

# Human IT

Tidskrift för studier av IT  
ur ett humanvetenskapligt perspektiv

## "I have a dream"...

av [Gunilla Jonsson](#)

---

### Abstract

*From a single point of departure the user shall be able to retrieve all kinds of information resources, be it in print or in electronic format, be it on a library shelf or available online, that is the dream. However, the first task for the library community with regards to online publications is saving them for the future. If we do not accomplish this, we cannot reuse or verify the facts of the online world. Manual cataloguing and classification will not achieve structure or well guided access on the Internet. The volume of documents is simply too big. To achieve an overview and structure in this environment the methods have to be automated, and we have to encourage metadata production performed by the authors. We should also try to use our existing help systems, authority files and subject term thesauri to achieve the basic structure which is now lacking on the Internet. We cannot solve the task by the traditional procedures, instead, we should concentrate our resources on method development.*

---

### Innehåll

[1. Introduktion](#)

[2. Skall biblioteken ha något med Internet att göra?](#)

[3. Vad skall biblioteken på Internet att göra?](#)

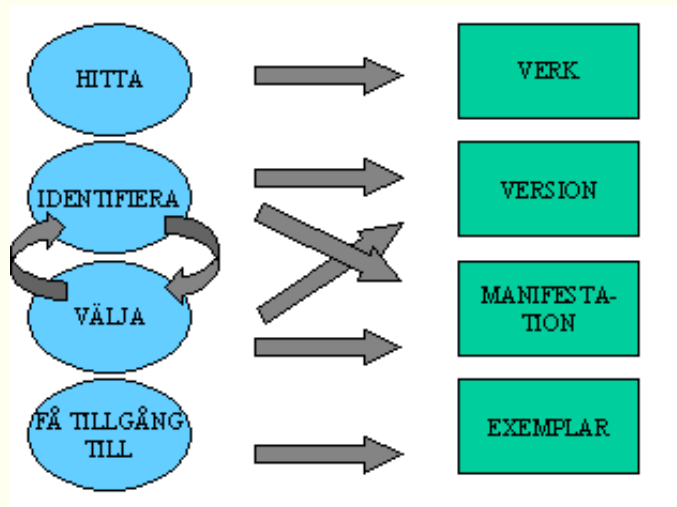
[4. Conclusion](#)

[Om författaren](#)

---

### 1. Introduktion

"I have a dream"... och den utgår från användarbehoven, så som de definieras i IFLA-rapporten *Functional Requirements for Bibliographic Records* (1997). Där skiljer man mellan att hitta verken, välja version (översättning, tolkning), välja manifestation (upplaga, inspelning) och slutligen få veta hur man får tillgång till ett exemplar. Nå, vad är det med det, säger förmodligen läsaren. Det där har vi väl alltid fixat, mer eller mindre, i våra bibliografiska databaser.



Drömmen är förstås att knyta ihop den bibliografiska basen med information om nätets innehåll, och att göra det på ett sådant sätt att användaren tydligt kan se vad som är tryck och vad som är online, och vad som eventuellt finns i båda formerna. I fråga om onlinedokument ersätts exemplarinformationen med direkt tillgång till fulltext, till själva dokumentet.

Mycket av diskussionen inom forskningsbiblioteken handlar för närvarande om huruvida man skall göra MARC-poster för webbdokument i Libris och i sina lokala kataloger eller ej, och vilka typer av dokument man i så fall skall rikta in sig på. Från användarhåll hakar man på och vill gärna ha "allt i Libris".

Jag menar inte att det är principiellt fel att katalogisera onlinedokument i Libris, och i vissa fall är det för närvarande också praktiskt (t.ex. beträffande material som man har tillgång till via avtal eller licens). Jag tror däremot inte att det leder någonvart när det gäller att åstadkomma den överblick över webbans innehåll som användaren väntar sig. Om vi i det korta perspektivet försöker göra användaren till viljes genom att lägga "allt" i Libris måste vi i det längre perspektivet avsäga oss rollen som informationsförmedlare i förhållande till det som finns allmänt tillgängligt online. – Internet eller world wide web är beteckningar för det allmänt tillgängliga nätverk vi arbetar med idag. Jag använder dessa beteckningar av praktiska skäl, men resonemanget har självfallet avseende på vilket allmänt tillgängligt nätverk som helst.

([Åter](#) till början av artikeln)

## 2. Skall biblioteken ha något med Internet att göra?

Det beror på hur man uppfattar bibliotekens uppgifter. I kondenserad form skulle de kunna beskrivas som informationsförsörjning och informationsförmedling. Med informationsförsörjning menar jag då uppgiften att välja, skaffa, härbärgera och hålla informationsbärare tillgängliga och med informationsförmedling uppgiften att ge vägledning till innehållet i informationsbärarna.

Ingen bestrider väl bibliotekens roll när det gäller traditionella informationsbärare, men gäller det Internet är svaret inte självklart, framförallt inte för de unga, vackra nätproffsen. Det kunde man märka på diskussionerna i vissa internettidningar och på Biblist under hösten. Samtidigt har IT-branschen börjat rekrytera katalogisatörer från biblioteken och boolesk sökning diskuteras i mer än en dagstidningskrönika (av vilka det numera går nästan en på varje text sida i tidningarna).

Att det som vi uppfattar som kärnkunskaper för en bibliotekarie plötsligt efterfrågas utanför biblioteken betyder i och för sig inte att biblioteken som organisationer är bäst lämpade att fungera som vägvisare till innehållet på Internet. När jag själv svarar obetingat ja på rubrikens fråga beror det också på att jag börjar med att nalkas problemet från informationsförsörjningens sida.

([Åter](#) till början av artikeln)

### **3. Vad skall biblioteken på Internet att göra?**

#### *Bevara*

En fullständig avgörande uppgift för biblioteksväsendet som helhet är att ansvara för den långsiktiga informationsförsörjningen, dvs. för att informationsbärarna bevaras så att innehållet kan göras tillgängligt både för samtid och framtid. Detta är en helt nödvändig samhällsuppgift även för det som publiceras online. Tar vi inte (eller får vi inte) det ansvaret hamnar vi vad gäller onlinepubliceringen nästan i en muntlig kultur med alla svårigheter det innebär i fråga om att återanvända eller verifiera uppgifter som varit allmänt spridda.

För bevarandepengiften krävs det en stabil organisation som kan förväntas leva kvar i århundraden, och som också står fri från olika branschintressen. Jag kan inte se att det finns några organisationer utanför biblioteks-, och möjligen arkivväsendet, som är i stånd att ta på sig jobbet.

#### *Vägleda*

Det material man ansvarar för måste också kunna organiseras så att man kan hitta i det. Här kommer nu äntligen informationsförmedlingen in. Katalogisering och klassificering är de viktigaste organisationsinstrumenten för det tryckta materialet och de mest grundläggande hjälpmedlen för informationsförmedling – alltid kompletterade dock av bibliografier, uppslagsböcker och andra sekundärpublikationer.

Nu kan vi inte utan vidare använda samma metoder för att organisera innehållet på Internet. Det första och största hindret är helt enkelt mängden online-dokument. För närvarande finns det i storleksordningen 15 miljoner filer på den svenska delen av Internet, och även om vi räknar bort alla bild- och ljudfiler och att många filer hänger

ihop i större innehållsliga enheter blir det ändå ett stort antal miljoner kvar. Det kan jämföras med att KB tar emot omkring 240.000 tryck (dagstidningar oräknade) som pliktleveranser varje år. Av detta katalogiseras drygt 25.000 (inkl. artikelindexering) medan resten, huvudsakligen det som förr kallades "småtryck", endast klassificeras grunt. Inte ens om alla bibliotek radikalt minskade på katalogiseringen av det tryckta skulle vi kunna katalogisera mer än en futtig bråkdel onlinepublikationer. Tro inte heller att det tryckta minskar, nej, det ökar. Kanske går ekvationen bättre ihop på ett universitetsbibliotek. Där har man förmodligen större möjlighet att växla papper mot onlinetillgång. På nationalbiblioteket går den inte ihop, åtminstone inte beträffande det svenska materialet.

Det finns ett antal skäl som talar emot de traditionella katalogiseringsmetoderna när det gäller onlinematerialet. Ett, som dock inte är avgörande, är att MARC-formatet inte klarar att på ett systematiskt användbart sätt härbärgera den tekniska information som behövs om dokumenten. Ytterligare ett skäl, inte heller avgörande, är att den nuvarande Libris-basen saknar mekanismer för automatisk kontroll av länkar. Det borde inte vara något svårslösligt problem och det måste under alla förhållanden lösas, eftersom det redan finns, och bör fortsätta att finnas, hyperlänkar till onlinedokument i Libris. Det manar emellertid till försiktighet och ansvarstagande när det gäller vad man för närvarande lägger in i Libris.

Viktigare är dock att Internetmaterialet a) i ökande utsträckning förses med kataloguppgifter i Dublin Core (eller annat metadataformat) av upphovsmännen, b) i stor utsträckning bör kunna automatkatalogiseras. Detta ser jag som den enda framkomliga vägen om vi skall kunna få en bättre överblick över vad som finns att hämta på Internet. Vi kommer aldrig att få så stora anslag att vi kan åstadkomma överblicken med hjälp av MARC-katalogisering, eller ens en grund, manuell klassificering. Kulturarw<sub>3</sub>-arkivet på KB är naturligtvis den självskrivna basen där man utnyttjar befintliga metadata (vilket görs redan i dag i Svesök-projektet, <http://www.svesok.kb.se/>) och där man kan få fram information om de mest använda textformaten och experimentera med automatgenerering av metadata.

Jag menar att den som förordar att vi skall göra MARC-poster i Libris för ett lämpligt urval webbdokument väljer en mycket passiv roll för biblioteken när det gäller informationsförmedling via Internet. Jag menar också att man på det sättet väljer bort möjligheten till överblick. Vi kommer att vara hänvisade till slumpmässiga urval, vilket är precis det läge vi har idag.

### *Urval*

Överblick är en förutsättning för ett förnuftigt urval. För det tryckta materialet finns förlagsreklam, ämnesindelade nationalbibliografier (nu tänker jag även på det utländska materialet) och inte minst ämnesklassificerade poster i Libris-basen som stöd för urval och förvärv. Detta saknas än så länge huvudsakligen på Internet. Yahoo, Evreka och en del andra kommersiella söktjänster har bra länkkataloger, men särskilt heltäckande är de inte. Det ger en snabb jämförelse olika söktjänster emellan besked om. Jan Hagerlid har hävdats att det för närvarande finns ungefär 1000 onlinedokument och länkar som är relevanta för hans bibliotek. Det är möjligt att han har rätt, och att man har fiskat upp det mesta av det bästa, men jag tillåter mig att tvivla. Svesöks ämnesredaktörer vittnar om en tilltagande svårighet att hitta material som är relevant

inom deras respektive områden, och jag tror inte det beror på dem, utan på att den grundläggande struktureringen saknas.

Jag vill nu inte förneka att det från ett universitetsbiblioteks utgångspunkt kan vara logiskt att avgränsa sin uppgift på ett annat sätt än jag gör från ett nationellt perspektiv. Universitetsbiblioteken skulle mycket väl kunna begränsa sin uppgift till att katalogisera den publicering som det egna lärosätet åstadkommer online, och i den mån akademisk onlinepublicering ersätter papperspublicering ser det väl i förstone inte ens ut som ett resursproblem för universitetsbiblioteket. Samtidigt erbjuder övergång till online möjligheten att få in metadata direkt från upphovsmännen. Har universitetsbiblioteken råd att avstå från det? Som universitetsbibliotekarie skulle nog jag hellre satsa på att få in metadata från upphovsmännen – för eventuell uppgradering av bibliotekarier – än på att göra om jobbet i MARC-format. (Det finns ju en självklar bas att arbeta vidare från i SAFARI, <http://safari.hsv.se/>.)

Svårigheten är naturligtvis att de tekniska system vi använder för informationsförmedlingen inte är ikapp utvecklingen och behoven. Vi kan inte hårbärgera Dublin Core-poster och MARC-poster i samma databas. Samsökning via Z39.50-gränssnitt skulle lösa många problem, men det finns idag inte några riktigt lämpliga baser att implementera ett sådant gränssnitt emot. Vi befinner oss i moment 22.

([Åter](#) till början av artikeln)

## **4. Hur skall biblioteken ge vägledning till Internet?**

Vi måste kunna hantera metadata, både på en individuell, manuell nivå och i söksystem. Jag tycker att vi har hyfsad kompetens när det gäller den manuella nivån. Den kunskapen är en förutsättning för att vi skall kunna vara med och påverka utvecklingen av metadataformat, och det tycker jag också vi har uppnått, eller åtminstone är på väg att uppnå. Vi försöker också sprida användningen av metadata.

Det är sämre på systemsidan. Vi är dåliga på att återutnyttja metadata. En upphovsman som vill finnas med i flera olika söktjänster kan tvingas fylla i ett nytt metadataformulär för varje tjänst. Det borde räcka att bara lägga på det som är specifikt för respektive tjänst.

Vi har heller inte någon heltäckande, allmänt tillgänglig sökmöjlighet vare sig för den levande eller för den "historiska" delen av webben, den som inte finns online längre. Framförallt kan vi inte samsöka Svesök eller andra söktjänster och Libris. Kunde vi det tror jag att den största delen av pågående eller planerat dubbelarbete i form av MARC-poster för onlinematerial skulle försvinna av sig själv.

Jag är emellertid övertygad om att vi också måste experimentera med automatgenerering av metadata. Naturligtvis skall vi också försöka styra, t.ex. genom att ta fram formatrekommendationer, eller DTD:er, för metadata delar till olika typer av dokument. Vi måste kunna gå ut och tala om att om man lämnar vissa uppgifter på ett angivet sätt så kommer metadata att genereras, och dokumentet kommer att återfinnas

lättare. Relevansrankning i svarslistorna är ett klart behov för webbmaterial. En sådan rankning skulle t.ex. kunna utgå ifrån hur posten framställts och där bibliotekarieframställda, eller uppgraderade, poster skulle rankas högst, manuellt framställda uppgifter av andra aktörer närmast och automatgenererade lägst. Andra rankningskriterier är önskvärda och tänkbara men svåra att åstadkomma i den ostrukturerade mängd det handlar om.

För att kunna göra åtminstone en grov strukturering av materialet måste vi se till att så många hjälpsystem som möjligt blir tillgängliga online på ett sådant sätt att de kan utnyttjas i metadatageneratorer. Det gäller auktoritetsfiler för person- och institutionsnamn, det gäller kontrollerade ämnesordslistor och klassifikationssystem. Auktoritetsfilen i nytt Libris får alltså inte vara "inlåst" i Libris-systemet, den måste kunna användas för metadataframställning på webben.

SAB:s ämnesord är KB på väg att lägga ut på webben, de är förmodligen redan tillgängliga när detta publiceras. Det som vi kan göra, eller redan gör, från KB räcker naturligtvis inte. Det finns andra ämnesordssystem än SAB:s, och det finns, som Lars Aronsson har påpekat, auktoritetsfiler för personnamn utanför biblioteksområdet. Samarbete mellan biblioteken och över institutionsgränser är nödvändigt, mappningar, åtminstone på generell nivå, mellan olika ämnessystem är viktiga. – Sådana mappningar bör förstås ske i riktning mot så stora, standardiserade system som möjligt. Under detta år har vi t.ex. i KB på uppdrag av SAB:s kommitté för katalogisering och klassifikation uppdaterat konkordansen mellan SAB:s klassifikationssystem och Dewey. Skall man mappa ämnesord bör det nog ske mot LCSH (Library of Congress Subject Headings), som är världens mest använda ämnesordssystem och som kan fungera som en "axel" för alla andra.

Jag tror också att ämnesordslistor kan vara en utgångspunkt för en grovstrukturering av webbens innehåll grundad på språkanalys. För att få ett meningsfullt resultat krävs förmodligen någon sorts taggning av klasser av ämnesord. Detta är ännu på idéstadiet, och det måste undersökas om idén är hållbar, men det är ett typexempel på den sorts arbetsuppgifter som jag menar att vi måste inrikta oss på om vi skall kunna bidra till överblick och orienteringsmöjligheter på webben.

Jag tror helt enkelt inte att vi kan lösa dagens informationsproblem med de medel som står till buds. Jag väljer att lägga resurserna på att förändra och utveckla medlen.

[\(Åter till början av artikeln\)](#)

## Om författaren

Gunilla Jonsson anställdes vid Kungl. biblioteket 1971. 1996 biblioteksråd, chef för Avdelningen för Insamling & Dokumentation. Tidigare chef för referenssektionen. Tillsammans med Ylva Tjeder upphovsman till databasen SB17, Svensk bibliografi 1700-1829.

[\(Åter till början av artikeln\)](#)

