

Human IT

Tidskrift för studier av IT
ur ett humanvetenskapligt perspektiv

Digitalisering och bevarande från projekt till långsiktiga program

av [Johan Mannerheim](#)

Innehåll

[1. Inledning](#)

[2. Plattform för bilddatabaser](#)

[3. Långsiktigt bevarande av digital information](#)

[4. Digitalisering och bevarande](#)

[Om författaren](#)

1. Inledning

I oktober 1999 var jag på en tredagars konferens i Utrecht om Digitisation of European Cultural Heritage. Det var påfallande vilken mångfald av projekt det finns, från enkla presentationer av godbitar ur samlingarna till filologiska projekt med avancerad tolkning av tecken och ord och samtidig presentation av textversioner från olika manuskript. Där fanns projekt, som integrerade digitalisering med lokal undervisning och forskning, och dokumenthanteringssystem med nationella ambitioner.

Kvaliteten varierade naturligtvis, men de stora problemen var att så många av projekten var rumphuggna, kortsiktiga eller oavslutade i brist på resurser och att teknik och mjukvaror snabbt åldras, så att resultaten redan håller på att bli oåtkomliga för studier.

Det bästa med konferensen var de rekommendationer som antogs. Man menade att nu var det dags att gå över från projekt till mer långsiktiga program:

All over Europe, experience in digitisation has been gained through

*pilot projects and innovative experiments in the field of cultural heritage. To bring about a transition to a stage of large-scale application, considerable effort and investments will be required to transfer existing knowledge to the operational level.*¹

Också på Kungliga biblioteket har det under senare år växt fram en mångfald av digitaliseringsprojekt, t.ex. porträtt ur kart- och bildsamlingen, affischer ur KB:s samling av över 500.000 affischer, ett urval svenska femton- och sextonhundredtalsböcker, en brevväxling mellan Selma Lagerlöf och Sophie Elkan och automatisk indexering av vår äldsta tidning, *Posttidender*, 1645–1721. Även här spretade projekten åt olika håll.

([Åter](#) till början av artikeln)

2. Plattform för bilddatabaser

För att råda bot på detta har KB under 1999 drivit ett projekt kallat Plattform för bilddatabaser. Avsikten har varit att skapa en gemensam plattform för produktion, lagring, tillhandahållande och registrering av digitala och även fotografiska (icke-digitala) avbildningar. Eftersom projektet berör många olika sidor av KB:s verksamhet har projekt- och referensgrupper varit sammansatta av anställda från många olika enheter och i den arbetsgrupp som behandlat registreringsproblematiken har även representanter från två andra bibliotek deltagit.

Ett område som visat sig vara viktigt har varit terminologi. Menar två bibliotekarier samma sak, när de använder ord som bild, original, bildfil och bilddatabas, eller pratar de förbi varandra? Och vad händer när en fotograf, en konservator och en dataexpert kastar sig in i diskussionen? Ett annat viktigt område har varit att formulera relevanta urvalskriterier för digitalisering. Hittills har de flesta digitaliseringsprojekt styrts av en önskan att tillhandahålla, medan bevarandenaspekten försumrats.

En väsentlig del av bilddatabasprojektet har rört kvalitet och standard. Hit hör frågan om vilka filformat vi ska använda för digitala bilder, för arkivering på lång sikt, för visning lokalt eller över Internet och för leverans till forskare, förlag och media. Likaså frågan om vilken bildkvalitet vi bör använda vid digitalisering av olika typer av samlingar. Här ingår upplösning, färgdjup, färgrymd och metoder för verklighetstrogen färgåtergivning. Hit hör också frågor om var olika metadata om bildfilen ska sparas: i bildfilen, i bilddatabasen eller i katalogposten.

Den mest omfattande delen av projektet har gällt bildregistrering. Det har gällt formatfrågor, främst arbete på att göra det svenska MARC-formatet bättre anpassat till bilder. Det har gällt indexering, där uppdraget varit att komplettera befintliga svenska ämnesord med ord för bildbeskrivning och att ta fram en gemensam teknik- och materialterminologi.²

Under arbetet med KB:s bilddatabasprojekt har vi mer och mer kommit till insikt om att en del av de frågor vi ställt oss inte kan lösas inom ramen för en institution. De behöver diskuteras i ett större sammanhang. KB har därför tagit initiativet till ett bredare projekt som syftar till samverkan mellan arkiv, bibliotek och museer, den s.k. ABM-sektorn. Tillsammans med Riksarkivet, Nationalmuseum och

Riksantikvarieämbetet har KB ansökt om medel hos Riksbankens Jubileumsfond och Stiftelsen framtidens kultur till ett ABM-projekt, som vi hoppas fler kulturarvsinstitutioner ska komma att delta i. Det går under arbetsnamnet Bilddatabaser och digitalisering – plattform för ABM-samverkan.

Registreringsdelen kommer här att få en annan inriktning än i KB-projektet, nämligen att hitta minsta gemensamma nämnare som alla kan omfatta, för att lägga grunden för samsökning inom arkivens, bibliotekens och museernas samlingar. Då kanske man inom en inte alltför lång framtid genom en enda sökning på Strindberg kan hitta arkivmaterial, tryckta böcker, handskrifter, målningar och fotografier av honom tillsammans med bilder föreställande honom inom hela ABM-sektorn.

([Åter](#) till början av artikeln)

3. Långsiktigt bevarande av digital information

Ett annat område som kommer att bli allt viktigare inom hela ABM-sektorn är det långsiktiga bevarandet av digital information. Om det som produceras digitalt idag ska vara tillgängligt för forskare och allmänhet i morgon och om hundra eller tusen år, måste vi skapa lika bra metoder för att bevara den digitala informationen som vi har för den pappersbundna.

KB kom på allvar in på detta område 1996 genom projekt Kulturarw³, som siktar till att bevara digitala svenska publikationer som finns online, framförallt svenska webbsidor. Eftersom antalet webbplatser (vilket ungefär motsvarar förläggare³) är stort, idag över 60.000, måste vi använda så automatiska metoder som möjligt. En insamlingsrobot söker av Internet efter svenska webbsidor och sparar de filer de är uppbyggda av. Den svenska webben omfattar för närvarande drygt sju miljoner sidor. Totalt har vi gjort sex nerladdningar sedan januari 1997 och har därigenom en historisk samling av svenska webbsidor på drygt 30 miljoner, vilket motsvarar 60 miljoner filer inklusive bilder, ljud m.m.

Svårigheterna med att bevara detta är inte att bevara den egentliga digitala informationen, dvs nollorna och ettorna, eftersom en kopia är exakt lik originalet. Det är nollorna och ettorna som ska bevaras, inte det medium som bär informationen. Problemet är att de program som tolkar informationen inte går att köra på morgondagens datorer och att de nya programmen inte kommer att kunna läsa gamla filer.

När det gäller webben minskar detta problem genom att de flesta filerna görs i relativt standardiserade filformat. Textformaten HTML och plain text och bildformaten GIF och JPEG utgör tillsammans 97 procent av filerna. Utsikterna för att kunna läsa dem är åtminstone under den närmaste framtiden goda. Det är bland övriga tre procent som problemen ligger: filer skapade av olika versioner av proprietära program som t.ex. MS Word, Excel, Acrobat och WordPerfect.

Det finns två sätt att angripa detta problem: migrering och emulering. Migrering innebär att man ständigt tar med sig dokumenten genom att konvertera dem till den aktuella mjuk- och hårdvarumiljön. Nackdelen är, som alla som konverterat texter mellan olika

ordbehandlingsprogram vet, att viss information, t.ex. layout, kan gå förlorad, även om textinnehållet i stort sett följer med ograverat. Vid upprepade konverteringar ökar risken för informationsförlust.

Vid emulering behåller man den ursprungliga filen. När man i framtiden vill läsa den skriver man ett nytt program som härmar, emulerar, det ursprungliga programmet i den nya miljön. Det pågår en del internationella projekt kring denna modell som vi följer med intresse. KB:s strategi kommer förmodligen att vara en kombination av de två modellerna. Vi kommer att migrera, men behåller de ursprungliga filerna för att kunna dra nytta av emuleringsprogram i framtiden.

Nu är webbsidorna inte den enda digitala information som vi på KB behöver bevara för framtiden. Vad händer t.ex. med dagens författarmanuskript? Förr tog KB och andra bibliotek ofta emot hand-skriftsarkiv från författarens dödsbo, ett material som är av stort värde för litteraturforskare som vill studera ett verks tillkomst och influenser. Idag finns detta material i författarnas persondatorer. Därför har vi tillsammans med KB:s handskriftsenhet och med stöd av författaren Sven Lindqvist, som är medlem av KB:s styrelse, startat ett litet projekt som siktar till att finna modeller att bevara digitala personarkiv. Vi försöker lära oss hur vi ska kunna ta hand om olika manusversioner, e-postkorrespondens, researchmaterial och andra spår från författarens verkstad.

Att bevara digital information på lång sikt är i själva verket ett generellt problem, inte bara inom ABM-sektorn, utan i samhället i stort, och det kommer bara att växa. Ännu nyligen var datorn ett redskap för att skapa tryckta eller skrivna dokument. Det digitala dokumentet är då en mellanform, som inte behöver bevaras när det slutliga pappersdokumentet är klart. Idag blir det allt vanligare att den digitala formen är den primära, som sedan kan spridas på olika sätt: på webben, med e-post, i form av en utskrift eller kanske rentav som en tryckt upplaga. Trots detta har de flesta institutioner och företag inte tagit itu med uppgiften att bevara denna primära information på sikt, utan förlitar sig på sekundära format som papper eller i vissa fall mikrofilm.

Att utveckla modeller och metoder för att bevara digital information på mycket lång sikt är ett område som utmärkt lämpar sig för samarbete inom ABM-sektorn.

([Åter](#) till början av artikeln)

4. Digitalisering och bevarande

De erfarenheter som vinnas av att bevara sådan information som i sin primära form är digital, och den infrastruktur och de resurser som skapas för detta, kan också användas för att ta hand om digitaliserat material. De gör det möjligt att undvika risken för att resultaten av olika digitaliseringsprojekt efter en tid blir oåtkomliga, eftersom den datormiljö de skapats i blivit föråldrad. Vid digitalisering kan man dessutom redan från början välja filformat och mjukvaror som är så standardiserade och utbytbara som möjligt. Därigenom blir det lättare att bevara sådan information i sin sekundära, digitala form på lång sikt.

Jag tänker avrunda med några tankar kring digitalisering och bevarande. Ofta tänker man slarvigt att digitalisering är en form av bevarande. Men det är det inte. Det kan

fungera precis tvärtom. Man digitaliserar för att öka tillgängligheten till en samling, en bok eller ett föremål. I de flesta digitaliseringsprojekt har man inte tänkt igenom vilka konsekvenser det kan få för originalen. En lyckad digitalisering ökar intresset för materialet. Är digitaliseringen av låg kvalitet vid exempelvis bilddigitalisering tillfredsställs inte forskarnas och allmänhetens behov av de grova bilder de erbjuds. Då söker de upp originalen, vilket ökar förslitningen av det kulturarv man ville presentera och kanske skydda.

Samtidigt gör man på många institutioner upp bevarandeplaner för sina samlingar. Sällan tar man i dessa planer upp de möjligheter som en digitalisering kan ge. Många lån av originalmaterialet skulle kunna undvikas om man kunde tillhandahålla det i digital form av tillräckligt hög kvalitet. Slitaget på originalen skulle minska. Problemet är att de som arbetar med bevarande och de som arbetar med digitalisering är specialister var och en på sitt område – och sällan mötas de två. Det behövs mer samarbete över gränserna både inom våra institutioner, mellan institutionerna inom hela ABM-området och mellan forskare och institutioner.

([Åter](#) till början av artikeln)

Om författaren

Johan Mannerheim, fil. kand. med både naturvetenskapliga och humanistiska ämnen, bibliotekarie, avdelningsdirektör, chef för Data och IT-enheten vid Kungl. biblioteket. Han har byggt upp den nationella mikrofilmningen av svensk dagspress i KB:s regi och medverkat i standardisering inom områdena mikrofilmning och "imaging". Han är boksamlare med inriktning på bokens historia och har undervisat vid Grafiska institutet i ämnet. Sedan mitten av 90-talet har han ägnat sig åt att utveckla KB:s infrastruktur på IT-området och är bl.a. ansvarig för den insamling av svenska webbsidor som KB bedriver för att bevara dem för framtiden.

([Åter](#) till början av artikeln)

Noter

1 Ur *Conclusions and recommendations of the international conference on Digitisation of European Cultural Heritage*, Utrecht, The Netherlands, 21–23 October 1999 ([Åter till texten](#))

2 En rapport kommer att publiceras under våren år 2000 i Kungliga bibliotekets rapportserie. ([Åter till texten](#))

3 Det exakta antalet förläggare på den svenska delen av Internet är svårt att fastställa pga 1. att en förläggare på Internet kan ha flera webbplatser, så siffran är för hög, 2. att webbhotell kan ha många förläggare inhyrda på samma webbplats, så siffran är för låg, 3. att någon statistik över svenska utgivare eller förläggare på Internet inte finns, så antalet webbplatser, som vi känner till, är den bästa uppskattning vi har. ([Åter till texten](#))

([Åter](#) till början av artikeln)

