



# **Studieresa till San Francisco för projektet Print on demand, 1-5 November 2010 : reserapport**

Urban Andersson  
Gösta Cramby  
Björn Odenbring



GÖTEBORGS UNIVERSITET  
GÖTEBORGS UNIVERSITETSBIBLIOTEK

## Innehåll

|  |    |
|--|----|
| Sammanfattning .....                     | 3  |
| Studiebesöken .....                      | 6  |
| 1. UC Press .....                        | 6  |
| UCPubS .....                             | 7  |
| 2. California Digital Library .....      | 8  |
| 3. Public Library of Science, PloS ..... | 9  |
| 4. Stanford University .....             | 10 |
| 5. Internet Archive .....                | 12 |
| Länkar till mer information .....        | 17 |

© Göteborgs universitetsbibliotek, 2010.

Digital version av denna rapport finns fritt tillgänglig i Göteborgs Universitets elektroniska arkiv, GUPEA, med adressen <http://hdl.handle.net/2077/24022>.

## Sammanfattning

Studieresan genomfördes under en vecka, i början av november 2010, som en del i projektet Print on demand.<sup>1</sup>

Med print on demand avses här tjänster som syftar till möjligheten att kunna trycka en publikation på begäran, till skillnad från – eller som ett komplement till – möjligheten att trycka en fast upplaga av samma publikation.

Denna definition rymmer ett antal scenarier, där man i det yttersta fallet överhuvudtaget inte trycker någon fast upplaga, utan enbart styckvis tryckning när efterfrågan uppstår. I ett mer realistiskt fall är dock denna form av utgivning ett komplement till traditionell tryckning av en fast upplaga, som då naturligtvis initialt kan vara mindre än i normalfallet.

Print on demand kan användas, och används, både när det gäller nytryck av äldre, sedan tidigare publicerat material och när det gäller nyutgivna publikationer.

Särskilda tjänster för att kunna erbjuda tryckning av enstaka exemplar av äldre, digitaliserade verk som inte längre faller under upphovsrätten har vuxit fram under senare år. Ofta som en direkt följd av andra projekt som syftar till att digitalisera stora mängder (äldre) litteratur.

Valet av San Francisco som mål för denna studieresa grundades främst på det faktum att några av de största och viktigaste aktörerna i världen är hemmahörande här. Både när det gäller den historiska liksom den framtida utvecklingen inom digital vetenskaplig publicering, inklusive lagring och tillgängliggörande av densamma.

Utöver print on demand ville vi också utnyttja tillfället att titta på olika aspekter av publiceringsverksamheten som helhet, liksom det pågående arbetet inom angränsande verksamheter som digitalisering och vetenskaplig kommunikation.

De vi besökte var UC Press, California Digital Library, Stanford University Library, Public Library of Science (PLoS) och Internet Archive.

Dessa institutioner, liksom de ansvariga vid respektive institution som vi valde att träffa, valdes för att få en så heltäckande bild som möjligt i förhållande till geografiska och tidsmässiga avgränsningar. Bland institutionerna hittar man såväl förlag (UC Press, PLoS), som forskningsbibliotek (California Digital Library, Stanford University library), samt en av de stora pionjärverksamheterna inom detta stora området (Internet Archive).

Den akademiska biblioteksverksamheten i USA präglas idag i hög grad av digitalisering och hantering av digitaliserade objekt, såsom lagring, beskrivning och tillgängliggörande.

I San Francisco finns också två institutioner som i hög grad har påskyndat – och även stakat ut vägen för – denna utveckling; Google, som genom sitt bokpartnerprogram sedan 2004 har digitaliserat mer än 15 miljoner böcker från, i första hand, nordamerikanska universitetsbibliotek, samt Internet Archive, som med det ursprungliga syftet att spara och bevara webbsidor, sedan ett antal år bedriver egen digitaliseringsverksamhet, i samarbete med bibliotek i USA och internationellt, och idag förfogar över ett arkiv bestående av många miljoner böcker, filmer, ljudfiler, programvaror och webbsidor.

---

<sup>1</sup> Ett pilotprojekt, med namnet "Print on demand vid Göteborgs Universitet" genomfördes under 2007/08 av universitetsbiblioteket, på uppdrag av Nämnden för Acta Universitatis Gothoburgensis.

Dessa två institutioner är olika till naturen, men har båda direkt medverkat till att skapa de växande digitala arkiv som de stora amerikanska universitetsbiblioteken idag förfogar över.

Som en direkt följd av denna digitaliseringsverksamhet har även projekt som HathiTrust uppstått.

Den kritik som tidigare har hörts när det gäller kvalitén på, framför allt Googles, digitaliseringar tonas allt mer ner, och man menar generellt att den idag är fullt acceptabel.

Diskussioner om datasäkerhet, liksom om format och deras beständighet, förs även i USA. Men vi har också mött ett tämligen genomgående pragmatiskt förhållningssätt och en pragmatisk inställning i dessa frågor, där vikten av att snabbt komma igång med och genomföra projekt ofta prioriteras före dessa diskussioner. Detta har naturligtvis bidragit till att arkiven har kunnat växa snabbt på förhållandevis kort tid.

UCPubS, vid University of California, skapades på grund av i stort sett samma problematik som de flesta universitet med publiceringsverksamhet upplever - behov av samordning, stöd till publicerande institutioner, gemensamma avtal (för bättre villkor), utnyttja kompetens på bästa sätt, samt bättre marknadsföring och större synlighet.

UCPubS arbetar utifrån en tredelad modell, där de enskilda skolorna ansvarar för vad som skall publiceras, innehåll, redigering och framställning av tryckfärdig fil. Universitetsförlaget, UCPress, ansvarar för utgivning, tryckning – inklusive print on demand (via externa leverantörer) - distribution och marknadsföring, men fungerar även som ett stöd åt skolorna, med praktiska råd och tips i processen.

California Digital Library, slutligen, ansvarar för det digitala bevarandet och – i förekommande fall – tillgängliggörandet open access, via plattformen eScholarship. Man utvecklar även verktyg för både författandet och peer review-processen.

Det långtgående syftet med UCPubS är att optimera publiceringsprocessen och, framför allt, ge skolorna möjlighet att fokusera på publikationernas innehåll, snarare än på de administrativa processerna. Erfarenheterna hittills är mycket goda.

Ett annat verktyg som man tror skall kunna bli ett stort stöd för den akademiska publiceringen är Open Monograph Systems (OMS), som utvecklas vid Stanford University Library.

Vad print on demand-tjänster beträffar så finns lite olika slutsatser att dra. Från förlagshåll (främst UC Press) ser man print on demand som en önskvärd utveckling. Som förlag vill man inte trycka fler exemplar än nödvändigt, och fungerande tjänster av den här typen garanterar tillgången till tryckta exemplar utan att dessa behöver lagerhållas av förlaget.

När det gäller möjligheten att beställa tryckta versioner av dokument som redan finns fritt tillgängliga digitalt så menar många att behovet av tryckta utgåvor av hög kvalitet - av såväl artiklar som böcker - kommer att kvarstå under överskådlig framtid. Dock går åsikterna isär om hur stort detta behov faktiskt är och kommer att vara.

Vi har naturligtvis också stött på den allmänna synpunkten att den tryckta boken är på väg bort från marknaden, kanske snabbare än man vanligtvis tror. Liksom en skepsis när det gäller tjänster som erbjuder möjligheten att beställa och köpa tryckta exemplar av publikationer som redan finns fritt tillgängliga i digital form.

Gemensamt för alla tjänster som vi har tittat på är att detta sköts av en tredje part, en s.k. "print on demand vendor". Denna sköter normalt såväl tryck som distribution och försäljning,

och kan i vissa fall vara samma företag som sköter motsvarande när det gäller den reguljära utgivningen.

Tilläggs här skall att man parallellt med print on demand också betonar vikten av digital tillgänglighet, i form av e-böcker, gärna mobilanpassade sådana, liksom goda avtal med stora nätbokhandlar som Amazon och Barnes & Noble, för både tillgänglighet och synlighet. Flera av de vi har besökt publicerar redan idag via dessa bokhandlar och/eller i anpassade format, såsom Kindle och EPUB och samtliga betonar vikten av att kunna tillhandahålla publikationer digitalt. Man ser också att efterfrågan på versioner i dessa format ökar kontinuerligt.



*University of California, Berkeley*

## Studiebesöken

Under studieresan besöktes, i tur och ordning, University of California Press, California Digital library, Public Library of Science, Stanford University och Internet Archive. Ett spontant studiebesök gjordes även hos Google.

### 1. University of California Press (UC Press)

UC Press är University of Californias universitetsförlag. Man grundades redan 1893 och är idag ett av de största universitetsförlagen i USA. Förlaget samordnar och administrerar merparten av de monografier och tidskrifter som publiceras vid universitet, med en årlig utgivning på ca 200 boktitlar och 40 löpande tidskriftstitlar. Hela katalogen omfattar ca 4,000 titlar, varav ca 2,000 är tillgängliga online. Huvudkontoret, som vi besökte, ligger i Berkeley, utanför San Francisco, men man har även kontor i New Jersey och England.

Här träffade vi **Laura Cerutti**, Director of Digital Content Development, som är den person på UC Press som har det yttersta ansvaret för digital publicering och anpassning av traditionell publicering i en digital värld. Hon är också ansvarig för förlaget del i samarbetet kring UCPubS (mer om detta nedan).

All tryckning sker via externa leverantörer. När det gäller print on demand så samarbetar med i dagsläget med en enda leverantör. Detta underlättar mycket, då man enklare kan komma överens om format och kommunicera eventuella problem. Det är också ur avtalssynpunkt en bra lösning.

Man ser också behovet av att ha bra avtal med Amazon och andra nätbokhandlar.

Bokhandlarna är, enligt Cerutti, viktigare än förlagen för slutkunden och boktitlar som inte är tillgängliga från nätbokhandlarna är också till stor del osynliga för kunderna.

I detta sammanhang nämns även digital utgivning och vikten av att kunna tillhandahålla nya böcker i format som kan läsas på eBokplattformar, såsom Kindle och ePUB.

Som exempel nämner hon försäljningen via Amazon Kindle, som under 2009 motsvarade 3% av den förlagets totala försäljning. 2008 var motsvarande siffra 1% och det är rimligt att tro att behovet kommer att öka kraftigt de närmaste åren, i takt med att utvecklingen och försäljningen av nya plattformar, som iPad och liknande, växer närmast explosionsartat.

Vi får, av olika skäl, inte veta exakt hur UCPress avtal med Amazon ser ut, men i korthet omfattar det ett rabattsystem, där Amazon får 30% av intäkterna från försäljningen, mot att man exponerar, marknadsför och säljer verken i både tryck och digital form i sin webbshop.

Nuvarande print on demand-tjänst är, ur kundens synvinkel, ett sömlöst system, där beställningar via förlagets webbplats automatiskt vidarebefordras, tillsammans med tryckfil, till leverantören som både trycker och distribuerar slutprodukten. Tjänsten omsätter ca \$200,000 per år.

Ofta ställs förlaget inför ställningstagandet huruvida man skall tillhandahålla vissa titlar print on demand eller trycka en helt ny upplaga. Detta är en ekonomisk fråga och det handlar om att uppskatta det faktiska behovet, vilket naturligtvis inte alltid är lätt.



*Laura Cerutti, UC Press, Berkeley*

Cerutti skulle gärna se en verksamhet och en utgivning som är helt baserad på print on demand och digital utgivning, då hon ser detta som en en garant mot marknadens oberäknlighet. Man vet sällan på förhand hur många exemplar av en viss titel som kommer att säljas. Hon betonar även att UC Press, liksom andra förlag, inte vill vara en lagerverksamhet.

## **UCPubS**

UCPubS är ett samarbetsprojekt som påbörjades 2008, på grund av i stort sett samma problematik som de flesta universitet med publiceringsverksamhet upplever - behov av samordning, stöd till publicerande institutioner, gemensamma avtal (för bättre villkor), utnyttja kompetens på bästa sätt, samt bättre marknadsföring och större synlighet.

UCPubS arbetar utifrån en tredelad modell, där de enskilda skolorna ansvarar för vad som skall publiceras, innehåll, redigering och framställning av tryckfärdig fil.

UCPress, ansvarar här för utgivning, tryckning – inklusive print on demand (via externa leverantörer) - distribution och marknadsföring, men fungerar även som ett stöd åt skolorna, med praktiska råd och tips i processen. Förlaget kan även förmedla kontakter med annan expertis utanför universitetet, exempelvis inom upphovsrätt, design och korrekturläsning. Man tillhandahåller även riktlinjer för publicering, i praktiken en komprimerad version av Chicago manual of Style<sup>2</sup>, liksom mallar för InDesign och andra verktyg. Man utför också slutkontroll av tryckfil.

För närvarande planeras upprättandet av ett virtuellt community för att bättre kunna kommunicera problem och synpunkter, samt för att ge inblandade parter möjlighet att enklare kommunicera med varandra och dra nytta av varandras kunskaper och erfarenheter.

---

<sup>2</sup> **The Chicago Manual of Style** är en handbok och en samling vedertagna riktlinjer för publicering. Den första upplagan kom redan 1906 och berörde främst typografiska regler. Den 16:e och senaste upplagan (2010) innehåller även riktlinjer för digital publicering. Den finns numer tillgänglig som prenumeration online - <http://www.chicagomanualofstyle.org>

California Digital Library, slutligen, ansvarar för det digitala bevarandet och – i förekommande fall – tillgängliggörandet open access, via plattformen eScholarship (se nedan). Man utvecklar även verktyg för både författandet och peer review-processen.

Det långtgående syftet med UCPubS är att optimera publiceringsprocessen och, framför allt, ge skolorna möjlighet att fokusera på publikationernas innehåll, snarare än på de administrativa processerna.

Under 2009 publicerades ca 60 boktitlar inom ramen för detta samarbete.

## 2. California Digital Library

California Digital Library (CDL) är en enhet inom University of California, med syfte att stödja universitetets samtliga enheter med olika it-tjänster. Inte helt olik den roll som DigIT har vid Göteborgs universitetsbibliotek, fast givetvis i betydligt större skala.

Detta omfattar såväl verksamheter som hantering av licensierade informationsresurser och drift av samkatalogen Melvyl, som digitalisering och publicering. Det var de senaste verksamheterna som vi fokuserade på vid vårt besök.

Personerna vi träffade här var **Heather Christenson**, Project Manager, Mass Digitization, **Cathrine Mitchell**, Director, Publishing & UCPubS co-ordinator, samt **Paul Fogel**, Technical Lead for Hathi Trust and Mass Digitization.

Sedan 2002 driver och utvecklar CDL *eScholarship*, som är University of Californias gemensamma plattform för open access-publicering. Från början var detta en extern tjänst, men använder nu det egenutvecklade XTF.

*XTF (eXtensible Text Framework)* är ett ramverk för publicering, baserat på XML/XSLT och java, utvecklat vid CDL och tillgängligt open source.

Utöver eScholarship så används XTF av ett antal olika publiceringssystem runt om i världen.



The screenshot shows the eScholarship website interface. At the top, there is a navigation menu with links for Home, About, Browse, Publish, Help, and My Items (0). A search bar is located on the right. The main content area displays a book listing for "Diaspora without Homeland: Being Korean in Japan", edited by Sonia Ryang and John Lie. The book cover is visible on the left, and the title and editors are on the right. A table of contents is shown on the left side of the page, listing five chapters. The page footer includes the text "eScholarship is powered by the California Digital Library with The Berkeley Electronic Press."

Exempel på publikation i eScholarship.



eScholarship används för lagring och tillgängliggörande av publikationer som publiceras inom ramen för UCPubS. Systemet innehåller även en funktion för att möjliggöra beställning av tryckta exemplar. Dessa beställningar vidarebefordras med automatik till UCPress, som i sin tur skickar tryckfil till leverantör av print on demand. Själva tryckfilen kan därefter, vid behov, lagras i eScholarship.

CDL har inga egna print on demand-tjänster.



*Cathrin Mitchell, Paul Fogel & Heather Christenson, California Digital Library, Oakland.*

CDL samordnar även digitaliseringsverksamheten vid University of California – varav en stor del utförs av Google och/eller Internet Archive, och man är också starkt engagerade i HathiTrust.<sup>3</sup>

Vi fick även en grundlig genomgång av detta mycket intressanta projekt.

### **3. Public Library of Science (PLOS)**

The Public Library of Science (PLOS) är ett open access-förlag som, sedan 2007, publicerar ett antal tidskrifter inom medicin och naturvetenskap. Första tidskriften, PLOS Biology lanserades oktober 2003. Allt innehåll är peer-reviewed och publicerat under Creative Commons licens.

Utgivningen finansieras till största delen genom författaravgifter och medlemskap, till mindre del genom fonder och donationer.

---

<sup>3</sup> **HathiTrust** är ett projekt som initierades av University of Michigan, 2008, med syftet att skapa ett nationellt, sökbart arkiv av digitaliserade objekt, primärt böcker. Arkivet består till stora delar av digitaliserade titlar från Googles och Internet Archives digitaliseringsprojekt, men även material som digitaliserats lokalt. Alla kan söka i arkivet, men titlar som inte är public domain är bara tillgängliga i fulltext för användare vid institutioner som "äger" boken. I dagsläget innehåller arkivet drygt 7 miljoner titlar, tillhörande ett 40-tal deltagande bibliotek. Såväl arkivet som antalet deltagande bibliotek utökas kontinuerligt och man välkomnar nya medlemmar, även utanför USA och Canada.

Huvudkontoret ligger i centrala San Francisco och här träffade vi representanter för olika delar av förlagets verksamhet.

**Rick Cave**, IT-chef och **Eddie Dickey**, Software Development Manager, visade Ambra – det egenutvecklade system som idag används för publicering av samtliga tidskrifter. Vi tittade även på Article Level Metrics, ALM, ett egenutvecklad statistikverktyg som läser och bearbetar statistik ur webbloggar. Båda dessa programvaror finns tillgängliga open source.

Förutom publiceringen på webben så har man även utvecklat applikationer för bland annat iPhone och iPad. Dessa är dock i dagsläget fortfarande på experimentstadiet.

**Liz Allen**, Director of Marketing and Business Development visade förlagets nuvarande print on demand-tjänst.

Man använder här en extern tjänst, EzReprint, som tillhandahålls av Odyssey Press, ett företag som ligger nära geografiskt och dessutom har en väl uppbyggd kompetens inom tryckning av vetenskapliga publikationer.

Systemet erbjuder beställaren en enkel och smidig rutin för beställning, som går direkt till tryckeriet som trycker levererar högkvalitativa artikelkopior enligt ett antal olika villkor som beställaren har angett. Det totala priset beräknas i realtid, baserat på antal sidor och exemplar, kvalitet, leveransvillkor och leveransadress (man levererar till hela världen).

Den minsta tillåtna beställningen är i dagsläget 25 exemplar, och vid mycket stora beställningar gäller särskilda villkor.

På frågan om huruvida det verkligen existerar ett behov av att kunna beställa tryckta kopior av artiklar som ändå finns fritt tillgängliga för nerladdning svarar Allen att det finns, och kommer att finnas, behov av tryckta artiklar av hög kvalitet. Man ser också att antalet beställningar snarare ökar än minskar.

Tjänsten vänder sig sannolikt till institutioner, snarare än till privatpersoner. Minimiantalet vid beställning motiveras med att tryckning av färre än 25 ex. inte skulle vara ekonomiskt försvarbart.

**Donna Okubo** talade slutligen lite om PLoS målsättning och ambitioner. Man ser naturligtvis den snabba utvecklingen inom open access som positiv, och välkomnar fler medlemmar.

Göteborgs universitet är för närvarande inte medlem i PLoS. Dock förekommer viss publicering i förlagets tidskrifter. PLoS är ett av två förlag som diskuteras i samband med en eventuell kommande, nationell upphandling av medlemskap hos förlag för open access-publicering i Sverige (det andra är BioMed Central).

I samband med besöket berörde vi möjligheten att, per automatik, kunna få artiklar av universitets forskare, som publicerats i någon av förlagets tidskrifter, deponerade i GUPEA. En sådan tjänst tillhandahålls idag av BioMed Central, men inte av PLoS. Donna Okubo såg dock positivt på att även PLoS skulle kunna tillhandahålla detta i framtiden.

#### 4. Stanford University

Vid Stanford träffade vi **Juan Pablo Alperin**, PhD vid Stanford Univ. samt lead developer/manager för OMP-projektet.

*Open Monograph Press (OMP)* är ett system för (open access) publicering av monografier,

med stöd för hela det redaktionella flödet. Det utnyttjar samma ramverk som Open Journals System (OJS) - som Göteborgs UB sedan några år använder för tidskriftspubliserig – liksom Open Conference System (OCS), som också har använts vid universitetet.

Dessa system utvecklas inom ramen för The Public Knowledge Project (PKP) - ett samarbetsprojekt som syftar till att utveckla verktyg för vetenskaplig publicering och kommunikation. Projektet leds av the University of British Columbia, Vancouver, med Stanford University som aktiv part.

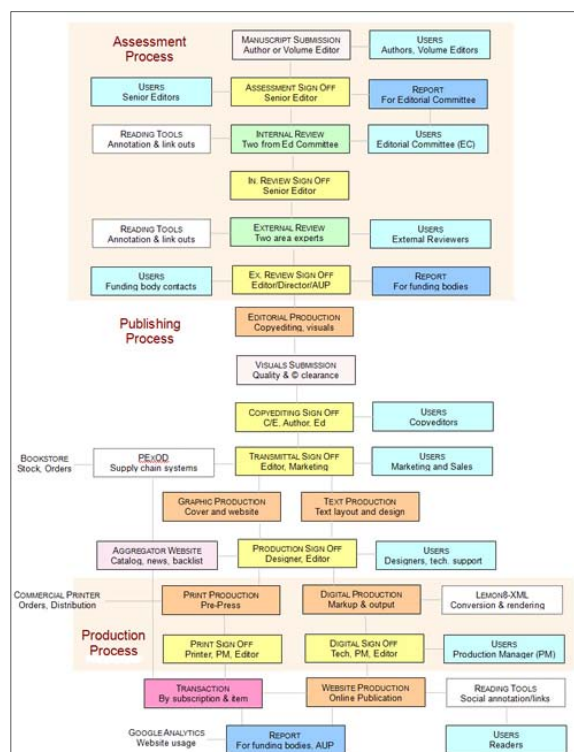
Samtliga programvaror som utvecklas tillhandahålls open source.

Den första produktionsversionen av OMP är beräknad till januari 2011.

Till skillnad från tidigare versioner av PKP:s program är OMP uppbyggt i moduler som i framtiden skall gå att flytta runt mellan de olika publiceringssystemen.

Lemon8, ett sedan länge efterfrågat verktyg för omvandling av PDF-filer, MS Word och andra format till XML kommer också att finnas som en modul till alla systemen.

Arbetsflödet i OMP bygger på att manuskript laddas upp i systemen, där olika tjänster hanterar peer-review, publicering, copywriting och andra flöden på en redaktion (se bild nedan).



Det nya OMP kommer att ha ett uppdaterat gränssnitt, som senare skall föras över till de andra systemen. Dessutom kommer det att finnas möjlighet att skapa egna roller och arbetsflöden för dessa inom systemet, om man har egna roller i ett publikationsflöde. Det går att logga in för korrektur, layout, etc och gränssnittet anpassar sig efter detta och visar de val/funktioner som är relevanta för denna typ roll.

OMP har, till skillnad från de andra systemen, stöd för Marketing Strategies.

OMP kommer med MODS och ONIX metadata  
andra format med egna xsl:er kan laddas inifrån systemet from OMP - kommer också att introduceras i de andra programmen.

Det grundläggande formatet i OMP är XML. Detta innehåller både själva texten + metadata och kan skickas i direkt till tryckeriet. Ingen konvertering till annat format behövs.  
Det finns mallar för formgivning av böcker, men inom dessa ramar är det fritt att skapa ett unikt utseende. (Till skillnad mot vanliga automatiskt genererade e-covers)

Ett visst stöd för onlinebetalningar finns, om man tänker sig att använda OMP i samband med en lokal hantering av distributionen och/eller print on demand.



*Juan Pablo Alperin, PKP & Stanford University.*

## 5. Internet Archive

Internet Archive bildades 1996 i San Francisco, och är en icke-kommersiell organisation, med syfte att arkivera och bevara ”allt” som finns tillgängligt och publicerat i digital form, inklusive texter, bilder, ljud, film, mjukvara och webbsidor. Det är en ingenjör driven organisation där tekniken står i centrum.

Arkivets mest kända tjänst är förmodligen *The Wayback Machine*) – en söktjänst där man kan söka och visa historiska versioner av webbsidor, baserat på URL.

En annan tjänst är *Archive-It*, som är en plattform där olika institutioner – inklusive ett stort antal bibliotek – kan arkivera sina digitala samlingar och på så sätt få dessa både bevarade, indexerade och sökbara och presenterade i ett och samma gränssnitt.

Arkivets verksamhet bygger på den tidiga insikten om att den snabba framväxten och utvecklingen av internetbaserade tjänster och den snabbt ökande mängden av digitalt

publicerat material kan leda till att viktig information och kunskap kan gå förlorad, då traditionella bevarare av detta inte är rustade för att hantera den typen av material. På arkivets hemsida kan man bl a läsa att "without cultural artifacts, civilization has no memory and no mechanism to learn from its successes and failures. And paradoxically, with the explosion of the Internet, we live in what Danny Hillis has referred to as our "digital dark age.""

Arkivets främsta inkomstkälla är digitalisering av objekt åt andra. Kostnaden för detta är normalt 10 cent/sida. Bland nuvarande samarbetspartners finns många av de stora nordamerikanska forskningsbiblioteken, såsom universitetsbiblioteken vid Yale, Princeton, UCLA och Boston University, men man har även en del stora internationella samarbeten, exempelvis med Bibliotheca Alexandrina.

Sedan ett år tillbaka har arkivet sin primära hemvist i en renoverad, och något ombyggd före detta kyrka i the Presidio, i nordvästra San Francisco.



*Internet Archives lokaler i the Presidio.*

Här träffade vi **Brewster Kahle**, Digital Librarian och en Internet Archives grundare, **Robert Miller**, Director of books och ansvarig för digitaliseringsverksamheten, samt **Ginger Bisharat**, Executive assistant. Vi fick även möjlighet att delta vid den veckovis återkommande kombinerade lunchen och personalmötet.

Brewster Kahle demonstrerade arkivets lagringskapacitet. Allt data lagras i stora diskpackar, s.k. PetaBoxes, där varje box kan lagra 1 Petabyte (PB) data, vilket motsvarar 1,000 TB, eller 1 miljon GB. Totalt använder arkivet ca 8 PB för lagring, varav webbarkivet (The Wayback Machine) disponerar hälften. Utöver detta finns ett antal kopior av hela arkivet, samt en exakt spegling, placerad i en annan lokal, utanför San Francisco. Den sistnämnda ersätter i dagsläget såväl kontinuerlig backup, som eventuella RAID- och UPS-system, då man anser

att speglingen i sig ger den säkerhet man behöver.



*En av Internet Archives s.k. PetaBoxes.*

Utöver konvertering till olika format utförs i praktiken ingen rutinmässig efterbehandling av de inskannade bilderna, de digitaliserade objekten som tillgängliggörs via deras olika söktjänster, kan sägas komma direkt från skannern. Skanningen sker via egenutvecklade maskiner (liknande den bokvaggan som används vid universitetsbiblioteket) och äger rum i en speciell lokal där en mängd maskiner är uppradade efter varandra. En del av maskinerna står i små "bås" klädda med svarta textilier, men själva rummet är varken specialmålat eller skyddat från solljus utifrån.

Robert Miller talade om digitaliseringsverksamheten och syftet med denna. Han demonstrerade även arkivets söktjänster.

Böcker skannas, bearbetas med OCR (Abbyy FineReader) och sparas, tillsammans med metadata (MARC XML), i JPEG2000, som är det format man har valt. Fokus ligger inte primärt på långtidslagring, utan lagringsformatet är snarare valt utifrån vad man i dagsläget bedömer vara det bästa. Man menar att man relativt snabbt kan konvertera till andra format om och när detta skulle bli aktuellt.


Utifrån den inskannade förlagan skapas olika versioner för presentation och nerladdning, inklusive PDF, EPUB, text (OCR), HTML, Kindle, DeJaVu och Daisy.

Gränssnittet innehåller även en inbyggd läsare, Internet Archive BookReader, för läsning online. Detta är en open source-lösning, byggd i PHP och javascript, med många användbara funktioner, som exempelvis full bläddring, zoom, olika vyer, utskriftsfunktion och fulltextsökning

Läsaren kan bäddas in på annan sida, vanligtvis i en iFrame, men programvaran kan också användas utanför Internet Archive, dvs för lokala samlingar på den egna webbplatsen. Detta kräver lite modifiering av källkoden, samt en plattform som stöder PHP och att bilderna lagras i en kompatibel struktur (och möjligen också i ett kompatibelt format).

Ebook and Texts Archive > University of Toronto - Robarts Library > Göteborg; skisserade skildringar af Sveriges andra stad i våra dagar, jämte en återblick på dess minnen, för saväl turister som hemmavarande, efter tryckta och otryckta källor. Med en karta öfver Göteborg

### View the book



[Read Online](#) (~304 pg)  
[PDF](#) (17.6 M)  
[B/W PDF](#) (15.7 M)  
[EPUB](#) (~304 pg)  
[Kindle](#) (~304 pg)  
[Daisy](#) (~304 pg)  
[Full Text](#) (653.0 K)  
[Divu](#) (9.7 M)

All Files: [HTTP](#)


[Help reading texts](#)

### Resources

[Bookmark](#)

### Göteborg; skisserade skildringar af Sveriges andra stad i våra dagar, jämte en återblick på dess minnen, för saväl turister som hemmavarande, efter tryckta och otryckta källor. Med en karta öfver Göteborg (1884)

**Author:** [Cederblad, Albert](#)  
**Subject:** [Gothenburg, Sweden](#)  
**Publisher:** [Göteborg D.F. Bonniers](#)  
**Language:** [Swedish](#)  
**Call number:** AHS-2526  
**Digitizing sponsor:** [University of Toronto](#)  
**Book contributor:** [Robarts - University of Toronto](#)  
**Collection:** [robarts](#); [toronto](#)

 This book has an [editable web page](#) on [Open Library](#).

#### Description

52

#### Reviews

[Be the first to write a review](#)  
Downloaded 6 times

#### Selected metadata

**Page-progression:** lr  
**Scanningcenter:** uoft  
**Mediatype:** texts  
**Identifier:** gteborgskisser00cede  
**Ppi:** 400  
**Camera:** Canon 5D  
**Operator:** scanner-carrie-lounsbury@...

Exempel på publikation i Internet Archive Text Archive. Den digitaliserade boken tillgängliggörs i det här fallet i en mängd olika format och kan även läsas online med full möjlighet till sökning och navigation i texten. Möjligheten till print on demand är här inaktiverad.

En print on demand-tjänst för arkivets material finns och är inkorporerad i den vanliga söktjänsten. Den är dock i nuläget inaktiv. Robert Miller menar att denna tjänst har visat sig vara en besvikelse, då den inte använts så mycket som man, inklusive de deltagande biblioteken, ursprungligen hade räknat med. Man använder idag ett externt tryckeri som får en beställning när någon använder funktionen. Man trycker då en kopia av den aktuella boken från PDF-förlagan, och levererar till beställaren.

Samtliga fritt tillgängliga titlar är också direkt tillgängliga för användare av Espresso Book machine<sup>4</sup>, via ett särskilt gränssnitt.

Robert Miller tror att print on demand för material som är fritt tillgängligt digitalt, i framtiden enbart kommer att användas i ett fåtal undantagsfall, exempelvis när en bok skall ges som gåva, eller användas i något annat speciellt sammanhang. I övrigt så tillhör framtiden, enligt honom, de digitala plattformarna.

Man samarbetar också med organisationer som ideellt bidrar med kompletterande information eller som utför kostnadsfria digitaliseringar åt arkivet.

Man samarbetar med bibliotek när det gäller digitaliserat men upphovsrättsskyddat. Denna problematik löser man genom att hantera de digitala filerna precis som biblioteket hanterar

<sup>4</sup> Espresso Book Machine (<http://www.ondemandbooks.com>) är en portabel lösning för att trycka, skära och binda böcker (paperback) från PDF. Den senaste versionen kan färdigställa en bok på upp till 830 sidor, inklusive omslag i färg, på mindre än 10 minuter. Den har också en programvara – EspressoNet – för att kunna hantera nätbaserade titlar från ett antal samarbetspartners, inklusive Internet Archive. Inköpskostnaden för en EBM är i dagsläget ca \$150,000 och det finns f.n. ett 40-tal maskiner i drift. Robert Miller nämner University of Alberta Bookstore (Edmonton, Canada) som exempel på en användare som utnyttjar denna teknik mycket, men är osäker på användningen i övrigt.

fysiska böcker: när en digital kopia är utlånad, blir den otillgänglig för en annan låntagare tills lånetiden gått ut. Fyra bibliotek använder sig av denna metod idag. De håller också på att utveckla metoder så att man som låntagare ska kunna dela mindre delar som t ex specifika kapitel av en lånat digitalt objekt med andra.

Merparten av det arkiverade materialet är av det slag som på grund av materialets ålder faller utanför gällande upphovsrätt, men man har även börjat digitalisera upphovsrättsligt skyddade verk. Även här har man en pragmatisk inställning, och den grundläggande principen är att man menar att en mycket stor del av det skyddade materialet förmodligen i praktiken kan digitaliseras och tillgängliggöras utan problem. Och i de fall där upphovsrättsinnehavare har invändningar kan man enkelt plocka bort det aktuella materialet från söktjänsten.

Slutintycket av Internet Archive är en progressiv organisation, som har betytt, och fortfarande betyder mycket för utvecklingen inom våra verksamhetsområden. Man har ett genomgående pragmatiskt förhållningssätt i många, i grunden svåra, frågor, och man utvecklar goda tekniska lösningar, vilket gör att verksamheten lever och stora mängder kunskap kan bevaras och tillgängliggöras på ett sätt som är både effektivt, användarvänligt och ekonomiskt.



## Länkar till mer information

Amazon Kindle - [http://en.wikipedia.org/wiki/Amazon\\_Kindle](http://en.wikipedia.org/wiki/Amazon_Kindle)

Archive-It - <http://www.archive-it.org>

Barnes&Noble eBooks - <http://www.barnesandnoble.com/ebooks/index.asp>

California Digital Library - <http://www.cdlib.org>

Chicago Manual of Style - <http://www.chicagomanualofstyle.org>

eScholarship - <http://escholarship.org>

EPUB - <http://en.wikipedia.org/wiki/EPUB>

Espresso Book Machine - <http://www.ondemandbooks.com>

eXtensible Text Format (XTF) - <http://www.cdlib.org/services/publishing/tools/xtf>

Hathi Trust - <http://www.hathitrust.org>

Internet Archive – <http://archive.org>

Internet Archive Text Archive - <http://www.archive.org/details/texts>

Odyssey Press - <http://www.odysseypress.com/>

Open Monograph Press (OMP) - <http://pkp.sfu.ca/omp>

Peta boxes - <http://www.archive.org/web/petabox.php>

PKP foundation - <http://pkp.sfu.ca>

Public Library of Science (PLoS) – <http://plos.org>

PLoS EzReprint - <http://www.plos.org/journals/print.php>

UC Press - <http://www.ucpress.edu>

UCPubS - <http://www.ucpress.edu/partners.php?p=ucpubs>

University of Alberta Bookstore -  
<http://www.bookstore.ualberta.ca/index.cfm?index=STATIC/ESPRESSOBOOKS>

The Wayback Machine - <http://www.archive.org/web/web.php>