

Umeå universitetet

Institution för sociologi

Bibliotekarieprogrammet HT 2007

Rapport moment 2

Handledare: Christer Karlsson

Student: Saašha Metsärantala

## RARA-projektet



## INLEDNING

Denna rapport ingår i bibliotekarieprogrammet vid Umeå universitet. Tre viktiga sidor som ingår i bibliotekarieyrket handlar om: • kunskap om böcker, • kunskap om (digital) informationsförmedling och • bemötande av låntagare. Jag ville gärna skriva någonting som berör flera av dessa sidor — och inte minst hur de samverkar. Därför kändes det speciellt angeläget att skriva min rapport kring RARA-projektet. RARA (eller rättare sagt RARA 2) handlar om att tillgängliggöra gammal litteratur — och i synnerhet rara böcker varav projektets namn — via nätet och projektet berör därmed både: • kunskap om (gamla) böcker, • kunskap om (digital) information och (i viss mån) • anpassning till låntagarnas förutsättningar för att göra informationen tillgänglig. Min känsla är att RARA-projektet berör två och ett halvt av dessa för bibliotekarieyrkets viktiga sidor.

När jag hänvisar till tryckta källor brukar jag använda mig av Harvard-systemet. Inom ramen för detta arbete anges källhänvisningar inom parentes och dessa sammanställs i källförteckningen sist i denna rapport. Vad filformat anbelangar hänvisas till <http://www.w3.org/TR/REC-png-multi.html> <http://en.wikipedia.org/wiki/PDF> <http://www.w3.org/Graphics/JPEG/jfif3.pdf> <http://www.w3.org/Graphics/SVG> och [http://en.wikipedia.org/wiki/Tagged\\_Image\\_File\\_Format](http://en.wikipedia.org/wiki/Tagged_Image_File_Format) för utförligare upplysningar. Alla webbkällor kontrollerades måndagen 2007-10-29 eller fredagen 2007-12-21, förutom de under avsnittet »några andra inskanningsprojekt» som besöktes vecka 10 år 2008.

Observera att viktig information finns tillgänglig på <http://www.acc.umu.se/~saasha/rara> där jag bland annat lagt en del filer, exemplifieringar, filter och animeringar, vilka kan bidra till en annan uppfattning än vad som kan åstadkommas på ett pappersark eller i ett pdf-dokument.

Arbetet med denna rapport har inneburit både textredigering, bildgranskning och filterknäpning. Allt arbete har jag utfört under operativsystemet Ubuntu, som finns att ladda ner (både gratis och lagligt) från <http://www.ubuntulinux.org> där även alla program som jag använt mig av ingår.

Ett varmt tack riktas till UB:s personal utan vilka denna rapport inte hade kunnat skrivas.

## BAKGRUND

Umeå universitet och dess bibliotek (UB) grundades på 1960-talet ([http://www.umu.se/umu/om\\_umu/historik.html](http://www.umu.se/umu/om_umu/historik.html)). Under de senaste decennierna har UB förvärvat material av olika slag, bland annat drygt en miljon böcker. Förvärv sker på olika sätt, bland annat pliktexemplar, inköp och donationer (<http://www.ub.umu.se/ombiblioteket/Littpolicy1-2-20041.pdf>).

Kortfattat kan sägas att de tryckerier som befinner sig i Sverige är skyldiga att skicka sju pliktexemplar till var och ett av följande forskningsbibliotek: KB, Lunds universitet, Uppsala universitet, Umeå universitet, Göteborgs universitet, Stockholms universitet och Linköpings universitet. I de fall ett förlag i Sverige låter trycka en bok utomlands är förlaget skyldigt att skicka pliktexemplar (<http://www.kb.se/forlag-tryckeri/pliktexemplar> <http://www.notisum.se/rnp/sls/lag/19931392.htm>).

Inköp innebär att UB köper in en del böcker som varken trycks eller ges ut i Sverige. Vidare köps i vissa fall in extra exemplar av de böcker som kommer som pliktexemplar.

Donationer innebär att personer som äger fler böcker än de vill ha ibland skänker bort en del av sina böcker till UB. Det är främst genom donationer som UB förvärvat gamla böcker. (<http://www.ub.umu.se/ombiblioteket/Littpolicy1-2-20041.pdf>)

UB har i sin policy att de böcker som är över hundra år gamla inte skall vara till hemlån. Av dessa gamla böcker är det endast en bråkdel som låntagarna över huvud taget får bläddra i. De allra flesta betraktas som så pass värdefulla och omtaliga att de förvaras i låsta magasin som endast ett fåtal bibliotekarier har tillgång till. En nackdel med detta är förstås att allmänheten förhindras från att studera dessa böcker. Å andra sidan kräver 10 § bibliotekslagen (SFS 1996:1596) att »Länsbibliotek, lånecentraler, högskolebibliotek, forskningsbibliotek och andra av staten finansierade bibliotek skall avgiftsfritt ställa litteratur ur de egna samlingarna till folkbibliotekens förfogande samt i övrigt samverka med folk- och skolbiblioteken och bistå dem i deras strävan att erbjuda låntagarna en god biblioteksservice.» (<http://www.notisum.se/rnp/sls/lag/19961596.htm>). Även om bibliotekslagen egentligen inte är överdrivet bindande framgår det tydligt att det är högst angeläget för ett forskningsbibliotek som UB att

tillgängliggöra sina skatter för allmänheten. Det är där som RARA-projektet kommer in! (<http://www.ub.umu.se/script/samlingar.htm>)

## **SYFTE**

Syftet med denna rapport var att analysera ett projekt för inskanning av äldre böcker vid ett bibliotek, nämligen RARA-projektet vid universitetsbiblioteket i Umeå. Däri ingick att kritiskt granska själva inskanningen, bearbetningen av denna, webbsidans design och kostnadseffektiviteten. Därtill ingick även att föreslå några förbättringar och att väldigt kort jämföra detta projekt med ett par andra inskanningsprojekt vid andra bibliotek.

## **HISTORIEN BAKOM RARA**

Tillsammans med några av de större samlingar som skänkts bort till UB kom även ett kortregister där böckerna är förtecknade. Det fanns en önskan om att datorisera dessa register. Att mata in uppgifterna för hand skulle innebära ett omfattande arbete, en risk för att läsa fel det som står skrivet på kortet samt en förlust av »känslan» av att se hur dessa registerkort såg ut. Då tillkom RARA 1 år 2002, som gick ut på att skanna in dessa registerkort och tillgängliggöra dem på nätet (<http://www.ub.umu.se/script/samlingar.htm>), exempelvis Östergrensamlingen (<http://130.239.76.160/oostergren/ipac/SearchForm.jsp>).

När detta projekt väl var avslutat beslöts om att starta RARA 2 år 2004 (<http://www.ub.umu.se/script/samlingar.htm>). Avsikten med RARA 2 är att skanna in äldre litteratur. Tre kriterier används vid prioritering av vilken litteratur som skall skannas in och göras tillgänglig på nätet: • boken berör Norrland, • boken är speciellt efterfrågad och / eller • boken är speciellt ömtålig (<http://www.ub.umu.se/loan/rara.htm>).

## **FÖRFARANDE OCH FÖRSLAG TILL FÖRBÄTTRINGAR**

De böcker som skannas in är speciellt gamla och ömtåliga. Vissa böcker är stora. Att vända böckerna upp och ner för varje inskannat uppslag skulle innebära en uppenbar

risk för att skada dessa. Därför skannas de utvalda böckerna i 300 dpi in med en A0-skanner där böckerna kan ligga på rygg. Endast ett blad vänds när nästa sida skall skannas in och boken behöver inte vändas upp och ner (intervju). Varje inskannat uppslag lagras som högupplöst bildfil i okomprimerat tif-format (tiff). Varje fil upptar tiotals megabyte och en bok kan lätt fylla en DVD (intervju).

Utifrån dessa tiff-filer skapas mindre jpg-filer, vilka kan visas på nätet. Självklart är jpg-filerna av betydligt sämre kvalitet än tiff-filerna men fördelen är att de tar storleksordningsvis hundra gånger mindre lagringsutrymme och är därmed betydligt lättare att överföra på nätet.

Med avsikt att åstadkomma en finess delas varje uppslag itu, vilket gör det möjligt för surfaren att »bläddra» i boken (se till exempel <http://rara.ub.umu.se/bookview/ipac/BookView.jsp?BookId=20>). Itu-delningen utförs av personalen på UB, medan bläddringsfunktionen ingår i ett köpt datorprogram.

I rapportskrivandet ingår att vara kritisk, vilket innebär den svåra uppgiften att utvärdera dagens RARA och komma med relevanta och konstruktiva förslag till förbättringar. För att underlätta arbetet med RARA, minska kostnaderna och minimera risken för fel kommer jag härmed med några förslag:

- Bakgrunden

Vid inskanning ligger boken på en bakgrund, vars färg inte nödvändigtvis skiljer sig speciellt mycket från bokpärmens eller boksidornas färger. Dessutom är bakgrundens färg inte helt jämn. Valet av en helt annan färg än de färger som förekommer på sidans marginaler och bokens pärm skulle underlätta en senare bakgrunds borttagning, vilket skulle innebära en snabbare och samtidigt mera verklighetstrogen bläddringsåtergivning på nätet. Vidare bör bakgrundens färg vara så jämn som möjligt (se nedan).

- Filernas efterleder

Filernas efterleder förekommer ibland med versaler och ibland med gemener. Detta innebär en uppenbar risk för olika fel. Om inga namnkrock riskeras kan detta åtgärdas väldigt enkelt med

```
\ls | gawk '{lo=tolower($0);if(lo!=$0){print "\mv -i " $0 " " lo}}' | sh
```

- Fillagring av de inskannade böckerna

De inskannade filerna lagras som sagt såsom okomprimerade tiff-filer, vilket innebär att varje sida kan ta tiotals megabyte. Eftersom filerna inte innehåller flera lager, CMYK-färgrymd (se länkarna nedan) eller annan information som inte kan lagras i png-filer vore i ett sådant sammanhang png-formatet att föredra. En konvertering till png kan enkelt åstakommas enligt följande:

```
\ls *.tif | gawk '{print "\convert " $0 " " substr($0,1,length($0)-4) ".png"}' | sh
```

I och med detta förfarande skulle uppskattningsvis 45 procent av lagringsutrymmet befrias. Om ett visst antal böcker tar exempelvis 100 gigabyte när de är lagrade med det okomprimerade tif-formatet kommer en konvertering till png-formatet att innebära att lagringen av dessa böcker kommer att uppta cirka 55 gigabyte (i stället för ett hundra) utan att bildkvaliteten på något sätt försämras. Om bakgrundens färg är förhållandevis jämn (se ovan) kommer utrymmesåtgången att minska ytterligare en liten aning.

- Bearbetning av filerna

UB:s personal bearbetar filerna med en kommersiell programvara — bland annat för att klippa itu uppslagen till sidor. För att undvika de kostnader som är förenade med inköp och uppgradering av programvara föreslår jag att övergå till gratisprogrammet gimp (<http://gimp.org>) som även ingår i Ubuntu.

- Visning på nätet

Visning — inklusive bläddring — av uppslag på nätet sker via ett datorprogram som UB köpt för 37.000:- av ett företag vars tjänster RARA-teamet överväger att överge bland annat på grund av dessa kostnader. Detta bläddrande blir mycket långsamt om det utförs vid aningen trögare datorer — tag gärna en promenad till studiesalen på UB och surfa till <http://rara.ub.umu.se/bookview/ipac/BookView.jsp?BookId=20> för att bläddra i en RARA-bok och bilda dig en uppfattning om detta. Jag har inte försökt hitta något bättre alternativ »på marknaden» eftersom fullgoda alternativ finns utanför den kommersiella världen. Bläddrandet är en programmeringsmässigt enkel funktion och priset för det program som UB använder är helt orimligt, speciellt med tanke på att företaget kräver att få en viss ersättning för varje publicerad bok (intervju). Härmed föreslår jag alternativ som inte kostar någonting (<http://www.acc.umu.se/~saasha/rara>). Jämför gärna de olika alternativ från olika nätuppkopplade datorer, vad gäller både

hastighet och kvalitet!

Ett märkligt krav som det av UB inköpta programmet ställer är att böckernas inskannade sidor skall lagras i filer vars förled skall bestå av varken fler eller färre än åtta stycken siffror, vilket förstås lätt kan leda till att filer förväxlas. Sådana krav är orimliga och skulle lätt kunna undvikas genom en smidigare programmering. Det alternativ jag föreslår på <http://www.acc.umu.se/~saasha/rara> godtar i de allra flesta fall (egentligen beroende på vilket filsystem som används) filnamn på upp till cirka 250 tecken, vilka inte är begränsade till siffror: även bokstäver (från a till z) godtas men jag rekommenderar att undvika bland annat é, å, ä och ö eftersom dessa bokstäver riskerar att återges på ett förvrängt sätt om undermåliga teckentabell-konverteringar utförs på filnamnen. Likaså avråder jag kraftigt från att använda mellanslag och interpunktion i filnamn.

Än så länge klarar endast ett fåtal (även bland de moderna) webbläsare av att hantera avancerade SVG-filer på rätt sätt — även om enstaka webbläsare lyckas hantera enkla SVG-filer nästan någorlunda ([http://en.wikipedia.org/wiki/Scalable\\_Vector\\_Graphics#Support\\_for\\_SVG\\_in\\_web\\_browsers](http://en.wikipedia.org/wiki/Scalable_Vector_Graphics#Support_for_SVG_in_web_browsers)). Mycket tyder dock på att detta kommer att förbättras inom kort. SVG-formatet var ursprungligen tänkt för att hantera vektorgrafik, men dess smidiga uppbyggnad har möjliggjort dess utveckling och de senaste specifikationerna ger utrymme för både animationer, tånjningar och länknings till punktgrafik. På sikt råder det nog föga tvekan om att SVG-formatet kommer att erbjuda riktigt användbara alternativ för att »bläddra» i böcker över nätet, bland annat därför att detta format är lätthanterligt samtidigt som det visuella intrycket är mera tillfredsställande än både det kommersiella alternativ som UB använder sig av idag och det alternativ jag föreslår. I stället för att bläddra »stegvis» med ett tiotal steg per bladvändning kan ett SVG-baserat alternativ »bläddra» steglöst. Vidare kommer det att bli lättare att automatiskt anpassa inzoomningsnivån till datorskärmens storlek.

- Bokens pärm följer med

Vid bläddring följer bokens pärm med. Att undvika detta är fullt möjligt men det är något som i många fall torde kräva en viss arbetsinsats. Inom ramen för denna rapport avser jag inte att föreslå någon lösning till detta problem.

- Pdf-filer

Det efterfrågas att kunna ladda ner böckerna som pdf-filer (intervju). Detta skulle exempelvis underlätta utskriftsproceduren. För att åstadkomma detta bör filerna bearbetas för att anpassas till utskrift. Problemet är inte så enkelt som det kan förefalla: »Hög kvalitet» innebär olika saker om avsikten är att ge ett så verklighetstroget intryck som möjligt eller att texten skall vara så läslig som möjligt. Vidare ställs olika krav på en pdf-fil om den är avsedd att skrivas ut på en färgskrivare eller på en svart-vit skrivare. Något annat som bör finnas i åtanke är den resulterande filens storlek. På <http://www.acc.umu.se/~saasha/rara> beskriver jag hur pdf-filer lätt kan skapas helt automatiskt, så att varje boksida skall motsvara en sida i en pdf-fil som motsvarar hela boken.

## NÅGRA ANDRA INSKANNINGSPROJEKT

Ett antal andra inskanningsprojekt har sett dagen under åren. Dessa är mer eller mindre ambitiösa, mer eller mindre kostsamma och mer eller mindre lyckade. Vissa bedrivs av ideella föreningar, andra av bibliotek och ytterligare andra kommersiellt. De mest ambitiösa och påkostade behöver inte alls vara de mest användbara. Här nedan tar jag upp några exempel mycket kortfattat.

- Skanning av affischer vid Kungliga Biblioteket

KB har bedrivit ett enkelt, föga ambitiöst men fungerande litet projekt som gick ut på att skanna in ett antal äldre affischer, vilka blivit tillgängliga på nätet. Bildernas kvalitet är visserligen begränsad, men webbsidan är någorlunda användarvänligt designat. (<http://www.kb.se/samlingarna/digitala/affischer>)

- Svenskt tryck före 1700

Vid kungliga biblioteket pågår projektet »Svenskt tryck före 1700» som går ut på att skanna in böcker tryckta före 1700 och tillgängliggöra dessa på nätet (<http://www.kb.se/F1700/start.htm>). Avsikten är säkert god, men användarvänligheten är inte den optimala. KB använder sig av ett kommersiellt program som jag inte hittar aktuell information om på nätet, vilket kan bero på att programmet kanske inte säljs längre. Hur som helst finns det gratisprogram som klarar av liknande bild-bearbetningar på ett betydligt smidigare och användarvänligare sätt (<http://www.imagemagick.org/script/>



convert.php). Det är förhållandevis enkelt att se små bilder på hela bokuppslag i en liten ruta i webbläsaren (där texten är så liten att den är oläslig) eller att zooma in ett litet område, så att texten blir läslig men då syns endast en minimal del av bokuppslaget. Följande bok tas som exempel:

»Lucida intervalla, Eller Goda och Gudeliga Tankar om Menniskians Dödelighet och Werldens Fåfängelighet» av Johansson, Lars (kallad Lucidor, 1638-1674) och tryckt 1685 (<http://www.kb.se/F1700/LL.htm>).

Det går att se småbilder på bokens uppslag (<http://ds.kb.se/?mapp=3&fil=Li-1>) och därifrån zooma in för att förstora små områden men att uppfylla en sådan enkel önskan som att få bilder på hela uppslag för en hel bok där texten är läslig kräver ett inte alltför intuitivt förfarande, exempelvis:

```
\gawk 'BEGIN{bsk=".tif.fpx?"  
cell=1000,1000&rgn=0,0,1.15,1&cvt=png\047";for(i=1;i<9;i++){print  
"wget \047http://orion.kb.se/de/f1700/Li-" i bsk "\nmv \047Li-" i bsk "  
lucidaintervalla" i ".png"}}'| sh
```

Mera intuitivt och användarvänligt än så borde det lätt kunna bli!

- Folkmusikkommissionens notsamling och Musikmuseets spelmansböcker

Sedan 2007-10-15 tillgängliggör Folkmusikkommissionen tusentals sidor av noter med visor (<http://www.smus.se/earkiv/fmk>). Detta projekt har resulterat i en tekniskt enkel och användarvänlig webbplats, där det är lätt att se hela och lättlästa bilder på de inskannade sidorna. Onödigt krångliga och påkostade funktioner lyser med sin frånvaro och det vore önskvärt om andra skrifter skulle tillgängliggöras på ett sådant sätt. Det är lätt att leta och hitta helt enkelt.

- Google books

Google books (<http://books.google.com>) hör till de allra mest ambitiösa projekten (<http://books.google.com/googlebooks/about.html>). Det är mycket omfattande men också ganska krångligt, vilket i vissa fall lär bottna i upphovsrätts skäl. Även när ingen upphovsrätt gäller på boken blir det betydligt krångligare än det borde vara att till exempel hämta en hel bild på en viss boksida.

## **EFTERTANKE**

För ett par år sedan var jag på en föreläsning av Inger Mossberg, som är en av de ytterst få kvinnliga kirurger i Sverige. Under sin föreläsning pratade Mossberg om begreppet »märkvärdisering». Med detta menar Mossberg ett sätt att framställa saker och ting som märkvärdigare än vad de egentligen är, för att därmed framstå som expert i sitt gebit, varmed det blir lättare att bevara sitt revir. Mossberg har blivit bemött med en hel del motstånd i sin strävan att bli och verka som kirurg. Som ett led i detta motstånd har märkvärdisering ofta använts. Enligt Mossberg har många kirurger framställt sin specialitet som »märkvärdigare» än vad den egentligen är, för att därmed framstå som experter och försöka hålla Mossberg undan det som de betraktade som sitt revir. Detta är ett exempel på hur märkvärdisering ibland används i patriarkatets tjänst.

Jag menar att något liknande kan sägas om det företag som säljer sitt program till UB. Att visa bilder på nätet hör till det allra mest grundläggande inom datavetenskap. Detta företag har dock märkvärdiserat bildvisningen och presenterat sig som enastående expert. Därmed har UB blivit lurat på en ofantlig summa, vilket jag — som liten student och flitig UB-besökare — tycker är djupt sorgligt. Bättre hade det varit om UB hade använt dessa pengar till att köpa in vettiga böcker — eller hinna längre med RARA-projektet.

## **SAMMANFATTNING**

RARA-projektet vid universitetsbiblioteket i Umeå är ett projekt för inskanning av äldre och ömtåliga böcker som har vuxit en del under de senaste åren. De inskannade böckerna är (delvis) tillgängliga via internet. RARA-projektet fungerar någorlunda, men vissa saker skulle lätt kunna förbättras, vilket jag visar i denna rapport. Ett smidigare val av programvara och filformat skulle kunna frigöra hårddiskutrymme, förbättra användarvänligheten och samtidigt minska kostnaderna, vilket skulle frigöra resurser för att skanna in böcker i snabbare takt.

## KÄLLFÖRTECKNING

- Föreläsning

Mossberg, I. Föreläsning vid Umeå universitet (2005-04-24)

- Intervju

Intervju av RARA-personal vid Umeå universitetsbiblioteket (2007-10-23).

- Länkförteckning

<http://130.239.76.160/oostergren/ipac/SearchForm.jsp> Östergrens katalog

<http://books.google.com>

<http://ds.kb.se/?mapp=3&fil=Li-1>

<http://en.wikipedia.org/wiki/CMYK> CMYK-färgrymd (Cyan, Magenta, Yellow, Key)

<http://en.wikipedia.org/wiki/PDF> Upplysningar kring pdf

[http://en.wikipedia.org/wiki/Scalable\\_Vector\\_Graphics](http://en.wikipedia.org/wiki/Scalable_Vector_Graphics) SVG-formatet

[http://en.wikipedia.org/wiki/Tagged\\_Image\\_File\\_Format](http://en.wikipedia.org/wiki/Tagged_Image_File_Format) TIF-formatet

<http://gimp.org> Ett bildredigeringsprogram

<http://rara.ub.umu.se/bookview/ipac/BookView.jsp?BookId=20> Exempel på en inskannad bok tillgänglig via nätet inom ramen för RARA-projektet

<http://susning.nu/F%E4rgrymd> Färgrymd

<http://www.acc.umu.se/~saasha/rara> Exempel på en alternativ RARA-sida

<http://www.imagemagick.org/script/convert.php>

<http://www.kb.se/F1700/LL.htm>

<http://www.kb.se/F1700/start.htm>

<http://www.kb.se/forlag-tryckeri/pliktexemplar> Pliktexemplar

<http://www.kb.se/samlingarna/digitala/affischer>

<http://www.notisum.se/rnp/sls/lag/19931392.htm> Lagen om pliktexemplar

<http://www.notisum.se/rnp/sls/lag/19961596.htm> Bibliotekslagen

<http://www.smus.se/earkiv/fmk>

<http://www.ub.umu.se> Umeå universitetsbibliotek

<http://www.ub.umu.se/loan/rara.htm> RARA:s webbsida

<http://www.ub.umu.se/ombiblioteket/Littpolicy1-2-20041.pdf> Förvärv av litteratur

<http://www.ub.umu.se/script/samlingar.htm> Information om UB:s specialsamlingar

<http://www.ubuntulinux.org> Ubuntu

<http://www.w3.org/Graphics/JPEG/jfif3.pdf> JPEG-formatet

<http://www.w3.org/Graphics/SVG>

<http://www.w3.org/TR/REC-png-multi.html> Specifikationer till PNG-formatet