

Handskrifter som bild på nätet - Ostindieprojektet

En studie i metoder för digitalisering av handskrifter

Av Helena Skoglund

Inledning

Handskrifter är ömtåligt material. De är också unika - de finns oftast bara i ett exemplar. Samtidigt finns det ett stort intresse och en stor efterfrågan på att få ta del av sådant material. En digitalisering ökar tillgängligheten till handskrifter för alla intresserade samtidigt som slitaget minimeras.

Göteborgs UB önskade öka sin kompetens inom digital bildhantering och pröva metoder att göra handskriftsmaterial tillgängligt över Internet. Svenska ostindiska kompaniets arkiv valdes för att det är mycket efterfrågat av många olika användargrupper.

Kontakter togs med Historiska institutionen och gemensamt inlämnades till BIBSAM en ansökan om bidrag för ett projekt med titeln "Handskrifter som bild på nätet: Ostindieprojektet: studie i metoder för digitalisering av handskrifter". I maj 1997 beviljades 193 tkr. till detta. Arbetet med projektet startade sommaren 1997, intensifierades under hösten 1998 och avslutades i juni 1999.

Projektgruppen har bestått av Helena Skoglund, Lennart Stark, Margareta Benner och från Historiska institutionen Nils-Göran Nilsson.

Syftet var

- att pröva olika metoder för bildfångst, bildbehandling, lagring och tillhandahållande
- att göra en expertbedömning ur användarperspektiv av kvalitetsskillnader vid olika metoder
- att öka Göteborgs UB:s kompetens på området
- att publicera ett substantiellt handskriftsmaterial på nätet

Ostindiska kompaniets arkiv

Svenska ostindiska kompaniet grundades 1731 och upphörde 1813. Dess historia brukar kronologiskt indelas efter de olika oktrojerna företaget fick. Med oktroj menas här, av kungen beviljat tillstånd att bedriva handel under viss tid på länder söder om Godahoppsudden.

1 oktrojen	1731-1746
2	1747-1766
3	1766-1786
4	1786-1806
5	1806-1813

Vid utgången av varje oktroj inköptes fartyg, byggnader etc. av de nya oktrojnehavarna. Därför kan företaget med lika rätt betecknas som en serie kompanier inbördes avlösande varandra.

Kompaniets huvudsakliga verksamhet var handel på Kina, där det ägde ett fabriks i Kanton. Under de fyra första oktrojerna var direktionen förlagd till Göteborg för att under den femte flyttas till Stockholm. I Göteborg ägde kompaniet bl.a. magasinshus (nuv. Göteborgs stadsmuseum) och ett skeppsvarv vid Klippan.

Vid Göteborgs UB förvaras de största bevarade resterna av kompaniets arkiv. Huvuddelen av handlingarna förstördes redan under kompaniets verksamhetstid. Vid Göteborgs UB finns 1.5 hyllmeter handlingar, rullor, matriklar och journaler, som uppskattningsvis motsvarar ca 10 000 textsidor, de flesta i form av inbundna volymer.

Intresset för Svenska ostindiska kompaniet har växt mycket under senare år. I flera år har Historiska institutionen vid Göteborgs universitet gett sina uppsatsskrivande studenter ämnen relaterade till kompaniet. Efter en tid av intensiv användning av arkivet var slitaget så kraftigt att institutionen och Handskriftsavdelningen vid Göteborgs universitetsbibliotek kom överens om att man skulle upphöra med att ge studenterna kompanirelaterade ämnen. Samtidigt började planerna på möjligheterna för digitalisering av delar av arkivet ta form.

Målgrupp

Målgruppen för detta projekt kan delas upp i flera olika kategorier.

En grupp är anställda inom bibliotek och arkiv som är intresserade av en lösning på hur man digitaliserar handskrifter, gör dem sökbara och tillgängliga via www. För denna grupp är det inte originalhandlingarna som är det primära utan det är bildkvalitet och metoden för tillgängliggörandet.

Forskare och uppsatsskrivande studenter är en viktig målgrupp. Denna grupp har ursprungligen tillgång till originalmaterialet på plats i Göteborg. Denna grupp av användare gav upphov till planerna på digitalisering och deras frekventa användning medför slitage.

Det finns en stor grupp utanför den akademiska världen som är intresserade av materialet från kompaniet. Det är släktforskare och historiskt intresserade privatpersoner vars tillgång till arkivet är begränsad på grund av geografiska avstånd och risken för slitage. Hit hör också intressenter i skolan.

Projektplan

I ansökan om projektbidrag gjordes följande arbetsplan:

"Av totalt 10 000 textsidor väljs 1750 ut. Materialet delas in i grupper som överförs till bild med olika metoder.

- Konventionell fotografering på 35 mm färgfilm, arkivbeständig färgdiafilm och svartvit film som skannas till foto-cd.
- Färgfoto med digital kamera av högsta standard.

- Skanning i svartvitt med s.k. bokskanner.

Fotografering utförs med bibliotekets utrustning och dess personal. Efter framkallning överförs bilderna till foto-cd.

För skanning används teknik med hög upplösning. Svartvit, gråskale- och färgskanning jämförs. Bilderna lagras i bildhanteringsformat. Redigering, lagring, komprimering och uppäckning prövas med några alternativa format. Bilderna indexeras i en separat databas. De läggs in på server i html-format för www -distribution men även för högre upplösning.

Återgivning i olika förstöringsgrader på skärm och i utskrift i skrivare med fotokvalitet jämförs. Som utprodukt prövas också cd-rom. Härigenom kan upplösningsgrad, tillgänglighet och överföringstider över www jämföras med lokal användning direkt på cd-rom-läsare. Beställning av valda bilder "på begäran" som cd-rom blir också möjlig från visningen på www.

Historiker som är specialister på dokumentstudier bedömer kvalitetsskillnader mellan återgivningar som åstadkommit med olika metoder. Som resultat av studien jämförs bedömning av kvalitetsskillnader med kostnader för de använda metoderna."

Förteckning över digitaliserade handlingar

Beskrifning på skeppet Göta Leijons resa...af C. H. Braad. Här ingår även journal för skeppet under samma tid, gällande för år 1751-52 (476 s.).

Rulla för skeppet Finland gällande år 1769 (15 s.).

Rulla för skeppet Cronprins Gustaf gällande år 1781 (12 s.)

Journal för skeppet Cronprins Gustaf gällande för år 1781-83 (462 s.)

Rulla för skeppet Gustaf III gällande för år 1791 (16 s.).

Rulla för skeppet Gustaf Adolph gällande för år 1797 (18 s.).

Journal för skeppet Gustaf Adolph gällande för år 1797-1800 (678 s.).

Rullorna

Rullorna listar alla besättningsmän, både de som var anställda för att ta hand om själva handeln och officerare och manskap. Från superkargören ner till jungmännen. Rullorna innehåller uppgifter om födelseort, hemvist, ålder, civilstånd och månadsgage under resan. Det finns även anteckningar om de besättningsmän som dog under resan och de som mönstrade på under färden för att ersätta avlidna besättningsmän.

Journaler

Journalen fördes dag för dag under tiden till sjöss efter en förtryckt mall. Noggranna uppgifter fördes om allt som hade med navigationen och väderleksförhållanden att göra. Man förde in händelser som rörde livet till sjöss, om man siktade andra skepp, om man

tvingades ransonera mat och vatten, proviantering, när besättningsmän avled, ibland noteras dödsorsak.

Dagboken av Christopher Hinric Braad

Dagboken som även innehåller journalen, förs under Götha Leijons resa till Indien. Den innehåller dagliga anteckningar om livet ombord, beskrivning av länder och befolkningar man mötte under vägen. Dagboken innehåller även teckningar, mest med motiv viktiga för navigeringen ex. kustlinjer.

Projektets genomförande

Av olika orsaker försenades projektet drygt ett år. Detta medförde vissa avvikelser från den ursprungliga projektplanen, eftersom nya program för visning av bilder på www och bildbehandling lanserades.

Metodvalen gjordes med tanke på att den tekniska lösningen skulle vara så driftsäker, stabil och enkel som möjligt. Lösningen kan provas och drivas hos andra intresserade, inte bara hos Göteborgs UB (förutsatt att man har tillgång till personal med kunskaper om programmering). Ambitionen var att få bildkvaliteten så bra att webbversionen kan ersätta originalet, samtidigt som bildfilerna inte får vara för stora. Filstorleken bör inte förorsaka långa väntetider vid användandet av basen.

Under hela projektet har vi fört en dialog med flera grupper av användare: kollegor inom bibliotek och arkiv, forskare från olika institutioner och en projektgrupp från skolvärlden (Läromedelsprojektet Ostindiska kompaniet, Utbildningsförvaltningen, Göteborg).

Förstudie

Studien började med en genomgång av aktuell information på området. Vi sökte information för att hitta exempel på liknande projekt. Det visade sig att det fanns flera exempel på digitalisering av tryckt material, men ifråga om handskrifter fann vi inga exempel på så omfattande digitalisering som vårt projekt syftade till. I allmänhet var det enstaka eller få bilder ur handskrifter. Vi fann inga exempel på att göra ett handskriftsarkiv sökbart med indexering gjord av användare.

Mycket av det vi hittade när vi sökte information i ämnet behandlade digitalisering av tryck. Det fanns mycket dåligt med information om projekt som hade arbetat med handskrifter. Handskrifter ställer helt andra krav på bildkvaliteten. Dels ger handstilen texten olika svärta allteftersom man skriver, dels är det nödvändigt med mycket god återgivning för att kunna tyda ut personliga handstilar.

Samtidigt som vi ville ha en bra bildkvalitet, var problemet att det digitala arkiv som byggdes upp inte fick bli för långsamt att arbeta med för den som satt vid en vanlig persondator med modemuppkoppling.

Vi valde att göra en första omgång testfoto-cd med skannade fotonegativ. Kontakt togs med två fotoföretag i Göteborg som gjorde varsin test-cd. Efter diskussioner med fotoföretagen valde vi att prova två bildformat, Kodak photo-cd och flashpix.

Fotografering och bildfångst

Analog fotografering

Projektarbetet började med fotografering av originalen. Detta arbete gjordes av universitetsbibliotekets egen fotograf. Fotograferingen försvårades av att materialet är bundna volymer. Det gör det svårare att få med hela texten och undvika skuggor utan att anstränga pärmen. Fotograferingen gjordes med analog kamera.

Planen var att fotografera både i svart/vitt och i färg, och både enkelsidor och uppslag. Härigenom skulle vi kunna jämföra dels tidsåtgången för fotograferingen, dels användbarhet och läslighet med olika metoder. Totalt blev det 907 exponeringar varav 128 enkelsidiga och 779 på uppslag. Av dessa var totalt 151 exponeringar i färg.

Fotograferingen i svartvitt gav mycket sämre resultat bildmässigt än färgfotona. En stor del, drygt 10 %, av de svartvita bilderna visade sig tyvärr ha defekter som gör dem oanvändbara. Defekterna uppkom vid framkallningen men upptäcktes inte vid granskningen av negativerna. Det mörkrumsarbete som en fotograf lägger ner på att förbättra papperskopiorna föll bort i och med att vi skannade negativerna direkt. Nu kom sådana defekter med på bilderna som skulle ha upptäckts, och eventuellt kunnat redigeras bort, vid den vanliga proceduren att göra papperskopior av bilderna.

Den film som användes, svartvit grafisk film, gav bilder med skarpa kontraster, lämpliga för återgivning av tryck men inte av handskrifter med stora nyansskillnader. Detta stod inte klart förrän vi sett den färdiga digitaliserade bilden. Färgbilderna gjordes som diabilder och gav betydligt bättre resultat.

Digital fotografering

Mot slutet av projekttiden gjorde vi försök med fem kommersiellt saluförda digitala kameror. De kameror som provades var av studiokvalitet (i prisklassen från 100 tkr och uppåt). Alla fem kameror gav utmärkt bildkvalitet. Det som skilde dem åt var den tekniska lösningen för att ta den digitala bilden. Den praktiska skillnaden var i stora drag: tid för exponering, möjligheterna till online-bild och medföljande programvaror.

Rullorna för skeppen Finland och Gustaf III (15 respektive 16 sidor) valdes ut till att fotograferas digitalt. Rullan för Gustaf III är sammanbunden med journalen, detta innebär att kameran måste kunna ta bra bilder även om sidan inte ligger helt plant.

Riksarkivet som tillhandahåller digitala fototjänster fick i uppdrag att fotografera de två rullorna med enkelsidor, 18 Mb bildfilstorlek och okomprimerade. Bilderna levererades på en cd. Den kamera som Riksarkivet använde var en av de kameror vi hade gjort försök med.

Bilderna skulle vara okomprimerade för att undvika att förlora ytterligare bildinformation vid en komprimering. Att göra bilderna helt okomprimerade visade sig vara opraktiskt, bilderna LCV-komprimerades och sparades sedan i tif-format. LCV-komprimeringen är inte förstörande och fördelen med tif är att det är ett öppet format. Efter komprimeringen är bilderna på ca. 18 Mb. På servern ersatte vi de tidigare

svartvita bilderna av rullorna med de digitalt fotograferade färgbilderna. Dessa färgbilder är arbetskopior i flashpix av de digitala originalbilderna.

De digitala bilderna hade en högre upplösning än de inskannade färgdiorna vilket ger ännu mer bildinformation. Här fanns det inget att anmärka på.

Avsteg från fotograferingsplanen

Enligt projektplanen skulle vi använda arkivbeständig färgdiafilm för några bilder. Detta genomfördes inte. Vi kom fram till att de digitala färgbilderna är av så god kvalitet att de kan bevaras som arkivoriginal. En framtida konvertering till nya format kan göras på kopior av de digitala 'originalbilderna' och dessa kan därför bevaras utan kvalitetsförluster.

Vi gjorde inte heller några bilder på s.k. bokskanner (en repromaskin som kopierar böcker rättvända). Det visade sig nämligen att upplösningen på dessa inskannade bilder är för låg.

Eftersom flashpix-formatet gjorde tillgängligheten över Internet mycket snabb, bedömde vi att det inte skulle finnas behov av cd-rom för bruk på distans. Till detta bidrog även utskriftskvalitén som är mycket bra.

Bildfilformat

Om en digital version av en handskrift ska kunna ersätta originalet krävs det att digitaliseringen är gjord med en så hög upplösning att det möjliggör förstorandet av svårlästa partier utan att bilden tappar i skärpa. Sådana bildfiler blir stora och mycket långsamma att skicka över www. Man kan givetvis komprimera filerna, men alltför hård komprimering ger synliga kvalitetsförluster i bilden. Det är heller inte säkert att användaren alltid behöver den högsta upplösningen och då måste man alltså tillhandahålla flera varianter av bilden, vilket är både tids- och utrymmeskrävande.

Det visade sig att bilderna i Kodak photo-cd format blev för stora för att vara lättarbetade. För att minska bildstorleken provade vi med att redigera bilderna i Photoshop. De svarta ramarna runt bilderna togs bort. Bilderna komprimerades sedan och sparades ner i jpeg - format. Detta arbetsätt tog lång tid och skulle göra arbetet med att få bilderna färdiga att länkas till basen omfattande. Vi ville även försöka att undvika den förlust av bildinformation som en jpeg-komprimering innebär. Bildformatet gif ansåg vi inte vara lämpligt för projektet. Det ger inte fullfärg och är inte ett öppet format.

Under projektets gång fick vi kännedom om flashpix, ett nyutvecklat bildformat som gav möjlighet att komma förbi dessa svårigheter. Flashpix är ett öppet bildfilformat som gemensamt utvecklats av Kodak, Live Picture, Microsoft och Hewlett-Packard. Flashpix skapar delbilder med lägre upplösning av det digitala originalet (högsta upplösningen). Varje upplösning är uppdelad i 64 x 64 block. Pixlarna i blocken kan antingen vara okomprimerade, jpeg- eller enfärgskomprimerade.

Man kan tänka sig flashpix som en pyramid, där den översta och minsta bilden ger den lägsta upplösningen. Denna bild används när man vill se hela eller en stor del av bilden.

När man förstorar någon detalj av bilden hämtas informationen från någon av de följande delbilderna där man har en högre upplösning. Detta ger stora fördelar när man skickar bilderna över internet, man behöver inte skicka en högre upplösning än vad användaren frågar efter.

Detta format gör det möjligt att välja själv var man vill zooma i bilden. Slut användaren blir inte hänvisad till redan gjorda förstoringar utan kan fritt välja själv. Det är även möjligt att skriva ut delar av, eller hela textsidan i den förstoring man valt.

Materialet intresserar skilda intressegrupper och att alla själva kan välja vad man vill titta närmare på gör att infallsvinkeln på materialet inte begränsas.

Hösten 1998 fanns det även möjligheter att själv spara bilder direkt i formatet flashpix med nya versionen av Photoshop.

De ovan nämnda företagen har också utvecklat IIP, Internet Imaging Protocol, som tar tillvara alla fördelarna i flashpixformatet. IIP används av den serverprogramvara (i vårt fall HP:s ImageIgniter) som krävs för att visa upp flashpixbilderna på webben. Med hjälp av IIP sänds bilderna effektivare kB för kB än en jpeg som skickas på vanligt sätt. För att visa upp bilderna i webbläsaren kan man låta servern skicka ut en java-applets eller så kan man använda en plugin. Vi har valt att använda oss av plugin-varianten eftersom det ger användaren fler möjligheter. Nackdelen med plugin varianten är att användaren måste ladda ner plugin-programmet (som är gratis), hur detta går till beror på vilken typ av webbläsare man använder. Men vi anser att fördelarna överväger med att ge användaren större möjligheter när bilden visas.

Kontakt togs med några större fotoföretag i Göteborgsregionen. Projektgruppen bad om en offert om de skulle vara intresserade av att åta sig uppdraget att

- skanna ca. 1000 negativ, både svart/vit och färgfilm.
- spara ner bilderna på cd i bildformatet flashpix.

Av sex större fotofirmor var det bara en som var intresserad och/eller ansåg sig ha kapacitet att skanna och överföra ca. tusen bilder till foto-cd i det bildformatet vi ville ha. De arbetade redan med flashpix i en egen bild databas. Vi gick igenom offerten och accepterade.

Skanning

När all fotografering var klar gick vi igenom bilderna och funderade över justeringar som skulle vara nödvändiga vid skanningen och utformningen av databasen. Vi hade valt att fotografera både enkelsidor och uppslag. Detta visade sig vara besvärligt med tanke på hur databasen skulle läggas upp.

I basen länkas information och indexering till varje bildfil. Det som är att föredra är om en textsida motsvaras av en bildfil. Det stora flertalet av det som fotograferats (779 st.), hade två textsidor per bildfil. Detta medförde stor risk för att göra misstaget att länka fel information till fel bildfil. Felsökning för ett fel av denna typ skulle ta mycket lång tid. Det enklaste sättet att komma runt detta problem var att vid skanningen dela de bilder som var tagna på ett uppslag. Detta innebar att alla bilder med textuppslag skannades två gånger och sedan delades bilderna i bildbehandlingsprogram och den ena hälften av

uppslaget togs bort. Att skanna bilderna dubbelt efter varandra var viktigt för att få bildfilerna i nummerordning för att underlätta när de skulle läggas in i basen. Detta medförde att projektet startade med 907 negativ och slutade med ca 1700 digitala bilder. Vid skanningen valde vi att fortsätta med att ta bort de svarta ramarna runt bilderna.

Dessa justeringar vid skanningen gjorde att den tog längre tid och blev dyrare. Det var nödvändigt att dela bilderna för att få en lättarbetad bas för oss och slutanvändarna. Detta är speciellt viktigt eftersom de deltagande historikerna planerar att som en fortsättning av det aktuella projektet, ämnesindexera bilderna.

De skannade bilderna levererades till biblioteket efterhand som de blev klara på en cd. Vi bytte namn på filerna och sparade ner dem på servern. I basen lades sedan in pekare till bildfilerna på servern. I basen skiljer vi inte på färg och svart/vita bilder.

Databasen

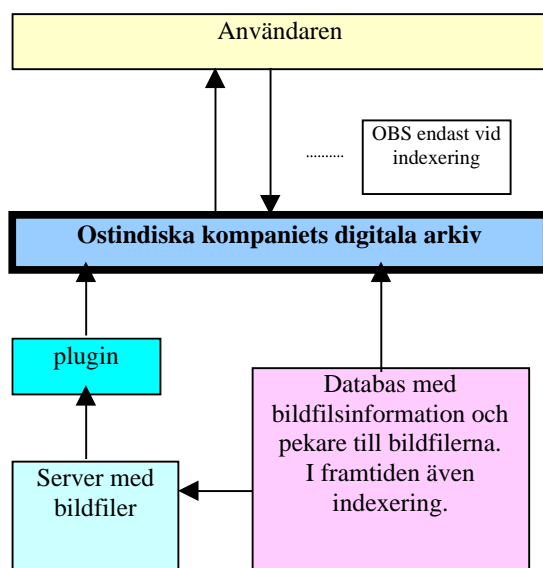
Denna bas har gjorts i Access, men i förlängningen räknar vi med att den bör migreras till Microsoft SQL-server eller Oracle. Detta blir nödvändigt om basen blir mycket använd eftersom Access blir trög om den ska serva för många användare samtidigt.

Det som hela basen bygger på är grundregistret. För att grundregistret skulle passa olika typer av material ska det vara så enkelt som möjligt och uppgifterna vara sidspecifika. Ett exempel på detta är journalerna som sträcker sig över två, tre år men året för sidan är ju alltid ett och det är det som finns med i grundregistret, det sidspecifika året. Andra uppgifter som finns i grundregistret är bildfilnamn, namn på det verk sidan kommer ifrån och pekare till bildfilen på servern.

Bildfilerna döptes så att man från sifferkombinationen direkt skulle kunna utläsa vilket verk och vilken sida av verket bildfilen innehåller, totalt sex siffror. Varje aktuellt verk fick ett eget tresiffrigt nummer som endast bygger på den ordning vi fick bilderna levererade till biblioteket. Det var enklast så för att då kunde vi efterhand vi fick bilderna länka dem till basen. Titelnumret följs av sidnummer, även det tresiffrigt. Rullan för skeppet Cronprins Gustaf har nummer 100 och bildfilnamnet för första sidan är 100001.fpx. Skälet var att underlätta arbetet med filerna i basen. Att ta fram själva bilden och se vad det är för något är en alldeles för långsam arbetsmetod. Textsidorna är vid en snabb anblick mycket lika varandra. Journalerna och rullorna fylldes i efter en mall, för journalernas del var mallen förtryckt.

Nästa register vi lade till efter grundregistret var ett titelregister, med en relation mellan varje titel i klartext och de tre siffrorna i varje titelnummer.

För att förtydliga hur databasen hör ihop med servern:



Indexering styrd av användare

Det finns planer på att fortsätta arbetet med databasen på följande sätt. Basen kommer med tiden att byggas ut med ett indexeringsformulär som är utformat efter de innehållsliga uppgifter som historikerna finner intressanta. Auktoriserade användare kommer själva att kunna berika basen med egna registreringar via formuläret. Databasen kommer att byggas ut för att kunna hantera indexeringen och även göra den sökbar. Allteftersom indexeringsmängden ökar blir de sökningar man kan göra i basen mycket intressanta.

Med en fullt utbyggd indexering skulle man kunna söka på sådana frågeställningar som: hur många av besättningsmännen var födda i Halland, i vilken stadsdel besättningsmännen bosatta i Göteborg bodde, finns det samband mellan befattningsgrad och var de var bosatta osv.

Att använda basen

För att kunna titta på bilderna krävs det att man använder sig av en webbläsare version 4.0 eller högre.

Använder man Explorer så går man in i det digitala arkivet och klickar på att visa en bild. Man kommer då att få en fråga om att installera Actice X Control - man svarar ja, sedan kan man se bilderna.

Netscape-användare behöver installera en plugin plus en extrafil som saknas i pluginen. Den plugin som vi använder finns att ladda hem fritt från vår server, adress: <http://tryggve.ub.gu.se/Openpix/> Här finns även en kort installationsbeskrivning på hur man ska göra.

Plugin finns idag tyvärr bara för PC, använder man Macintosh får man använda java-applets för att kunna se bilderna. Nackdelen med detta är att det begränsar användaren vid visning av bilden. Här räcker det med att kontrollera att webbläsaren är inställd på att acceptera java.

Göteborgs universitet använder SUNET så för användarna inom universitetet är laddningstiderna för bildfilerna inget problem (med förbehåll att de använder moderna datorer).

Test har utförts av studenter att arbeta med Ostindiska kompaniets digitala arkiv med modem (både 28.8 och 56 bps). Med modemuppkoppling tar det lite längre tid men att arbeta med det digitala arkivet tar inte längre tid än att titta på andra sidor. Tidsåtgången för att ladda hem bildfilerna beror även på datorkapacitet och övrig trafik på nätet. Bedömningen är att användare med modem inte behöver vänta besvärande länge för att ladda hem vårt material jämfört med annat.

Utskriftskvalitén på en skrivare av standardmodell blir mycket tydlig. Har man tillgång till en färgskrivare så kan man skriva ut sidor som utseendemässigt ligger mycket nära originalet. I vissa fall blir texten tydligare i utskriften än hos originalet p.g.a. att man kan manipulera kontrasterna mellan text och papper. Detta kan vara en fördel vid mycket svårlästa eller skadade handlingar. Vid utskrift har man möjlighet att skriva ut hela sidan eller ett utsnitt från sidan.

Historikernas bedömning av bilder gjorda med olika metoder

Historiska institutionens roll inom projektet är att representera de framtida användarna av Ostindiska kompaniets digitaliserade arkiv, såväl för forsknings- som för undervisningsändamål. Arbetsgruppen har bestått av FK Nils-Göran Nilsson, doc. Christer Ahlberger, FK Adam von Schéele och FM Kenneth Nyberg.

Arbetsgruppen har främst diskuterat frågor som rör läsbarhet med olika fotograferings- och skanningsmetoder, möjligheter att indexera innehållet efter olika forskningsbehov samt hur materialet bäst används i undervisningen på olika nivåer. Dessa diskussioner har förts inom gruppen och tillsammans med personalen vid Göteborgs universitetsbibliotek.

För att finna den fotograferings- eller skanningsmetod som ger den bästa läsbarheten har jämförelser gjorts mellan det digitaliserade materialet och originalhandlingarna samt de vanligaste sätten att reproducera arkivmaterial: fotokopia och mikrofilm. Jämförelsen innefattar läsning av det digitaliserade materialet direkt på dataskärmen samt av pappersutskrifter i färg och svart/vit. Vid läsningen har arbetsgruppen speciellt uppmärksammat möjligheterna att avgöra senare tillägg och att tyda s.k. genomblödningar, dvs. när bläck från motstående sida har trängt igenom papperet och försvårar läsningen. Jämförelsen baseras på rullorna för skeppen Cronprins Gustaf 1781 och Gustaf III 1791 eftersom originalen är likvärdiga. Den första rullan har fotograferats med färgdia som därefter digitaliserats och den senare har både fotograferats analogt med svart/vit film och digitalt i färg.

Rulla för skeppet Cronprins Gustaf, analogt fotograferad i färg och skannad

Läsningen av rullan på dataskärmen upplevs som likvärdig med att läsa originalhandlingen, det är lätt att identifiera tillägg och att med hjälp av delförstoring lösa upp genomblödningar. Just möjligheten att förstora delar av texten gör det till och med lättare att läsa otydliga avsnitt på skärmen än det är att läsa originalet med hjälp av vanligt förstoringsglas. Även färgutskriften av rullan ligger mycket nära originalet i läsbarhet, klart överlägsen de traditionella reproduktionssätten mikrofilm och fotokopia. Utskrift på svart/vit skrivare ger dock en otydlig och svårläst text eftersom bakgrundsfärgen medför dålig kontrastverkan. Den når inte upp till den vanliga fotokopians läsbarhet.

Rulla för skeppet Gustaf III, analogt fotograferad med svart/vit film och skannad

Läsningen av rullan på dataskärmen måste anses som likvärdig med att läsa mikrofilm, med undantag för bättre förstoringmöjligheter på dataskärmen. Dessa räcker dock inte alltid till för att avgöra senare tillägg och att lösa upp genomblödningar, delvis på grund av den skarpa kontrasten mellan text och bakgrund. Utskrifterna i svart/vitt är delvis svåra att läsa på grund av att svagare nyanser försvinner och de kan därför inte sägas nå upp till fotokopians läsbarhet.

Rulla för skeppet Gustaf III, digitalt fotograferad i färg

Dataskärmläsning av rullan är helt likvärdig med att läsa originalhandlingarna. En mycket god upplösning av texten vid förstoring för detaljgranskning medför att senare tillägg i texten kan identifieras och att genomblödningar löses upp utan svårigheter. Färgutskriften ligger också mycket nära originalhandlingarna, då till och med vattenstämpeln kan utläsas på kopiorna. Även i färgutskriften kan senare texttillägg med lätthet avgöras. De svart/vita utskriften är lättlästa och helt jämförbara med vanliga fotokopior i nyanser och tydlighet.

Sammanfattningsvis kan konstateras att för användaren framstår den digitalt fotograferade rullan som den som bäst uppfyller kraven på läsbarhet, oavsett om den läses på dataskärm, färg- eller svart/vit utskrift.

Kostnad för fotografering och skanning

Färgdia

Tidsåtgången för fotografering i färg var 14 min/exponering. Vid en timkostnad av 150 kr ger detta ett styckkostnad på 35 kronor. Till detta kom kostnad för framkallning, film och leverans på foto-cd, totalt 18:25/exponering. Skanning från foto-cd till flashpix-format, inklusive dubbelskanning och delning av de bilder som tagits av uppslag, gjordes på en fotofirma. Kostnaden per bild blev 24:31. Totalkostnad för en digital bild från färgdia blev 77:56.

Svartvitt foto

Tidsåtgången för fotografering på svartvit film var 7 min/exponering inkl. framkallning. Detta ger en arbetskostnad per bild på 17:50. Materialkostnaden var 1:25/bild. Kostnaden för skanning till flashpix samma som för färgdia, 24:31. Totalkostnad för en digital bild från svartvit film blev 43:06.

Digitalt foto

För de 31 bilder som gjordes av Riksarkivet med digital kamera och som levererades på en cd-skiva, blev styckpriset 74:20.

Dessa siffror ger en ungefärlig kostnadsbild för de olika metoderna. Priset från fotofirman på skanning till flashpix var ett introduktionspris för ett nytt format. Vid ett förnyat uppdrag i större skala skulle det antagligen bli högre. Priset från Riksarkivet var också ett introduktionspris, men samtidigt beräknat efter relativt stor tidsåtgång därför att just dessa originalhandlingar var särskilt svåra att fotografera. Den mycket lägre kostnaden för svartvita bilder måste ställas i relation till den lägre kvalitén på dessa. Om de hade granskats och efterbehandlats för att undvika defekter, skulle skillnaden i arbetskostnad ha jämnats ut.

Resultat

Projektet har resulterat i "Ostindiska kompaniets digitala arkiv" placerat på Göteborgs UB:s webbplats. Det digitala arkivet är fritt tillgängligt för alla intresserade.

Göteborgs UB har utvecklat en lösning på uppgiften med hjälp av kända programvaror och programmeringsspråk att visa digitaliserade handskrifter på webben. Valet av bildformat har gett en flexibilitet åt användaren som kan motsvaras av hur man arbetar med originalet. Användaren väljer själv vad som är intressant att titta närmare på. Laddningstiderna för bildfilerna är tillräckligt korta för att det ska upplevas som enkelt att arbeta med de digitala handlingarna.

De digitalt tagna bilderna har en sådan kvalitet att de kan betraktas som digitala original och ersätter originalhandlingarna även i ett forskningssammanhang.

De dryga 10 % av bilderna som är defekta kommer att ersättas. I förlängningen kommer även alla svart/vita bilder att bytas ut. De svart/vita bilderna har inte den kvalitet som gör att de kan ersätta originalen i ett forskningssammanhang.

Kontakter inom och utanför universitetet gör att kännedomen om det digitala arkivet sprids väl. Vi räknar med att det kommer att börja användas så fort det läggs ut. Under sommaren kommer det digitaliserade arkivmaterialet att användas för första gången av Historiska institutionen i undervisning av studenter på Paleografi, 5p., som ges inom ramen för sommaruniversitetet.

Sammanfattning och slutsatser

Efter att ha testat digitala kameror av studiokvalitet är vår bedömning att det enklaste sättet att digitalisera handskrifter eller andra handlingar som ställer stora krav på bildkvalitet är att använda digitalt fotografi från början.

Med denna lösning slipper man processen med att först fotografera originalen, skanna negativerna och sedan döpa om bilderna och lägga in information om bildfilerna i databasen. Eftersom skanningen i vårt fall skedde utanför universitetsbiblioteket var det ett led i processen vi i praktiken hade lite inflytande över. Processen tog lång tid och nu när det visade sig att vissa bilder var defekta hade det gått lång tid från fotografering till upptäckt och lokaliseringen av de defekta bilderna. Arbetet med att lägga in information om bildfilerna i databasen underlättas om man kan få bilderna i serie och bilder som tillhör samma verk samtidigt.

Det finns inte skäl att använda fotonegativ som arkivoriginal.

De bilder som togs i färg har en sådan kvalitet att de kan betraktas som digitala original. Dessa bilder kan ersätta originalen i ett forskningssammanhang. De svartvita bilderna har inte den kvalitén att de fullt ut kan ersätta originalen. Möjligheterna till konvertering av dessa bilder till framtida bildformat anser vi vara goda. Konvertering av dessa bildfiler medför inte någon informationsförlust. Att även fotografera materialet med vanlig (analog) kamera skulle vara dubbelarbete.

Våra erfarenheter i punktform:

- * Finns det möjlighet - använd digital kamera med en gång. Man slipper dubbelarbetet med att fotografera originalen för att sedan skanna negativerna. Digitalt tagna bilder som sparas i ett öppet bildformat okomprimerade eller i en oförstörande komprimering kan betraktas som digitala arkivexemplar.
- * Använd färgbilder. Användbarheten är likvärdig med originalet. Digitalt tagna färgbilder ger en mycket bra utskrift både i färg och svart/vit.
- * Fotografera enkelsidor. Undvik att få för mycket av underlaget med på bilden, det stör intrycket av bilden.
- * Var noggrann vid fotograferingen, defekter som inte hade syns på en papperskopia syns direkt på en digital bild.
- * Ha höga krav på bildformat och bildupplösning.
- * Tänk på överföringshastigheterna på webben.

Källförteckning

Eftersom vi i vårt arbete inte har använt andra digitaliseringsprojekt eller litteratur om handskriftsdigitalisering hänvisar vi inte till sådana källor.

Om Svenska ostindiska kompaniet

Kjellberg, Sven T., Svenska ostindiska compagnierna 1731-1813: kryddor, te, porslin, siden. Malmö, Allhems förlag, 1975.

Digitalisering

Det finns få tryckta verk som är aktuella och som behandlar digitalisering, det vi har använt är:

Besser, Howard och Trant, Jennifer, Introduktion till bilddigitalisering. Stockholm, Nordiska museets förlag, 1998.

Flashpix.com
<http://www.flashpix.com>

Hewlett Packard, Internet imaging
<http://www.image.hp.com>

Kodak, Digital products
<http://www.kodak.com/US/en/digital/flashPix/index.shtml>

Live Picture
<http://www.livepicture.com>

Microsoft
<http://www.microsoft.com>

"Pictor - mötesplats för den digitala bilden", projektet drivs av Riksantikvarieämbetet och Den digitala salongen.
<http://www.dds.se/pictor/>

Göteborgs UB har under projektet samarbetat med representanter från:

Historiska institutionen, Göteborgs universitet
<http://www.hum.gu.se/~hiswww/>

Projekt Ostindiska Kompaniet
Ett läromedelsprojekt som ger elever möjlighet, att med IT-teknikens hjälp, tränga in i en viktig västsvensk epok. Materialet skall kunna användas såväl i grundskolan som gymnasiet.
<http://ostindiska.educ.goteborg.se/index.htm>