

# Förslag till policy för digitaliseringsprojekt inom Lunds Universitets Bibliotek

*Sigfrid Lundberg (sigfrid.lundberg@lub.lu.se)*

Lunds Universitets Bibliotek  
PO Box 134  
SE-221 00 Lund  
Sweden

## ABSTRACT

Digital arkivering är att bereda serverplats för information om oss, vår tid och vår historia för våra barn och deras efterkommande. Detta dokument beskriver förutsättningarna för digitalisering, lagring och arkivering samt elektronisk publicering av olika samlingar inom Lunds universitet och dess bibliotek. Särskild vikt läggs vid att utveckla en finansieringsmodell för lagring, tillgängliggörande och bevarande.

## 1. Inledning

**S**YFTET MED DETTA dokument<sup>1</sup> är inte att motivera behovet av digitala arkiv; det räcker att konstatera att vi lever med en tidsinställd bomb. När den detonerar skattar viktig information åt förgängelsen. Vilken information som är mest hotad lämnar jag därhän. Inte heller diskuterar jag alla viktiga bibliografiska eller tekniska spörsmål om hur ett arkiv bör göras sök- och navigeringsbart eller vilka filformat som bör användas. Detta är i sig centrala frågor som måste dokumenteras för varje samling.

I stället koncentrerar jag mig på bevarandet som affärsidé, och hur möjliga strategier för detta slags verksamhet kan eller bör beskrivas. Inledningsvis spenderar jag en del utrymme på att klargöra några självklarheter som rör det digitala bevarandets väsen och dessutom ytterligare lite på vad bevarande inte är och aldrig kan bli. Till slut formulerar jag en finansieringsmodell för verksamheten utifrån de premisser jag formulerat.

Jag beskriver också, i Appendix A, hur Biblioteksdirektionens infrastruktur ser ut idag, och hur vi tänker oss att kunna etablera nya arkiv till rimlig kostnad.

### 1.1. Att etablera ett arkiv innebär ett förvärv

På samma sätt som förvärvet av en boksamling innebär att den måste katalogiseras, och sättas på en hylla där informationen i samlingen kan återvinnas, måste informationen i ett digitalt arkiv katalogiseras och ges en serverplats.

### 1.2. Arkiverat material måste förses med metadata

Att bevara något som inte i någon mening är katalogiserat är digitalt självmord. Som ett minimum måste det i likhet med okatalogiserat tryck, sorteras, dateras

<sup>1</sup> **Tillgängligt som:** [http://www.lub.lu.se/~siglun/laurentius/digitaliserings\\_policy.pdf](http://www.lub.lu.se/~siglun/laurentius/digitaliserings_policy.pdf) **Version:** \$Revision: 1.5 \$ **Date:** 2002/05/28 23:15:14 \$

och förses med någon form av etikett.

### **1.3. Förvärv innebär att en förvärvspolicy måste etableras och att ett digitaliserat bestånd måste underhållas**

I en värld med begränsade resurser, både när det gäller tid och pengar måste man välja. Vad är angeläget? Sannolikt bör en referensgrupp etableras för detta. Några punkter att ta i beaktande är:

- Anknytning till Lunds universitet.<sup>2</sup>
- Kondition. Känsligt material äger företräde; existerar digitala surrogat behöver inte originalen utsättas för slitage och smutsigt tummande.
- Förväntat intresse från allmänhet såväl som från forskarsamfund och akademisk utbildning. Material som kan tänkas vara intressant även utanför den akademiska bondbyn Lund har företräde.
- Prioritet ges åt överföring av material till maskinläsbar text snarare än enbart bevarande av bilder.
- Intresse ur tekniskt eller bibliografiskt perspektiv. Projekt som utvecklar personalens kompetens och skänker arbetsglädje.

Slutligen måste sannolikt allt förvärv, trots allt, vid någon tidpunkt kompletteras med gallring.

### **1.4. Själva arkiveringen kostar mer än driften**

Det är etableringen av ett nytt arkiv, eller när ett objekt införlivas i en samling som är förknippat med kostnader. Resten är i princip att byta magnetband vid säkerhetskopiering. Och att med någon regelbundenhet byta ut obsolet hårdvara.

### **1.5. Administration av "lånekort" är inte önskvärd**

En administrativ apparat runt våra arkivtjänster är en oacceptabel kostnad. Jag kan man tänka mig att vi begränsar offentlig tillgänglighet till ett fåtal IP-nummer, någon domän eller via Universitetets standard LDAP-lösningar, men inte något mer sofistikerat än så.

Biblioteken inom vårt nätverk har material av olika slag som med olika behov av bevarande, och en del av det kan behöva digitalt bevarande. Material som av olika upphovsrättsliga eller integritetsrelaterade anledningar kräver skydd, bör antingen inte digitaliseras eller bevaras tillsammans med sin kataloginformation på magnetband eller på en avstängd dator. Vilket också är en policy.

### **1.6. Ett arkiv är inte ett webbhotell**

Ett digitalt arkiv har som ett av sina viktigaste mål att bevara något. Hörs på namnet, skulle jag vilja säga. Vem, om någon, som har behov av det bevarade, och när, kan man inte säkert veta i förväg.

Ett webbhotell kännetecknas av att kunder hyr serverplats för att tillgängliggöra information *i dag*. Härav följer att ett arkiv inte kan bli ett webbhotell. Sådana debiterar per månad, under en tidsbegränsad abonnemangsperiod. *Vi måste debitera per evighet. Och göra vårt bästa för att försöka leva upp till vad*

<sup>2</sup> Det är sannolikt att ingen annan kommer att bry sig mer om vårt material än vi själva. Anknytningen kan bestå i att objekten ingår i våra samlingar eller att de har med vår verksamhet i dag eller i förgången tid att göra.

*det innebär.*

## 2. Förslag till policy

Att digitalisera och bevara material i digital form bör bli en vardagssyssla. När vi bereder plats för en samling skall den sedan finnas tillgänglig (förhoppningsvis) i evärdelig tid.

Vi kan särskilja två typer av verksamheter, *projekt* respektive *program*. I detta sammanhang har ett projekt som mål att digitalisera ett givet kvantum information, och publicera detta. När det är gjort återstår att se till att arkivet hålls tillgängligt. Ett program är ett projekt utan slut. Det vill säga, det material som skall digitaliseras är stadd i kontinuerlig tillväxt.

Gemensamt för båda är att vi söker pengar tillsammans med våra partners (eller debiterar dem) för den *extra* kostnad det innebär att vi lägger upp samlingen. För projekt tar vi betalt en gång, och objekten införlivas i *våra* samlingar. Ett program innebär en kontinuerlig merkostnad, och måste, så länge samlingen växer, ha en budget för att bekosta inläggningen. Annars finns i princip inga skillnader mellan projekt och program. De resulterande samlingarna underhålls enligt samma principer, det må vara medeltida handskrifter eller flygfoton. *Vi bör inte acceptera någon form av tidsbegränsade kontrakt när det gäller tillgänglighet för materialet.* Här gäller antingen målet evigt liv, eller någon framtida gallringspolicy. Fast det senare är en annan historia.

I skrivande stund kan vi inte förvänta oss att få finansiering för ett stadigvarande digitaliseringsprogram. Därför bör vi debitera tre års drift för varje projekt, och därvid hoppas på att bidragen från alla digitaliseringsprojekt i viss mån kan betala det allmänna underhållet för den infrastruktur vi bygger upp.

Katalogisering är en tidskrävande syssla. Men principen är klar enligt min mening: Utan relevant bibliografiskt arbete inkluderat i ett projekt, inget arkiv.

Vi utvecklar även användargränssnitt för dessa tjänster. I dem framgår det med hjälp av ikoner, bakgrundstexter och dylikt vilka som äger och förvaltar originalsamlingen. Därutöver utvecklar vi även databassystem som möjliggör navigering och sökning. Hur databaserna utformas beror på vilken typ av material som digitaliserats. Filosofin härvidlag är att systemen måste vara enkla och funktionsdugliga, men vi avstår helst från alla typer av gränssnitt som utnyttjar alltför futuristiska egenskaper hos webbklienter. Vi måste ha hållbara lösningar som kan förväntas fungera på egen hand under många år.

## Appendix A. Infrastruktur för arkiv och digitalisering

Inom Laurentiusprojektet<sup>3</sup>

har vi anskaffat en snabb SUN Enterprise serverdator med dubbla processorer och rejäla diskar. Utrymmet är runt 1 TB ( $\approx$  1000 GB). Systemet är snabbt och optimerat för ändamålet. För säkerhetskopiering finns en juke-box med magnetband och en robot som sköter byte av band. Det är min bedömning att Laurentius' infrastruktur har många års återstående livslängd.

Ett antal nya bevaringsprojekt är tilltänkta för denna infrastruktur. Några program har inte formulerats.

---

<sup>3</sup> Nylander, Eva; Borell, Mattias and Lundberg, Sigfrid:  
*S:t Laurentius Digital Manuscript Library*, Lund University Libraries2002.  
<<http://laurentius.lub.lu.se>>.

- Handskrifterna kör där redan.
- Jarringsamlingen. Digitalisering av en stor samling handskrifter på arabiska, persiska men framför allt på det uighuriska språket insamlade av ambassadör Gunnar Jarring.
- Esse Ericssons arkeologiska flygfoton.
- Eventuellt drottning Kristinas handskrifter (enbart katalogen) vid Vatikanbiblioteket.
- Eventuellt andra digitaliserade flygbilder och lantmäterikartor från GIS -centrum.
- annat ...

Allt detta kan inte få plats på denna server. Tanken är att komplettera den med ytterligare, billigare lagring av de digitala originalen. Dessa förvaras okomprimerat direkt i datorns filsystem i TIFF-format och kräver mellan 20 och 60 MB per bild. De enskilda bildernas filnamn knyter dem till metadata i XML , som ju som ett strukturerat textformat är beständigt.

Från rätt inköpskälla kostar ett bestyckat diskkabinett för att hålla 1.5 – 2.0 TB runt 150.000:–. Laurentiusdatorn går enligt uppgift inte att koppla till ytterligare diskar, däremot kan man låta varje nytt större arkiv skötas av en enklare PC, och på de billiga diskarna lagra de stora originalen vars material är viktigt men sällan efterfrågat i jämförelse med de komprimerade bilderna som kommer att förvaras på Laurentius, vars kapacitet är snart sagt obegränsad för detta ändamål.

Förutsättningarna är med andra ord mycket goda för att såväl lagra, som hantera stora elektroniska arkiv.

För Jarringsamlingen har vi sökt 400.000:– för såväl bildhantering och hårdvara här på (LUB).

Digitaliseringen sköts på en separat post i budgeten nere på Helgonabacken. Projektet har också poster för katalogisator (Helgonabacken) och XML -stöd till denne (LUB).

Vi har tänkt en liknande prislapp för lagring i fallet Esse Ericsson. IK Foundation kommer själva att ombesörja katalogiseringen av bilderna, men vi vill samordna sökmaskinen med GIS -centrum.