

Det danska digitaliserade kulturarvet

Fra arktiske fotos over Matador til Barbie Girl: Texter, böcker, bilder, film og radio bliver i disse årne digitaliserat i anseelige mængder. Men hvad er status? En undersökning av användning, sökstrategier och findability.

Af Jonas Fransson jf@iva.dk, Det Informationsvidenskabelige Akademi

Kulturarvet som arkiverats av olika institutioner har under senare år börjat digitaliseras och samlingar gjorts tillgängliga på Internet. Ofta sägs syftet vara att tillgängliggöra och sprida kulturarvet i demokratisk anda till medborgarna. Frågan är om de tilltänkta användarna hittar till kulturarvet på internet och om det används. Användningen beror delvis på om medborgarna har intresse för kulturarvet, men användningen beror också på om det digitaliserade kulturarvet är lätt eller svårt att hitta när man söker information.

Under 2000-talet har kulturarvet börjat digitaliseras i större omfattning i Danmark. Det är än så länge små mängder kulturarv som har digitaliserats i förhållande till den totala sparade kulturarvet, men samlingarna med digitaliserat material omfattar anseelige mängder texter, böcker, bilder, filmer och radio. En stor del av samlingarna är tillgängliga på internet, men hur används de? Det är den centrala frågan i mitt phd-projekt vid Det Informationsvidenskabelige Akademi.

Projektet handlar om medborgarnas användning av det digitaliserade kulturarvet. Fokus är alltså användning i vardagslivet och inte professionell användning av t.ex. forskare. Att studera en så hetrogen användning är en utmaning men det leder förhoppningsvis till nya spännande upptäckter och insikter om vardagslivets användare. *Everyday life information seeking* är en nytt och växande forskningsområde inom *human information behaviour*. I vardagslivet är ofta informationsbehoven inte lika klart formulerade av användarna själva som vid informationssökning i arbetet, och beteendet kan vara mer rutinbaserat, som tv-tittande, eller vara reaktioner på uppkomna vardagsproblem.

Vad är kulturarv?

Vad är kulturarv och vad är digitaliserat kulturarv? Jag utgår ifrån minnesinstitutionernas digitaliserade samlingar och definierar digitalt kulturarv som det som ABM-institutionerna tillsammans med audiovisuella samlingar (t.ex. Danmarks Radio) har sparat och sedan digitaliserat. Detta för att kunna avgränsa undersökningen på ett sätt som fungerar operationellt. Begreppet kulturarv analyseras på ett bra sätt av Bernard Eric Jensen i boken *Kulturarv: et identitetpolitisk konfliktfelt* och definitionen av kulturarv handlar om politik, kultur och identitet. Även Aquas musikvideo Barbie Girl på

Youtube kan ses som danskt kulturarv tillsammans med exempelvis Matador.

Utifrån min operationella begränsning av det danska digitala kulturarvet så finns det ett antal samlingar tillgängliga på internet. Den kanske mest populära är Bonanza som innehåller radio- och tv-klipp från Danmarks Radios arkiv. På Det Kongelige Biblioteks webbplats finns flera centrala kulturarvsresurser tillgängliga. En är Arkiv för Dansk Litteratur (ADL) med över 160.000 sökbara sidor av den klassiska danska litteraturen, en annan är Guaman Pomas Inka-krönika som är en del av världskulturarvet. I Kunstindex Danmark återfinns mer än 50.000 fotografier av konstverk från olika danska museer. Andra stora digitaliserade bildsamlingar är Arktisk Instituts Arktiske bilder med 44.000 foton, och Marinens Biblioteks digitale fotoarkiv som innehåller runt 28.000 foton. Samlingarna är ofta med i Bibliotek.dk, på då som samlingar. På KB:s webb finns också *KulturPerler*, en strukturerad ingång till det växande antalet samlingar med digitaliserat kulturarv.

Samlingarna med digitaliserat kulturarv kan delas in två nivåer: resurs och objekt. Resurserna är samlingen av objekt i ett informationssystem, i en lokal databas eller i ett webbpubliceringsystem. Objekt är de enskilda filerna i format som html och pdf eller bild-, ljud- och videofiler. Objekten görs tillgängliga i resurserna. Nivåindelningen är viktig eftersom användaren interagerar med systemet (resurs) och med informationen (objekt) samtidigt.

Findability i centrum

Relationen mellan användare och digitaliserat kulturarv kan illustreras av en triangel mellan användare, information och informationssystem. Användarna interagerar både med informationssystemet och med information när hon söker, hitta och utvärderar information. Men det är komplexare än så. Internet är en avgörande del av både användarnas informationssökning och tillgängliggörandet av kulturarvsresurserna. Internet, eller rättare sagt webben, är det informationsrum där det digitaliserade kulturarvet blir tillgängligt och webben är ett informationsrum med sina lagar precis som ett fysiskt informationsrum.

Informationen på webben är lagrad i ett lokalt informationssystem, på en webbplats eller i en databas, som i sin tur att



Creative Commons: www.flickr.com/photos/brooklyn_museum

Der finns nu en tendens att göra kulturarvet tillgängligt i externa, idéella eller kommersiella, webbtjänster. Brooklyn Museum Archives har tillgängliggjort bilder på Flickr Commons så bilderna går att återvända.

tillgänglig på webben. En användare måste interagera med informationssystemet på webben för att sedan få tillgång till informationen. Vanliga webbsidor ligger ytligt och är lättillgängliga, medan kulturarvsobjekt som digitaliserade bilder ofta ligger djupare, tillgängliga i en databas. De fyra centrala delar är alltså användarna, informationen, informationssystemen och webben, och den gemensamma nämnaren mellan dem är hittbarhet (eng. *findability*). För att användarna ska kunna hitta och använda informationen måste den vara hittbar. Informationsarkitekten Peter Morville definierar *findability* enligt följande i boken *Ambient findability*:

The quality of being locatable and navigable.

The degree to which a particular object is easy to find or locate.

The degree to which a system or environment supports navigation and retrieval.

Som Morvilles definition visar kan hittbarhet gälla både på objektnivå (b) och på systemnivå (c). Eftersom informationen ofta är lagrad i ett lokalt informationssystem som i sin tur är tillgängligt på webben, ett globalt informationssystem, finns det många aspekter att beakta och därmed olika grader av hittbarhet. Om man tar en ett inskannat fotografi som exempel så kan bildfilens olika grader av hittbarhet i förhållande till en sökmotor vara:

Bilden är indexerad i sökmotor (t.ex. Google) och kan hittas direkt vid en sökning i sökmotorn på relevanta termer i metadatan (hög hittbarhet).

Webbsidan bilden är publicerad på är indexerad i sökmotor och kan hittas indirekt via webbsidan genom sökning i sökmotorn på relevanta begrepp i anslutning till bilden på webbsidan (mellan hittbarhet).

Första sidan eller söksidan på samling som bilden tillhör är

indexerad i sökmotor och kan hittas genom sökning i sökmotorn på samlingens namn eller i beskrivningen av den (låg hittbarhet).

Hur hittbart är egentligen det digitaliserade kulturarvet? Vanligtvis publiceras det i ett lokalt informationssystem, i en databas eller i ett webbpubliceringssystem. Ibland begravs det digitaliserade materialet långt ned på institutionernas webbplatser och är därmed svårt att hitta. Ibland har kulturarvet lanserats på temawebbplatser, vilka lätt bortprioreras när projektpengarna är slut och förfaller. Samtidigt finns det nu en tendens att göra kulturarvet tillgängligt i externa, idéella eller kommersiella, webbtjänster. Library of Congress har t.ex. tillgängliggjort bilder på Flickr Commons, en del av bilddelningstjänsten Flickr där bilderna får fria creative commonslicenser så bilderna går att återanvända. British Museum har haft ett samarbetsprojekt med Wikimedia, föreningen bakom Wikipedia. Ett av projektets mål har varit att sprida information om museets samlingar till allmänheten.

Hur hänger British Museum-projektet ihop med hittbarhet? Jo, med fler och bättre artiklar i Wikipedia om museets samlingar ökar antalet besökare på museets webbplats, fler Wikipedialäsare klickar sig vidare. Finns information om det digitaliserade kulturarvet där användarna är (Wikipedia, Google, osv.) så ökar sannolikheten för att kulturarvet når användarna i allmänheten.

Artiklar på Wikipedia är bara ett sätt att öka sin digitala synlighet på. Ett annat sätt är att optimera webbplatsen där kulturarvssamlingarna publicerats för sökmotorerna, så kallad *search engine optimization (SEO)*. Första steget är att tillåta att materialet får indexeras av sökmotorer, och att publiceringssystemet eller databaslösningen inte hindrar indexeringen. Utöver grundkraven på indexerbarhet finns det många aspekter inom SEO att arbeta för att hamna högre i träfflistorna, så som länkar, nyckelordsfrekvens i texter och metadatan. Om inte det digitaliserade kulturarvet kan hittas

vid en googlesökning blir den digitala synligheten låg då Google är den viktigaste navigeringstjänsten på webben idag.

Sociala medier är en annan viktig kanal för ökad digital synlighet och för ökat användande. Om de digitaliserade bilderna, ljuden, filmerna och texterna är lätta för användarna att dela med sig av till andra, t.ex. med share-knappar kopplade till Facebook och Twitter, kan genvägar in till samlingarna skapas. Genvägar som leder direkt till kulturarv som berört eller varit relevant för en tidigare användare.

Hittbarheten är viktig för att användarna ska hitta fram till informationen, i detta fall det digitaliserade kulturarvet. Men minst lika avgörande är användarnas informationsbehov. Användarnas sökning efter digitaliserat kulturarv kan vara svårt att entydigt benämna informationsbehov, utan intentionerna bakom sökningen kan vara väldigt skiftande. En grundläggande indelning av sökningarnas syften är mellan nytta och nöje. Nyttan inkluderar arbete, studier och skolarbete, men även fritidsaktiviteter med inslag av organisation och krav. Ett exempel på det sistnämnda är att man som medlem i en lokalhistorisk förening måste hitta en lämplig historisk bild som illustration till föreningens medlemsblad. Nöje är det som görs för den egna upplevelsen, njutningen eller kunskapsutvecklingen, och inte av yttre tvång. Exempel på nöje är att surfa på internet planlöst, att leta efter underhållande videos på Youtube eller att söka explorativt av nyfikenhet.

Vad gör användarna i kulturarvsresursen?

När användarna kommit till kulturarvsresursen vad gör de då? Vänder de i dörren eftersom träffen i sökmotorn inte motsvarade förväntningarna eller utforskar användarna samlingen grundligt? En del av användarbeteendet kan extrahearas genom analys av resursens webbloggar där varje kommunikationen mellan användarens webbläsare och webbplatsen lagras. En del av webbeteendet utförs i enbart webbläsaren och syns inte i loggarna på serversidan, t.ex. att text markeras och klipps ut eller tryck på webbläsarens bakåt-knapp. Men de två viktigaste aspekterna lagras i loggfiler, klickning på länkarna och sökning i den lokala sökmotorn.

Varje söksträng i resursen säger något om användarnas informationsbehov, varje sökord har valts utifrån intentioner och förhoppningar. Studeras en användares söksträngar under en söksession kronologiskt kan man se hur sökningen utvecklas och eventuellt anpassas till systemets förutsättningar. På samma sätt säger varje klick på en länk något om användarens mål med sin aktivitet i resursen. Navigationen genom en ämnesstruktur kan följas fram till ett objekt, t.ex. ett foto, eller vägen mellan två videoklipp. Om resursen har funktioner som dela-knappar (publicera länken på t.ex. Facebook) eller möjlighet till att kommentera kan det vara intressant att studera vilka objekt som utlöser sådana aktiviteter och när i söksessionen det görs. Även när användarnas aktivitet avslutas i resursen syns i loggfilen i form av utebliven fortsättning på sessionen.

Informationssökningsbeteendet i en enskild resurs är beroende på sökmöjligheterna i resursen. Det krävs en länkad ämnesstruktur för att det ska gå att browsa på ett enkelt sätt. Finns inte det så är användaren utlämnad till att söka med ord. På samma sätt om det inte finns någon intern sökmotor

eller någon metadata att söka i om det t.ex. gäller bilder, då måste användaren browsa bland objekten. Den digitala miljön avgör vilka sökbeteenden som är möjliga.

Varje samling är unik och varje resurs har sina speciella förutsättningar, men finns det likheter i användarnas beteende i olika kulturarvsresurser?

Hur hittar användarna till kulturarvsresursen?

På vilka sätt kommer användarna till det digitaliserade kulturarvet? Är det genom portaler som KB:s KulturPerler, via bibliotekskataloger som Bibliotek.dk och REX eller är det via generella sökmotorer som Google och Bing? En del av användarnas navigationsvägar återfinns i kulturarvsresursernas loggfiler där webbadressen från hänvisningssidan, den sidan användaren var på precis före besöket, sparas tillsammans med eventuella sökord om de kommer från en sökmotor.

Vid tillgång till webbloggar även från sök- eller navigationstjänster, t.ex. från KulturPerler, kan det gå att spåra användaren bakåt i dennes sökning. Från tidpunkten användaren kom till en kulturarvsresurs från KulturPerler går det att följa dennes väg bakåt genom KulturPerler som studera hur sökvägen har sett ut. Har kulturarvsresursen valts utifrån utgivare eller medietyp? Har andra kulturarvsresurser beskrivningar studerats av användaren innan den aktuella valts?

Sökningarna som sker i sökmotorer som Google kan inte följas då all information som genereras vid den omfattande användningen hemlighålls av företagen. Alla relevanta sökningar som inte leder användaren fram till de undersökta kulturarvsresurserna kan inte studeras, eller vad användarna prioriterar i en valsituation mellan en kulturarvsresurs och en annan resurs i samma träfflista.

Hur tillgänglig/hittbar är kulturarvsresursen och dess objekt?

Hur hittbar är kulturarvsobjekten på webben för användarna? Är informationen inläst i informationssystemen? Hur hittbara är informationssystemen i sig på webben för användarna? De undersökta kulturarvsresurserna analyseras enligt schema på både resurs- och objektnivå. Exempel på analysfrågor:

Är resursen katalogiserad i t.ex. REX eller Bibliotek.dk?

Är de enskilda objekten i katalogiserade någonstans?

Är objekten indexerade i sökmotorer?

Är objektet nedladdningsbart om det är en multimediafil eller är det urklippbart om det är text?

Har objektet en statisk eller unik URL som går att länka till?

Vilka rättigheter är kopplade till objekten (full copyright, Creative commons-licenser)?

Finns det webb 2.0-funktioner som t.ex. share-knappar eller möjlighet att kommentera de enskilda objekten?

Hur mycket metadata finns till de enskilda objekten?

Förutom findability-problematiken som ska kartläggas



Creative Commons: www.flickr.com/photos/brooklyn_museum

Brooklyn Museum Archives: Austrian Pavillion, Paris Exhibition (år 1900).

genom analysens schemat kan det kopplas till Digitaliseringsurvallets rekommendationer (Kulturministeriet, 2009):

”Det er vigtigt, at materialet kan eksponeres så bredt som muligt, herunder Europeana (som understøtter begge tilgange), Google og andre søgemaskiner.”

Vilka är användarnas tankar kring kulturarvsresurserna?

Vad vill användarna hitta i de digitaliserade kulturarvsresurserna och hur vill de använda objekten? Tycker användarna att deras behov blir uppfyllda eller har de andra önskemål? Brukar de söka i kulturarvsresurserna eller efter digitaliserat kulturarv på internet? Användarnas tankar, åsikter och vanor kring de undersökta resurserna och kring digitaliserat kulturarv i allmänhet är den fjärde studerade aspekten. Planen är att koppla en webbenkät till en eller flera av de undersökta resurserna och fånga tankarna hos aktiva användare.


Nackdelen med att koppla webbenkäten till de undersökta resurserna är att undersökningen inte når potentiella användare, men en indelning av användarna är mellan de som har valt att söka upp resursen och söka efter information, och de andra som har använt externa söktjänster, t.ex. Google, och där genom hamnat i kulturarvsresursen.

Vilka informationsbehov har användarna och hur det omsätter de behovet i praktiken, t.ex. val av sökord och sökmotod?

Implikationer

Med avhandlingen hoppas jag kunna bidra med ökad kunskap om hur det digitaliserade kulturarvet hittas av använ-

darna, men också mer kunskap om digitalt informationsbeteende mer generellt. Att sätta användarbeteendet i en funktionalitetskontext är ett nytt förhållningssätt som jag hoppas blir fruktbart, miljön kan vara en avgörande faktor i olika valsituationer. Kanske kan olika typer av användare urskiljas beroende på olika informationsbehov eller sökmönster? I avhandlingen kommer också hittbarhet som begrepp utvecklas och tillämpas, samt det digitala kulturarvets hittbarhet kartläggas bland webbens stora informationsmängder. Kanske finns det ett samband mellan hur mycket enskilda kulturarvsresurser och -objekt används och graden av hittbarhet?

Studiens resultat kan ha implikationer på många olika områden som systemdesign, informationsarkitektur och optimering gentemot sökmotorer. Även metadata, rättigheter och urval inför digitalisering viktiga aspekter, tillsammans med guider och målgruppsanpassning. Men först och främst hoppas jag studien ger ökad kunskap om användarna av det digitaliserade kulturarvet, både generellt och i relation till de undersökta kulturarvsresurserna och söktjänsterna. 

Referenser

- Fransson, J. (2007). *Effektivare informationssökning på webben: en handbok i konsten att söka information*. Ronneby: HEXA.
- Jensen, B. E. (2008). *Kulturarv: Et identitetspolitisk konfliktfelt*. Kbh.: Gad.
- Kulturministeriet, & Digitaliseringsudvalget. (2009). *Digitalisering af kulturarven: Endelig rapport fra Digitaliseringsudvalget*. København: Kulturministeriet.
- <http://kum.dk/ServiceMenu/Publikationer/2009/Digitalisering-af-kulturarven>
- Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.