselquist redovisar i sin magisteruppsats i Biblioteks- och informationsvetenskap två anledningar till varför SAFARI valt Dublin Core som indexeringsformat. Den första är att ”formatet ansågs ha goda möjligheter att bli en standard” och den andra att Dublin Cores ”femton element täckte --- SAFARI:s behov.” Ett kompletterande fält för målgrupp fick dock läggas till.<sup>116</sup> Målgrupperna (som anges i fältet *TargetGroup/TargetAudience*) är forskare, högskolestuderande,

---

<sup>111</sup> Lawrence och Giles, 1999, s 108

<sup>112</sup> Thomas, Charles F. och Griffin, Linda S., ”Who Will Create Metadata For The Internet?”, *First Monday*, 1998

<sup>113</sup> Day, 1997

<sup>114</sup> DCMI:s hemsida, *Dublin Core Metadata Initiative*, rubrik ”The Dublin Core: A Simple Content Description Model for Electronic Resources”

<sup>115</sup> SAFARI-tjänsten är gratis och finns tillgänglig på URL: <http://www.safari.hsv.se/>, det finns även en mindre lundautgåva *SAFARI i Lund* på URL: <http://www.lu.se/lu/forskn/sil/>

<sup>116</sup> Fagerlind, Marita och Gisselquist, Gunilla, *Metadata enligt Dublin Core*, 1999, s 36

allmänhet/skola och företag/organisationer. I SAFARI finns därmed möjlighet för exempelvis forskaren att hitta resurser på önskvärd nivå.

För att kunna publiceras på SAFARI ställs följande krav på sidorna:

- namn på institution eller motsvarande
- lättfattlig text (målgrupp gymnasieungdom) som övergripande beskriver vilken forskning som bedrivs
- namn på institutionens samtliga forskare samt deras e-postadress (denna information behöver inte ligga på sidan utan kan finnas med som länk)
- namn på ansvarig för hemsidan/ansvarig utgivare
- datum för senaste ändring av det intellektuella innehållet
- teknisk ansvarig, till exempel webmaster
- länkar uppåt till överordnad organisation

Innehållet på sidorna skall vara övergripande information om projekt och pågående forskning, eller information om publikationer eller personer (kompetens).<sup>117</sup>

I SAFARI kan man göra två typer av sökningar; enkel och avancerad. Med avancerad sökning kan man söka efter exempelvis ett visst namn, en viss typ av dokument eller efter all forskning vid en viss organisation.

I träfflistan visar dokumentens Dublin Core-element *DC.Title*, *DC.Description* och *DC.Identifier*, *DC.Publisher* och *DC.Publisher.Address* - alltså titel, kort beskrivning, utgivare och länk till dokumentet.

dc.title hyperlänkat med dc.identifier

dc.description

dc.publisher hyperlänkat med dc.publisher.address

Figur 11. Dublin Core-fält som döljer sig i SAFARI:s träfflista

<sup>117</sup> SAFARI i Lund, *Vad kan registreras i SiL?*, <http://www.lub.lu.se/~colm/safari/typer.html.sv>

Från träfflistans länk går du vidare till källan ute på den forskande myndigheten. SAFARI klassificeras med CERIF, ett hierarkiskt klassifikationsschema som även tillåter en bläddringsfunktion i en trädstruktur. CERIF innehåller 350 termer i tre olika nivåer och finns översatt till flera europeiska språk. Klassifikationen skall i första hand tjäna som underlag för bläddringsstrukturen.<sup>118</sup>

Informationen produceras decentraliserat genom att man låter de olika forskande myndigheterna runt om i landet ansvara för att förse sin www-baserade forskningsinformation med Dublin Core. Posten läggs i HTML-dokumentet som inbäddad metadata. Metadataposten skapas med hjälp av SAFARI:s inmatningsformulär (se Figur 12).

The image shows a screenshot of a Netscape browser window displaying the 'SAFARI Dublin Core metadata editor' web form. The browser's address bar shows the URL 'http://nothelub.lu.se/safari/metadata/safari/plangsv'. The form contains several sections with labels and input fields: 'TITEL' (Title) with a text input field; 'FÖRFATTARE/TEXTANSVARIG' (Author/Responsible) with a dropdown menu for 'Person som upphovsrätt: "Efternamn, Förnamn"', a text input for 'Författarens (Email) adress', and a red '+' icon; 'URL' (URL) with a text input field and a red '+' icon; 'ÄMNE: Klassifikation CERIF' (Subject: CERIF Classification) with a dropdown menu showing 'Ingen CERIF-kod vald' and a red '+' icon; 'ÄMNE: Nyckelord' (Subject: Keywords) with a text input field and a red '+' icon; 'Målgrupp' (Target Audience) with a dropdown menu showing 'oläsrhetensskola' and a red '+' icon; and 'BESKRIVNING' (Description) with a text input field. The browser's status bar at the bottom shows 'Document: Done'.

Figur 12. Inmatningsformulär för SAFARI<sup>119</sup>

Efter inmatning i formuläret begär man att få informationen genererad som en Dublin Core-post uttryckt i HTML, vilken man klistrar in i källdokumentet mellan attributen <HEAD> och </HEAD>. SAFARI:s robot sparar dokumentets URL och hämtar därefter metadataposten från källdokumentet vid ett senare tillfälle.

Möjligheten för interoperabilitet är viktig för en sådan här typ av tjänst, för sökning finns stöd för Z39.50:s GILS-profil.<sup>120</sup> Marita Fagerlind och Gunilla Gisselquist menar

<sup>118</sup> SAFARI i Lund, *SiL och klassifikationssystemet CERIF*, <http://www.lub.lu.se/~jorgen/safari/cerif.html>

<sup>119</sup> SAFARI, "Dublin Core metadata editor"

<sup>120</sup> Lundberg, Sigfrid, *En arkitektur för ett distribuerat system för spridning av forskningsinformation*, 1999, s 7

i sin undersökning av tjänsten att denna Z39.50-profil skulle kunna ”underlätta parallella körningar mellan t.ex. SAFARI och LIBRIS.”<sup>121</sup>

Resursadresserna kontrolleras en gång i veckan. Då kontrollerar sökroboten även om resursernas metadata förändrats och överför förändringarna till SAFARI:s ”katalog”.

### 3.3 Samordning mellan metadataformaten

Att utveckla ett format speciellt för den egna kontexten kan inledningsvis förefalla ekonomiskt, men är i ett längre perspektiv svårförsvårbart. Ett mer framsynt och ekonomiskt katalogiseringsformat erbjuder möjligheten att kopiera, importera och dela existerande poster med andra. Detta skall fungera även om dessa andra kataloger använder ett annat katalogiseringsformat. I de föregående avsnitten har vi behandlat två olika metadataformat och bland annat granskat deras förmåga till interoperabilitet. Möjligheterna att använda dem i olika miljöer och för att beskriva olika dokumenttyper bottenar i formatkreatörernas strävan att låta formaten följa internationella standarder och protokoll. Detta för även med sig att framtida översättning till andra format eller bruk i gemensamma databaser är möjligt. En post skapad med Dublin Core skall alltså helst kunna konverteras till en MARC-post utan att man skall behöva katalogisera källan på nytt.

I detta avsnitt kommer vi att översiktligt presentera protokollet Z39.50, ”metadataramverket” RDF samt CORC-projektet (*Cooperative Online Resource Catalog*). Vi vill med detta skapa en vidare förståelse för vikten av de metadataformatets funktionaliteter som kan vara angelägna för det bibliotek som skall bygga upp en virtuell samling. Paul Miller menar att ”Z39.50 is often attacked --- [för att vara, vår anm.] no match for the latest 'great idea' (currently this 'great idea' is usually cited as being some combination of XML and RDF).”<sup>122</sup> Vi kommer inte att diskutera vilken samordningsmöjlighet som egentligen är den ”bästa idén”. Vi menar att det ena verktyget inte behöver utesluta det andra och fokuserar de möjligheter till interoperabilitet de bägge skänker.

#### 3.3.1 Z39.50

Z39.50 är en standardiserat protokoll för datakommunikation mellan olika databaser. Z39.50 är första hand bibliotekens skötebarn och protokollets funktionalitet har utvecklats speciellt med biblioteksvärldens behov i åtanke. Protokollet kan användas dels för utbyte av information mellan bibliotekskataloger världen över, men också för sökning i kommersiella databaser och virtuella samlingar på Internet. Utvecklingen av standarden görs av Library of Congress *Z39.50 Implementor Group* (ZIG). Z39.50 har existerat och utvecklats sedan andra hälften av 1980-talet. Den tredje och senaste version heter *ANSI/NISO Z39.50-1995, Information Retrieval (Z39.50): Application Service Definition and Protocol Specification*.<sup>123</sup>

---

<sup>121</sup> Fagerlind och Gisselquist, *Metadata enligt Dublin Core*, 1999, s 49

<sup>122</sup> Miller, Paul, ”Z39.50 for All”, *Ariadne*, 1999

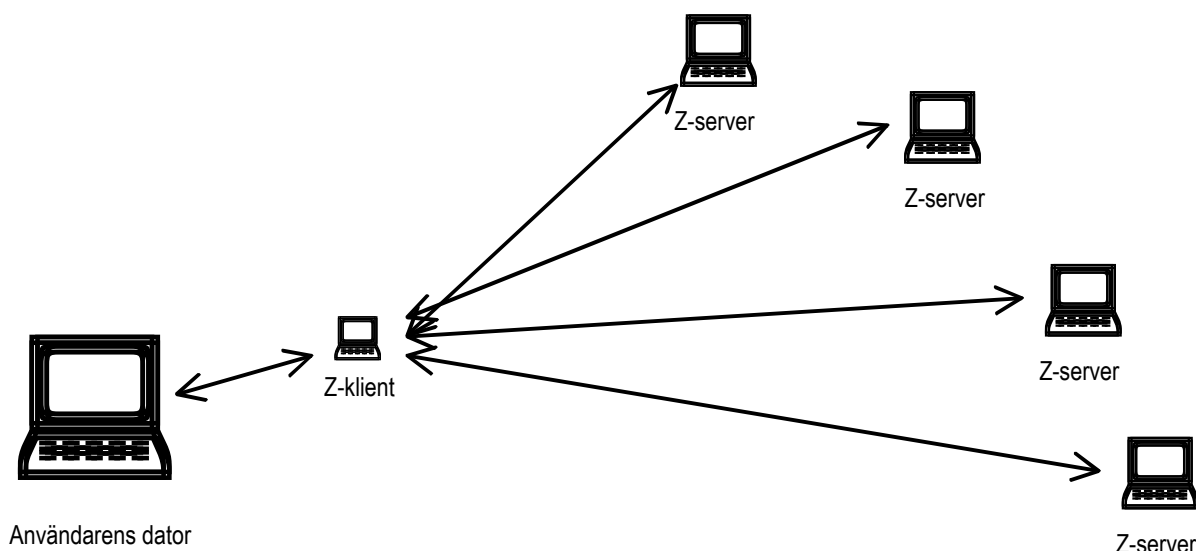
<sup>123</sup> *ANSI/NISO Z39.50-1995, Information Retrieval (Z39.50): Application Service Definition and Protocol Specification* är en amerikansk standard. Det finns en motsvarande internationell standard, *ISO 23950:1998, Information and documentation – Information Retrieval (Z39.50) – Application service definition and protocol specification*. Mer om detta kan man läsa om i exempelvis Paul Millers ”Z39.50 for All”, *Ariadne*, 1999



Z39.50 gör det möjligt för användaren att använda det egna systemets sökspråk även då man söker i (en eller flera) externa databaser. Protokollet kan även användas exempelvis för överföring av poster från en central katalog till olika lokala bibliotekskataloger. Z39.50 följer en klient-servermodell, där den ena datorn, Z-klienten (*origin*), skickar en förfrågan till Z-servern (*target*) som behandlar frågan och skickar tillbaka sitt svar. För att kommunikationen skall fungera krävs stöd för protokollet hos både server och användarens sökprogramvara/klient. För detta utvecklar man speciella profiler, som översätter det egna formatet och sökbegreppen till Z39.50-förståeligt "språk". Protokollet arbetar allt som oftast osynligt för användaren. "Developers have simply become good – possibly *too* good, given their need to raise awareness in order to sell future products? - at hiding it."<sup>124</sup>

---

<sup>124</sup> Miller, 1999



Enkelt uttryckt går sökning med Z39.50 till ungefär så här:

- Användaren vid exempelvis bibliotekets OPAC väljer vilken/vilka kataloger (Z-servers) han/hon vill söka i och anger söksträng
- OPAC skickar detta vidare till Z-klienten (ofta en del av bibliotekets mjukvara, vilket för tydlighetens skull inte framgår av figuren)
- Z-klienten översätter förfrågan till Z39.50-förståeligt "språk"
- Z-klienten upprättar kontakt med den eller de Z-servers användaren valt och skickar förfrågan vidare
- Z-servern översätter förfrågan till det sökspråk som används i den aktuella katalogen, utför sökningen, mottar svar, samt skickar detta vidare till Z-klienten i det format klienten angivit
- Z-klienten mottar träffarna och vidarebefordrar dessa till OPAC
- Träfflistan visas snyggt och prydligt i bibliotekets OPAC

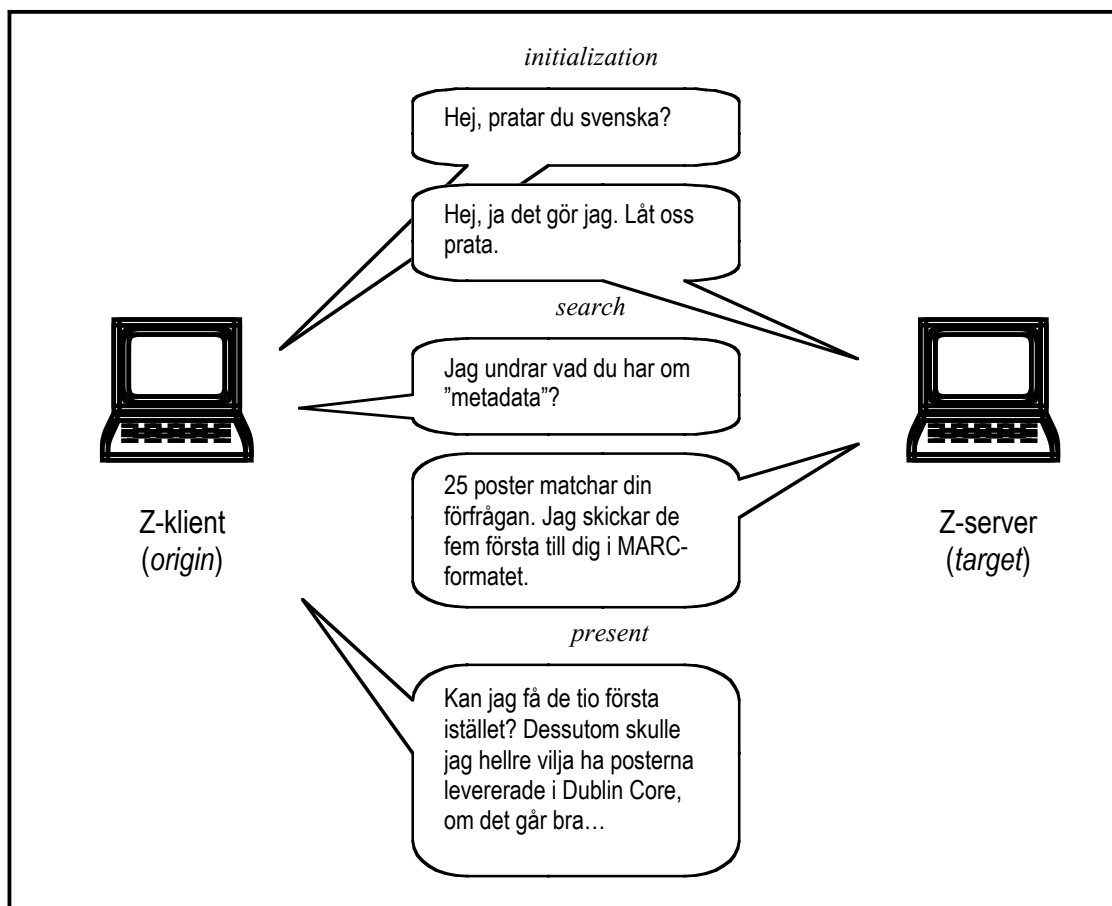
### Figur 13 Sökning med Z39.50

Direkt från sin terminal kan man alltså söka i ett annat biblioteks katalog, externa databaser eller i den lokala bibliotekskatalogen (vilken då fungerar som Z-server). Man kan också, som i exemplet ovan, göra simultana sökningar i flera olika Z-servers samtidigt. *Arts & Humanities Data Service* (AHDS) är en virtuell samling som använder sig av Z39.50 för sökning i fem olika databaser, vilka innehåller information inom olika ämnen och bygger på olika katalogiseringsformat. Paul Miller vid Hulls universitet menar att, trots databasernas olikhet, "the combination of Z39.50 and Dublin Core elements is sufficient to enable meaningful searches across five sites".<sup>125</sup>

Kommunikationen mellan Z39.50-datorer består av elva block (*facilities*) som var och en delas upp i en eller flera *services*. De tre grundläggande och vanligaste tjänsterna är; *initialization*, *search* och *present*. *Initialization* är det skede i vilket systemen upprättar kontakt och enas om vilka regler som skall gälla för kommunikationen. Om klient och server implementerat olika tjänster sätts standard på sessionen enligt principen om minsta gemensamma nämnare. Det är nämligen möjligt att

<sup>125</sup> Ibidem

implementera endast vissa av de *facilities* protokollet stöder. Under *initialization* eller *init* hanteras också eventuell lösenordsöverföring. Under *search* överförs och utförs klientens fråga, och slutligen levereras svaret och presenteras med hjälp av tjänsten *present*. Med *present* styr man alltså hur man vill ha svaren levererade, för bibliografiska poster exempelvis i vilket format man vill ha dem. Ett kraftigt förenklat exempel på z-kommunikation presenteras nedan (Figur 14):



**Figur 14 Förenklad Z39.50-kommunikation mellan två datorer<sup>126</sup>**

För att Z-kommunikationen skall kunna komma till stånd krävs, som vi ovan noterat, att både klient och server har stöd för protokollet. Båda parter använder sig i kommunikationen av olika uppsättningar attribut (*attribute sets*) för att styra hur sökningen skall gå till. Ett *attribute set* består av en samling gemensamma attribut som skall användas för ett visst ändamål. Den i biblioteksvärlden mest kända uppsättningen attribut är kanske Bib-1. Bib-1 utvecklades ursprungligen för att hantera bibliografisk information, men används även i andra sammanhang. Med de attribut som ingår i Bib-1 kan man styra vilka fält (exempelvis titel- eller författarfält) sökningen skall ske i, vilka värden som skall gälla för trunkering et cetera.

Z39.50 är inte självklart knutet till någon syntax, eller format att uttrycka exempelvis bibliografisk information i. Då formatet framför allt använts i bibliotekssammanhang stöder dock de flesta z-servers någon medlem i MARC-familjen. Data kan emellertid

<sup>126</sup> Figuren bygger på ett exempel i Paul Millers "Z39.50 for All", 1999

också överförs i SUTRS (*Simple Unstructured Text Record Syntax*, eller vanlig ostrukturerad text), XML (*eXtensible Markup Language*, vilket vi kommer att diskutera närmare längre fram) eller på annat sätt. I olika profiler (*profiles*) samlar man *attribute sets* och syntaxer som skall passa särskilda behov. Aktuell för biblioteksvärlden är just nu *Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery*. I denna har man försökt jämka samman redan existerande profiler och uppnå en jämförelsevis enkel enhet där värdena för alla *attributes* är strängt definierade. Bath-profilen är tänkt att fungera som ett slags gemensam plattform som ökar funktionaliteten i mötet mellan Z-klient och -server.

Vi har hittills talat endast om de möjligheter till sökning Z39.50 erbjuder. Protokollet kan dock användas till mycket annat. Med hjälp av protokollets olika *services* kan man på ett bekvämt sätt hantera fjärrlån, filtrering av information, databasuppdatering et cetera. Många söktjänster och virtuella samlingar använder sig också av Z39.50 för dessa syften.<sup>127</sup>

*Er lykken gjort?* frågar sig den danske bibliotekariestudenten Rasmus Bruun retoriskt apropå Z39.50.<sup>128</sup> Naturligtvis är lyckan inte gjord i och med utvecklingen av protokollet, men man kan mycket väl vara en bit på väg. 1998 skriver Ralph LeVan i ett diskussionsunderlag från Dublin Core-initiativet att många frågor återstår att ta ställning till. Som vi sett enas klient och server i *init* om en gemensam nivå för kommunikationen. Är flera servers inblandade i en session kan svaren från dessa vara olika precist definierade. Som i den gamla motsättningen mellan minimalister och maximalister måste man kanske ibland vara beredd att "trade precision for recall", det vill säga att godta mindre precisa svar i interoperabilitetens namn.<sup>129</sup> Förhoppningsvis kan gemensamma ansträngningar som *the Bath profile* tjäna som ett slags *common ground* för mötet mellan klienter och servers från olika miljöer och med olika katalogiseringsformat i botten

År 2000 beskrivs Z39.50 i *Biblio Tech Review* som alltmer viktigt för den framtida utvecklingen av sammanlänkade bibliotekssystem. Man förutspår här också att bibliotekssystemen kommer att bli lika "Z39.50 enabled" som deras *Online Public Access Catalogues* (OPACs) blivit "Web enabled".<sup>130</sup> Enligt Paul Miller har Z39.50 ofta anklagats för att på flera nivåer vara överdrivet komplext (vilket kanske inte framgått av denna vår mycket förenklade genomgång), gammaldags och inte tillräckligt anpassat till www-miljön.<sup>131</sup> Miller menar också att det är sant att

Z39.50 has quirks and limitations --- [men, vår anm.] the very fact that Z39.50 has been extensively used for long enough to be criticized as old fashioned is surely a testament both to its robustness and to the lack of any viable alternative.<sup>132</sup>

Att protokollets framgångar delvis grundar i att *gammal är äldst* kan vara sant, men standarden breddas och förnyas också hela tiden, och olika *Profiles* och *Attribute sets* skapas för att möta behoven hos nya användare. Det har fästs oerhörda förhoppningar

---

<sup>127</sup> Biblio Tech Review, *Z39.50: Part 1 – an overview*, 2000

<sup>128</sup> Bruun, 1997

<sup>129</sup> LeVan, Ralph, *Dublin Core and Z39.50, Draft Version 1.2*, 1998

<sup>130</sup> Biblio Tech Review, *Z39.50: Part 1 – an overview*, 2000

<sup>131</sup> Miller, 1999

<sup>132</sup> Ibidem

till Z39.50, och det förutspås komma att betyda lika mycket som MARC-formatet på sin tid gjorde för biblioteksvärlden.<sup>133</sup> Miller hävdar att

It appears for the moment [september 1999, vår anm.] that, whatever its limitations, Z39.50 remains the only effective means of enabling simultaneous queries upon distributed heterogeneous databases, and this remains something that the broader user community wants to be able to do.<sup>134</sup>

### 3.3.2 RDF & XML

Clifford Lynch skriver i en artikel i *Scientific American* från 1997 att

The Web, moreover, still lacks standards that would facilitate automated indexing. As a result, documents on the Web are not structured so that programs can reliably extract the routine information that a human indexer might find through a cursory inspection: author, date of publication, length of text and subject matter.<sup>135</sup>

Tre år efter Lynchs artikel har vi kommit en bit på vägen mot ett strukturerat www. RDF (*Resource Description Framework*) är en nykomling som till skillnad från MARC, Dublin Core och Z39.50 inte har någon självklar anknytning till biblioteksvärlden.

RDF utvecklas av W3C, en sammanslutning av aktörer som arbetar för att skapa enhetliga standarder och regler för Internetteknik. RDF är en infrastruktur eller en stomme inom vilken beskrivning och utbyte av metadata kan göras.

Arbetet med RDF befinner sig i ett utvecklingskede, men det finns redan mycket tilltro till strukturen från både biblioteksvärlden och Internetsamfund. För autencitetskontroll vid exempelvis handel över nätet förutspår Dan Brickley och R.V. Guha att RDF i kombination med digitala signaturer ”will be key to building the 'Web of Trust' for electronic commerce, collaboration, and other applications.”<sup>136</sup>

RDF uttrycks i bland annat markeringsspråket XML, som också utvecklas av W3C. XML är ett mer strukturerat språk för att hantera information på www än HTML. XML är en delmängd av SGML (*Standardized Markup Language*), ett markeringsspråk med ISO-standardiserade regler (ISO 8879:1986). HTML är en enklare variant; de attribut markeringsspråket stödjer handlar huvudsakligen om grafisk presentation, exempelvis ger `<I>text</I>` kursiverad *text*.

Kraftigt förenklat kan man säga att skillnaden mellan HTML och XML är att det sistnämnda ger dokumentproducenten möjlighet att skapa och koppla egna attribut till delar av dokumentet. På så vis kan man dela upp dokumentet i olika beskrivningsdelar, exempelvis `<author></author>`, ”author” kallas inom RDF för en *PropertyType*. RDF sätter inte upp regler för dessa semantiska *PropertyTypes*, utan istället för den struktur - ett slags containerarkitektur - som semantiken skall verka i.

W3C:s arbete med XML och RDF kan förbättra återvinning i sökmotorer genom att förbättra beskrivningen av en enkel hemsida, en site eller ett digitalt biblioteks

---

<sup>133</sup> Se exempelvis Göran Skogmars ”Från onlinekataloger till superkataloger”, *LIBRIS-meddelanden*, 1994, s 6

<sup>134</sup> Miller, 1999

<sup>135</sup> Lynch, 1997

<sup>136</sup> Brickley, Dan och Guha, R. V., ”Resource Description Framework (RDF) Schema Specification”, 1999

innehåll och dess relationer. Inom RDF-strukturen kan man blanda olika metadataformat i en gemensam RDF-kontainer, och i denna ange att exempelvis värdena för både ”DC.Creator” och MARC:s ”100 a”, två tänkbara *PropertyTypes*, representerar var information om resursens författare kan återfinnas. På så sätt skapas interoperabilitet mellan formaten och möjligheterna för att utnyttja den strukturerade informationen blir nästan gränslösa. Tänkbart är att sökmotorn hämtar olika *PropertyTypes* och vid kombination av dessa - likt uttagna satsdelar - kan ge en presentation av vilka *titlar* som skrivits av en viss *författare*. Metadataproducenten kan därmed välja att använda det format som passar bäst i den egna tjänsten.

XML:s flexibilitet i kombination med RDF-strukturen utlovar såväl bättre interoperabilitetsmöjligheter som bättre möjligheter för dokument att användas i databasmiljö, och i linje med detta även förbättrade möjligheter för användaren att återvinna information från www. Utsikterna för det digitala biblioteket förbättras således om RDF:s strukturerade metadataamverk kommer i omfattande bruk. Det kvarstår dock fortfarande semantiska problem som det behövs mänskligt intellekt för att lösa.

### 3.3.3 Samordning i praktiken: exemplet CORC

*Cooperative Online Resource Catalog* (CORC) är ett av bibliotek världen över kooperativt ägt system för katalogisering av elektroniska resurser.<sup>137</sup> Eller som man själv uttrycker det ”CORC empowers librarians with automated tools for the cooperative creation, selection, organization, and maintenance of web-based resources.”<sup>138</sup> Projektet drivs av OCLC och har funnits sedan januari 1999. För närvarande är man ännu på pilotstadiet, men i framtiden ämnar man ha tusentals bibliotek anslutna.<sup>139</sup> Katalogen är till för biblioteken, och är inte tillgänglig för allmänheten. CORC bygger vidare på tidigare OCLC-projekt, som *InterCat Project* och *NetFirst*. Som kuriositet kan vi nämna att *InterCat* i samarbete med Library of Congress ledde fram till skapandet av MARC-fält 856.

Med CORC ges möjligheten att integrera resurser från CORC:s databas i bibliotekens lokala kataloger. Databasen skapas genom användning av ”automated tools and librarian talent”<sup>140</sup>. Biblioteken söker upp resurser som sedan indexerats automatiskt med fristående metadata i den gemensamma databasen. Biblioteken kontrollerar och berikar CORC-beskrivningen.

Inom CORC sätts inga strikta urvals- och kvalitetskriterier för resurserna upp, utan OCLC låter istället de enskilda biblioteken ställa upp dessa kriterier. Tanken är att i såväl detta som i arbetet med kontroll och beskrivning av de elektroniska resurserna skall arbetsbördan bli hanterlig eftersom många bibliotek delar katalogen. Metadata gurun Traugott Kochs begrepp *quality controlled subject gateways* kan inte användas om CORC, då samlingen varken är ämnesbaserad eller aktivt kvalitetskontrollerad. Koch, LUB NetLab, ser CORC som extremen av en *subject*

---

<sup>137</sup> På CORC:s hemsida (<http://corc.oclc.org/>) erbjuds möjligheten att logga in som gäst.

<sup>138</sup> OCLC Online Computer Library Center Inc., *CORC Honor Roll of Participants*

<sup>139</sup> ”Currently there are some 200 libraries participating and we plan to gradually open it to the group of 7,000 cataloging libraries after it becomes a production system in the summer of 2000”, Hickey, Thomas B. ”CORC: a system for gateway creation” *Online Information Review*, 2000, s 53

<sup>140</sup> OCLC Online Computer Library Center Inc., *The OCLC Cooperation Online Resource Catalog project*

*gateway* ”a truly universal service, all/many subjects, global, multilingual, all types of resources”.<sup>141</sup> Styrkan i CORC är att man vill samla många olika samlingar i samma databas och att stödja både bibliotekens traditionella MARC-format och ge stöd även för Dublin Core-beskrivningar, och fullt stöd för standarder som Z39.50 och RDF.

I avsnittet ”Format för objektbeskrivning” redovisade vi DESIRE:s typologi för metadataformat. I band 2 hamnade de strukturerade formaten, med sina framväxande standarder och grundläggande fältstrukturer. Dublin Core fick exemplifiera ett metadataformat av band 2-typ. I CORC-katalogen används både Dublin Core och MARC för metadatabeskrivningarna, och OCLC har inom CORC valt en beskrivningsnivå som motsvaras av den de femton elementen med *qualifiers* Dublin Core erbjuder. Posterna inom CORC mappas mellan Dublin Core och MARC, vilket gör att katalogisatören kan välja ett behändigt format att beskriva med, och att andra därefter kan välja att importera posten i det format som önskas till den egna katalogen. Posterna kan importeras till och exporteras från basen i XML/RDF, HTML och Unicode.

På detta sätt kan katalogisatörens kunskaper inom MARC-katalogisering tillvaratas. Juha Hakalas tanke om att inklistrad Dublin Core-metadata kan tjäna som en utgångspunkt för MARC-katalogisering ”given that conversion to MARC is possible”, blir här vida överträffat genom att mappning mellan formaten inom CORC ska ske utan att det ena formatet behöver justeras för att fungera.<sup>142</sup>

CORC:s modell är ett av många sätt att lösa objektbeskrivningen av de elektroniska resurserna. I en ganska närliggande framtid måste våra bibliotek parallellt hantera pappersdokument och elektroniska resurser, inte minst dokument på www. I det följande kommer vi att betona vikten av att biblioteksvärlden beaktar detta, och utvecklar strategier för sin hantering av www-resurser.

---

<sup>141</sup> Koch, Traugott, ”Quality-controlled subject gateways: definitions, typologies, empirical overview”, *Online Information Review*, volym 24, nr 1, 2000, s 27

<sup>142</sup> Hakala, 1999, s 22

## METADATA I BIBLIOTEKET

---

I detta avsnitt kommer vi att presentera och diskutera några olika modeller för samlingar av elektroniskt material, ”det digitala biblioteket”. Vi utgår i detta mer tydligt från biblioteksperspektivet och pekar på vilka konsekvenser de olika modellerna medför för den enskilda instansen. Särskilt avseende fästs vid de elektroniska och de tryckta dokumentens olika karaktär, metadataformatens möjligheter samt vilken grad av kontroll biblioteket vill ha över sina virtuella samlingar. Ytterst handlar det ju om vilken hållning man intar visavi de elektroniska resurserna. Nöjer sig biblioteket med att endast erbjuda användarna åtkomst till de elektroniska resurserna har man självfallet inget inflytande över vilken form av och hur noggrann objektbeskrivning resurserna får. Men biblioteket kan även välja en mer offensiv linje, med aktivt ”förvärv” av elektroniskt material. I detta fall blir frågor om objektbeskrivning, val av beskrivningsformat, arkivering och eventuellt samarbete med andra katalogiserande instanser aktuella. Ett logiskt steg är även att digitalisera de förfintliga samlingarna för att underlätta spridning och åtkomst. Även i detta arbete spelar beskrivningsformatet stor roll.

### 4.1 Det elektroniska dilemma

Överlag var biblioteken tidiga Internetanvändare. Bruket av www-resurser kan emellertid variera mycket mellan ett litet landsortsbibliotek och ett större forskningsbibliotek. Anledningar till detta kan spåras i exempelvis biblioteksanvändarnas behov och bibliotekets ekonomiska ramar. I sämsta fall kan det vara så att biblioteket endast *tillhandahåller* uppkoppling mot Internet, men inte utnyttjar dess möjligheter som informationsresurs. Internets möjligheter reduceras då till en bra-att-ha-grej, något biblioteket vill kunna erbjuda de användare som vill surfa och ha skoj.

Biblioteken kan också nöja sig med att lägga in ett antal bokmärken i sin webläsare till olika, i bästa fall kvalitetskontrollerade samlingar. Vad biblioteksanvändaren i detta fall kommer att hitta beror på en mängd olika omständigheter: hur effektiv den valda söktjänsten är (hur pass väl den använder sig av metadata), hur pass datavan och kunnig (och villig att hjälpa!) bibliotekarien är, och sist men inte minst hur lång tid den informationssökande har på sig. Det borde finnas något som tillgodoser hans/hennes behov bland de mer än hundra miljoner dokument exempelvis AltaVista samlat in. Svårigheten är bara att hitta det. I de bägge ovanstående ”modellerna” är dessutom de resurser som skulle kunna vara intressanta för biblioteket och dess användare uppblandat med hemsidor av i bästa fall privat intresse, reklamsiter et cetera.

Men hur skall biblioteket då ta sig an uppgiften? Vi har redan diskuterat problemen de elektroniska resurserna ställer katalogisatorerna inför. Då man skall beskriva den enskilda resursen slås man av att den gestaltar alla problem man brottats med då man skulle beskriva det ”traditionella” materialet – och några till. De elektroniska dokumenten kan mycket lätt flyttas, förändras eller försvinna. Till skillnad från de traditionella resurserna är de inte knutna till något särskilt bibliotek, tvärtom kan de lätt hämtas och kopieras av vem som helst. En elektronisk resurs kan vidare bestå av många sinsemellan olika typer, och blandas på www med objekt som över huvud



taget inte kan betecknas som publikationer. Då nya former av elektroniska resurser också uppträder hela tiden kan det vara svårt att avgöra om den aktuella resursen ens borde katalogiseras. Och det kanske mest akuta problemet med de elektroniska dokumenten: de äro legio. Sten Hedberg konstaterar uppgivet att de är ”minst hundra gånger fler än de traditionella dokumenten, och dessa minskar inte nämnvärt i antal.”<sup>143</sup> Att utöver arbetet med det traditionella materialet bedöma och eventuellt katalogisera denna dokumentmassa är en jätteuppgift. Juha Hakala påpekar också att trots att bibliotekariernas ansvarsområde, de traditionella resurserna, utökats med det elektroniska materialet har man ”no extra resources to do the job.”<sup>144</sup>

Som biblioteksanvändare förväntar man sig att biblioteket tillhandahåller även elektroniskt material, och kraven på biblioteken förväntas öka i framtiden eftersom exempelvis utgivning av elektroniska tidskrifter endast sker på www.

## 4.2 Virtuella samlingar och digitala bibliotek

Dokumenterna på www finns där allihop, och förutsatt att man har den rätta URL:en är de flesta av dem direkt åtkomliga. Är man inte riktigt säker på vad man söker har man dock inte stora chanser att hitta rätt bland miljontals resurser. Riktigt meningsfulla blir åtkomstmöjligheterna först då dokumenten samlats in och beskrivits i en katalog. I den virtuella samlingen finns ett urval av resurser samlade efter olika kriterier - allt från ”vi tar emot allt vi kan hitta” till ämnes- eller målgruppspecifika samlingar. Det är den valda plattformens beskrivningsnivå som gör det möjligt att kvalitetsbedöma resursen. Internetkatalogerna kan vara antingen kommersiella som NetSök eller offentligt finansierade som Länkskafferiet eller SAFARI.<sup>145</sup>

Den *value-adding* eller det ytterligare värde en virtuell samling innebär är förutom urvalet just de metadataposter man beskrivit resurserna med. Bibliotekstjänsts NetSök har vi tidigare berört i avsnittet ”MARC i praktiken: exemplet NetSök”. NetSök använder sig som vi sett av korta MARC-poster för att beskriva de resurser man valt ut, något som verkar rimligt utifrån de förutsättningar man har; importen av poster från NetFirst och den nära kopplingen till deras stora databas BURK. Dessa bägge använder sig av MARC-formatet i olika varianter, och kravet på interoperabilitet med främst den sistnämnda basen har medfört att man valt MARC-formatet även i sin katalog över elektroniska resurser. För SAFARI verkar valet av Dublin Core självklart. I SAFARI sköts katalogiseringen huvudsakligen ute på de forskande instanserna, och det är viktigt att formatet är lätt att uttrycka sig i. Inmatningsformuläret underlättar också registreringen av dokument i SAFARI. Att katalogposten ligger inbäddad i dokumentet gör att det ligger närmare till hands för hemsidesproducenten att uppdatera informationen då sidans innehåll förändras. Inom en vecka har SAFARI:s sökrobot samlat in posten och uppdaterat samlingen.

En intressant aspekt av de virtuella samlingarna är att man med dessa alltmer kommit att marknadsföra förtroende. Ansvar för dessa samlingar, deras urval och underhåll, är helt utanför användarnas kontroll. I exemplet NetSök ser vi hur Bibliotekstjänst (BTJ) genom att betona sina höga urvalskriterier rörande struktur, språkbruk och

---

<sup>143</sup> Hedberg, Sten, *Bruket av metadata enligt Dublin Core : Principer, teknik och tillämpningar utanför Sverige*, senast uppdaterad 000512

<sup>144</sup> Hakala, 1999, s 22

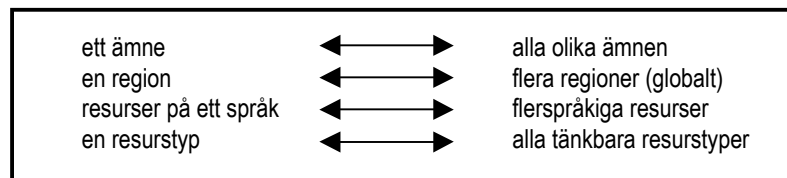
<sup>145</sup> Länkskafferiet, <http://länkskafferiet.skolverket.se>

tillförlitlighet försöker bygga upp ett förtroende för NetSök och BTJ. Kunderna är tvungna att förlita sig på NetSök-redaktörernas värdering även i de mer diffusa frågorna om kvalitetsurval. På så sätt blir valet att använda sig av NetSök, eller någon annan kvalitetskontrollerad samling, en fråga om förtroende.

Traugott Koch definierar i artikeln ”Quality-controlled subject gateways: definitions, typologies, empirical overview” *subject gateways* på samma sätt som vi definierat vad vi kallar *virtuella samlingar*. Virtuella samlingar är

... Internet services which support systematic resource discovery. They provide links to resources (documents, objects, sites or services), predominantly accessible via the Internet. The service is based on resource description. Browsing access to the resources via a subject structure is an important feature.<sup>146</sup>

Vår virtuella samling och Kochs *subject gateway*'s täckningsgrad spänner mellan



**Figur 15 Subject gateways enligt Koch<sup>147</sup>**

De virtuella samlingarna kan alltså skilja sig åt på många olika sätt. Men framför allt är de ett utmärkt exempel på det kommersiella värde katalogiseringens *value-adding* har, utöver sin betydelse för den bibliografiska kontrollen.

Generellt kan vi säga att det ställs annorlunda krav på katalogiseringen av de elektroniska dokumenten än på de mer traditionella dito. Inte skulle man acceptera en MARC-post där inte titelfältet fanns med? För de elektroniska resurserna förlitar man sig emellertid alltför ofta helt på den använda söktjänstens bristfälliga beskrivningar, liksom hela Internet ursäktas med att det är så uselt strukturerat. Vid uppbyggnaden av ett digitalt bibliotek måste ribban sättas högre. Vad skiljer då ett digitalt bibliotek från en virtuell samling? I litteraturen stöter vi på många olika definitioner av vad ”det digitala biblioteket” egentligen är och vad det skall göra. Många är också de förmenta synonymerna; ”hybridbibliotek”, ”virtuellt bibliotek”, ”elektronisk samling” et cetera.

Robin Murray vid engelska Fretwell-Downing Informatics definierar ett digitalt bibliotek som ”a managed environment providing integrated access to an extensible range of information services independent of location, format and curatorial domain within the appropriate business and legal framework.”<sup>148</sup> Med ”an extensible range of information services” avser Murray tillgång till både traditionella och elektroniska resurser.

<sup>146</sup> Koch, 2000, s 24 f

<sup>147</sup> Ibid, s 27

<sup>148</sup> Murray, Robin, ”The digital library jigsaw: fitting the pieces together”, *Online information '99*, 1999, s 175

Library of Congress redogör i sin *Challenges to building an effective digital library* för tio olika utmaningar i arbetet med att bygga upp digitala bibliotek. Punkt fyra i denna handlingsplan behandlar interoperabilitet; utmaningen är att "Establish protocols and standards to facilitate the assembly of distributed digital libraries."<sup>149</sup>

I punkt två konstaterar man att "In order to apply scarce resources to the digitalization of significant quantities of content, it is often necessary to reduce the level of detail offered in accompanying catalogs or indexes." Detta skall pareras med utvecklingen av "search and retrieval tools that compensate for abbreviated or incomplete cataloging or descriptive information."<sup>150</sup>

Riktigt intressant blir det i punkt sex, där utmaningen är att "integrate access to both digital and physical materials". I uppbyggnaden av ett digitalt bibliotek måste man ta hänsyn till att användarens fokus är på dokumenten, inte på vilken form de råkar befinna sig. Eller som Arms uttrycker det "Users want intellectual works, not digital objects". Han fortsätter "The underlying architecture has to support two main needs. It must provide methods for grouping digital library objects and must provide means for retrieval."<sup>151</sup>

För att återvända till Library of Congress och deras utmaningar (punkt sju) måste materialet också presenteras på ett enhetligt vis och trots samlingarnas heterogena natur får man inte förlita sig på ett minsta-gemensamma-nämnare-tänkande vad beträffar objektbeskrivningarna.<sup>152</sup>

Vår definition av ett digitalt bibliotek kommer ganska nära Traugott Kochs beskrivning av en *Quality-controlled subject gateway*, även om han i sitt begrepp inte tar hänsyn till integration av olika typer av samlingar. Koch definierar *quality-controlled subject gateways* som

*... Internet services which apply a rich set of quality measures to support systematic resource discovery. Considerable manual effort is used to secure a selection of resources which meet quality criteria and to display a rich description of these resources with standards-based metadata. Regular checking and updating ensure good collection management. A main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing.*<sup>153</sup>

Ett digitalt bibliotek tillhandahåller elektroniska resurser på biblioteksmanér; det vill säga med avseende på aspekter som kvalitetskontroll, samlingsunderhåll och dito utveckling, samt elaborerade och standardiserade beskrivningsformat (metadata). Det digitala biblioteket skall använda sig av kontrollerade vokabulärer (exempelvis klassifikationssystem), vilka även används till att ordna samlingen hierarkiskt. Det digitala biblioteket förhåller sig till aktuella standarder (för exempelvis objektbeskrivningen) för att främja interoperabilitet med andra samlingar. Vi anser vidare att det digitala biblioteket skall sträva efter integrerad tillgång till såväl elektroniskt som tryckt, eller mer "traditionellt", material i de fall det redan finns en

---

<sup>149</sup> Library of Congress, *Challenges to building an effective digital library*, 1998

<sup>150</sup> Ibidem

<sup>151</sup> Arms, William Y., "Key concepts in the architecture of the digital library", *D-Lib Magazine*, 1995

<sup>152</sup> Library of Congress, *Challenges to building an effective digital library*, 1998

<sup>153</sup> Koch, 2000, s 26

samling sådant. Att integration inte ingår i Kochs ovan citerade definition beror på att denna utgår från en samling av uteslutande elektroniskt material.

Robin Murray vid Fretwell-Downing Informatics kallar denna integration mellan nytt och mer traditionellt material ”the key challenge for the digital library”. I samma artikel finns en fin översikt över de olika standarder som är aktuella i samband med det digitala biblioteket och des breda spektra av uppgifter. Bland dessa finner vi såväl MARC och Dublin Core (för objektbeskrivning), som Z39.50 (för sökning).<sup>154</sup>

Begreppsförvirringen till trots hoppas vi nu ha rätt ut begreppen virtuella samlingar och digitala bibliotek något. Vi övergår nu till det enskilda biblioteket och de olika strategier som står det till buds.

### 4.3 Digitala strategier

Biblioteket kan som vi nämnt välja att knyta sig till någon eller några av de virtuella samlingar som finns. Här återkommer vi till våra tidigare exempel NetSök och SAFARI, vilka vi tidigare diskuterat främst med formaten i fokus. Vi studerar nu de praktiska konsekvenser det medför för biblioteken att knyta sig till samlingar av detta slag.

Att använda sig av en extern samling är ett bekvämt alternativ, sedan man väl valt vilken/vilka samlingar man skall knyta sig till. I praktiken behöver man sedan inte själv tänka på eller ta ställning till frågor som urval, underhåll, sökmöjligheter et cetera. Den lokala bibliotekskatalogen kommer i detta fall också att vara helt avskild från den i vilken de elektroniska resurserna finns. Medaljens baksida är att man har mycket litet eller inget inflytande över samlingen, vad den skall innehålla och hur noggrant resurserna skall beskrivas. Det enskilda biblioteket har då det knyter sig till en utomstående virtuell samling fjärrmat sig från möjligheterna att samordna de elektroniska resurserna med de traditionella. Möjligheterna att göra de elektroniska resurserna samsökbara med de som förvaras i den ordinarie katalogen är att betrakta som obefintliga.

I de exempel på virtuella samlingar vi diskuterat finns inget självklart band till dokumentproducenterna (eller dokumentens ansvariga utgivare). Detta gör samlingarnas uppdatering både mer essentiell och svårare att genomföra. Varken SAFARI eller NetSök har något bevarandeansvar, utan man länkar endast till de i basen beskrivna resurserna. Detta medför att samlingarna slits hårt över tid i den dynamiska www-miljön. Risken att länkar brister eller dör är något som ständigt måste pareras. I NetSöks fall sker detta med hjälp av felsökningsprogram (ambitionen är att köra dessa en gång i månaden) och dels genom manuell kontroll av hela basens innehåll, vilket genomförs någon gång om året. SAFARI:s sökrobot besöker alla de indexerade sidorna en gång i veckan, men om innehållet radikalt förändrats utan att metadatamärkningen gjort det kan sökroboten givetvis inte avgöra detta. För de samlingar som endast länkar till resurserna är en elaborerad kundfeedback ett sätt att komplettera felsökningsprogram och dylikt.

---

<sup>154</sup> Murray, 1999, s 175. Genomgången av digitala biblioteksstandarder finns på sidorna 178-179.

Det metadataformat databasproducenten valt hamnar något i bakgrunden för biblioteket. Det framsynta biblioteket kanske tar hänsyn till vilket metadataformat den virtuella samlingen bygger på, med framtida samordning i tankarna.

Men generellt kan vi nog säga att sådana bekymmer är just det man vill undvika genom att knyta sig till någon annans samling. I den folkbiblioteksriktade samlingen NetSök framgår det inte ens av BTJ:s beskrivning vilket metadataformat som ligger bakom de tjugusiga fullposterna.

Teoretiskt skulle man kunna tänka sig att det enskilda biblioteket skapar sig en egen virtuell samling. Detta hör troligtvis till undantagen. Våra folkbibliotek är ofta ekonomiskt hårt pressade, och få har kompetens och resurser att själv katalogisera ens det "traditionella" materialet. En del bibliotek bygger upp sina egna länksamlingar, men detta genomförs utan att länksamlingen görs sökbar. Känd och mycket använd är exempelvis den ambitiösa länksamling Mölndals bibliotek byggt upp.<sup>155</sup>

I de fall då biblioteken tycker sig ha tid och råd att göra detta borde frågan om metadata vara extra aktuell, då man i realiteten är i färd med att bygga upp en separat bibliotekskatalog.

Ett problem med de elektroniska resurserna som vi hittills endast flyktigt berört är bevarandefrågan. Det kan synas smidigt att bevara elektroniska resurser då de, till skillnad från exempelvis pappersburna resurser, varken samlar damm eller blir uppätta av mal. Den elektroniska resursen ställer dock även arkivarien inför nya frågeställningar. De flesta filformat elektroniska dokument sparas i kräver datorer och programvara för att kunna förstås, och vad händer när det sista exemplaret av Microsofts *Word* eller Adobes *Acrobat Reader* har raderats? Det går inte att utfärda några garantier för ett elektroniskt dokumentets läsbarhet om inte bevarandeuppdraget även innefattar att bevara en dator med fungerande programvara. Inom arbetet med arkivering av elektroniska resurser finns en viss tilltro till tekniken bakom strukturerade markeringsspråk som SGML och XML, som visserligen är beroende av en dator som tolk, men som har den stora fördelen att de är plattform- och programvaruoberoende.

Margaret Phillips beskriver i en artikel om det australienska PANDORA-projektet nationalbibliotekets arbete med att bevara de elektroniska resurserna för eftervärlden. PANDORA står för *Preserving and Accessing Networked Documentary Resources of Australia*, och här samlar man in och kopierar de resurser man bedömt som lämpliga för samlingen. Nationalbibliotekets brist på resurser medför att man måste vara än strängare i urvalet.<sup>156</sup> Eftersom resursbristen är ett predikament man delar med i stort sett varenda annat bibliotek i hela världen är PANDORA:s lösning något att begrunda.

Fördelarna med att på detta sätt bevara en egen upplaga av de resurser som ingår i samlingen är att man vet att den alltid finns där (om kanske inte alltid i sin senaste version). Även en sådan samling kräver dock mycket underhåll. Phillips återger en kalkyl från British Library i vilken kostnaden för att bevara en elektronisk resurs beräknats till mer än tjugo gånger högre än motsvarande siffra för mer traditionellt material (beräknat på en 25-årsperiod). Kostnaden sjunker drastiskt då det rör sig om

---

<sup>155</sup> Mölndals stadsbiblioteks länkkatalog, <http://www.molndal.se/bibl/>

<sup>156</sup> Phillips, Margaret E., "The National Library of Australia: Ensuring long-term access to online publications", *the Journal of Electronic Publishing*, 1999

större mängder dokument (mer än 10 000 insamlade resurser om året), men är fortfarande fem gånger högre än priset för att bevara tryckt material.

#### 4.4 MARC eller Dublin Core?

Ett tankefel som ofta möter oss i litteraturen är att MARC och Dublin Core på något sätt skulle stå i motsättning till varandra, som ett slags metadataformatens David och Goliat. Med Alan Hopkinson menar vi snarare att ”it is probably better to say that the protagonists, the traditional and the more recent formats, are complementary not adversary.”<sup>157</sup>

MARC kan inte anses vara ett mer komplext format än Dublin Core för beskrivning av elektroniska resurser. Den outtalade försäkran om att beskrivningen gjorts av en specialutbildad bibliotekarie MARC-posten innebär, samt att formatet är knutet till ett vedertaget och beprövat regelverk för katalogisering, kan vara bedräglig. Även Dublin Core kan beskriva ett dokument enligt AACR2. I figuren nedan ser vi dock formaten jämförda i sin ”traditionella” tillämpning.

	<b>DUBLIN CORE</b>	<b>MARC</b>
<b>Objektbeskrivning</b>		Fullständig
	Strukturerad	
<b>Formatattribut</b>		Många
	Utvalda	
<b>Metadataproduktion</b>		Manuell, kräver specialistkompetens
	Manuell	
<b>Katalogiseringskostnad</b>		Hög
	Låg	

**Figur 16 En jämförelse mellan de två formaten i deras vanligaste tillämpning**

Mot bakgrund av de interoperabilitetsmöjligheter bägge formaten erbjuder kan det tyckas likgiltigt vilket format man väljer. MARC erbjuder större möjligheter än Dublin Core att beskriva exemplar (*items*) i enlighet med *Functional Requirements for Bibliographic Records*. I det digitala biblioteket är det emellertid mer intressant att kunna ge tillfredsställande beskrivningar av uttryck (*expressions*) och manifestationer (*manifestations*) av ett verk (*work*), bland annat eftersom det i Internetmiljön är blott alltför enkelt att skapa olika elektroniska manifestationer av ett verk. Att Dublin Cores beskrivningar är inriktade mot det innehållsmässiga verket snarare än det enskilda exemplaret kan i belysning av detta ses som en styrka hos formatet.

I Task Force on Metadata's ”Summary Report” från 1999 vill man dessutom lägga ett femte användarbehov, *manage/administer*, till de fyra behov IFLA koncentrerat sig

<sup>157</sup> Hopkinson, 1999, s 17

på i *Functional requirements* (sökning, identifikation, urval och åkomst).<sup>158</sup> Man medger visserligen att detta femte kriterium främst fyller administrativa behov, och måste anses vara underordnat de fyra andra, vilka riktar sig direkt mot användaren och hans/hennes krav på katalogen. Som framgått i detta avsnitt kan dock aspekter som hurpass enkelt, flexibelt och billigt formatet är att använda och administrera knappast överskattas för den som bygger upp och skall underhålla katalogen.

Vi har tidigare talat om MARC som specialisternas format. Dublin Core är å sin sida så enkelt att dokumentproducenten själv kan beskriva sin resurs med formatet. För biblioteken har just detta alternativ dock aldrig varit intressant. Den data som krävs i dessa sammanhang måste vara konsekvent och följa standarden för att kunna fungera i katalogen. Dublin Cores ambition att vara lättanvänt har emellertid gjort att man försökt hålla formatet så flexibelt som möjligt och redan från början satsat på att utveckla bekväma inmatningsformulär, *templates*. Dessa verktyg måste anpassas för att passa även de professionella katalogisatörernas behov och arbetssätt.

I det australiensiska PANDORA-projektet planerar man att använda sig av fullständiga MARC-poster för objektbeskrivning i samlingen. För tillfället använder man sig av ett enklare interrimformat man själv utformat, men Margaret Phillips skriver

We also strongly believe that because electronic publications form part of our national imprint, they should be included in our national bibliography, described using full level MARC (Machine Readable Cataloguing) records, as we do for traditional library materials. We are encouraging other Australian deposit libraries to take the same approach as we work toward a distributed national collection of Australian electronic publications.<sup>159</sup>

Den uppenbara fördelen med att använda MARC-formatet för att katalogisera elektroniska resurser är att MARC redan nu är det gängse biblioteksformatet. De ”elektroniska” posterna skulle lätt kunna integreras i de befintliga bibliotekskatalogerna.<sup>160</sup> Att fysiskt inlemma de poster som beskriver elektroniska resurser i den ordinarie katalogen är dock inte helt oproblemiskt. Sten Hedberg betonar att det ”i dagsläget” måste

betraktas som synnerligen tveksamt om det finns någon egentlig anledning att skapa MARC-poster ur DC-poster och lagra dessa i samma databas som egentliga MARC-poster, vilket skulle innebära att elektroniska resurser skulle föras in i en katalogtjänst som ursprungligen lagts upp för konventionellt biblioteksmaterial.<sup>161</sup>

Det olämpliga med integrerade kataloger är ju inte endast att olika metadataformat passar olika bra för elektroniska resurser och mer ”traditionellt” biblioteksmaterial. På grund av de elektroniska dokumentens specifika natur kräver deras katalogposter ojämförligt mycket mer underhåll.

---

<sup>158</sup> Task force on metadata: Committee on cataloging: Description and access, ”Summary report June 1999”, 1999

<sup>159</sup> Phillips, 1999

<sup>160</sup> Dempsey och Heery, 1997, s 56

<sup>161</sup> Hedberg, *Bruket av metadata enligt Dublin Core : Principer, teknik och tillämpningar utänför Sverige*

Fördelen MARC har i och med biblioteken och katalogisatörernas vana vid formatet och de goda möjligheter till interoperabilitet formatet erbjuder uppvägs dock av en lika uppenbar nackdel: MARC-katalogisering är tidskrävande och har sedan dag ett utformats för att passa statiska dokument.<sup>162</sup> Trots att de dynamiska elektroniska resurserna varit en utmaning i flera år nu, har man inte heller lyckats tillfredsställande lösa hur de skall MARC-katalogiseras. Detta borde egentligen heller inte vara nödvändigt. Som svampar ur jorden har under senare år nya metadataformat dykt upp, format som är särskilt utformade för att beskriva elektroniska resurser. Dempsey och Heery skriver att det är föga troligt att något metadataformat kommer att få monopol på objektbeskrivningen över huvud taget, ”for a number of more or less well known reasons, not least the investment represented by legacy systems in terms of technology and human effort.”<sup>163</sup> Frågan om MARC eller Dublin Core är det ”bästa” formatet tjänar kanske att dölja den verkliga frågan för biblioteken – ska vi förmedla även elektroniskt material till våra användare, med samma garantier som vi erbjuder för det tryckta?

## 4.5 Objektbeskrivning i heterogena samlingar

Som vi visat i avsnittet ”Samordningsverktyg” kan man uppnå samsökbarhet utan att behöva integrera olika kataloger. Posterna förvaras då i separata databaser, eventuellt med olika metadataformat som grund. Sökningar utförs dock utifrån ett gemensamt gränssnitt, med exempelvis Z39.50 som länk mellan databaserna. För användaren syns bara en katalog. Z39.50-samhället har accepterat Dublin Cores attributuppsättning, och detta gör att ”DC-elementen ur söksynpunkt blir likvärdiga med de element som definieras med MARC-formatens fältkoder”, en utveckling som i sin tur kommer att underlätta bibliotekens sökmöjligheter i Dublin Core-baserade kataloger.<sup>164</sup>

I vår jakt på ett svenskt bibliotek som arbetar med utveckling av ett digitalt bibliotek fann vi att det inom Lunds Universitetsbibliotek (LUB) just nu utvecklas en modell för beskrivning av LUB:s elektroniska material. Detta är i sin tur en del av arbetet med uppbyggnaden av ett digitalt bibliotek. Arbetet med att ta fram en policy för beskrivning av elektroniska resurser görs av katalogisatörer och IT-bibliotekarier från olika delar av verksamheten. I förslagsutkastet skriver Jörgen Eriksson att

modellen möjliggör ett distribuerat arbetssätt, där biblioteken vid Lunds universitet kan samarbeta om att samla in och beskriva webbresurser. Vi tror att det finns både kvalitets- och rationaliseringsvinster att göra om man i stället för att bygga upp lokala länksamlingar kan enas om att arbeta mot en gemensam databas.<sup>165</sup>

---

<sup>162</sup> ”Most of the library automation systems which are operational today rely heavily on the standards that were used for the production of the card catalogue. It may be that there was an interim stage when this was justified, but this has long past. However, there is such an enormous investment in databases which use some form of the MARC format for their organisation that there is no easy or economic way at the moment in which they can be superseded or revised.” Ayres, 1996, s 3

<sup>163</sup> Dempsey och Heery, 1997, s 5

<sup>164</sup> Hedberg, *Bruket av metadata enligt Dublin Core : Principer, teknik och tillämpningar utanför Sverige*

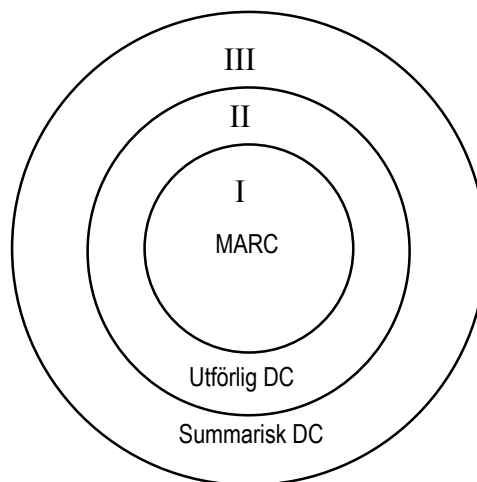
<sup>165</sup> Eriksson, Jörgen, Gustafsson, Jan och Zethson, Britt-Sofi, *Metadataanvändning vid Lunds Universitets bibliotek: ett förslag*, Lunds universitetsbibliotek (internt), 1999



Det material som skall ingå i tjänsten är tänkt att hålla samma vetenskapliga karaktär och kvalitet som LUB:s övriga resurser. De resurser LUB vill förmedla är främst avhandlingar, elektroniska tidskrifter, forskningsrapporter, artiklar och examensarbeten. Detta är resurser som biblioteket hade förvärvat om de publicerats i tryckt form, men biblioteksanvändaren riskerar att gå miste om om biblioteket inte katalogiserar och bevarar det. För material som publiceras i både tryckt och elektronisk form kan tillgång till det elektroniska exemplaret underlätta användarens åtkomstmöjligheter.

I modellen ser man till olika typer av dokument och delar in dem i olika ambitions- och katalogiseringsnivåer. Vilken nivå ett dokument sorteras in i påverkar också hur lång tid LUB kommer att lagra resursen.

Modellen är uppbyggd med utgångspunkt från Roger Brissons *The world discovers cataloging*, vilken han skrev för att försöka förankra uttrycket ”digitala bibliotek”.<sup>166</sup> Han diskuterar utförligt metadata, vilket han menar är en hörnsten i uppbyggnaden av digitala bibliotek, och rekommenderar olika beskrivningsnivåer för olika typer av material. Materialet kan även indelas i nivåer efter sin grad av relevans för det aktuella biblioteket. LUB:s modell för katalogisering av det elektroniska materialet utnyttjar bägge dessa urvalskriterier och innehåller följande tre nivåer:



**Figur 17 LUB-modellen**<sup>167</sup>

I nivå I ingår elektroniskt publicerade avhandlingar med full MARC-beskrivning. För elektroniska avhandlingar gäller att LUB är ansvariga för att tillgängliggöra dem i minst tio år. Nivå I kostar man på fullständig MARC-katalogisering, utförd av katalogisatörer inom LUB. Avhandlingarna skall bevaras i en databas och posten införlivas med den lokala bibliotekskatalogen.

I nivå II hamnar exempelvis forskningsrapporter. En utförlig Dublin Core-post upprättas för dessa resurser. Resurserna skall förses med inbäddad metadata. Dokumenten inom nivå II hämtas också in för fulltextbevarande på LUB, men

<sup>166</sup> Brisson, Roger, ”The World Discovers Cataloging - A Conceptual Introduction to Digital Libraries, Metadata and the Implications for Library Administrations”, *Journal of Internet Cataloging*, 1999

<sup>167</sup> Eriksson, Gustafsson och Zethson, 1999

bevaras kortare tid än de resurser som ingår i nivå I. Även på denna nivå är det en katalogisatör som skapar dokumentbeskrivningen.

Inom den tredje nivån tänker man använda en mer summarisk Dublin Core-beskrivning där ett urval av de fält som används för material i cirkel II används. De fält som skall användas för denna beskrivningen borde vara de obligatoriska fälten från nivå II. Posterna skapas som fristående metadata med pekare ut mot internetdokumentet. Då katalogisatören lagt in beskrivningen i dokumentet samlas metadataposten in av en robot och läggs i en databas tillsammans med posterna från nivå II.

Styrkan i denna modell är att man ser till vilka dokument som skall beskrivas och därefter placerar dem i en struktur där de olika nivåerna kräver olika stora arbetsinsatser. På alla nivåer görs objektbeskrivningarna av utbildade katalogisatörer. Materialet tillåts alltså i viss mening styra katalogstrukturen. Skiktningen i olika katalogiseringsnivåer skapar utrymme för de många olika resurstyper som skall rymmas i en och samma modell. Samsökbarheten mellan de olika databaserna är viktig, och kopplingarna dem emellan ser ut att kunna bli ganska avancerade, då de olika resurserna dessutom skall samsas med det traditionella tryckta materialet och fristående kommersiella tidskrifter och databaser.

Trots att denna modell i en genomgång förefaller ganska invecklad, förenar den flexibilitet (olika katalogiseringsformat, olika katalogiseringsnivåer) med interoperabilitet (samsökbarhet i de olika baserna) och utgör en bra grund för ett digitalt bibliotek.

## SAMMANFATTNING OCH SLUTSATSER

---

Som ett led i vårt uppsatsarbete har vi allt som oftast gått på fest och plågat våra bordsgrannar och andra med utförliga svar på frågan ”Men vad är det egentligen ni skriver om?”. Vi står i stor skuld till de oskyldiga förbipasserande som tåligt avhört våra redogörelser för detta. Vår ambition, ytterligare stärkt av dessa erfarenheter, har varit att klart och redigt besvara frågorna Vad är metadata? Varför är metadata och arbetet för att effektivisera metadataformaten viktigt? Vilka problem ställs metadata kreatörerna inför idag? och Hur förhåller sig biblioteken till det elektroniska materialet?

Vidare har vi velat åskådliggöra några olika modeller för bibliotekens förhållningssätt visavi det elektroniska materialet och rekommendera strategier inför uppbyggnaden av digitala bibliotek. För att besvara frågeställningarna är en del utvecklingar oundvikliga. Vår förhoppning är att dispositionen ändå har gjort vår tankegång lätt att följa även om, eller kanske i synnerhet om läsaren inte redan är väl förtrogen med alla begrepp och termer.

För bibliotekarieskrået är vikten av fungerande metadataformat inget nytt. Om vi någon gång slagit in till synes öppna dörrar har vi gjort det med tanke på de våra medgäster som skulle ha somnat över bordet om de inte fått höra den ”långa versionen”, och sålunda kanske för första gången fått upp ögonen för begreppen ”metadata” och ”objektbeskrivning av elektroniska resurser”. Priscilla Caplan har döpt sin rapport från DC-1 till ”You call it corn, we call it syntax-independent metadata for document-like objects”.<sup>168</sup> Vi har många gånger erfarit vikten av att tala om majs istället för om syntaktiskt oberoende metadataformat för beskrivning av DLO. Även utanför bibliotekariedisciplinen finns nämligen drömmen om en ”biblioteksordning” på Internet, vilken skulle reducera bruset och göra Internet riktigt användbart.

För de traditionella resurserna kan vi nog räkna med att MARC rullar på som den gigantiska maskin det är. Att informationsöverflödet i allmänhet och de elektroniska resurserna i synnerhet tvingar fram alternativa lösningar och kompletterande format är heller inte överraskande. Dublin Core är idag ohotat som standard, men formatet måste marknadsföras mer.<sup>169</sup> Först då det indexerade materialet når en kritisk massa kommer man att kunna sprida en medvetenhet om hur viktigt det är, och vad det kan göra för informationssökaren. Man brukar säga att ”ett jobb som ingen gör, blir allas angelägenhet” och sällan har det varit mer tillämpligt än då vi talar om objektbeskrivning av elektroniska resurser.

### 5.1 Formatets betydelse

Det torde stå klart att metadata kommit att spela en allt större roll, och att valet av katalogiseringsformat kan få mycket stor betydelse för samlingarnas framtida åtkomlighet. De olika samordningsverktyg som utvecklats marginaliserar inte formatet, utan gör det istället än mer viktigt. Det format biblioteket väljer måste vara flexibelt och tillåta interoperabilitet. Då katalogisering över huvud taget är en

---

<sup>168</sup> Caplan, 1995, s 19-23

<sup>169</sup> Hakala et al., 1998, s 7

avsevärd investering måste formatet också väljas med hänsyn till framtida användning och behov. Biblioteken måste redan på det här stadiet ställa sig frågan huruvida man kommer att vilja integrera de ”elektroniska” posterna med bibliotekets reguljära katalog. Det är klokt och långsiktigt att välja lösningar som stöds av redskap som Z39.50 och RDF. I bibliotekskatalogen som ingen annan stans är konformism viktig och kan få en avgörande betydelse för katalogens användbarhet. Vi påminner oss Alan Hopkinsons ord då han inskräper att ”conformity and uniformity are required rather than agreement.”<sup>170</sup> Sten Hedberg varnar på BIBLIST för implementationen av ”lokala fiffigheter” i katalogiseringsarbetet.<sup>171</sup> I samma andetag förordar dock Hedberg underligt nog en egen lokal fiffighet: en nationell, svensk Dublin Core-variant.<sup>172</sup> Överlag vill vi rekommendera biblioteken att tänka efter både en och två gånger innan de förändrar och anpassar de standardiserade formaten. Variationer i praxis skall vara övertänkta, befogade och konsekvent genomförda för att inte leda till isolering av den aktuella katalogen.

Vi talade inledningsvis om mer avancerade och intelligenta sökverktyg som ett alternativt/kompletterande medel att bringa det elektroniska materialet i ordning. Även här är de två vägarna mer av en figur än ett realistiskt alternativ. Riktigt ”intelligenta” sökstrategier kräver noggranna beskrivningar. Den avancerade sökmotorns kalkyler är dock ofta ett sätt att i viss mån kompensera bristfällig objektbeskrivning. Man tvingas då fråga sig om det är ekonomiskt försvarbart att satsa på mer avancerade sökprocesser som ersättning för noggrann katalogisering? Vi har redan citerat Dempsey/Heery då de konstaterar att

Web indexes based on robot extraction of (currently unstructured) metadata are cheap to create, are automatic. Documentation of a particular collection by specialists are expensive. ”Information gateway” services add value through intellectual effort, and are correspondingly expensive. These factors will drive the creation of author-produced metadata and more sophisticated automatic extraction techniques. However, the creation of full, structured metadata will remain expensive, wherever along the use line that cost falls.<sup>173</sup>

Vi vet inte heller vilka resurser framtiden bär i sitt sköte. Med utgångspunkt från de elektroniska versionerna av vetenskapliga artiklar förutspår Mats Lindquist att de nya möjligheter hypertexten öppnar för författarna kommer att förändra själva dokumentens utformning. Han talar i detta sammanhang dels om de möjligheter mediet erbjuder författaren att direktlänka till referenser och på detta sätt låta hypertexten göra artikeln till något som mer liknar en fysisk del av den större helheten, och dels om hur en artikels olika delar kan existera på vitt skilda ställen och i lika vitt skilda former vilka endast tillfälligt länkas samman vid presentation,

Multimedia documents that are ”born digital” naturally appear as parts that are linked together, since the parts are usually produced by different tools, or are collected from different sources or data-capture devices.<sup>174</sup>

---

<sup>170</sup> Hopkinson, 1999, s 19

<sup>171</sup> Hedberg Sten, ”Vad gör vi egentligen? (Ex: Outsourcing)”, inlägg i BIBLIST 000407

<sup>172</sup> Hedberg, Sten, ”Utkast till katalogiseringsregler och formatinstruktion för en svensk Dublin Core-tillämpning”, *Nya vägar för Boken*, 1999

<sup>173</sup> Dempsey och Heery, 1997, s 7

<sup>174</sup> Lindquist, 1999

Den enskilda artikeln ges alltså möjlighet att å ena sidan bli än mer specialiserad, å andra sidan splittrad på olika adresser och i olika former. Hur framtidens elektroniska resurser kommer att se ut kan vi endast sia om. Kanske är den utveckling Lindquist talar om något som kommer att accelerera, kanske dyker helt nya sätt att presentera vetenskapliga landvinningar upp.

Man frågar sig om det inte skulle vara möjligt att utveckla ett standardformat som passar alla och sedan hålla sig till det, istället för att som vi sett hålla på att trassla med olika format, Z39.50 och översättning av färdiga metadataposter? Det torde dock vara uppenbart att ett och samma format inte kan vara tillräckligt flexibelt samtidigt som det äger den höga grad av specialisering den bibliografiska kontrollen kräver. Eller som Robin Murray uttrycker det "Given this blend of the specific and the generic, it is clear that emerging digital and hybrid library environments require solutions developed from the integration of a large number of open system components."<sup>175</sup> Av samma skäl är det, som vi redan noterat "unlikely that some monolithic metadata format will be universally used."<sup>176</sup>

## 5.2 Det digitala biblioteket: Babels torn *revisited*

Visst skulle man kunna gå på i ullstrumporna och helt enkelt nöja sig med att katalogisera så mycket av det elektroniska materialet man bara hann och stoppa in det i den befintliga bibliotekskatalogen. Mycket snart skulle bibliotekets situation dock bli ohållbar och deras katalog värdelös. Det är angeläget att biblioteksvärlden, även det enskilda biblioteket, medvetet väljer vilken hållning man skall inta i sin hantering av de elektroniska resurserna, för inte hamna i detta eller något ännu värre dilemma.

De senaste åren har informationsrevolutionen tagit fart. Det känns allt svårare att uttala sig om bibliotekens framtida roll, användarnas behov, nya dokumenttyper samt vilka beskrivningsformat som kan passa dessa. Att biblioteket kan, och vill, göra mycket för att tillgängliggöra det elektroniska materialet är emellertid ställt utom allt tvivel. Dessa faktorer sammantagna, parade med den hårda verkligheten, där biblioteken måste hushålla med sina resurser, talar för övertänkta strategier och samarbete i uppbyggnaden av digitala bibliotek. Det är viktigt att vid valet av vilket format en samling skall använda sig av lägga sig vinn om flexibla lösningar och öppna standarder. Dessa måste ge goda möjligheter till skalbarhet och framtida interoperabilitet. Vid uppbyggnaden av ett digitalt bibliotek måste man överväga både vad som passar i dagsläget och vad som kan tänkas bli aktuellt i framtiden. Vad vill man att samlingen skall kunna användas till? Med vilka andra samlingar är det intressant att kunna samarbeta? Vilken grad av bibliografisk kontroll är önskvärd och möjlig? De digitala bibliotek vi nu lägger grunden till är företeelser vi kommer att få leva länge med, och alltför nischade och specialiserade lösningar är något som kan visa sig dyrt i längden för det enskilda biblioteket.

---

<sup>175</sup> Murray, 1999, s 175

<sup>176</sup> Dempsey och Heery, 1997, s 5

<b>MODELL FÖR HANTERING AV DET ELEKTRONISKA MATERIALET</b>	<b>KONSEKVENSER FÖR SAMLINGEN / FORMATETS BETYDELSE</b>
Biblioteket erbjuder <b>åtkomst</b> till elektroniska resurser	<b>Konsekvenser:</b> Biblioteket har ingen kontroll över urvalskriterier, samlingsunderhåll, uppdatering eller bevarande. De elektroniska resurserna är avskilda från den lokala bibliotekskatalogen. Samsökbarhet är inte möjlig.
	<b>Formatets betydelse:</b> Biblioteket får hålla till godo med det beskrivningsformat och därmed de sökmöjligheter den valda samlingen erbjuder. Import av poster är inte aktuell.
Biblioteket <b>förvärvar aktivt</b> elektroniska resurser och gör dem tillgängliga för användarna	<b>Konsekvenser:</b> Biblioteket har full kontroll över sin samling elektroniska resurser. Man kan välja huruvida man vill fysiskt integrera katalogposterna i den lokala katalogen. Det är möjligt att gå vidare och göra samlingen samsökbar med den lokala katalogen.
	<b>Formatets betydelse:</b> Valet av beskrivningsformat görs av biblioteket själv. Formatet skall tillfredställa användarbehoven i enlighet med <i>Functional Requirements</i> , samt tillåta enkel uppdatering och underhåll av posterna för det elektroniska materialet. Att ha framtida samordning med andra samlingar och/eller den egna lokala katalogen i tankarna är viktigt.
Biblioteket försöker bygga upp ett <b>digitalt bibliotek med integrerad tillgång</b> till såväl elektroniskt som mer traditionellt material	<b>Konsekvenser:</b> Biblioteket har full kontroll över sin samling elektroniska resurser. Man kan välja huruvida man vill fysiskt integrera katalogposterna i den lokala katalogen, eller om man vill förvara dem i en egen katalog. Samsökbarheten utesluter inte möjligheten att använda sig av flera olika kataloger, olika metadataformat eller olika beskrivningsnivåer. För användaren syns bara en katalog.
	<b>Formatets betydelse:</b> Valet av format är avgörande. Detta görs av biblioteket själv. Formatet/formaten skall tillfredställa användarens behov enligt <i>Functional Requirements</i> , samt tillåta enkel uppdatering och underhåll av posterna för det elektroniska materialet. Formatet måste stödjas av standarder som Z39.50.

**Figur 18 Bibliotekets olika förhållningssätt visavi det elektroniska materialet**

Vi har diskuterat möjligheten att knyta sig till en extern samling, och i sammanhanget diskuterat det folkbiblioteksriktade NetSök. Denna samling är egentligen försvinnande liten jämfört med de många gratisjänster som ligger fritt åtkomliga på Internet. NetSöks styrka jämfört med flertalet av dessa är de strikta kvalitetskriterierna, samt att man producerar (fristående) metadata, i vilka gör att kunderna (biblioteken) lätt kan göra mer meningsfulla fritextsökningar, söka i de olikafälten, eller till och med söka på SAB-klassning. NetSök är skapat särskilt för de svenska folkbiblioteken. Trots detta är kanske det som är avsett att passa alla inte alltid varken den enda eller den bästa lösningen för det enskilda biblioteket.

För att ha full kontroll över urvalskriterier, objektbeskrivning och sökmöjligheter kan biblioteket själv börja katalogisera de elektroniska resurserna. Vi har diskuterat det olämpliga att placera dessa poster i den reguljära katalogen, på grund av de speciella krav på underhåll och funktionalitet dessa poster har. Sökprotokoll som Z39.50 gör det dock möjligt att uppnå samsökbarhet utan att posterna är fysiskt samlade i en katalog. Tvärtom öppnas möjligheten att erbjuda användaren sökmöjlighet inte endast i den egna/de egna katalogerna, utan också i andra kataloger som stöder protokollet.

Vi har även diskuterat möjligheten att skikta objektbeskrivningen, så att olika typer av material, eller material med olika relevans för det aktuella biblioteket, ”kostas på” olika beskrivningsnivåer. LUB-modellen som tar utgångspunkt från vilka resurser

som skall inlemmas i katalogen och låter deras beskrivningskrav påverka hur samlingen skall byggas upp, ser vi som ett bra förhållningssätt för biblioteket. Samsökbarheten finns kvar utan att man för den skull behöver blanda resurser med olika bevarandetid, beskrivningsnivå et cetera i samma katalog. Att LUB är noga med att låta objektbeskrivningen utföras av vana katalogisatörer är något vi tror kommer att bestå. Biblioteksvärlden kommer aldrig att intressera sig för glad, men amatörmässig metadata. Som Milstead/Feldman uttrycker det

unless the author of a book is a trained indexer, she is not likely to produce a high-quality index. True, the author/creator knows her creation better than anyone else, but this does not mean that she is able to step back to see how it fits into the universe of knowledge or objects, so as to best apply useful tags.<sup>177</sup>

Det finns som vi sett många olika ansträngningar som vill ge biblioteksvärlden bättre och mer kontrollerad tillgång till webbresurserna, exempelvis CORC och samordningsverktyg som RDF och Z39.50. NetSök och SAFARI är fristående kataloger, men med de möjligheter till interoperabilitet som MARC och Dublin Core erbjuder kan även dessa samlingar komma att bli fungerande delar av de kataloger biblioteken redan använder sig av. De format vi undersökt och de samordningsverktyg vi diskuterat strävar alla efter interoperabilitet med varandra. Eftersom sådan är möjlig borde detta vara i fokus för biblioteken.

Att biblioteken samarbetar kring utvecklingen av det digitala biblioteket är kanske det bästa sättet att effektivisera användningen av de elektroniska resurserna. I CORC samlas många bibliotek kring katalogiseringen av de elektroniska resurserna. Men CORC är ett gigantiskt projekt med tämligen centraliserad beslutsfattning. Det lilla biblioteket, eller specialbiblioteket med sina särskilda behov nog kan känna att de kommer bort och nivelleras i den globala samlingens allmänna bästa. Olika specialbibliotek kunde dock kunna dela upp insamlandet och katalogiseringen av de elektroniska resurserna mellan sig och sedan sammanföra dem till ämnesspecifika samlingar. Exempel på detta är Renardus akademiska resurssamling och *European Treasury Browser* (ETB), som skall skapa samsökbarhet och utbyte av metadataposter mellan olika europeiska kvalitetssamlingar för skolelever och lärare. På ett liknande sätt skulle andra specialsamlingar ute i världen kunna dra nytta av varandras urvals- och katalogiseringsarbete.

Samverkan mellan de olika formaten och de olika biblioteken erbjuder möjligheter att hålla kostnaderna nere och höja effektiviteten. Visst är det lovvärt att bygga upp stora länksamlingar som även andra bibliotek kan använda sig av, men interoperabilitetsmöjligheterna mellan två länksamlingar är att betrakta som obefintliga, och vi skulle hellre se att de svenska biblioteken samlades i gemensamma strategier kring metadatahantering och förhållningssätt till de elektroniska resurserna över huvud, istället för att skapa egna "mölndalare".

I uppbyggnaden av ett digitalt bibliotek är det många byggstenar som ska fogas samman, inte bara hantering av metadata. I kommande studier skulle det vara intressant att se undersökningar av det digitala biblioteket i sin helhet ur ett kostnadsperspektiv, eller inom andra delområden som interoperabilitet mellan olika

---

<sup>177</sup> Milstead och Feldman, 1999a

klassifikationssystem (*Knowledge Organisation Systems*) och upphovsrätt vid bevarande av fulltexter i det digitala bibliotekets databaser.

CORC startar upp först i juli 2000 och vidare studier i huruvida en sådan modell skulle passa för svenska förhållanden vore intressant. Flera av våra studieobjekt, RDF, XML, CORC, LUB-modellen, är just i färd med att testas och utvecklas - utvärderingar av dessa eller andra initiativ som inte rymts inom denna uppsats ramar - exempelvis EULER, Renardus eller ETB - vore också intressant att se. Och om framtida forskare inte fått nog av metadata efter detta, så skulle även en djupdykning i objektbeskrivning för arkiv och museer säkert vara spännande.



För folket i Babel var inget omöjligt, för de talade samma språk och kunde så enas om och genomföra de mest fantastiska projekt. Olyckligtvis satte Gud stopp för detta. Med de kunskaper vi har nu, och med ett vettigt användande av funktionella metadataformat, kan vi lära åtminstone våra bibliotekskataloger att förstå varandra. Och om vi bryter den långa språkföribistringen vore det inte omöjligt att vi skulle kunna återuppta vårt byggarbete.



## KÄLLFÖRTECKNING

---

*Datum inom parentes motsvarar vår senaste kontroll av den elektroniska resursens URL.*

Arms, William Y., "Key concepts in the architecture of the digital library", *D-Lib magazine*, July 1995

URL: <http://www.dlib.org/dlib/July95/07arms.html>  
(000105)

Ayres, Fred, "What is the future for catalogues and cataloguers?", *Catalogue & Index*, nr 122, 1996, s 1-5

Benito, Miguel, *Bibliografisk kontroll. En introduktion till katalogisering, klassifikation och indexering. Andra upplagan*, Tarancos bokförlag, Borås 1995

Biblio Tech Review, "Z39.50: Part 1 – an overview", uppdaterad senast 000129

URL: [http://www.biblio-tech.com/html/z39\\_50.html](http://www.biblio-tech.com/html/z39_50.html)  
(000207)

Bray, Tim, *RDF and Metadata*, 980609

URL: <http://www.xml.com/xml/pub/98/06/rdf.html>  
(990830)

Brickley, Dan och Guha, R. V., *Resource Description Framework (RDF) Schema Specification*, W3C Proposed Recommendation 03 March 1999

URL: <http://www.w3.org/TR/PR-rdf-schema-19990303/>  
(990929)

Brisson, Roger, "The World Discovers Cataloging - A Conceptual Introduction to Digital Libraries, Metadata and the Implications for Library Administrations", *Journal of Internet Cataloging*, [Vol. 1 (4). (1999), p. 3-31]

URL: <http://www.personal.psu.edu/faculty/r/o/rob1/publications/JICmetadata.pdf>  
(PDF-fil)  
(990929)

Bruun, Rasmus, *Er lykken gjort? – om metadata som indekseringsgrundlag ved bibliografisk kontrol af webdokumenter*, Köpenhamn 1997

URL: [http://www.db.dk/student/k96/4/rb/ogaver/2semopg\\_tekst.html](http://www.db.dk/student/k96/4/rb/ogaver/2semopg_tekst.html)  
(990906)

Byrne, Deborah J., *MARC manual. Understanding and using MARC records*, 2 ed., Libraries Unlimited Inc., Englewood, Colorado 1998 (Library of Congress Cataloging-in-Publication Data)

Caplan, Priscilla, "Cataloging Internet resources", *The public-access computer systems review* 4, 1993:2, s 61-66

URL: <http://info.lib.uh.edu/pr/v4/n2/caplan.4n2> (ASCII-fil)  
(991021)

Caplan, Priscilla, "You call it corn, we call it syntax-independent metadata for document-like objects", *The public-access computer systems review* 6, 1995:4, s 19-23

URL: <http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>  
(990929)

Chapman, Ann, Day, Michael och Hiom, Debra, "Metadata: Cataloguing practice and Internet subject-based information gateways", *Ariadne*, nr 18, 980504

URL: <http://www.ariadne.ac.uk/issue18/metadata/intro.html>  
(000105)

Day, Michael, "Extending metadata for digital preservation", *Ariadne*, nr 9, 1997

URL: <http://www.ariadne.ac.uk/issue9/metadata/intro.html>  
(000105)

DCMI, "Dublin Core Metadata Initiative"

URL: <http://purl.oclc.org/dc/>  
(000213)

Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set, Version 1.1: Reference Description*, OCLC, 990702

URL: <http://purl.oclc.org/docs/core/documents/rec-dces-19990702.htm>  
(991109)

Dempsey, Lorcan och Heery, Rachel, *Specification for resource description methods. Part 1. A review of metadata: a survey of current resource description formats*, version 1.0, 970319

URL: <http://ukoln.ac.uk/metadata/desire/overview/overview.pdf> (PDF-fil)  
(991021)

Eriksson, Jörgen, *Metadata – the description of electronic objects, documents and services*, föredrag Nordic Net Centre, senaste uppdatering 971027

URL: <http://www.lub.lu.se/~jorgen/foredrag/meta.html>  
(990929)

Eriksson, Jörgen, Gustafsson, Jan och Zethson, Britt-Sofi, *Metadataanvändning vid Lunds Universitets bibliotek: ett förslag*, Lunds universitetsbibliotek (internt), 990930

URL: <http://www.lub.lu.se/~jorgen/metadata/bakgrund.html>  
(991007)

Fagerlind, Marita och Gisselquist, Gunilla, *Metadata enligt Dublin Core*, magisteruppsats i Biblioteks- och informationsvetenskap, BIVILS skriftserie 1999:1, Lund, 1999

Furrie, Betty (in conjunction with the Data Base Development Department of the Follett Software Company), *Understanding MARC Bibliographic: Machine-Readable Cataloging*

URL: <http://lcweb.loc.gov/marc/umb/>  
(000119)

Gradmann, Stefan, *Cataloguing vs. metadata: Old wine in new bottles?*, 64<sup>th</sup> IFLA general conference, augusti 1998  
URL: <http://www.ifla.org/IV/ifla64/007-126e.htm>  
(990929)

Hakala, Juha, "Internet metadata and library cataloguing", *International cataloguing and bibliographic control*, 28 (1) 1999, s 21-25

Hakala, Juha et al., *The Nordic metadata project. Final report*. NORDINFO-rapport, 1998:1,  
NORDINFO, Helsingfors 1998  
(finns även i elektronisk version URL: <http://linnea.helsinki.fi/meta/nmfinal.htm>  
(000521))

Hedberg, Sten, "Bruket av metadata enligt Dublin Core : Principer, teknik och tillämpningar utanför Sverige", *Nya vägar för Boken*, senast uppdaterad 000512  
URL: <http://www.kb.se/NVB/Metadata/meta1.htm>  
(000521)

Hedberg, Sten, "Utkast till katalogiseringsregler och formatinstruktion för en svensk Dublin Core-tillämpning", *Nya vägar för Boken*, 990409  
URL: <http://www.kb.se/nvb/pdf/DC-katalogisering.pdf> (pdf-version)  
Alt. <http://www.kb.se/nvb/DC-katalogisering/format1.htm> (html-version, dock med ogiltiga interna länkar)  
(000518)

Hedberg Sten, "Vad gör vi egentligen? (Ex: Outsourcing)", inlägg i BIBLIST 200407  
URL: <http://segate.sunet.se/htbin/wa?A2=ind0004&L=BIBLIST&P=R2002>  
(000520)

Heery, Rachel, "Review of metadata formats", *Program*, Vol 30, No. 4, oktober 1996, s 345-373  
URL: <http://www.ukoln.ac.uk/metadata/review.html>  
(991021)

Henriksson, Alf, Törngren, Disa och Hansson, Lars, *Bra Böckers Hexikon. En sagolik uppslagsbok*, andra utgåvan, Höganäs 1988

Hickey, Thomas B. "CORC: a system for gateway creation" *Online Information Review*, volym 24, nr 1, 2000, s 49-53

Hopkinson, Alan, "Traditional communication formats: MARC is far from dead", *International cataloguing and bibliographic control*, 28 (1) 1999, s 17-21

IFLA, *Functional Requirements for Bibliographic Records. Final Report*, 1997  
URL: <http://www.ifla.org/VII/s13/frbr/frbr.htm>  
(Finns även i tryckt version: ISBN 3-598-11382-X)  
(000612)

IFLA, *Digital Libraries: Metadata Resources*  
URL: <http://www.ifla.org/II/metadata.htm>  
(000521)

Koch, Traugott, "Quality-controlled subject gateways: definitions, typologies, empirical overview", *Online Information Review*, volym 24, nr 1, 2000, s 24-34

Koch, Traugott och Ardö, Anders, "Automatic classification of full-text HTML-documents from one specific subject area", *DESIRE II D3.6a*, working paper 2, senaste uppdatering 000225  
URL: <http://www.lub.lu.se/desire/DESIRE6a-WP2.html>  
(000521)

Kungliga bibliotekets hemsida, *LIBRIS: Enkel sökning*  
URL: <http://www.libris.kb.se/enkel.sokning.html>  
(000120)

Kungliga bibliotekets hemsida, *Libris: Frågor och svar*  
URL: [http://info.libris.kb.se/infosvensk/fragor\\_svar.html](http://info.libris.kb.se/infosvensk/fragor_svar.html), senast uppdaterad 000417  
(000521)

Kungliga bibliotekets hemsida, *SweMARC – formatöversikter*, senast uppdaterad 000112  
URL: <http://info.libris.kb.se/swemarc/presentation.htm>  
(000119)

Lawrence, Steve och Giles, C. Lee, "Accessibility of information on the web", *Nature*, volym 400, 990708.  
(Artikeln kan erhållas som PDF-fil via URL: <http://www.wwwmetrics.com/>)  
(000213)

LeVan, Ralph, *Dublin Core and Z39.50*, Draft Version 1.2, 980202  
URL: <http://purl.oclc.org/dc/documents/notes/notes-levan-19980202.htm>  
(990830)

Library of Congress, *Challenges of building an effective digital library*, 1998  
URL: <http://memory.loc.gov/ammam/dli2/html/cdebl.html>  
(000214)

Library of Congress Network Development and MARC Standards Office, *Dublin Core/MARC/GILS Crosswalk*  
URL: <http://lcweb.loc.gov/marc/dccross.html>  
(000120)

Library of Congress Network Development and MARC Standards Office, *Guidelines for Coding Electronic Resources in Leader/06*  
URL: <http://lcweb.loc.gov/marc/lldr06guide.html>  
(000119)

Library of Congress Network Development and MARC Standards Office, *Guidelines for the Use of Field 856: Revised August 1999*

URL: <http://lcweb.loc.gov/marc/856guide.html>

(000120)

Library of Congress Network Development and MARC Standards Office, "MARC 21 concise format for bibliographic data"

URL: <http://lcweb.loc.gov/marc/bibliographic/ecbdhome.html>

(000119)

Library of Congress Network Development and MARC Standards Office, "MARC Advisory Committee"

URL: <http://lcweb.loc.gov/marc/advisory.html>

(000119)

Lindquist, Mats G., "Not your father's references: Citations in the digital space", *The journal of electronic publishing*, volym 4, nr 3, 1999

URL: <http://www.press.umich.edu/jep/04-03/lindquist.html>

(991021)

Lundberg, Sigfrid, *En arkitektur för ett distribuerat system för spridning av forskningsinformation*, 990318

URL: <http://www.safari.hsv.se/cris.pdf> (PDF-fil)

(000520)

Lynch, Clifford, "Searching the Internet", *Scientific American*, mars 1997

URL: <http://www.sciam.com/0397issue/0397lynch.html>

(000105)

Miller, Eric, "An Introduction to the Resource Description Framework" *D-Lib Magazine*, nr 5, 1998

URL: <http://www.dlib.org/dlib/may98/miller/05miller.html>

(000424)

Miller, Eric, Miller, Paul och Brickley, Dan, *Guidance on expressing the Dublin Core within the Resource Description Framework (RDF)*, Draft proposal 990701, Dublin Core Metadata Initiative

URL: <http://ukoln.ac.uk/metadata/resources/dc/datamodel/WD-dc-rdf/>

(990929)

Miller, Paul, "Z39.50 for All", *Ariadne*, nr 21, 1999

URL: <http://www.ariadne.ac.uk/issue21/z3950/intro.html>

(991003)

Milstead, Jessica och Feldman, Susan, "Metadata: Cataloging by Any Other Name", *ONLINE*, 1999a, s 25-31

URL: <http://www.online.inc.com/onlinemag/OL1999/milstead1.html>

(991021)

Milstead, Jessica och Feldman, Susan, "Metadata Projects and Standards", *ONLINE*, 1999b, s 32-40

URL: <http://www.online.inc.com/onlinemag/OL1999/milstead1.html#projects>  
(991021)

Murray, Robin, "The digital library jigsaw: fitting the pieces together", *Online Information '99. Proceedings. 23<sup>rd</sup> international online information meeting. London, 7-9 December 1999*, s 175 – 180, Oxford 1999

(elektronisk version på URL: <http://www.ingenta.com> (000214))

OCLC Online Computer Library Center Inc., *CORC Honor Roll of Participants*

URL: <http://www.oclc.org/oclc/corc/news/honor.htm>  
(000519)

OCLC Online Computer Library Center Inc., *The OCLC Cooperation Online Resource Catalog project*

URL: <http://www.oclc.org/oclc/promo/10520corc/>  
(000519)

Phillips, Margaret E., "The National Library of Australia: Ensuring long-term access to online publications", *the Journal of Electronic Publishing*, volume 4, issue 4, June 1999

URL: <http://www.press.umich.edu/jep/04-04/phillips.html>  
(000105)

Skogmar, Göran, "Från onlinekataloger till superkataloger", *LIBRIS-meddelanden*, nr 70 (1994, mars), s 4-14

SOU 1998:111, *E-plikt. Att säkra det elektroniska kulturarvet*, Norstedts Tryckeri AB, Stockholm, 1998

Task Force on Metadata and the Cataloging Rules, "Final Report", 980821

URL: <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei3.html>  
(990929)

Task force on metadata: Committee on cataloging: Description and access, "Summary report June 1999", 1999

URL: <http://www.ala.org/alcts/organization/ccs/ccda/tf-meta3.html>  
(000522)

Thomas, Charles F. och Griffin, Linda S., "Who Will Create The Metadata For The Internet?", *First Monday*, vol. 3 no. 12, december 1998

URL: [http://www.firstmonday.dk/issues/issue3\\_12/thomas/](http://www.firstmonday.dk/issues/issue3_12/thomas/)  
(000105)

Weibel, Stuart, et al., *OCLC/NCSA metadata workshop report*, OCLC, mars 1995

URL: <http://purl.org/dc/workshops/dc1conference/report.htm>  
(991109)

Weibel, Stuart, et al., "The 4<sup>th</sup> Dublin Core Metadata Workshop Report", *D-Lib Magazine*, nr 6, juni 1997

URL: <http://www.dlib.org/dlib/june/metadata/06weibel.html>

(991109)

Weibel, Stuart och Miller, Eric, "Image Description on the Internet. A Summary of the CNI/OCLC Image Metadata Workshop September 24-25, 1996. Dublin, Ohio", *D-Lib Magazine*, nr 1, januari 1997

URL: <http://www.dlib.org/dlib/january97/oclc/01weibel.html>

(000110)

Weibel, Stuart, "The state of the Dublin Core Metadata Initiative April 1999", *D-Lib Magazine*, nr 4, april 1999

URL: <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/april99/04weibel.html>

(991102)

Xu, Amanda, *Metadata conversion and the library OPAC*, 1997

URL: <http://web.mit.edu/waynej/www/xu.htm>

(000125)

## FÖRKORTNINGAR

---

AACR2	<i>Anglo-American Cataloguing Rules, 2<sup>nd</sup> edition</i>
AHDS	<i>Arts &amp; Humanities Data Service</i>
ALA	<i>American Library Association</i>
ANSI	<i>American National Standards Institute</i>
Arpanet	<i>Advanced Research Projects Agency Network</i>
ASCII	<i>American Standard Code for Information Interchange</i>
BTJ	<i>Bibliotekstjänst AB</i>
CNI	<i>Coalition for Networked Information</i>
CORC	<i>Cooperative Online Resource Catalog</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DESIRE	<i>Development of a European Service for Information on Research and Education</i>
DLO	<i>Document-Like Object</i>
DOI	<i>Digital Object Identifier</i>
DTD	<i>Document Type Definition</i>
EAD	<i>Encoding Archival Description</i>
ETB	<i>European Treasury Browser</i>
FTP	<i>File Transfer Protocol</i>
GILS	<i>Government Information Locator Service</i>
HTML	<i>Hyper Text Markup Language</i>
HTTP	<i>Hyper Text Transport Protocol</i>
IAFA/Whois++	<i>Internet Anonymous ftp Archives/Whois++</i>
IFLA	<i>International Federation of Library Associations and Institutions</i>
IMT	<i>Internet Media Types</i>
IP	<i>Internet Protocol</i>
ISBN	<i>International Standard Book Number</i>
ISO	<i>International Standard Organisation</i>
KRS	<i>Katalogiseringsregler för svenska bibliotek.</i>
LUB	<i>Lunds Universitetsbibliotek</i>
MARBI	<i>(ALA committee for) MACHine-Readable Bibliographic Information</i>
MARC	<i>MACHine Readable Cataloguing</i>
MCF	<i>Meta Content Framework</i>
NCSA	<i>the National Center for Supercomputing Applications</i>
NISO	<i>National Information Standards Organization</i>
OCLC	<i>Online Computer Library Center office of research</i>
OPAC	<i>Online Public Access Catalogue</i>
PICA	<i>Project for Integrated Catalogue Automation</i>
PICS	<i>Platform for Internet Content Selection</i>
RDF	<i>Resource Description Framework</i>
SAFARI	<i>Spridning av Forskningsinformation till Allmänheten över Internet</i>
SGML	<i>Standard Generalised Markup Language</i>
SHOE	<i>Simple HTML Ontology Extensions</i>
SUTRS	<i>Simple Unstructured Text Record Syntax</i>
TCP	<i>Transmission Control Protocol Protokoll</i>



TCP/IP	<i>Transmission Control Protocol/Internet Protocol</i>
TEI	<i>Text Encoding Initiative</i>
UDLO	<i>Unidentified Document-Like Object</i>
UKOLN	<i>UK Office for Library and Information Networking</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
URN	<i>Uniform Resource Name</i>
W3C	<i>World Wide Web Consortium.</i>
WWW	<i>World Wide Web</i>
XML	<i>eXtensible Markup Language.</i>
ZIG	<i>Z39.50 Implementors Group</i>

*Denna uppsats färdigställande hade varit en omöjlighet utan  
våra handledare, Mats Lindquist och Colm Doyles, råd  
och stöd. De skulle bägge förtjäna en dedikation om  
uppsatsen inte redan varit tillägnad Björn Tell.  
Tack även till Jan Gustafsson, Johanna  
Sandberg och Traugott Koch.  
Metadata rules!*

”In the middle of the 19th century there were 7 distinct track gauges for railroads operating in North America. This infrastructural impediment to the flow of goods had demonstrable effects on economic development, to say nothing of the additional costs of supporting such a rail network. It is important to note that nobody stopped the trains to wait for the tracks to become the same width. Over a period of 20 years a variety of work-arounds were deployed to ease the transition to a single track gauge. Meanwhile, the mail got delivered, the milk went to market, and people rode trains to populate the American West.

We are currently participating in the laying of track for the transport of metadata on the Web. We will not have gotten every aspect of the job just right the first time. That does not mean we should stop the metadata trains, nor that the cars will all be carrying the same things. But we should be looking for ways to improve interoperability, to reconcile semantics, structure, and syntax where it makes sense to do so. Failure to do so will force parallel evolution of non-interoperable systems that share many functional requirements.”

(Weibel, Stuart, ”The state of the Dublin Core Metadata Initiative April 1999”)