



27 februari 2004

Riksbankens Jubileumsfond
 Box 5675
 114 86 Stockholm

Redovisning av projektet Ett virtuellt Ostindiskt kompani-arkiv

Margareta Benner och Helena Havner

Sammanfattning

De 1700-talshandlingar från det svenska ostindiska kompaniet som bevarats vid Göteborgs universitetsbibliotek har i sin helhet publicerats som bilder på webben, tillsammans med ett mindre antal från andra institutioner. Totalt har ca 15 000 bildsidor från 6 ägarinstitutioner lagts ut för fri åtkomst. Originalbilderna har lagrats som tiff eller jpeg på en filserver. Som webbvisningsformat har använts DjVu. Drygt 7000 sidor har laddats ner av användare oktober 2002 – januari 2004. Användare har berikat webbplatsen med egen innehållsindexering. Webbplatsen kommer att underhållas och erbjudas i fortsatt drift vid Göteborgs universitetsbibliotek.

<http://www.ub.gu.se/samlingar/handskrift/ostindie/>

Inledning

Projektet har pågått vid Göteborgs universitetsbibliotek under åren 2001 – 2003. Medel beviljades i oktober 2000.

Projektet hade flera syften. Det första var att skapa och webbpublicera bilder i hög kvalitet av arkivhandlingar som bevarats från det svenska ostindiska kompaniet, verksamt 1731 – 1813. Huvuddelen av dessa handlingar finns idag på Göteborgs UB. Ambitionen var att publicera samlingarna i sin helhet. Digitaliseringsprojekt har annars ofta inriktats på särskilt efterfrågade och intressanta handlingar. Att ge tillgång till dokumenten på distans, och att minimera slitaget av originalhandlingarna är naturligtvis ett viktigt syfte med digitaliseringen.

Avsikten var att också inventera och publicera kompanidokument som finns på andra institutioner. Därigenom skulle ett webbarkiv skapas som gav användaren *en* ingång till handlingar som i original befinner sig på olika institutioner.

Ett andra syfte var att utveckla webbplatsen tillsammans med användare och erbjuda dem att berika den med sakindexering och forskningsresultat. Medsökande i projektet var prof. Thomas Lindkvist vid Historiska institutionen, Göteborgs universitet.

En sakindexering som hänvisade till bilder av textsidor skulle innebära en enkel och billig metod att ge ingång till innehållet i handskrifterna. En vedertagen men dyr metod att göra handskrifter sökbara är att koda en avskrift med t.ex. SGML. En annan metod är optisk läsning. Detta har hittills inte varit realistiskt för handskrifter men kan bli möjligt i framtiden.

Projektet är en fortsättning på ett pilotprojekt vid Göteborgs UB 1997-1999, finansierat av BIBSAM¹.)

Genomförande

Databas och webbplats

Vi valde att skapa en egen databas därför att vi ville arbeta nära Historiska institutionen och få en databas som var anpassad efter historikernas behov. Inom Göteborgs universitetsbibliotek finns kompetensen för databasbyggande.

I databasen finns det tre användarnivåer som alla nås via webbplatsen. En publik del som ligger helt fritt tillgänglig, en del som är avsedd för studenter som indexerar (denna del är lösenordsskyddad). Det finns även ett administratörsgränssnitt som är avsett för dem som arbetade med att bygga upp basen.

Databasen har tre stora sökingångar: dokument, personer och ämnen.

Programvaror och utrustning

Bildproduktion

Djvu encoder från Lizardtech

Photoshop från Adobe

LupasRename (freewareprogram för att numrera om filer)

Kameror

Kontron ProgRes 3012 med tillhörande mjukvara

Canon EOS-1Ds med tillhörande mjukvara

Skanner HP Epson expression 1640XL med tillhörande mjukvara

Databashanterare och webbplats

Roxen CMS 3.4, webbserver/CMS

PostgreSQL 7.2, databashanterare

Perl 5.8, för att importera och bearbeta data som har funnits i Excel-ark

Bildarkiv, bildproduktion och lagring

Som lagringsmedium för arkivbilderna valde vi att lagra dem på en filserver. Servern ingår i ett nätverk med backup-rutiner. Den används endast för lagring av bilder, och ett begränsat antal personer har rättigheter till den. Kraven på säkerhet och långtidsbevarande av arkivbilderna är därigenom uppfyllda.

¹ Projektet Handskrift som bild på nätet 1998-99. Projektrapporten finns att läsa som pdf-fil på <http://www.ub.gu.se/samlingar/handskrift/ostindie/ostind.pdf>

Vi ville inte spara bilderna på cd- eller dvd-skivor. Anledningen var att vi ville undvika problemet att hålla ordning på massor av skivor. Dessutom har skivorna begränsad livslängd och man kommer att bli tvungen att föra över innehållet till nytt lagringsmedium efter en tid.

Vi har huvudsakligen tagit bilderna med kamera eftersom de allra flesta originalen är bundna eller häftade volymer. Lösa blad har skannats (främst makulerade aktiebrev). Vid fotograferingen har vi sparat bilderna direkt på filservern och vid bearbetning arbetat direkt mot denna.

Under projekttiden har vi lagt mycket tid och energi på att få den digitala fotograferingen att fungera. Vi ville ha en snabb digitalkamera med tanke på att vi hade stora mängder arkivmaterial att digitalisera. Valet föll på en Kontron ProgRes 3012 från Jenoptik i Tyskland.² Denna kamera använde vi fram till sommaren 2003 då programvaran till kameran blev obrukbar vid en uppgradering av operativsystemet på datorn. Vi hade då hela tiden haft stora problem med att få bilderna att hålla den kvalitet vi ville ha. Största enskilda problemet var att bilderna innehöll färgstick i rött och grönt. Eftersom kameran skannade av originalet var den relativt långsam; det tog ca en halv minut innan bilden hade förts över till datorn.. Sommaren 2003 köpte Göteborgs UB en ny digitalkamera, en Canon EOS-1Ds. Den har fungerat utan dessa problem.

Bildformat

Vi bestämde oss för att spara originalbilderna i ett icke-licensbundet standardformat och valde tiff och jpeg. Från början använde vi tiff genomgående. Av praktiska skäl har vi övergått till att spara det material som vi fotograferat med Canon-kameran i högkvalitets-jpeg. Alla bilder, både från kamera och skanner, är tagna i 300 dpi.

Som visningsformat för den bildkopia som visas på webben valde vi Djvu från Lizardtech.³ I pilotprojektet hade vi använt Flashpix som visningsformat, men eftersom detta format kräver en egen speciell bildserver och utvecklingsplanen för formatet var oklar valde vi att byta⁴. En viktig fördel med Djvu är att man på ett enkelt sätt kan hålla ihop bilderna av alla sidorna i ett dokument och visa dem i en följd, så som man gör i ett pdf-dokument. Zoomning och detaljzoomning fungerar väl. Användaren kan skriva ut och ladda ner bilderna. Formatet hade tidigare valts och testats av Lantmäteriet. En nackdel är däremot att användaren måste ladda ner en (gratis) plugin för att kunna se bilderna.

Söksystem och indexering

En forskare, FD Reine Rydén, anställdes i projektet när databas och söksystem byggdes upp. Han gjorde bl.a. provindexering av 53 dokument för att formulera termer och gruppera dem i sammanhang. Resultatet presenterades vid ett seminarium på Historiska institutionen. Beslutet blev att skapa en ämnesordlista i två nivåer. Termerna hämtades ur dokumenttexterna men med modern stavning och ordform. Vid projektets slut finns 2136

² Denna kamera fanns då i flera exemplar på Riksarkivet och användes av det då mycket omskrivna samarbetsprojektet Saganet <http://saga.library.cornell.edu/>

³ För mer information, se <http://www.lizardtech.com> och <http://www.djvuzone.org/>

För en genomgång av olika bildformat lämpliga vid digitalisering se Projektet *Bilddatabaser och digitalisering - plattform för ABM-samverkan* webbplats för digitalisering <http://abm.kb.se/akt4cd/default.htm>

⁴ Kungliga biblioteket använder FlashPix-bilder i sitt webbaserade digitala bibliotek med hjälp av en plugin, se <http://www.kb.se/F1700/default.htm>

termer inkl ortnamn. De grupperades i 16 ämnesområden för att ge användaren överblick. Vi valde att göra så istället för att konstruera ett thesaurussystem. Skälet är att materialet har ganska enhetligt innehåll och termerna är relativt få.

En annan del av databasen skapades för att registrera personnamn som förekommer i rullor och matriklar. Den innehåller även sökbara ortnamn och befattningar. Vid projektets slut är 1260 personnamn sökbara.

Databasen innehåller också formulär för att skapa poster för dokument.

Digitaliserade dokument

Eftersom Svenska ostindiska kompaniet inte bevarade något samlat företagsarkiv, ställdes vi ibland inför frågan om ett dokument skulle anses passa in i det digitala arkivet. Det finns handlingar som är privata dokument, t.ex. skrivna för eget bruk av deltagare i någon resa. Vi har inte haft tillfälle att inkludera något sådant. En handling med bara personanknytning till arkivet ingår, nämligen en hittills opublicerad notbok som tillhört Patrick Alströmer, en av direktörerna. Projektet fick tillgång till den genom vänligt erbjudande av prof. Jan Ling.

I projektet fullföljdes fotograferingen av alla de dokument som tillhör Göteborgs UB, inkl. en deposition från Göteborgs Stadsmuseum. En annan deposition (Colin Campbells dagbok från kompaniets första resa 1732-33) fotograferades inte. Den ligger permanent utställd på muséet. Eftersom den finns utgiven i tryck 1996 är den lätt tillgänglig. Därefter fortsatte fotograferingen med 2 dokument på Göteborgs Stadsmuseum, 5 på Sjöfartsmuséet i Göteborg, 5 på Uppsala universitetsbibliotek, 6 på Kungl. Vetenskapsakademien och 1 dokument på Kalmar länsbibliotek. Totalt har 159 poster över digitaliserade dokument upprättats. Ca 15 000 bilder finns utlagda vid projektets slut.

Några tryckta texter som kan vara svåråtkomliga har också digitaliserats, Carl Gustav Ekebergs ostindiska resa åren 1770-1771, som utgavs 1773, och några debattbroschyrer om kompaniet av Johan Arckenholtz, 1731 - 1735. Detta är ett bihang utan anspråk på djupare tankar om urval.

Ett virtuellt arkiv

Att sammanföra arkivspillror från olika håll och publicera dem på webben till ett någorlunda fullständigt virtuellt arkiv, har bara delvis lyckats. Som ett komplement har arkivförteckningar från fyra institutioner kopierats och publicerats som pdf-dokument på webbplatsen. Det ger en hjälp för användaren att orientera sig bland flera institutioners kompanidokument på ett ställe.

Berikning av materialet genom användaraktiviteter

Det seminarium som hölls med Historiska institutionen inför uppläggningsen av databasen gav goda uppslag. Den forskarassistent som var anställd i projektet har gjort huvuddelen av indexposterna då han testade olika modeller av indextermer och ämnesingångar.

Däremot har det hittills blivit mindre resultat än planerat av idén att forskare och uppsatsskrivande C-studenter skulle fortsätta att berika arkivet med indexeringar. Fyra studenter har åtagit sig uppgifter, två av dem har haft kontakt med handledaren och levererat registreringar.

Det kan finnas flera skäl till att sakindexering hittills inte visat sig så lockande som uppsatsämne. Det kan ha uppfattats både som för svårt och för lite meriterande. Däremot kan man vänta sig att källmaterialet framöver blir användbart för andra typer av uppsatser. Man kan också hoppas att indexering av t.ex. personnamn kan bli en biprodukt från uppsatser som behandlar ämnen kring Ostindiska kompaniet. Webbplatsen kan fortfarande erbjudas som publiceringsplats för uppsatser och annat otryckt material som handlar om Ostindiska kompaniet. Den är också användbar för paleografikurser.

För att administrera användardeltagandet lades rutiner upp för avtal, behörighet, access och lösenord, godkännande av registrering etc.

Användningen

De användare vi kan identifiera är de som kontaktar projektgruppen personligen. Hittills är det huvudsakligen tre grupper som har gjort det: skolelever, historiskt intresserad allmänhet (ofta släktforskare) och kollegor inom ABM-området.

Projektets främsta målgrupp är studenter och forskare. Inga sådana personer utanför projektet har kontaktat oss. Det behöver inte betyda att de inte finns och använder webbplatsen. Forskningsanvändning kan dessutom inte förväntas förrän volymen har blivit tillräckligt stor. De flesta dokumenten har publicerats först under hösten 2003.

Skoleleverna har kontaktat projektet genom e-post, frekvensen är i snitt en fråga i veckan. Det eleverna frågar om är allmänna frågor om kompaniets verksamhet, oftast i samband med att man ska skriva grupparbete och specialarbete. Historiskt intresserad allmänhet har använt både e-post och telefon med samma frekvens som skoleleverna. Det man har varit intresserad av är främst möjligheter att få fram information om vissa personer och detaljfrågor om kompaniet. Kollegor har ställt frågor om programvaror, resurser och kostnader. De som har ställt frågor till projektet har fått svar oftast dagen efter och som längst inom ett par dagar.

Användarstatistik

Det finns ingen fullständig webbstatistik för webbplatsen men det vi har är tillräckligt för att ge en bild av användningen. För 2001 har vi 1100 träffar i månaden utslaget över året, mer under terminstid och mindre över sommaren. I oktober 2002 lade vi till en statistikmodul i databasen och fick då fram antalet visade djvu-sidor per titel. Totalt har det visats drygt 7000 djvu-sidor från oktober 2002 till i början av januari 2004.

Den mest visade titeln hittills är Brevkopiebok för skeppet Hoppet med 487 visade sidor; originalet har 96 sidor. Denna titel var den första som vi digitaliserade och som inte tidigare publicerats inom pilotprojektet "Handschrift som bild på nätet". Den var också den första som blev ämnesindexerad.

Informationsspridning om projektet

Generellt kan man säga att vi verkar ha nått ut med information om projektet. Det har märkts i kontakterna med våra användare och kollegor, för kunskapen om vad vi har gjort har varit mycket stor. När vi har gått ut och bett om att få låna in arkivmaterial som normalt inte alls lånas ut eller bett att få kopior på arkivförteckningar för publicering, har vi mött stort intresse och generositet.

Information om projektet har kontinuerligt skickat ut till Biblist och Arkforum. Dessa två e-postlistor vänder sig till personer verksamma inom forskningsbibliotek och arkiv.

Webbplatsen har anmälts till alla större länksamlingar och webbplatser både inom Sverige och utomlands som har varit av intresse. Den har också anmälts till alla de upprop som kommit från internationella organisationer och projekt som inventerar digitaliseringsarbete inom ABM-området.

En broschyr som framställdes 2002 i svensk och engelsk version har delats ut vid alla presentationstillfällen både av projektet och i andra sammanhang då bibliotekets verksamhet har presenterats. Den har även funnits att ta för alla bibliotekets besökare.

Det har förekommit inslag i radio (P1), TV (Västnytt) och lokalpress.

Följande presentationer har gjorts:

Alla utom IFLA-föreläsningen har ägt rum i Göteborg.

2001

Februari: Presentation för Marieholms Rotaryklubb

Mars: Presentation för föreningen Datorn i släktforskningen, region väst. Visning för kollegor från högskolorna i Växjö, Karlstad och Örebro

April: Visning för kollegor från Stockholms UB

Maj: Presentation i samband med Vetenskapsfestivalen tillsammans med Landsarkivet i Göteborg

September: Medverkan på Bokmässan

Visning för kollegor från Kungliga bibliotekets handskriftsavdelning

Oktober: Presentation på Humanistdagarna, UB:s dag

2002

April: Presentation för föreningen Seniornätet

Augusti: Projektledaren och projektbibliotekarien deltog i 68th IFLA General Conference and Council 2002 i Glasgow. Vi höll en presentation vid sektionen för Social Science Libraries med rubriken The digitisation of the records of the Swedish East India Company

September: Visning för Göteborgs folkhögskola

En informationsfolder i engelsk och svensk version framställdes.

2003

September: Presentation på Bokmässan i Göteborgs universitets monter

Oktober: Föreläsningar på Universitetets vecka både på Humanistdagarna /UB:s dag och på Naturvetenskapens dag

November: Föreläsning på Föreningen för informationsspecialisters höstkonferens

2004

Maj: Medverkan i Vetenskapsfestivalen, Göteborgs universitet

För att sprida information om projektet till forskarsamhället tänker vi presentera det i Historisk tidskrift och vid Nordiska historikerdagarna i Stockholm 2004.

Fortsatt drift

Efter projektets slut finns inga aktuella planer för fortsatt digitalisering, men webbplatsen kommer att hållas i löpande drift på Göteborgs universitetsbiblioteks webbplats. Ansvar finns utpekad för tekniskt och innehållsligt underhåll och för användarkontakter. Arbetet med berikande genom användaraktiviteter ligger tills vidare i malpåse men kan lätt återupptas. Det finns en beredskap för att bevaka möjlig nyutveckling.

Avslutning

Det viktigaste resultatet av projektet är att ett stort centralt källmaterial har gjorts tillgängligt på webben. Detta är en långsiktig investering som kan användas för forskning och universitetsutbildning framöver. Webbplatsen har hittills visat sig vara attraktiv för användare från skola och allmänhet.

Erfarenheterna av användarnas berikning är hittills små. Hela materialet har inte varit tillgängligt särskilt länge, och det är för tidigt att bedöma hur användbart det är för forskning. Webbplatsen kan erbjudas för eventuell kommande utökning med nytt material från användare.

Publikation

Benner, M.: The digital archive of the Swedish East India Company, 1731-1813: a joint project of a university library and a history department. *Online information review*, vol. 27 nr.5, 2003, s.328-332.