

<i>Skapat:</i>	<i>Skriven av:</i>	
080207	David Hansson, Lennart Stark och Gunilla Wiberg	

# Digitalisering av kulturarvet

## Projektplan

---

---

### Sammanfattning

Dokumentet beskriver projektarbetet för kursen Digitalisering av kulturarvet, 15 högskolepoäng vid Högskolan i Borås (distanskurs under VT 08). Dokumentet innehåller beskrivning av projektarbetets omfattning och utförande. Observera att dokumentet kan underkastas revision under projektarbetets gång. Detta styrs med versionsnummer. Version 1.0 är första utgåvan av denna projektplan.

## Innehåll

1	Bakgrund .....	3
2	Syfte och mål .....	3
2.1	Mål .....	3
2.2	Målsättningar .....	3
2.2.1	Målgruppsdefinition .....	3
3	Projektidé .....	3
3.1	Produkter .....	4
3.1.1	Webb .....	4
3.1.2	Bild .....	4
3.1.3	E-text .....	4
4	Genomförande .....	4
4.1	Material .....	4
4.1.1	Beskrivning .....	4
4.1.2	Placering, befintlighet, skick och läsbarhet .....	5
4.1.3	Upphovsrättsläget .....	5
4.2	Projektorganisation .....	6
4.2.1	Kommunikation .....	6
4.2.2	Tidplan .....	6
4.2.3	Aktiviteter .....	6
4.2.4	Ansvarsplan .....	6
5	Riskanalys .....	7
5.1	Kunskaper utanför kursen .....	7

## 1 Bakgrund

Projektet löper under VT-08 med start 080125 och avslut 080823. Gruppens medlemmar studerar vid Borås Högskola på distans. Gruppmedlemmar är David Hansson, Lennart Stark och Gunilla Wiberg. I gruppform kommer ett faktiskt digitaliseringsprojekt att simuleras. Handledare för kursen "Digitalisering av kulturarvet" är Jan Buse, Mats Dahlström och Mikael Gunnarsson.

## 2 Syfte och mål

Det generella syftet med projektet är att vinna erfarenhet av det totala digitaliseringsflödet från urval till publicering.

### 2.1 Mål

Vårt mål med projektet är att sätta upp en miljö/system skalbart för storskalig digitalisering av kulturarvet samt utforska uppmärkning av text och finna en realistisk nivå för storskalighet.

### 2.2 Målsättningar

Projektets målsättningar utifrån kursdefinition är:

Att bildkällfilerna skall ha en faksimilversion (i pyramidtiff) som skapas och tillgängliggörs via webben genom användning av eRez presentationsverktyg.

Att textkällfil (TEI) skall användas för presentation av minst två olika textversioner passande målgrupp.

Projektets slutliga textfiler skall vara validerade.

Webbplatsen skall vara funktionell och användarvänlig.

Det digitala materialet (både TEI-fil och XHTML-filer) skall förses med dolda och synliga metadata. Adekvata uppgifter om det analoga originalmaterialet skall tillhandahållas, åtminstone i en TEI Header.

Källfiler skall göras tillgängliga för kursens deltagare på projektgruppens webbplats.

#### 2.2.1 Målgruppsdefinition

Tre tänkbara målgrupper

- Barn, förskolor, grundskolor och bibliotek.
- Den akademiska världen; t ex historiker och konstvetare.
- Allmänheten.

## 3 Projektidé

Tanken är att tillgängliggöra ett urval ur känsliga samlingar digitalt via Internet för att i största möjliga mån undvika vardagligt slitage och kunna erbjuda nuvarande och kommande

generationer tillgång till materialet. För att möjliggöra detta storskaligt behövs ett "repository" som strukturerar den digitala representationen av materialet.

## 3.1 Produkter

Vi planerar att använda en programvara som heter Fedora (Flexible Extensible Digital Object and Repository Architecture, <http://www.fedora.info/>). Detta är en öppen källkod-programvara för att skapa ett digitalt av en mängd olika typer av digitala objekt. Fedora kan hantera upp till en miljon objekt.<sup>1</sup>

Vi kommer i mån av tid även att utforska möjligheterna för andra programvaror såsom Greenstone och ContentDM.

### 3.1.1 Webb

Adobe Dreamweaver CS3 används för utformning av webbpresentationen, som utformas i enlighet med 24-timmarswebben 2.0<sup>2</sup> Den grafiska bildvisningen sker med den kommersiella programvaran eRez och FSI Viewer.

### 3.1.2 Bild

Redigering av bilder sker i Photoshop CS3. Boken skannas i sin helhet

### 3.1.3 E-text

Boken OCR-tolkas med hjälp av en utvärderingsversion av Abbyy Fine Reader OCR 7.0 Professional Edition och märks upp av projektgruppens medlemmar.

## 4 Genomförande

### 4.1 Material

Materialet är en barnbok, tryckt 1850.

#### 4.1.1 Beskrivning

Materialet som är valt är boken *Den nyttige bilderboken eller Vandringen genom verkstäderna*. Utgiven i Stockholm 1850 av J. C. Hedbom & Co:s förlag. Nedanstående marc-post är ett exemplar från Svenska barnboksinstitutet.

Fält	Indikatorer	Metadata
000		00633cam a22001817a 4500
001		10389573
003		SE-LIBR
005		20070322094657.0
008		070322s1850 sw j 000 0 swe c
040		a Sbi

---

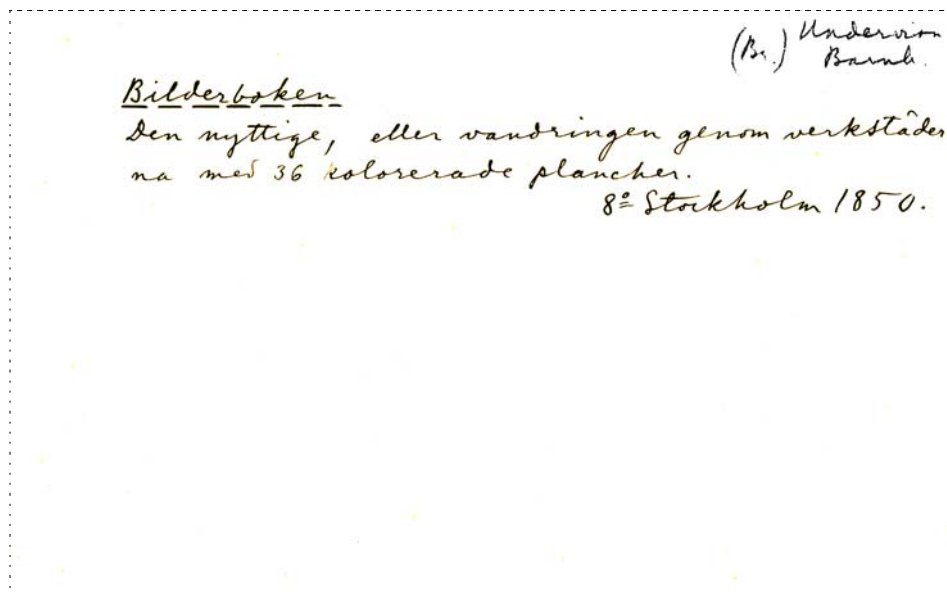
<sup>1</sup> <http://www.searchguide.se/oa/?cat=18>

<sup>2</sup> <http://www.verva.se/verksamhetsstod/webb/v124/>

Fält	Indikatorer	Metadata
245	0 4	a Den nyttige bilderboken eller Vandringen genom verkstäderna/c med 36 kolorerade planser
260		a Stockholm :b J. C. Hedbom,c 1850e (Stockholm :f Joh. Beckman)
300		a 16 s. [9] pl.bl. ;c 17x15 cm
500		a Litorgri, handkolorerad
510	2	a K-nr: 1065
841		5 Sbia x ab 0703224u 8 1001uu 0901128e 4
852		5 Sbib Sbih Peyron refi läsesalslån
852		5 Sbib Sbiz I Sbi marmorerat pappband

## 4.1.2 Placering, befintlighet, skick och läsbarhet

Nedan visas ett katalogkort (Kat-57) från Universitetsbiblioteket i Lund (skannad och retuscherad). Boken är placerad i ett magasin på Universitetsbiblioteket i Lund. Skick och läsbarhet är god.



Omfattningen av boken är 16 textsidor och 9 planschsidor med 36 bilder. Bokens faktiska storlek är 16x13,5 cm.

Boken innehåller en beskrivning av 36 olika yrken beskrivna från fader till hans tre barn.

Boken skick är gott och läsbarheten är tydlig så OCR-tolkning är genomförbart.

## 4.1.3 Upphovsrättsläget

Upphovsmannen till denna barnbok och upphovsmannen till de handkolorerade planscherna är okänd. Eftersom boken är tryckt och utgiven 1850 är upphovsrättsläget gynnsamt. Utredningen redovisas i projektrapporten.

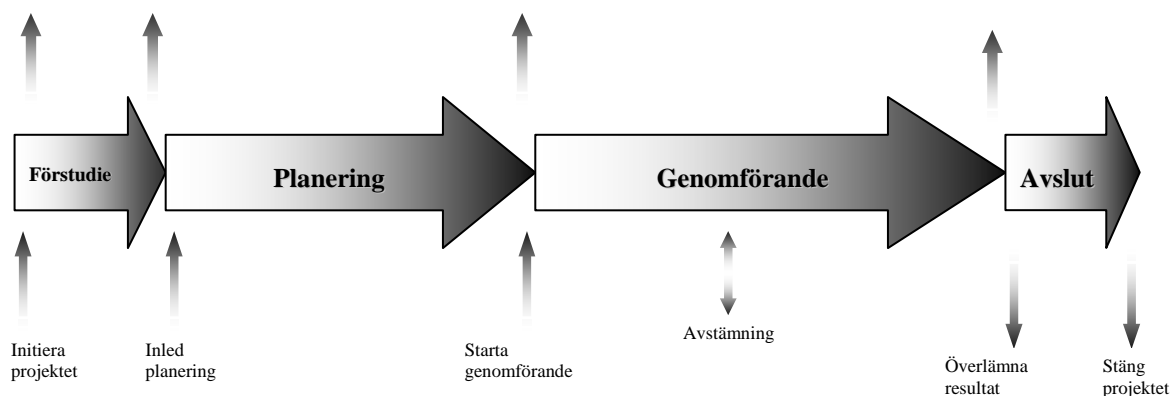
## 4.2 Projektorganisation

Samtliga i gruppen arbetar inom statlig regi och har erfarenhet av hierarki. I detta projekt arbetar vi med en platt organisation där alla har ansvarsområden och på så sätt bitvis agerar projektledare. Samtliga gruppmedlemmar kommer att vara delaktiga i alla moment oberoende av ansvarsområde. Dokumentationen görs löpande genom projektet av samtliga medlemmar. Detta sker genom versionshantering via google groups.

### 4.2.1 Kommunikation

Kommunikation mellan gruppmedlemmar hålls kontinuerligt och informellt. Kontakt via e-post och möten via webb en gång i månaden och vid behov. Gruppen har via google groups en samlingsplats för all gemensam information och dokumentation.

### 4.2.2 Tidplan



### 4.2.3 Aktiviteter

ID	Aktivitet	Varaktighet	Start	Slut	2008-01-21				2008-01-26				2008-02-04				2008-02-11							
					ti	o	f	s	m	ti	o	f	s	m	ti	o	f	s	m	ti	o	f	s	
1	Projektplanering	29 dagar?	on 08-01-23	må 08-03-03	[Gantt bar]																			
2	Upphovsrättsliga frågor	17 dagar?	on 08-01-23	to 08-02-14	[Gantt bar]																			
3	Bildbehandling	49 dagar?	on 08-01-23	må 08-03-31	[Gantt bar]																			
4	Skanning	7 dagar?	on 08-01-23	to 08-01-31	[Gantt bar]																			
5	OCR-tolkning	11 dagar?	må 08-01-28	må 08-02-11	[Gantt bar]																			
6	Repository implementering	28 dagar?	on 08-01-23	fr 08-02-29	[Gantt bar]																			
7	Textbehandling	60 dagar?	må 08-03-10	fr 08-05-30	[Gantt bar]																			
8	Webbpresentation	96 dagar?	må 08-02-18	må 08-06-30	[Gantt bar]																			
9	Dokumentering, löpande	138 dagar?	on 08-01-23	fr 08-08-01	[Gantt bar]																			
10	Publicering	2 dagar?	on 08-07-30	to 08-07-31	[Gantt bar]																			

### 4.2.4 Ansvarsplan

# Digitalisering av kulturarvet

VEM	VAD
David Hansson	Fedora, XML, TEI, dokumentation
Lennart Stark	Bildbehandling, OCR, XML, TEI, dokumentation
Gunilla Wiberg	Webbpresentation, Flash, XML, TEI, dokumentation

## 5 Riskanalys

Händelse som kan påverka projektet negativt	Sannolikhet	Konsekvens	VÄGNING
Bortfall av gruppmedlem	1	8	8
Uppmärkning tar längre tid än beräknat	3	4	12
Repository mer tidskrävande än beräknat	3	8	24
Hårdvaru- resp mjukvaruproblem	1	9	9
Materialet håller ej	1	5	5

Förebyggande åtgärd	Åtgärd	Ansvarig	Deadline
Test av textuppmärkning på tidigt stadium	Begränsad del av boken blir uppmärkt	David	Siste mars
Repository implementeras tidigt	Samtliga involveras	David	Siste februari
Hårdvaru- resp mjukvaruproblem	Kontakter med leverantörer	Gunilla och David	löpande
Materialet håller ej	Val av nytt material	Lennart	Siste februari

### 5.1 Kunskaper utanför kursen

Kursen innehåller inte information angående ”repository”. Detta är ett tillägg för anpassningen till storskalig digitalisering.