

Digitaliseringsprojekt – rapport gällande digitalisering av:

*FOSTERLÄNSK SÅNG
VID
GÖTHA CANALS ÖPNANDE
I
HANS MAJ:T KONUNGENS
OCH
DEN KONGL. FAMILLENS
NÄRVARO,
DEN 26 SEPTEMBER 1832.*

Författare: Marie Gyll, Jessica Hultengren, Susanne Landén och Magdalena af Trolle

Vt 2025



HÖGSKOLAN
I BORÅS

Digitalisering av kulturarvsmaterial som uppgift inom kursen: *Digitalisering för bevarande och tillgänglighet*, 15 hp. Biblioteks- och informationsvetenskap, högskolan i Borås.

Materialet som digitaliserats ägs av Linköpings universitet.

Innehållsförteckning

Introduktion.....	3
Mål och syfte	4
Bakgrund	4
Projektets början, urval och relevans.....	4
Rättsliga förutsättningar.....	5
Dataskyddsförordning (GDPR).....	5
Upphovsrätt.....	5
Licenser	5
Metod	6
Inspektion	6
Bildfångst.....	6
Textfångst	8
GitHub, filer, metadata och webbplats	8
Tidsestimering, tidsrapportering.....	10
Analys	10
Sätta materialet i kontext	11
Digital bearbetning	11
Digitalisering för bevarande och presentation.....	12
Transkribering och tillgänglighet	13
Slutsats och vidare digitalisering	14
Referenser	17
Bilaga 1	20
Bilaga 2	21
Bilaga 3	22
Bilaga 4	25
Bilaga 5	29

Introduktion

På Linköpings universitetsbibliotek finns ett till synes anspråkslöst häfte, men när häftet öppnas möts läsaren av en pampig och patriotisk text: *Fosterländsk Sång vid Götha Canals Öpnande i Hans Maj:t Konungens och den Kungl. Famillens Närvaro, den 26 september 1832*. Texten, skriven av Magnus Martin af Pontin, skrevs till invigningen av Göta kanal, som markerade slutet på över två decenniers arbete med att förbinda Östersjön med Nordsjön. Detta häfte har stått i centrum för det digitaliseringsprojekt som redovisas i denna rapport. Syftet har varit att tillgängliggöra materialet digitalt, bidra till dess bevarande och att undersöka hur digitala återgivningar påverkar förståelsen av originalet, inte bara som dokument, utan även som historisk artefakt.

Digitalisering har, enligt Björk (2015), blivit en central metod för att bevara och tillgängliggöra kulturarv. Björk argumenterar för att om material inte är digitaliserat så är det i princip som att det inte finns. Genom digitalisering kan fler få tillgång till materialet, oavsett geografisk begränsning. Sånghäftet som digitaliseras i detta projekt finns i flera exemplar på ett fåtal bibliotek, men den unika version som använts här, med signaturer och etiketter, kräver fysisk närvaro i Linköping. I och med digitaliseringen öppnas andra möjligheter för tillgång och användning. Björk samt Conway (2015) lyfter att digitala reproduktioner inte är neutrala kopior, utan tolkningar av originaldokumentet. Digitalisering innebär ett urval och en gestaltning, vilket i sin tur påverkar hur materialet uppfattas och används. Björk poängterar att potentialen hos digitala dokument inte ligger i bevarande och tillgänglighet, utan i hur resursen används i förhållande till det fysiska dokumentet.

I detta projekt har digitaliseringen genomförts med flera användningssätt i åtanke. Sånghäftet finns dels som en bläddringsbar version för att efterlikna upplevelsen av originalet och dels som en sökbar och utskriftsvänlig textversion. Samtidigt är vi medvetna om att användarupplevelsen skiljer sig, att hålla det fysiska häftet i handen är inte detsamma som att navigera en digital version. För att sätta materialet i kontext och stödja förståelsen har vi också inkluderat en historisk bakgrund och begreppsförklaringar på webbplatsen.

Mål och syfte

Målet med detta projekt är att digitalisera sånghäftet som användes på invigningen av Göta kanal år 1832 i syfte att bidra till resursens bevarande och öka dess tillgänglighet, samt sätta resursen i en större kontext. För att uppnå detta genomförs följande:

- Kvalitativ bildfångst.
- Kvalitativ textfångst.
- Långsiktig lagring av källfiler och metadata med hållbara tekniker.
- Utveckling av användarvänlig webbplats.
- Publicering av det digitaliserade materialet på webbplatsen med öppen tillgång.

Bakgrund

Projektets början, urval och relevans

Projektiden växte fram när en av projektmedlemmarna, som är anställd vid Linköpings universitetsbibliotek, nämnde digitaliseringsuppgiften för sina kollegor. Vid biblioteket finns samlingar tillgängliga både via ett öppet bokmagasin och det gemensamma bokmagasinet Bokladan, där äldre tidskrifter och böcker förvaras och kan beställas (Linköpings universitet, u.å.). I Bokladan finns en samling med material om Göta kanal från 1800-talet, där ett sånghäfte från kanalens invigning väckte intresse för digitalisering, både hos gruppmedlemmen och kollegorna på Linköpings universitetsbibliotek. Det var detta som låg till grund för urvalet och är anledningen till att just detta sånghäfte digitaliserades. Dahlström (2011) kallar detta kritisk digitalisering, där material väljs ut medvetet till skillnad från massdigitalisering där så mycket som möjligt digitaliseras. Urval innebär alltid en viss bias eftersom vad som digitaliseras beror både på syftet och på den som gör urvalet.

Conway (2015) påstår att det har blivit norm att digitalisera material med kulturellt värde, både som ett sätt att bevara dokumenten och att möjliggöra vidare användning. Detta återspeglas i samtal med vår kontaktperson på Linköpings universitetsbibliotek, där sånghäftet bedömdes vara av intresse för forskare, lokalhistoriker och föreningar med fokus på Göta kanals historia. Biblioteket ser även digitalisering som ett sätt att minska risken för slitage på ömtåligt material. Samtidigt synliggörs en motsättning mellan norm och praktik då merparten av bibliotekets material inte är digitaliserat. Enligt vår kontaktperson beror detta bland annat på bristande resurser och begränsad kompetens för att hantera äldre trycksaker.

Rättsliga förutsättningar

Innan ett material digitaliseras och uppgifter publiceras på en webbplats är det viktigt att beakta rättsliga aspekter, så som hantering av personuppgifter, upphovsrätt och licenser.

Dataskyddsförordning (GDPR)

Den ökade digitaliseringen har lett till ett ökat behov av att skydda personuppgifter. I Sverige regleras det i form av GDPR och gäller all information som kan knytas till en levande person, (Integritetsskyddsmyndigheten, 2021). Då sånghäftet som digitaliseras är från år 1832 omfattas det inte av GDPR.

Upphovsrätt

I och med den digitala teknikens utveckling har tillgängligheten till upphovsrättsliga verk ändrats. Reglerna kan upplevas som snåriga och skapa osäkerhet. Vår kontaktperson på Linköpings universitetsbibliotek lyfter fram upphovsrätt som en av utmaningarna gällande digitalisering överlag. Litterära och konstnärliga verk skyddas enligt upphovsrättslagen och ger upphovspersonen rättigheter (Upphovsrättslag, 1960). Upphovsrätten gäller, enligt 4 kap 43 § upphovsrättslagen, till 70 år efter upphovspersonens död. När upphovsrätten har löpt ut är det fritt för allmänheten att använda verket utan de tidigare upphovsrättsliga begränsningarna (Patent- och Registreringsverket, 2025). Författare och upphovsperson till materialet som digitaliseras är Magnus Martin af Pontin, som levde 1781–1858 (Palmborg, 1954), vilket innebär att upphovsrätten har löpt ut.

Licenser

Även icke-upphovsrättsligt skyddat material behöver licenser för att alla ska få tillgång till, använda och dela data, i enlighet med upphovsrätts- eller databaslagarna (European Commission, u.å.). För öppet innehåll används Creative Commons-licenser frekvent, en licensmodell där det på ett klart sätt framgår hur upphovsrättsligt material får hanteras (Riksantikvarieämbetet, u.å.). Den mest öppna licensen är CC BY, där BY står för att korrekt erkännande krävs. Det gör det möjligt för återanvändare att, även i kommersiella sammanhang, sprida vidare, ändra, bearbeta och bygga vidare på verket så länge upphovspersonen anges. Eftersom ambitionen, i enlighet med Linköpings universitetsbiblioteks önskan, är att digitaliseringsprojektet ska vara så öppet och tillgängligt

som möjligt så har Creative Commons-licensen CC BY (version 4.0) valts. För bilden publicerad på webbplatsen under avsnittet för historia hänvisas till licensen CC BY-SA (version 4.0), där SA står för “dela lika”, vilket har deklarerats av upphovsrättsinnehavaren.

Metod

Inspektion

Det första steget i digitaliseringsprocessen var inspektion av den fysiska resursen. Häftet har tryckt textsnitt och är trådbundet med limmat omslag. Fram- och baksidan är orange och utan text. Inne i häftet finns etiketter och handskrivna ord. Det går att tyda upphovsmannens signatur “M.M. af Pontin” men övrig handskrift går inte att tyda. Sångtexten löper över totalt åtta sidor. För att bevara helhetsintrycket av originalhäftet valde vi att skanna hela uppslag, förutom framsidan och baksidan som skannades separat. Det beskriver Cowick (2018) som att hänsyn tas till originalets integritet, det vill säga att behålla originaldokumentets karaktär.

För att försöka tyda de handskrivna orden tog gruppen kontakt med arkivarier på Lunds universitet och Facebook-gruppen “Vi som läser äldre handskrift”, i samråd med Linköpings universitetsbibliotek. Tyvärr uppvisade svaren olikheter. Vi testade även att använda Transkribus, Abbyy FineReader och Riksarkivets AI-baserade transkriberingsverktyg HTRFlow för att tolka de handskrivna orden, vilket inte gav resultat.

Bildfångst

Vid val av bildfångstmetod är det viktigt att utgå från syftet med digitaliseringen (Zhang & Gourley, 2008). Syftet i detta projekt var att färgerna skulle spegla originalet, att texten skulle framträda så tydligt som möjligt samt att bilderna skulle bevaras långsiktigt. Johansson et al. (2006) lyfter fotografering och skanning som olika former av bildfångst. Vi valde att skanna resursen eftersom Johansson et al. påpekar att det vid fotografering finns risk för att texten kan bli skev eller otydlig. För bildfångsten användes flatbäddsskannern, RICOH MP C5504ex, som finns på Linköpings universitetsbibliotek. Det som kan vara negativt med skanning framhäver Johansson et al. är att det bland annat blir ett visst brus som gör att bilderna kan bli korniga. Trots det upplevdes ändå texten bli tydlig vid skanning, vilket var ett huvudsyfte med bildfångsten. Eftersom ett annat syfte med skanningen var att återge häftet så visuellt likt originalet som möjligt testade vi att skanna häftet med bakgrundspapper

i olika färger. Bakgrundspapper i mörkare grå ton gjorde färger och kontraster på det skannade materialet så likt originalet som vi kunde komma. Johansson et al. påpekar att färgjusteringar kan göras i efterhand i bildredigeringsprogram, men eftersom vi fick önskad färg av bakgrundspappret så fanns det inget behov av färgredigering.

Deutsche Forschungsgemeinschaft (2013) rekommenderar att skanning görs med minst 300 dpi (punkter per tum). Zhang och Gourley (2008) skriver att det är bra att skanna med så hög upplösning som möjligt, bland annat för att långsiktigt bevara detaljer. Därför skannade vi med den högsta möjliga dpi som fanns på skannern. Vi skannade i tre olika filformat som fyller olika funktioner:

- TIFF, 400 dpi i svartvitt, som används för långsiktigt bevarande.
- JPEG, 600 dpi i färg, som används för presentation på webbplatsen.
- PDF, 600 dpi i färg, som i efterhand konverterades till PDF/A genom onlineverktyget PDF24 Tools (<https://tools.pdf24.org/en/>) för långsiktigt bevarande.

Skanningen genomfördes i liggande A3-format och bildfilerna mejlades till oss. Under den processen komprimerades filerna, vilket inte var önskvärt men var den tekniska lösning som fanns att tillgå. JPEG-filerna komprimerades dock avsiktligt i efterhand för att laddas upp snabbare på webbplatsen. I och med bakgrundspappret skapades en naturlig ram runt häftet vid skanningen. Det upplevde vi vara estetiskt tilltalande för faksimilen på webbplatsen. Det visar också att hela resursen har fångats, i enlighet med Deutsche Forschungsgemeinschafts (2013) riktlinjer. Ramen var dock aningen stor för visning på webbplatsen, varpå vi använde bildredigeringsverktyget Photopea (<https://www.photopea.com>) för att beskära JPEG-filerna och justera där häftet hamnat snett. TIFF- och PDF/A-versionerna har medvetet inte redigerats, för att bevara dem i så opåverkad form som möjligt.

Skanningen utfördes vid två olika tillfällen. Det första tillfället i enbart PDF-format, men eftersom vi önskade ram runt häftet och fler filformat för bättre bevarande skannades materialet om. I samråd med vår kontaktperson på universitetsbiblioteket ansågs häftet vara i skick att skannas om, med målet att med olika filformat inte behöva skanna om i framtiden. Zhang och Gourley (2008) framhåller vikten av att undvika omskanning genom noggrann planering och hög kvalitet i den initiala bildfångsten. Vår erfarenhet illustrerar dock hur praktiska begränsningar, tidsramar och tillgång till utrustning påverkar digitaliseringsarbetet i

mindre projekt. Omskanningen blev nödvändig för att uppnå projektets mål, men i efterhand framstår det som tydligt att en mer genomtänkt plan i inledningsskedet hade kunnat minska behovet av att hantera originalmaterialet flera gånger.

Textfångst

För att sångtexten på webbplatsen skulle bli mer lätthanterlig och sökbar omvandlades den till maskinläsbar text, till skillnad från den inskannade faksimilen som är i bildformat. Tanner (2004) beskriver att Optical Character Recognition (OCR) kan användas för att omvandla bilder av text till maskinläsbar och redigerbar text. Vi testade olika OCR-verktyg och det som gav bäst resultat var OnlineOCR (<https://www.onlineocr.net>). När vi använde utdataformatet .txt fick vi flest antal korrekta ord och överskådligt gränssnitt. Vijayarani och Sakila (2015) framhåller att OCR-verktyg ska kunna läsa filer som innehåller flera sidor. Trots det märkte vi vid första OCR-konverteringen att inte alla sidor transkriberades, varpå konverteringen behövde göras ett uppslag i taget. För att få en bild av hur väl OCR fungerar skriver Hegghammer (2022) samt Vijayarani och Sakila att noggrannheten kan beräknas på olika sätt. Se bilaga 1 för mer beskrivning och resultat.

GitHub, filer, metadata och webbplats

En översikt över projektets filer, deras funktion och hur de hänger samman med GitHub och webbplats finns i bilaga 2. Arbetet har följt riktlinjer från Deutsche Forschungsgemeinschaft (2013). För att strukturera vår data har vi använt oss av TEI (Text Encoding Initiative), baserat på XML (Extensible Markup Language), vilket möjliggör både långsiktigt bevarande och semantisk sökbarhet. En transformation från TEI-XML till HTML har genomförts med XSLT (Extensible Stylesheet Language Transformations). Med hänsyn till semantik har webbplatsen byggts upp med fokus på HTML-elementet <article> (w3schools, u.å.).

Själva webbplatsen består av fem separata HTML-filer, en för varje sida. Styling hanteras via en gemensam CSS-fil (Cascading Style Sheets) och via bootstraplänkar i HTML-filerna. Layouten bygger på flexbox och grid-strukturer för att skapa en balanserad och responsiv design. Genom att inkludera <meta name="viewport" content="width=device-width, initial-scale=1.0"> säkerställs att innehållet anpassas efter olika skärmstorlekar. Tillgänglighet och webbstandarder har varit vägledande genom hela processen. WCAG-riktlinjer (Web Content Accessibility Guidelines, <https://www.w3.org/TR/WCAG21>) har följts genom användning av

semantisk HTML, konsekvent rubrikstruktur (<h1>, <h2>, osv.), aria-labels för att underlätta navigering med skärmläsare, tydliga fokusmarkeringar i CSS och användning av klassen visually-hidden för skärmläsarspecifika element.

Bildmaterialet presenteras i en karusell och inkluderar alt-texter och <figcaption> för att kontextualisera innehållet. För att säkerställa att bilderna är responsiva används Bootstrap-klasser som img-fluid och w-50, vilket gör att bilderna anpassas efter olika skärmbredder. JavaScript har använts för att skapa interaktiva komponenter, såsom bildkarusell och zoomfunktion. Dessa har utformats med fokus på tillgänglighet, exempelvis genom att säkerställa att de är navigerbara med tangentbord och skärmläsarstöd för dynamiskt innehåll.

En viktig del i digitaliseringsarbete är även hantering av metadata, vilket Cowick (2018) beskriver som information om materialet med syfte att underlätta långsiktigt bevarande och förståelse för hur materialet ska kunna användas. Författaren framhåller vidare att det finns olika typer av metadata: beskrivande, administrativ, bevarande och strukturell. Det har vi anammat genom att bland annat beskriva resursens ursprung, teknisk information om bilderna, sånghäftets uppbyggnad och dess layout samt rättighetsinformation gällande resursen och den digitala faksimilen.

I enlighet med Deutsche Forschungsgemeinschaft (2013) har metadata samlats i TEI-filen. Filerna med bilder på faksimilen innehåller teknisk metadata. Ytterligare metadata har adderats till filerna för bevarande och tillgänglighet. Metadatan i TIFF-filerna har sedan extraherats med ExifTool till åtta XML-filer. XML-filerna har sedan kopplats till TEI-filen. Att använda etablerade standarder för metadata är något som både Deutsche Forschungsgemeinschaft (2013) samt Dappert et al. (2016) rekommenderar, bland annat ur bevarandesynpunkt. För mer detaljer se bilaga 3.

För att tillföra information till sånghäftets text har vi använt TEI för att markera strukturella och semantiska element. Enligt Deutsche Forschungsgemeinschaft (2013) kräver detta att överväganden görs gällande i vilken utsträckning texten ska markeras. Dahlström (2011) benämner detta som intellektuell redigering. Även om texten i sig inte redigerats har medvetna val gjorts gällande hur texten ska presenteras och tolkas. Exempelvis har anvisningar som "recitativ" markerats för att särskilja dem från den egentliga sångtexten.

Interpunktion och radbrytningar har markerats där de tjänar som vägledning för framförandet, vilket bidrar till semantik och rytm. Även namn och platser har märkts upp i TEI för att förbättra både förståelse och sökbarhet, medan osäkra tolkningar lämnats omarkerade för att undvika felaktigheter. För att tydliggöra uppmärkningen har vi i TEI-metadata förklarat vad uppmärkningarna betyder och skapat en taxonomi, där vi listar vad olika värden för attribut står för.

Tidsestimering, tidsrapportering

Projektgruppens deltagare har löpande noterat vilken tid som lagts på vad i projektet. Upplägget för den rapporteringen har varit inspirerad av de faser som Björk (2015) identifierar för digitaliseringsprocessen. I bilaga 4 finns översikt, diagram och tabell som visar förhållandet mellan digitaliseringsprocessens olika faser och delar samt mellan projektgruppsmedlemmarna.

Vi uppskattade att varje projektmedlem skulle lägga cirka 200 timmar på projektarbetet. Arbetsuppgifterna fördelades succesivt under projektets gång. Det var svårt att på förhand estimerade hur lång tid de olika delarna i projektet skulle kräva. En stor del av tiden ägnades åt att sätta sig in i olika digitala verktyg, filtyper och märkspråk. När vi fick god förståelse för dessa blev arbetet mer effektivt.

Vi estimerade att mycket tid behövde läggas på den visuella bearbetningen av webbplatsen samt på transkriptionen av sångtexten, vilket vi också gjort. Arbetet med TEI och metadata samt arbetet med kodning tog längre tid än förväntat. Däremot gick momenten som rörde GitHub och XSL-transformeringar smidigare än planerat.

Analys

Det finns delar av digitaliseringsprojektet som utmärkt sig, där vi har gjort ställningstaganden och fattat beslut som påverkar digitaliseringsprojektets resultat. Dessa delar står i fokus för analysen.

Sätta materialet i kontext

Cowick (2018) betonar vikten av att vid digitalisering förstå vilket värde ett material har för allmänheten. Eftersom det i samtal med vår kontaktperson på Linköpings universitetsbibliotek framkom att materialet kan vara av intresse för bland annat lokalhistoriker valde vi att placera sånghäftet i en historisk kontext. Resursen är av betydande historiskt och kulturellt värde då det uppmärksammar en viktig infrastrukturell utveckling i svensk historia (Göta kanalbolag, 2023). På webbplatsen finns en sida med historia där det var viktigt att hitta en balans så att inte kontexten om Göta kanal tog överhand, utan låta faksimilen vara huvudfokus.

Björk (2015) uppmärksammar att fokus kan läggas på både innehåll och det fysiska utseendet för att få helhetsförståelse för materialet. Vår önskan med det digitaliserade materialet är att det i framtiden ska kunna användas i flera forum. Faksimilen visas på webbplatsen genom bilder och sångtexten är transkriberad för att kunna skrivas ut och användas i andra sammanhang. Genom att läsa texten går det att få en upplevelse av språket och livet på 1800-talet. Bland annat innehåller sånghäftet flera metaforer med hänvisning till nordisk mytologi samt flera retoriska stilgrepp, till exempel: "Då flög en tanke, Mimers like, Utöfver Svithiods hela rike (...)". I sångtexten förekommer även flera hänvisningar till svenska kungar, vilket vi tolkar visar på den framträdande roll och betydelse som monarkin hade under denna tid. Sammanfattningsvis transkriberades texten för att lyfta fram att sångtexten har ett eget värde.

Digital bearbetning

Björk (2015) och Dahlström (2019) argumenterar för att digitalt material enbart är en representation av det fysiska materialet. Dahlström fördjupar resonemanget och poängterar att autenticiteten i det digitala materialet kan ifrågasättas beroende på hur mycket materialet redigeras och bearbetas. Vi valde därför att bevara materialet så autentiskt som möjligt genom att undvika omfattande bearbetning. En digital bearbetning utfördes när vi beskar ramen och rätade upp häftet i de inskannade bilderna som visas på webbplatsen, vilket endast påverkar presentationen. Vi är medvetna om att vi, som både Björk och Dahlström påpekar, skapar ett material som kan få ny betydelse för användarna jämfört med det fysiska exemplaret. Björk påpekar att textens struktur kan återskapas med verktyg som TEI, men eftersom det är svårt att återge alla detaljer blir det ändå endast en spegling av det fysiska materialet. Själva upplevelsen av originalet kan aldrig helt ersättas, trots vår ambition att via

bilder, metadata och textuppmärkning förmedla dess karaktär. Digitaliseringen har dock, menar vi, ökat både tillgången och tillgängligheten till resursen.

Digitalisering för bevarande och presentation

Björk (2015) och Cowick (2018) påstår att enbart digitalisering av ett material inte är det som säkerställer bevarandet av materialet. Det handlar snarare om hur det digitala materialet används, vilket enligt vår tolkning innebär att materialet behöver vara välstrukturerat och tillgängligt. Vi har exempelvis samlat det digitala materialet i GitHub för att göra det så tillgängligt som möjligt. Cowick utvecklar resonemanget kring bevarande och påstår att både analogt och digitalt material behöver omsorg, men att digitalt kan behöva mer eftersom det behöver övervägas exempelvis hur filer utifrån format ska kunna öppnas och användas även i framtiden. Vår kontaktperson på universitetsbiblioteket lyfte att det finns risk för att analogt material tar skada vid för mycket hantering, men det finns även risk gällande digitalt material eftersom det kan gå förlorat eller vara sårbart på grund av tekniska problem. Cowick förtydligar vidare att det är viktigt att komma ihåg att även metadata, inte bara filer, ska bevaras på ett långsiktigt sätt. Det har vi anammat genom att lagra metadata i våra bildfiler och XML-filer. Vi märkte under processens gång att det var en balansgång gällande hur mycket metadata som skulle inkluderas då vi ville ha med så mycket som möjligt men samtidigt göra det överskådligt.

Johansson et al. (2006) samt Zhang och Gourley (2008) framhäver TIFF som ett flexibelt och hållbart filformat för lagring av bilder. Bildformatet är anpassningsbart och kan användas i olika typer av operativsystem. Det är effektivt för att bevara och arkivera högkvalitativa bilder för framtida användning. Eftersom TIFF-filer har dessa egenskaper valde vi, i bevarandesyfte, att skanna och spara bilder på sånghäftet i detta format, även om flatbäddsskannern bara kunde skanna i svartvitt. Vi förlorade därmed en del av den estetiska helheten, men behöll långsiktig kvalitet i bilderna.

Nackdelen med TIFF-filer förklarar Johansson et al. (2006) samt Zhang och Gourley (2008) är att filerna blir stora och därför inte optimala för att visas på webben. Vid skanningen komprimerades TIFF-filerna när de skickades via mejl, vilket innebar att bildfilerna blev mindre, även om de fortfarande var för stora för att visas på webbplatsen. Zhang och Gourley

förtydligar dessutom att TIFF-filer är förlustfria, vilket innebär att bilddata inte går förlorad när bildfilen komprimeras.

För att bilderna på webbplatsen skulle vara i optimalt filformat och i färg använde vi JPEG-filer, vilket Zhang och Gourley (2008) poängterar inte tar lika mycket plats och är lämpligt för webben. Vi sparade även bilderna i ett tredje filformat, PDF. Evans och Moore (2014) beskriver PDF som ett vanligt format, men påpekar att det inte möter behovet av långsiktigt bevarande. Däremot hävdar de att PDF/A är tekniskt mer avancerat och möjliggör att bilderna kan öppnas även vid teknikutveckling. Därför valde vi att konvertera PDF-filerna till PDF/A. På så sätt kunde vi spara bilderna långsiktigt, vilket inte JPEG uppfyller. Genom PDF/A-formatet kunde vi dessutom spara bilderna i färg, vilket inte var möjligt med flatbäddskannerns inställningar för TIFF.

Transkribering och tillgänglighet

Linköping universitetsbibliotek framhåller det positiva med att transkribera texten i sånghäftet, bland annat eftersom det då går att få texten uppläst, vilket ökar tillgängligheten. Hegghammer (2022) argumenterar för vikten av transkribering i form av att det kan öppna upp för hantering av historiska dokument som tidigare inte varit möjligt om objektet bara funnits analogt. Dock påstår Hegghammer att kvalitén för OCR-konverteringar kan variera. Tanner (2004) håller med men går djupare i sitt resonemang gällande att material tryckt före 1950 förmodligen inte kommer ha mer än 98% i noggrannhet, det vill säga att inte mer än 98% kommer att vara korrekt omvandlat från skannad bild till redigerbar text via OCR. Tanner hävdar att text skriven före 1900, vilket sånghäftet är, kan innehålla ord som inte är vanligt förekommande idag, vilket kan orsaka fel vid omvandlingen. Detta var även något som vi observerade, nämligen att de flesta felaktigt transkriberade orden var tidstypiska. Ett exempel är ordet "ständar" som OCR-verktyget tolkade som "stunder". Därför är det enligt Tanner viktigt att granska texten och manuellt rätta de ord och tecken som blivit fel, men att det kan vara tidskrävande. Vår granskning var ett tidskrävande arbete i form av korrekturläsning, men att transkribera hela texten manuellt hade varit ännu mer tidskrävande då det både hade tagit tid att skriva och korrekturläsa. Även beräkning av noggrannhet gjordes manuellt. Hegghammer påpekar att beräkning av noggrannheten kan genomföras automatiskt men eftersom vårt material har relativt få sidor var detta inte aktuellt.

Gällande OCR så märkte vi att resultatet kan variera. Vi använde nämligen samma OCR-verktyg och bildfiler fast vid olika tillfällen och fick olika resultat. Vi upplever ändå att vi genom beräkningen får en uppfattning kring OCR-verktygets noggrannhet. Bilaga 1 visar att precisionen var 97% för tecken, 94% för specialtecken och 96% för ord. Även om dessa siffror vid första anblick kan verka något låga, stämmer de överens med Tanners (2004) bedömning att precisionen för äldre dokument ofta ligger på 98% eller lägre. Det som drog ner precisionen i sånghäftet var första sidan där texten har olika typografiska stilar, vilket var svårt för OCR-verktyget att transkribera. Det visar på vikten av korrekturläsning, särskilt med fokus på specialtecken, som oftare tenderar att bli felaktigt transkriberade. Vår tolkning är att antalet felaktigt transkriberade specialtecken kan bero på papprets färg och kvalitet, vilket påverkar igenkänning av dessa tecken.

Slutsats och vidare digitalisering

Cowick (2018) påpekar att analogt material digitaliseras både i bevarandesyfte och för att främja tillgänglighet. Genom att digitalisera sånghäftet kan fler personer få tillgång till det, vilket också var en av anledningarna till att Linköpings universitetsbibliotek ville digitalisera materialet.

På biblioteket finns ytterligare material gällande Göta kanal. Biblioteket framhåller att resten av detta material skulle kunna digitaliseras på liknande sätt som vi gjort med sånghäftet, så att även det blir tillgängligt. Björk (2015) framställer digitaliseringens potential som obegränsad. Dock betonar vår kontaktperson på universitetsbiblioteket vikten av att sätta materialet i en kontext. I bilaga 5 finns bilder på material som hade kunnat vara intressant att digitalisera tillsammans med sånghäftet. Det materialet innehåller bland annat kostnadsestimeringar för bygget av Göta kanal samt reseskildringar som ger inblick i vad som var populärt under denna tid. Samlingen kan tillsammans ge ytterligare kontext och historisk betydelse i relation till varandra.

Digitaliseringens skalbarhet är begränsad, särskilt i en kontext som ett universitetsbibliotek, där mängden material är stor men resurserna för digitalisering är begränsade. Det krävs därför prioriteringar utifrån efterfrågan och materialets lämplighet (Dahlström, 2011). Vissa material, som sånghäftet, lämpar sig bättre för en noggrann och kontextualiserad digitalisering, snarare än massdigitalisering. Ett sådant arbete kräver tid för att skapa

metadata, förstå det historiska sammanhanget och hantera det fysiska originalet varsamt. Därför är det inte alltid praktiskt eller önskvärt att skala upp digitalisering utan att samtidigt riskera att kompromissa med kvalitet och kontext.

I projektet har fokus lagts på att genomföra en högkvalitativ bildfångst i syfte att bevara en kulturhistoriskt värdefull artefakt. Fokus har också varit att öka tillgången och tillgängligheten till resursen genom att publicera materialet med öppen tillgång på en användarvänlig webbplats. För att sätta kontexten och ge mervärde publiceras resursen tillsammans med information om dess historia och förklaringar av begrepp, vilket också speglar valet av att markera upp interpunktion, radbrytning samt namn och platser i TEI. Uppmarkering med hjälp av TEI riskerar enligt vår erfarenhet i detta digitaliseringsprojekt bli bias och subjektivt. Det finns exempelvis risk för att ett ord inte är uppmärksammat som ett namn, eftersom det är vår tolkning som gör att vi tror att det är en liknelse och inte nödvändigtvis ett namn. Det gör att den mänskliga faktorn i form av vår tolkning kan påverka och det blir en avvägning. Det går exempelvis inte att tolka alla ord med versaler i början som ett namn, eftersom det blandas i sångtexten.

Likt Björks (2015) tankar kring att digitaliseringens potential är obegränsad så går det att arbeta vidare med sånghäftet. Vår upplevelse är att det ändå finns vissa begränsningar utifrån de standarder som följs, i detta fall utifrån TEI. Dock finns det mer som hade varit intressant att arbeta med, som exempelvis att göra tolkningar av texten då den innehåller flera metaforiska uttryck. Det hade i det fallet varit intressant att markera stilistiska uttryck även i TEI.

Vi märkte överlag att det tog lång tid att sätta sig in i olika digitala verktyg, filtyper och märkspråk. Hade vi gjort om digitaliseringen eller fortsatt på samma sätt med ytterligare material hade det inte tagit lika lång tid, men det är fortfarande ett tidskrävande arbete. Dock anser vi att om digitaliseringsarbetet är välgjort från början behöver det inte göras om igen och det digitaliserade materialet finns tillgängligt för fler samtidigt som det analoga materialet inte slits ut. Genomtänkt digitalisering är med andra ord värdefullt, i alla fall så länge det används. Precis som Björk (2015) argumenterar för att analogt material som inte finns digitaliserat kan glömmas bort, behöver även digitala faksimiler synliggöras så att

allmänheten vet om att de finns, kan ta del av dem och använda dem i andra sammanhang som exempelvis forskning, även i framtiden.

Referenser

- Björk, L. (2015). *How reproductive is a reproduction? Digital transmission of text-based documents*. [Doktorsavhandling, Borås universitet]. <https://hb.diva-portal.org/smash/get/diva2:860844/INSIDE01.pdf>
- Conway, P. (2015). Digital transformations and the archival nature of surrogates. *Archival Science*, 15(1), 51–69. <https://doi.org/10.1007/s10502-014-9219-z>
- Cowick, C. (2018). *Digital curation projects made easy A step-by-step guide for libraries, archives, and museums*. Rowman & Littlefield.
- Dahlström, M. (2011). Editing libraries. I C. Fritze, F. Fischer, P. Sahle & M. Rehbein (Red.), *Bibliothek und Wissenschaft: Digitale Edition und Forschungsbibliothek*, 44, 91-106. Harrassowitz. <https://hb.diva-portal.org/smash/get/diva2:871326/FULLTEXT01.pdf>
- Dahlström, M. (2019). Copies and facsimiles. *International Journal of Digital Humanities*. 1, 195-208. <https://doi.org/10.1007/s42803-019-00017-5>
- Dappert, A., Guenther, R. S., & Peyrard, S. (2016). An introduction to implementing digital preservation metadata. I A. Dappert, R. S. Guenther & S. Peyrard (Red.), *Digital preservation metadata for practitioners implementing PREMIS* (s. 1-9). Springer. <https://doi.org/10.1007/978-3-319-43763-7>
- Deutsche Forschungsgemeinschaft. (2013). *DFG Practical guidelines on digitisation*. <https://www.dfg.de/resource/blob/176110/76abec10bdc30b41f18695145003d6db/12-151-v1216-en-data.pdf>
- European Commission. (u.å.). *Varför behöver vi licenser?* <https://data.europa.eu/elearning/sv/module4/#/id/co-01>

- Evans, T. N. L., & Moore, R. H. (2014). The use of PDF/A in digital archives: A case study from archaeology. *International Journal of Digital Curation*, 9(2), 123-138.
<https://doi.org/10.2218/ijdc.v9i2.267>
- Göta kanalbolag. (2023). *Den stora invigningen*. Göta kanal. <https://www.gotakanal.se/sv/en-vacker-historia/den-stora-invigningen>
- Hegghammer, T. (2022). OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. *Journal of Computational Social Science*, 5, 861–882.
<https://doi.org/10.1007/s42001-021-00149-1>
- Integritetsskyddsmyndigheten. (31 mars 2021). *Personuppgifter*.
<https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/introduktion-till-gdpr/personuppgifter/>
- Johansson, K., Lundberg, P., & Ryberg, R. (2006). *Grafisk kokbok 3.0 Guiden till grafisk produktion*. Bokförlaget Arena.
- Linköpings universitet. (u.å.). *Bibliotekets magasin*. <https://liu.se/artikel/biblioteksmagasin>
- Palmborg, N. (1954). Biografiska upplysningar om brevens adressater. I N. Palmborg (Red.), *Esaias Tegnér's brev II 1818–1823* (s. 457–465). Allhems förlag.
<https://urn.kb.se/resolve?urn=urn:nbn:se:lb-lb124739-faksimil>
- Patent- och Registreringsverket. (5 mars 2025). *Kulturarvsinstitutioner*.
<https://www.prv.se/sv/upphovsratt/for-myndigheter>
- Riksantikvarieämbetet. (u.å.). *Licensmärkning och beskrivningar*. <https://www.raa.se/om-riksantikvarieambetet/om-webbplatsen/licenser/>
- Tanner, S. (2004). *Deciding whether optical character recognition is feasible*. King's Digital Consultancy Services.

https://www.researchgate.net/publication/265748433_Deciding_whether_Optical_Character_Recognition_is_feasible

Upphovsrättslag (SFS 1960:729). Justitiedepartementet L3.

https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-1960729-om-upphovsratt-till-litterara-och_sfs-1960-729/

Vijayarani, S., & Sakila, A. (2015). Performance comparison of OCR tools. *International Journal of UbiComp (IJU)*, 6(3), 19–30.

https://www.researchgate.net/publication/281583162_Performance_Comparison_of_OCR_Tools

w3schools. (u.å.). *HTML semantic elements*.

https://www.w3schools.com/Html/html5_semantic_elements.asp

Zhang, A. B., & Gourley, D. (2008). *Creating digital collections A practical guide*. Chandos Publishing.

Bilaga 1

Noggrannhetsberäkning av OCR

Vi använde oss av tre olika sätt för att beräkna noggrannheten av OCR. Hegghammer (2022) samt Vijayarani och Sakila (2015) beskriver teckenprecision där det totala antalet korrekta tecken i OCR-konverteringen divideras med totalt antal tecken i texten i sånghäftet.

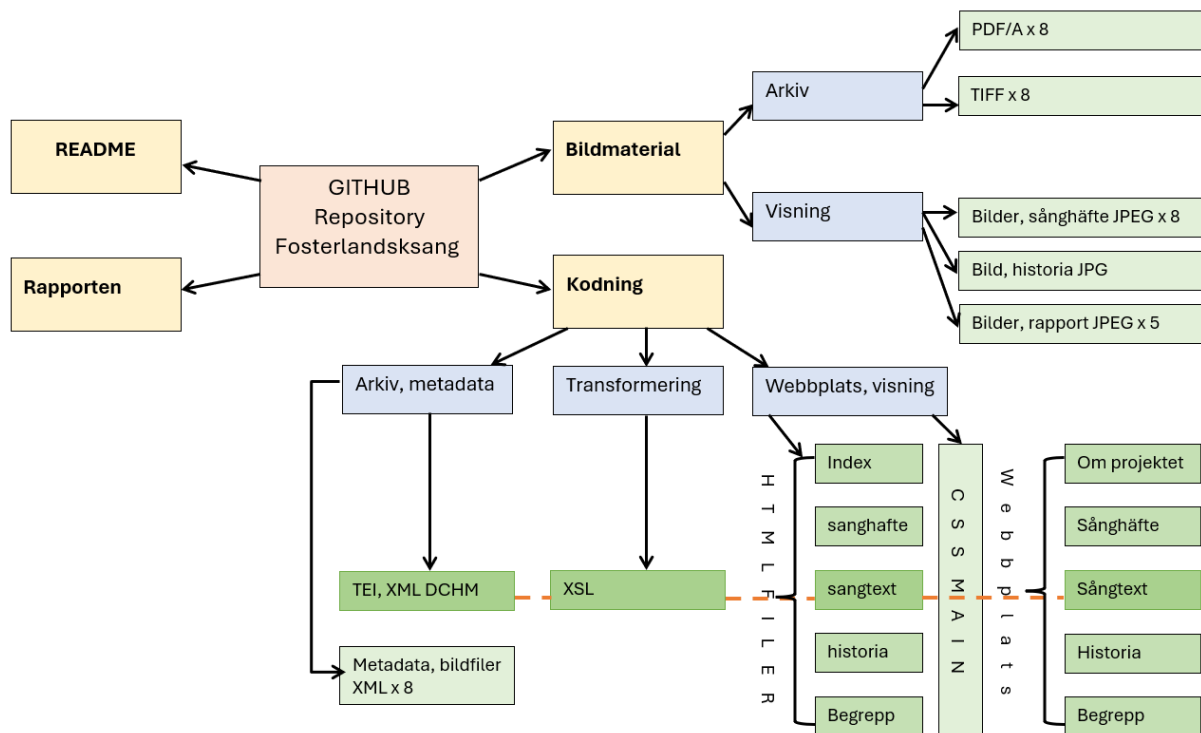
Resultatet multipliceras sedan med 100 för att få noggrannheten i procent. Då fick vi följande resultat: 97%. Exempel på tecken är bokstäver och vanlig interpunktion.

Vijayarani och Sakila (2015) beskriver att det även går att beräkna noggrannheten genom specialtecken där det totala antalet korrekta specialtecken i OCR-konverteringen divideras med totalt antal specialtecken i texten i sånghäftet. Resultatet multipliceras sedan med 100 för att få noggrannheten i procent. Som specialtecken räknas exempelvis interpunktion och bindestreck. Då fick vi följande resultat: 94%. Exempel på specialtecken är asterisk och längre bindestreck som används för att markera pauser.

Hegghammer (2022) beskriver att det går att beräkna noggrannheten genom ordprecision där det totala antalet korrekta ord i OCR-konverteringen divideras med totalt antal ord i texten i sånghäftet. Resultatet multipliceras sedan med 100 för att få noggrannheten i procent. Ett ord räknas som korrekt om det exempelvis är stavat likadant som i originalhäftet. Då fick vi följande resultat: 96%.

Bilaga 2

Processororienterad översikt över projektets filer, deras funktion och hur de hänger samman med GitHub och webbplats.



Bilaga 3

Metadata i TIFF-, PDF/A-, JPG- och JPEG-filer

Den metadata vi arbetat med omfattar teknisk, administrativ och beskrivande metadata. Den tekniska metadatan skapades i samband med skanningen av materialet. Utifrån projektets omfattning har vi dessutom valt att använda ISO-standardens Extensible Metadata Platform (XMP). Inom XMP har vi tillämpat Dublin Core Metadata Element Set och, via ExifTool, adderat metadata till både TIFF- och PDF/A-filerna. Vi har också använt standarden Information Interchange Model (IIM) och, i Adobe Bridge, lagt till ytterligare metadata såsom upphovsperson, uppgifter om upphovsrätt, textalternativ (alt text) för tillgänglighet samt nyckelord hämtade från den kontrollerade vokabulären "Svenska ämnesord". För ökad tillgänglighet adderade vi också alt text till de åtta JPEG-filer som motsvarar våra TIFF- och PDF/A-filer, samt till den JPG-fil vi använder i historiedelen på webbplatsen.

Metadata inbäddad i TIFF- och PDF/A-filerna:

XMP-dc

Contributor: Marie Gyll; Jessica Hultengren; Susanne Landén; Magdalena af Trolle (enheter som bidragit till resursens innehåll).

Creator: Magnus Martin af Pontin (den som i första hand är ansvarig för resursens intellektuella innehåll).

Date: 1832 (ett datum knutet till en händelse i resursens levnadstid – till exempel skapande, publicering eller digitalisering).

Source: Fosterländsk sång vid Götha Canals öppnande (en relaterad resurs från vilken den beskrivna resursen är härledd).

IPTC (IIM, legacy)

Copyright: CC BY (version 4.0). (Upphovsrätt).

Copyright Info URL: <https://creativecommons.org/licenses/by/4.0/>

IPTC Core

Headline: (Kort sammanfattning av bildens innehåll).

Description: Skannad bild av ett uppslag ur Fosterländsk sång vid Götha Canals öppnande av Magnus Martin af Pontin (beskrivning av bilden och dess innehåll).

Alt Text (Accessibility): (En kort textbeskrivning av syftet och innebörden av en bild, som kan nås med hjälpmedelsteknik).

Keywords:

Kanalbyggnad; <https://id.kb.se/term/sao/Kanalbyggnad>;

Invigningar; <https://id.kb.se/term/sao/Invigningar>;

Sångtexter; <https://id.kb.se/term/sao/S%C3%A5ngtexter>;

Sverige - Göta kanal; <https://id.kb.se/term/sao/Sverige--G%C3%B6ta%20kanal>;

1800-talet; <https://id.kb.se/term/sao/1800-talet>;

Infrastruktur; <https://id.kb.se/term/sao/Infrastruktur>;

Kanaler; <https://id.kb.se/term/sao/Kanaler>;

Slussar; <https://id.kb.se/term/sao/Slussar>

(nyckelord för att beskriva det synliga och abstrakta innehållet i bilden).

Description Writer: Marie Gyll; Jessica Hultengren; Susanne Landén; Magdalena af Trolle (de personer som skrivit beskrivningstext av bilden).

Copyright Notice: CC BY (version 4.0). (Information om upphovsrätten till bilderna).

Copyright Status: Copyrighted (upphovsrättslig status).

Rights Usage Terms: I enlighet med Creative Commons-licens CC BY (version 4.0). (Om hur bilden kan användas lagligt, kan också användas för att indikera Creative Commons-licens).

IPTC Extension

Artwork or Object in the Image:

Title: Fosterländsk sång vid Götha Canals öppnande (verkets titel).

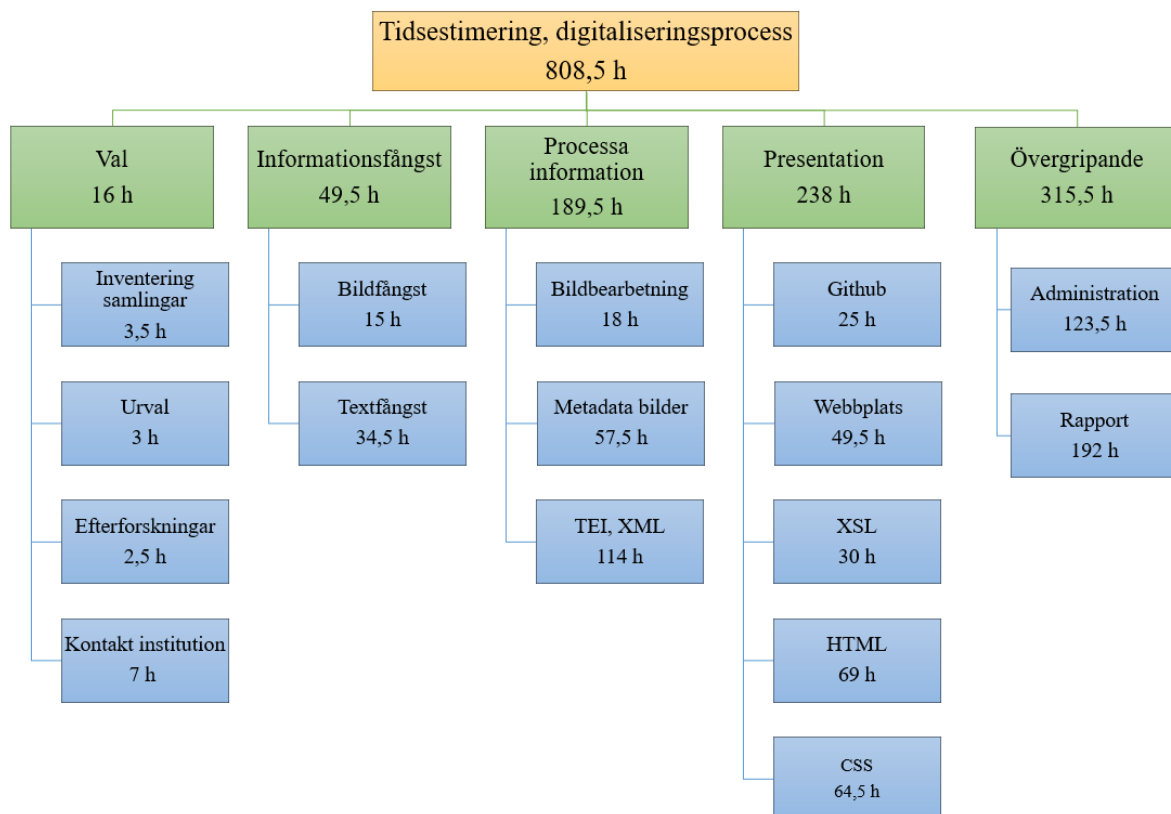
Creator: Magnus Martin af Pontin (verkets skapare).

Source: Linköpings universitetsbibliotek (namnet på organisationen som innehar verket).

Bilaga 4

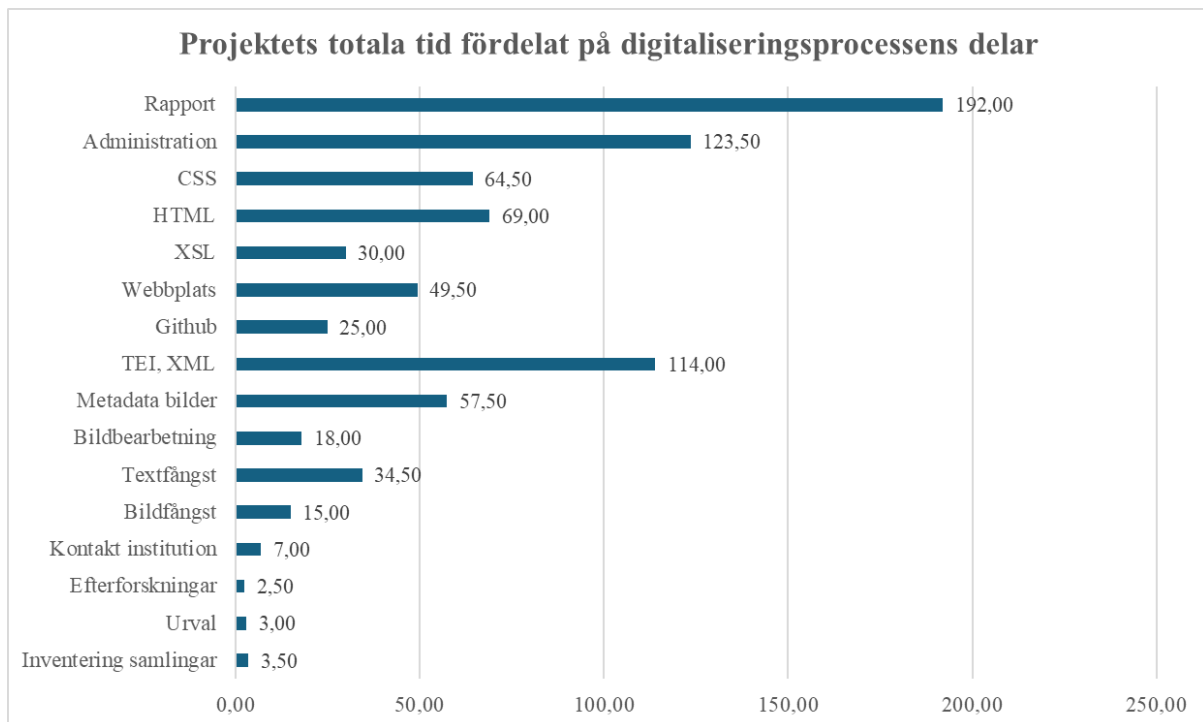
Tidsestimering

Som underlag för att kunna göra en relevant tidsestimering har projektgruppens deltagare löpande noterat vilken tid som lagts på vad i projektet. Uppställningen vi har arbetat utifrån och som illustreras med hjälp av den hierarkiska grafiken nedan är inspirerad av de faser som Björk (2015) identifierar hör till digitaliseringsprocessen.



Omfattningen hos några av rubrikerna på den tredje nivån (den blåa) kan vara relevanta att förtydliga. **Efterforskningar** omfattar den tid som lagts ner på att efterforska signaturer och stämplat. **Kontakt institution** omfattar den tid kontakten med institutionen tagit, kontakter som rört exempelvis foljebrev, material, licenser, mötesbokning och möten. **Bildfångst** omfattar tid för skanning, tester, ställningstaganden samt omskanning. **Textfångst** omfattar tid för OCR, beräkning av noggrannhet för OCR samt manuell transkribering. **TEI, XML** omfattar tid för både den metadata och kodning som arbetet med XML-filen inneburit. **GitHub** omfattar bland annat tid för upprättandet av readmefil, arkivering av TEI och bildfiler. **Webbplats** omfattar tid vi lagt på utformandet av själva webbplatsen utöver arbetet

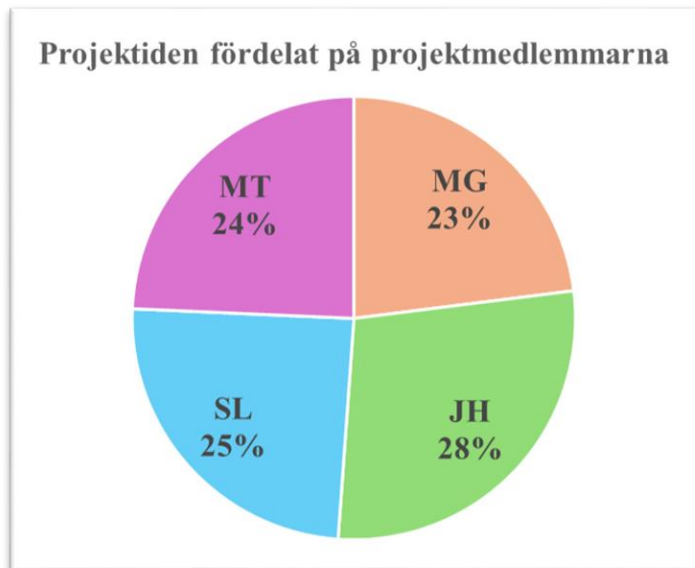
med kodningen och GitHub. Det inkluderar tid för planering, diskussioner, skisser samt författandet av texten på webbplatsen. **Administration** omfattar tid för projektgruppsmöten, författandet av följebrev, projektanmälan, mötesbokningar, anteckningar, skapandet av gemensamma dokument, tidsestimering/rapport, fördelning av arbetsuppgifter, frågeformulering samt förberedelser inför möten. **Rapport** omfattar tid för skrivning, bearbetning och korrekturläsning. **Utöver** tiden som redovisats tillkommer tid för inläsning, föreläsningar, handledartid, restid med mera. För att på ett mer översiktligt sätt visa hur den rapporterade projekttiden fördelats på projektets ingående delar infogas stapeldiagrammet nedan. Det är antal nedlagda timmar som visas med hjälp av den vågräta axeln samt anges i ytterkant av respektive stapel.



Fördelningen inom gruppen

Vi är fyra personer som har arbetat i projektet. Marie Gyll (MG), Jessica Hultengren (JH), Susanne Landén (SL) och Magdalena af Trolle (MT). Vår utgångspunkt i projektarbetet har varit att driva projektet gemensamt. Vi har i regel haft projektgruppsmöten två gånger per vecka för att utföra vissa uppgifter gemensamt, fördela arbetsuppgifter och stämma av hur arbetet fortskrider. Vissa uppgifter har alltså utförts gemensamt, vissa uppgifter har vi tillsammans kommit överens om att dela upp samt hur de ska fördelas mellan oss. Med vissa

mer omfattande delar har vi arbetat två och två eller två parallellt och med vissa mer specifika delar en och en. Vi har dock genomgående stämt av och redovisat fördelade uppgifter gemensamt på våra möten där vi också ofta tillsammans gjort de sista finjusteringarna. Det har varit viktigt för oss alla att vara och hålla varandra informerade i projektets alla delar.



Här redovisas ett cirkeldiagram över projektet som helhet, där den enskilde projektmedlemmens rapporterade nedlagda tid på projektet redovisas procentuellt.

För att slutligen koppla samman dessa delar, det vill säga hur mycket tid som har lagts på projektets olika delar av vem, så infogas en sista tabell för att illustrera förhållandet mellan dessa variabler.

Projektdel/projektmedlem	MG	JH	SL	MT	R
Inventering samlingar	0,50	0,50	0,50	2,00	3,50
Urval	0,00	0,00	0,00	3,00	3,00
Efterforskningar	0,50	0,50	1,00	0,50	2,50
Kontakt institution	1,00	1,00	1,00	4,00	7,00
Bildfångst	3,00	3,00	3,00	6,00	15,00
Textfångst	20,50	1,50	11,00	1,50	34,50
Bildbearbetning	0,00	0,00	18,00	0,00	18,00
Metadata bilder	53,50	1,50	1,00	1,50	57,50
TEI, XML	22,75	30,00	41,75	19,50	114,00

GitHub	5,50	8,00	4,50	7,00	25,00
Webbplats	14,25	10,75	16,25	8,25	49,50
XSL	1,50	16,75	1,50	10,25	30,00
HTML	1,50	23,50	10,00	34,00	69,00
CSS	0,75	31,25	4,75	27,75	64,50
Administration	25,25	45,25	28,25	24,75	123,50
Rapport	35,50	54,00	56,00	46,50	192,00
Tid lagd på projektet:	186,00	227,50	198,50	196,50	808,50

Bilaga 5

Vidare digitalisering

Material från Linköpings universitetsbibliotek som förslag för vidare digitalisering, rörande Göta kanal.

