

Digitalisering av utgrävningsrapport från 1920-talet

Marie Anneling

Amanda Torbjörnsdottir

Emma Walldén

1. Bakgrund

Inom ramen för kursen Digitalisering för bevarande och tillgänglighet har vi digitaliserat en utgrävningsrapport gällande en grävning år 1922 på Nosaby kyrkogård, belägen i dåvarande Villands härad. Idag tillhör Nosaby kyrka Kristianstad kommun (Svenska kyrkan, u.å.). I samband med en gravgrävning på kyrkogården, hösten 1921, hittade en kyrkväktare fyra större lerkärl. Nästan ett år senare, 6-7 oktober 1922, gjorde Olof Källström, då anställd som amanuens vid Lunds universitet (Krister Gierow, 1940), en efterundersökning och det är denna som dokumenterats i form av en utgrävningsrapport. Rapporten finns bevarad i ett magasin tillhörande Historiska museet vid Lunds universitet. Innan vi valde vad gruppen skulle digitalisera, sökte vi kontakt med olika institutioner och efterfrågade material som inte omfattas av upphovsrätten. Den valda rapporten är en av flera utgrävningsrapporter som lades fram som förslag av Lunds historiska museum, och är upphovsrättsfri. Rapporten är handskriven och har tillhörande fotografier från utgrävningsplatsen och fynden. Vårt syfte med digitaliseringen är dels att bevara rapporten, men också att göra innehållet mer lättillgängligt. Gruppen har framför allt haft gruppmöten digitalt via Zoom, och under projektets gång även samtalat genom chat i Messenger.

2. Bildfångst

Marie och Amanda besökte historiska museets magasin där de valde en passande utgrävningsrapport för ett mindre kulturarvsbevarande digitaliseringsprojekt. Fotograferingen genomfördes i magasinets egna fotostudio, som var utrustad med två softboxes, ett fotograferingsbord, samt en systemkamera (Nikon D7100) som efter våra önskemål var inställd på RAW-inställning. För att säkerställa kvalitén togs flera fotografier av var sida i syfte att kunna göra ett urval i efterhand. På grund av begränsningarna i kamerans visningsläge var det dock svårt att direkt bedöma bildkvalitén. En medtagen laptop användes för att granska fotografierna på plats, men begränsad zoomfunktion försvårade granskningen,

vilket i sin tur gjorde det svårt att avgöra hur bra bildkvaliteten blev. Efter genomförd fotograferingen sorterades det insamlade bildmaterialet, och testbilder som inte innehöll relevant innehåll exkluderades. Fotografierna samlades och organiserades i en gemensam molnbaserad lagringsmapp, på Google Drive, för att möjliggöra vidare bearbetning av Emma.

3. Ansvarsområden

Projektarbetets olika arbetsuppgifter delades upp i huvudansvar utifrån gruppmedlemmarnas respektive kompetensområden. Redigering av fotografierna påbörjades parallellt med transkribering av materialet. Sedan etablerades ett GitHub-repository, vilket baserades på en anpassad mall utformad av Wout Dillen. Marie har varit huvudansvarig för teknisk hantering av GitHub och kodning i TEI och XSL. Emma har hanterat fotoredigeringen och visualisering av hemsidan genom CSS, medan Amanda hade särskilt ansvar för rapporten och koordination av innehållet för den, samt skapade två tidslinjer och dokument för tidsestimering för projektet (se bilaga 1 och 2). Vi har gemensamt fattat beslut över vad vi vill uppnå med digitaliseringen och hur hemsidan ska se ut och bidra till varandras huvudsakliga ansvarsområden på olika sätt.

4. Transkribering

Transkriberingen genomfördes manuellt då textmängden var relativt begränsad och tillgängliga transkriberingsverktyg visade sig otillfredsställande för ändamålet. Den äldre skrivstilen och språkbruket innebar dock vissa tolkningssvårigheter. Efter att samtliga transkriberingar färdigställts granskade vi varandras texter och rättade de feltolkningar som identifierades.

5. Fotoredigering

Fotografierna av utgrävningsrapporten togs under begränsade förhållanden med utrustning tillgänglig på plats. Utrustning för digitalisering var begränsad på organisationen där vi fick utföra fotografering av utgrävningsrapporter. Till exempel saknades stativ för kameran och mer ljussättning, vilket resulterade i extra arbete för att i efterhand få till optimal digitalisering. Rapporten fotograferades från ovanstående position med hjälp av en pall, vilket medförde en sned vinkel och ojämn ljussättning trots användning av två

softbox-lampor. Underlaget (ett bord) visade spår av smuts och märken som råkat fångas på bild. Carmen Cowick (2018, s. 79) framhåller vikten av att använda en ren utrustning och arbetsyta för att säkerställa en så trogen återgivning som möjligt av originalmaterialet. Digital redigering av bakgrundsytan genomfördes för att avlägsna oregelbundenheter såsom damm, fläckar och ojämn belysning, i syfte att förbättra läsbarheten och skapa visuell enhetlighet mellan bilderna.

Bildfilerna sparades i RAW-format (NEF) och konverterades därefter till TIFF (16 bit, 300 dpi, sRGB) med hjälp av Darktable 5.0.1. I samband med denna konvertering uppstod oönskade färgförändringar: bilderna blev mörkare och färgmättnaden ökade vilket avvek från originalmaterialet. Detta upptäcktes senare i arbetsprocessen och åtgärdades genom färgjusteringar i Procreate 5.1.3 för JPEG-versionerna i syfte att återställa bildernas utseende så nära originalet som möjligt.

Det initiala efterarbetet påbörjades i gratisprogrammet GIMP. Dock visade det sig snart att programmets egenskaper inte motsvarade de behov som fanns för att uppnå önskad visuell kvalitet. För att kunna genomföra mer precisa färgjusteringar, perspektivkorrigering och retuscherings valdes istället Procreate. Procreate är en app av Apple som vanligtvis används för digital illustration. Trots att Procreate främst är avsedd för att illustrera, visade den sig ha de redigeringsverktyg som krävdes för detta projekt. Eftersom programmet redan fanns tillgängligt för privat bruk kunde det användas utan extra kostnad, och bidrog till att optimera bildernas visuella kvalitet inför publicering.

Eftersom belysningen vid fotograferingstillfället var ojämn behövdes manuell justering av ljusstyrka och färgtemperatur för att skapa visuell enhetlighet mellan bilderna. Skärpan förbättrades där det var möjligt och endast en bild uppvisar fortfarande viss oskärpa, men denna bedömdes som mindre störande efter redigering.

Vid införandet av bilderna i Github noterades att de digitaliserade bilderna, med sin kallvita bakgrund, var i stark kontrast mot den varmare vita tonen på webbplatsens gränssnitt. För att mildra denna visuella krock konverterades därför både de högupplösta JPEG-versionerna och de mindre JPEG-thumbnails till svartvitt. Detta minskade kontrasten och skapade en mer harmonisk helhet i den digitala presentationen.

I digitaliseringsarbetet framkom att metadata inte automatiskt följde med vid konvertering från NEF till TIFF i programmet Darktable. Den tekniska metadata, såsom kameramodell (NIKON D7100), datum samt grundläggande exponeringsinställningar behövde därför återskapas manuellt.

6. TEI XML, XSL och CSS

Inledningsvis låg fokus främst på *headern*, som är den del av dokumentet där metadata kring dokumentet och dess källor lagras för eventuell framtida användning. Utifrån projektets behov valdes element som bedömdes vara relevanta, `fileDesc` och `profileDesc`, båda tilldelades lämpliga underelement. Därefter genomfördes den första XSL-transformeringen av TEI XML-filen till HTML-filer för webbpresentation. Ett grovt utkast av den transkriberade texten i TEI-dokumentet skapades med styckeindelning samt ett utkast på line breaks.

Olika former av radbrytning provades under arbetets gång och slutligen beslutade gruppen att inte ha elementet `linebreak` i stor del av den löpande texten då det blev lättare att läsa. Amandas TEI-kod för understrykning lades in men fungerade ej först, efter kommunikation med kursledaren framgick det att alla element i TEI XML behöver ha en motsvarighet i de korresponderande XSL-filerna och det var där det hade blivit fel.

När det grova utkastet var färdigt kunde bilderna, JPEG samt JPEG-thumbnails, läggas in i TEI XML. Nästa steg var att med hjälp av XSL se till att materialet som fanns i XML-filen presenterades på ett önskvärt sätt på webbsidan. Då den förskapta mallen inte var fullt tillfredsställande utifrån vårt projekts behov, gjordes en hel del förändringar. Vi ville att olika delar av materialet skulle presenteras på olika separata webbflikar, samt inkludera den bakgrundsinformation som vi fann relevant för projektet på en egen flik. Vissa koder återfinns i alla XSL-filer då de är relevanta för alla flikar på webbsidan medan vissa koder är specifika för varje enskilt dokument, då de bara är lämpliga för enskilda webbflikar.

I slutskedet inkluderades även elementet `encodingDesc` med underelementet `projectDesc` i TEI headern. Detta då en beskrivning av syftet med projektet kändes relevant att både bevara i TEI XML samt göra tillgängligt på webbsidan. Likaså inkluderades copyright information i berörda dokument så att den informationen finns i TEI XML, Github repository samt lättåtkomligt på hemsidan. En länk till vårt Github repository lades även till på hemsidan.

Ändringar överfördes löpande till Github via Github Desktop, så Emma och Amanda kunde bidra med sina delar samt komma med förslag på ändringar och förbättringar gällande innehållet och hur det presenterades. I skrivande stund har över 350 commits gjorts med en eller flera ändringar som överförts till Github, vilket åskådliggör vilket omfattande arbete digitaliseringsprocessen har varit. Amanda har lagt in understrykningar och indrag i styckena

texten, samt korrekturläst och tittat på texten rent språkligt. Emma har haft ett särskilt ansvar över visualiseringen av hemsidan genom styling i CSS.

7. Upphovsrätt

Då utgrävningsrapporten inte omfattas av upphovsrätten, betraktas detta material som ett kulturarv öppen för spridning, vi har därför valt licensen CC BY 4.0, vilket är en Creative Commons-licens som tillåter andra att använda, distribuera, remixa, modifiera och bygga vidare på verket för vilket syfte som helst, så länge de anger upphovsmännen (Creative Commons, u.å.). Dock ville vår grupp även ge en bakgrund till Villands härad och Nosaby kyrka samt Olof Källström, och göra informationen tillgänglig i en egen flik på vår webbplats. Amanda skapade den informerade texten, samt fann ett porträttfotografi på Källström. Porträttfotografiet är upphovsrättsligt skyddat och används med tillstånd från Arkiv Gävleborg. Enligt överenskommelsen har vi försett fotografiet med vattenstämpel och arkivet anges som källa. Fotografiet av Källström används som ett komplement till bakgrunden för utgrävningsrapporten.

8. Analys

8.1 Motiv för digitalisering

Lars Björk (2015, s. 16-23) skriver att man vid digitalisering behöver ta ställning till vad som är viktigast att lyfta fram, textens innehåll eller de visuella aspekterna av materialet. Författaren menar dock att det är möjligt att kombinera båda fokus genom att både visa fotografier av materialet med en datorskriven text vid sidan om. Vår grupp har ett blandat fokus på materialitet och tillgänglighet. Vid redigeringen har vår ambition varit att betraktaren tydligt ska kunna se handskriften och fotografierna. Därför har vi tagit bort onödig bakgrund som inte ger betraktaren så mycket. Eftersom fotografierna i bilagorna är klistrade på sidorna i rapporten har vi valt att beskära bilden så att enbart fotografierna syns. Vi kan sägas ha gjort lite intrång på autenticiteten genom att ändra färgsättning för att det ska se bättre ut på webbsidan. Hade vi haft att göra med ett material där självaste verket och inte innehållet var det viktiga så hade vi tänkt annorlunda. Vi har lagt en hel del tid på det som Mats Dahlström (2011, s. 97-99) benämner intellektuell redigering, vilket innebär att innehållet i materialet förändras i syfte att förtydliga budskapet. Vi har gjort ändringar i vår

kodade text, till exempel är inte styckena indelade på samma sätt som i originaltexten. Vi har också tillfört information i form av bildtexter och bakgrundsinformation om platsen och författaren till rapporten, för att ge betraktaren ett tydligare sammanhang. Det innebär att vi flyttat oss en bit ifrån den ursprungliga rapporten och att våra egna tolkningar påverkar vad som blir viktigt att lyfta fram. Å andra sidan finns fotona på de ursprungliga dokumenten kvar.

Melissa Terras (2015, s. 737) skriver att digitalisering av ett material inte automatiskt bidrar till en ökad tillgänglighet. Författaren lyfter fram att materialet behöver distribueras på ett sätt som möjliggör återanvändning. Detta är något vi har haft i åtanke när vi har gjort förändringar i den transkriberade texten, och lagt in mer information om platsen och utgrävningsrapportens författare. Vi vill genom detta bidra till att rapporten blir mer lättläst och får ett tydligt sammanhang.

I en rapport om ett digitaliseringsprojekt inom kultursektorn faller det på sin plats att belysa projektets relevans för kulturarvsbevarande. Vi menar att en utgrävningsrapport från 1922 berättar något om svensk historia, i det här fallet om en plats i Skåne. Fynden i sig berättas det inte särskilt mycket om i rapporten och därför är det mer själva platsen och utgrävningen som blir viktig. Hur texten är formulerad, och vilka ord som används, säger mycket om det svenska språkets historia. Med lite efterforskning kan vi gräva i historien och ta reda på mer om platsen och den person som skrivit rapporten. På så sätt kan vi lyfta fram ännu mer historia.

Cowick (2018, s. 3-6) belyser flera viktiga fördelar med digitalisering av kulturarvsmaterial. En central aspekt är tillgången, då digitala objekt kan göras tillgänglig oberoende av fysisk plats eller öppettider. Digitalisering av materialet möjliggör åtkomst för användare i andra städer, länder eller tidszoner. Det skapar också nya möjligheter till samverkan mellan institutioner och kan öka synligheten och legitimiteten för en organisation i ansökningsprocesser om finansiering. Digitalisering kan dessutom fungera som ett skydd mot skador på det fysiska originalet. Om exempelvis ett föremål är skört, eller om en katastrof inträffar, som brand eller översvämning, kan ett digitalt exemplar säkerställa fortsatt åtkomst till informationen. Utgrävningsrapporten är ett praktexemplar för material som är skört. För att skydda det från eventuella skador blev valet att fotografera sidorna och inte skanna. Digitalisering möjliggör mer användning av rapportens innehåll utan ökad hantering och rapporten kan bevaras. Cowick (a.a., s. 3-6) lyfter just hur digitalisering minskar behovet av fysisk lagring och hantering av originalmaterial, vilket är särskilt värdefullt när

lagringsutrymme är begränsat. Genom att tillgängliggöra digitala versioner av material som annars riskerar att skadas vid användning, bidrar projekt som detta till att bevara innehållet för framtida forskning, utbildning och kulturförståelse.

Terras (2015, s. 738) skriver om hur digitalisering av kulturarv kan bidra till forskningen, då forskaren istället för att besöka platsen där materialet finns kan ta del av en digitaliserad version hemifrån. Författaren skriver även att lärare skulle kunna använda digitaliserat material i skolundervisningen. Även Deutsche Forschungsgemeinschaft (2013, s. 10), förkortat DFG, betonar att digitalisering av kulturarv har blivit ett verktyg som underlättar forskningen, och bidrar till en infrastruktur för en alltmer digital forskningsvärld. Enligt DFG är det viktigt att länka samman de olika resurserna och på så sätt skapa en virtuell infrastruktur för forskning. I vårt fall skulle en fortsättning av digitalisering av utgrävningsrapporter resultera i en större samling som blev möjlig att forska i, och att använda i syfte att lära om historia. För Historiska museet kan digitalisering av utgrävningsrapporter bli särskilt relevant som studiematerial, då Historiska museet tillhör Lunds universitet och kan bli användbara för såväl studenter i Lund samt studenter runt om i Skandinavien. Vi tror att en sådan samling skulle kunna få stor relevans inom ämnen som arkeologi, historia, geografi och historisk lingvistik. Terras (2015, s. 734) framför att upphovsrätten kan utgöra ett hinder vid digitalisering av kulturarv. Fördelen med äldre verk, likt den vi valde, är att de faller utanför upphovsrätt.

Björk (2015, s. 117-118) framför att digitalisering är en kostsam process, där det är viktigt att sätta det långsiktiga värdet i relation till kostnaden. Dessutom betonar författaren att teknisk utrustning har en kort livslängd, och att det behövs fungerande strategier för att underhålla tekniken och säkra långsiktig tillgänglighet (Björk, 2015, s. 153-154). Vår grupp har lagt väldigt mycket tid på att digitalisera (totalt 141 timmar och 55 min) en utgrävningsrapport på 9 sidor. För ett projekt i större skala hade ännu mer tid lagts. När digitalisering ingår i ett uppdrag på en arbetsplats, som till exempel en kulturinstitution, hade det kostat finansiella medel och personalresurser för arbetsgivaren. Även om vår grupp arbetat ideellt har projektet varit mycket tidskrävande. Med rätt förutsättningar hade dock processen kunnat effektiviseras. Exempelvis hade en mer avancerad fotograferingsutrustning, utgjord för att digitalisera verk av papper, med stativjustering och bildkvalitetskontroller kunnat minskat behovet av efterarbete. Vidare hade tillgång till mer avancerade bildredigeringsprogram, såsom Adobe Photoshop, kunnat underlätta det redaktionella arbetet, särskilt vid filhantering och färgjustering. Eftersom vi enbart hade tillgång till mindre

effektiva gratisprogram har mycket arbete behövts göras manuellt, vilket ytterligare förlängde projektiden.

8.2 Långsiktigt bevarande

DFG (2013, s. 8) påpekar att det inte finns någon universell lösning för långsiktigt bevarande som är tillämplig för alla typer av digitalt material. Istället betonas vikten av att lagra filer i säkra och stabila lagringssystem, där filerna är lätt överförbara mellan plattformar. En central utmaning med bevarande av digital data är att säkerställa att datan finns tillgänglig och möjlig att läsa och använda. Val av format för data och metadata är avgörande för den långsiktiga tillgängligheten. Enligt DFG:s riktlinjer (a.a. s. 33) ska fullständiga tryckta texter och manuskript vara kodade i TEI XML, då det ger de bästa förutsättningarna för långtidsarkivering, förutsatt att det är väldokumenterat. Den metadata som skapas i digitaliseringsprojektet bör vara en integrerad del av projektets arbetsflöde på så sätt att metadatan läggs in och uppdateras löpande. Det innebär att om projektet av någon anledning skulle avslutas i förtid kommer en komplett uppsättning metadata att finnas tillgänglig. (a.a. s. 26). Vidare betonar DFG (a.a. s. 34) vikten av att layouten ska kunna bevaras på lång sikt. Det rekommenderas därför att använda lämpligt formatspråk som exempelvis XSLT eller CSS. Vi har i vårt digitaliseringsprojekt följt dessa riktlinjer och tillämpat en språkstruktur för att möjliggöra framtida återanvändning och visning av material på webben. All text och metadata har strukturerats i TEI XML, och transformerats till HTML-filer via XSL och kommandot "format". Att hårdkoda delar av materialet direkt i HTML-filerna hade varit enklare, men det hade inte främjat långsiktigt bevarande. Avsteg har dock gjorts från DFG:s rekommendationer. Exempelvis anger DFG (a.a. s. 7) att metadata bör finnas tillgänglig på engelska. I detta projekt har endast svenska använts, då det är ett mindre projekt och all källmaterial är på svenska. Det bedöms därför rimligt att utgå från att även målgruppen behärskar det svenska språket.

DFG (2013, s. 41) radar upp grundläggande krav för att kunna navigera i digitaliserade dokument. Hemsidan ska enligt manualen erbjuda möjlighet att gå till önskad bild, bläddra mellan sidorna, hoppa tillbaka till början av dokumentet och även till slutet av dokumentet. Fulltextsökning ska vara möjligt och aktuell information om dokumentet (metadata) ska synas i beskrivningsfält lagrade i informationssystem. Det ska finnas en hjälpmeny som ger tydliga beskrivningar för navigering och sökning i den digitala samlingen. Om lämpligt kan man inkludera innehållsförteckning, strukturträd eller liknande och göra

dessas sökbara. Navigationshjälpmedel får gärna finnas med, som visar vilken del av dokumentet som användaren läser för tillfället. Vår grupp har frångått en hel del från dessa krav, med anledningen att vi endast digitaliserat ett dokument, som dessutom har ett begränsat antal sidor. Vi upplever att det ändå är lätt att navigera i dokumentet. Vi har lyft fram de delar av metadatan från XML-filen som känns relevanta för allmänheten på hemsidan, såsom `titleStmnt` och `publicationStmnt` och informationen om projektets syfte i `encodingDesc`. De som önskar mer detaljerad information om materialet eller processen kan gå in på vår Github som vi länkar till på webbsidan.

DFG (2013, s. 27) konstaterar att det inte är obligatoriskt att bädda in metadata direkt i bildfiler, men att detta kan vara ett värdefullt komplement. Eftersom olika program hanterar bildfiler på olika sätt, finns det en risk att metadata som lagras i headers kan försvinna eller förändras. Därför rekommenderas att metadatan även sparas i bildfilerna, exempelvis upphovsrätt, datum, plats, vilket kan öka tillgängligheten och användbarheten. Detta är särskilt produktivt om bilderna skulle användas fristående från systemet. Genom att placera ut metadata på flera platser möjliggörs en minskad risk för informationsförlust samt görs materialet hållbart över tid. Vår grupp har därför valt att inkludera så mycket information som möjligt direkt i bildfilerna, med fokus på kontext kring fotograferingen samt bildinnehåll. Detta bedöms som särskilt viktigt då bildfilerna kan komma att spridas eller användas utanför projektet, i samband med spridning enligt de principer som gäller för Creative Commons (u.å.).

Filerna i vårt projekt är sparade i Github Archive Program, vilket ger en försäkran att de går att få tag i och återskapa även om det inte skulle fungera att läsa själva hemsidan i webbläsare. GitHub Archive Program (u.å.) arbetar för att säkerställa långsiktigt bevarande, och källkoden är öppen där. Github skyddar data genom lagring av kopior i flera olika format och på olika platser. Github lagrar bl.a. kod på filmrullar i Arctic Code Vault, ett arkiv som ingår i Arctic Word Archive, en arkiveringsanläggning som ligger 250 m. in i ett arktiskt berg på Svalbard och är avsedd att bevara material i minst 1000 år (a.a.). Github kommer troligen även gå ur tiden någon gång, så troligtvis behöver man överföra filerna till något annat i framtiden.

8.3 OCR-tolkning

Simon Tanner (2004) beskriver hur OCR-tolkning används för att omvandla fotograferade texter till digital, maskinläsbar form som möjliggör redigering och sökning. Författaren

belyser dock att OCR har begränsningar när det gäller handskrivna dokument, eftersom variationer i handstil och otydliga bokstäver försvårar tolkningen. På grund av dessa tekniska hinder valde vår grupp att istället transkribera texten manuellt. Den transkriberade datorskrivna textversionen publicerades därefter på projektets webbplats där den är sökbar för användaren. Att materialet då är maskinläsbart underlättar navigationen, exempelvis genom webbläsarens inbyggda sökfunktion (vanligtvis aktiverad med tangentkombinationen Ctrl+F eller Cmd+F), vilket gör det möjligt att snabbt hitta specifika ord eller fraser i texten.

9. Synpunkter från uppdragsgivare

Historiska museet vid Lunds universitet gav oss ganska fria tyglar gällande hur materialet skulle digitaliseras och presenteras. Inför vårt besök i magasinet, visade vi några exempel på digitaliseringsprojekt som gjorts av studenter inom kursen tidigare år. Detta gjorde vi för att visa ungefär vad de kunde förvänta sig. Personalen på museet önskade utöver webbsidan, få tillgång till råmaterialet såsom TIFF bilderna, vilket enkelt gick att tillmötesgå då allt det finns öppet och tillgängligt i Github.

Vår kontaktperson på Historiska museet vid Lunds universitet fick ta del av materialet innan vi skickade in det och tyckte att det såg väldigt bra ut. Hon uppskattade att vi lagt in bakgrundsinformation och länkat till museets hemsida.

10. Utmaningar

Med facit i hand finns det genom hela projektets gång sådant som vi borde ha haft i åtanke tidigare, men det hade varit svårt för oss att veta på förhand då vi inte har gjort något liknande förut. Till exempel borde vi, innan vi besökte magasinet, kontrollerat så att vi hade ett fungerande program på datorn för att kunna zooma in och kontrollera fotografierna mer noggrant på plats. Vi borde även försökt få tag i ett stativ till fotograferingen. Det var inte förrän i efterhand vi insåg att det hade kunnat vara av vikt att vid fotografering av fotografierna lägga en linjal bredvid för att kunna visa storleken.

Vi la även mer tid på transkribering än vad som hade varit nödvändigt, genom att leta efter och testa olika transkriberingsprogram. Marie provade tre olika transkriberingsprogram: *Onlineocr.net*, *Transkribus* och *Google Drive*. Av de tre visade sig Google Drive vara bäst, trots att Transkribus riktar sig mer mot handskrivna texter. Även om Google Drive var det mest användbara blev endast ungefär en tredjedel rätt. Vi landade till slut i att manuell

transkribering skulle bli mest effektiv, särskilt med tanke på att det rör sig om en begränsad mängd text. I DFG (2013, s. 33) står det att manuell transkribering har fördelen av hög precision, men är ekonomiskt kostsamt. Eftersom det i vår fall rörde sig om ett mindre omfattande material inom en universitetskurs, utgjorde de ekonomiska kostnaderna inte något hinder.

Ett annat dilemma vi stötte på under arbetets gång rörde bildbeskrivningen av de bifogade fotografierna. Fotografierna är svartvita med varierande upplösning, vilket försvårade identifieringen av avbildade objekt. Då Källström inte skrivit någon beskrivande text till fotografierna ökade osäkerheten kring vad de föreställer. På ett fotografi (Källström, 1922, s. 8) uppstod osäkerhet kring om det är en och samma kruka som fotograferats från olika vinklar, eller två skilda krukor. I hopp om att hitta mer information som skulle kunna fastställa vilket det rörde sig om, undersöktes Riksarkivets digitala samlingar utan resultat. Efter noggrann granskning av fotografierna, och diskussioner inom gruppen, där vi lyfte skillnader i ytstruktur samt bevarandestatus, valde vi att tolka fotografiet som två separata krukor. Vi är medvetna om att vår tolkning är preliminär och att mer säker kunskap hade kunnat erhållas i samråd med en expert inom området. På grund av projektets begränsade tidsramar var dock detta inte möjligt att genomföra. Ett annat fotografi (a.a., s. 9) visar en person som undersöker fynd från en utgrävningsplats. Personens ansikte är dolt och därför blev identifiering av personen inte möjlig.

Marie, som har haft det huvudsakliga ansvaret över kodningen, stötte på en hel del utmaningar längs vägen. Att lära sig koda i TEI och XSL, samt hur Github fungerar tog förstås en hel del tid i anspråk och att förstå sig på XSL visade sig vara den största utmaningen. Marie fick där först sätta sig in i hur XSL fungerade, dels genom videor och exempel från kursens litteratur, men även genom att själv söka svar på nätet. Det bästa sättet visade sig dock vara att prova sig fram själv genom att intuitivt ändra och modifiera koderna i XSL-filerna och testa olika funktioner tills önskat resultat uppstod. Men det var långt ifrån alla gånger det blev som tänkt. Hon spenderade till exempel större delen av en arbetsdag åt att prova olika funktioner i Github och XSL, och råkade genom detta orsaka en bug, som hon fick lägga tid på att åtgärda. Vid ett annat tillfälle i projektets slutskede försökte hon byta namn på mappen som allt material låg i, vilket ledde till att Github hängde sig, och hon godkände då av misstag ett kommando som gjorde att Github återställde sig till en punkt där projektet befann sig för mer än en månad tidigare. Lyckligtvis fanns en uppdaterad lokal kopia av mappen, så det gick att skapa ett nytt repository med de filerna.

Ett annat problem som uppstod var att hemsidan såg annorlunda ut i webbläsaren på mobilen jämfört med hur det såg ut på datorn. I mobilläget överlappade texten på hemfliken ett fotografi och indragen i transkriberingen blev fel. Efter att Emma insett att bildformatet var fel rättade Marie till det i XSL och Emma kunde då snygga till det ytterligare i CSS. Indragen konstaterades bero på att radbrytningen är olika i webbläge och mobilläge och vi hade anpassat indent elementet i TEI XML efter radbrytningen i webbsideläget.

11. Slutsatser

Vi har förstått att digitalisering är en mycket tidskrävande, och i vanliga fall även kostsam process. Det har varit väldigt lärorik att få vara med och skapa ett sådant projekt. Vid startpunkten för projektet hade vi ingen tidigare erfarenhet av att digitalisera ett kulturarv och vi hade begränsade kunskaper i kodning. På mindre än en termin har vi lärt oss tillräckligt för att skapa en fungerande hemsida där en bortglömd utgrävningsrapport, som är över hundra år gammal, kan lyftas och göras tillgänglig långt utanför magasinets gränser.

(Antal ord: 4323)

12. Referenser

Björk, L. (2015). *How reproductive is a reproduction? Digital transmission of text-based documents*. University of Borås.

<https://hb.diva-portal.org/smash/get/diva2:860844/INSIDE01.pdf>

Cowick, C. (2018). *Digital curations projects made easy : a step-by-step guide for libraries, archives, and museums*. Rowman & Littlefield.

Creative Commons. (u.å.). *Erkännande 4.0 Internationell sammanfattning*. Hämtad 2025-05-23, från <https://creativecommons.org/licenses/by/4.0/deed.sv>

Dahlström, M. (2011). Editing libraries. I: C. Fritze, F. Fischer, P. Sahle & M. Rehbein (Hrsgg.), *Bibliothek und Wissenschaft. Vol. 44: Digitale Edition und Forschungsbibliothek*. Harrassowitz. 91-106.

Deutsche Forschungsgemeinschaft (2013). *Practical Guidelines on Digitisation*.
https://www.dfg.de/resource/blob/176110/76abec10bdc30b41f18695145003d6db/12-151-v12_16-en-data.pdf

Gierow, K. (1940). Skåningar i Stockholm. [Elektronisk resurs] / med en inledning: Skåningar i Stockholm i gången tid. Skånetryckeriets förlag, Malmö.

GitHub, (u.å.). *The Arctic Code Vault*. <https://archiveprogram.github.com/arctic-vault/>

Härad. (u.å.). I *Nationalencyklopedin*. Hämtad 2025-04-28, från
<https://www.ne.se/uppslagsverk/encyklopedi/l%C3%A5ng/h%C3%A4rad>

Källström, O. (1922). *Grävning vid NOSABY, Villards härad*. Historiska museet vid Lunds universitet.

Svenska kyrkan, (u.å.). *Om Nosaby kyrka*. Hämtad 25/5-2025, från
<https://www.svenskakyrkan.se/platser/7587-nosaby-forsamling-nosaby-kyrka>

Tanner, Simon (2004). *Deciding whether Optical Character Recognition is feasible*. King's College. https://www.researchgate.net/profile/Simon-Tanner/publication/265748433_Deciding_whether_Optical_Character_Recognition_is_feasible/links/5bd7475e92851c6b27972809/Deciding-whether-Optical-Character-Recognition-is-feasible.pdf

Terras, M. (2015). Opening Access to collections: the making and using of open digitised cultural content. *Online Information Review*, 39(5), 733-752.

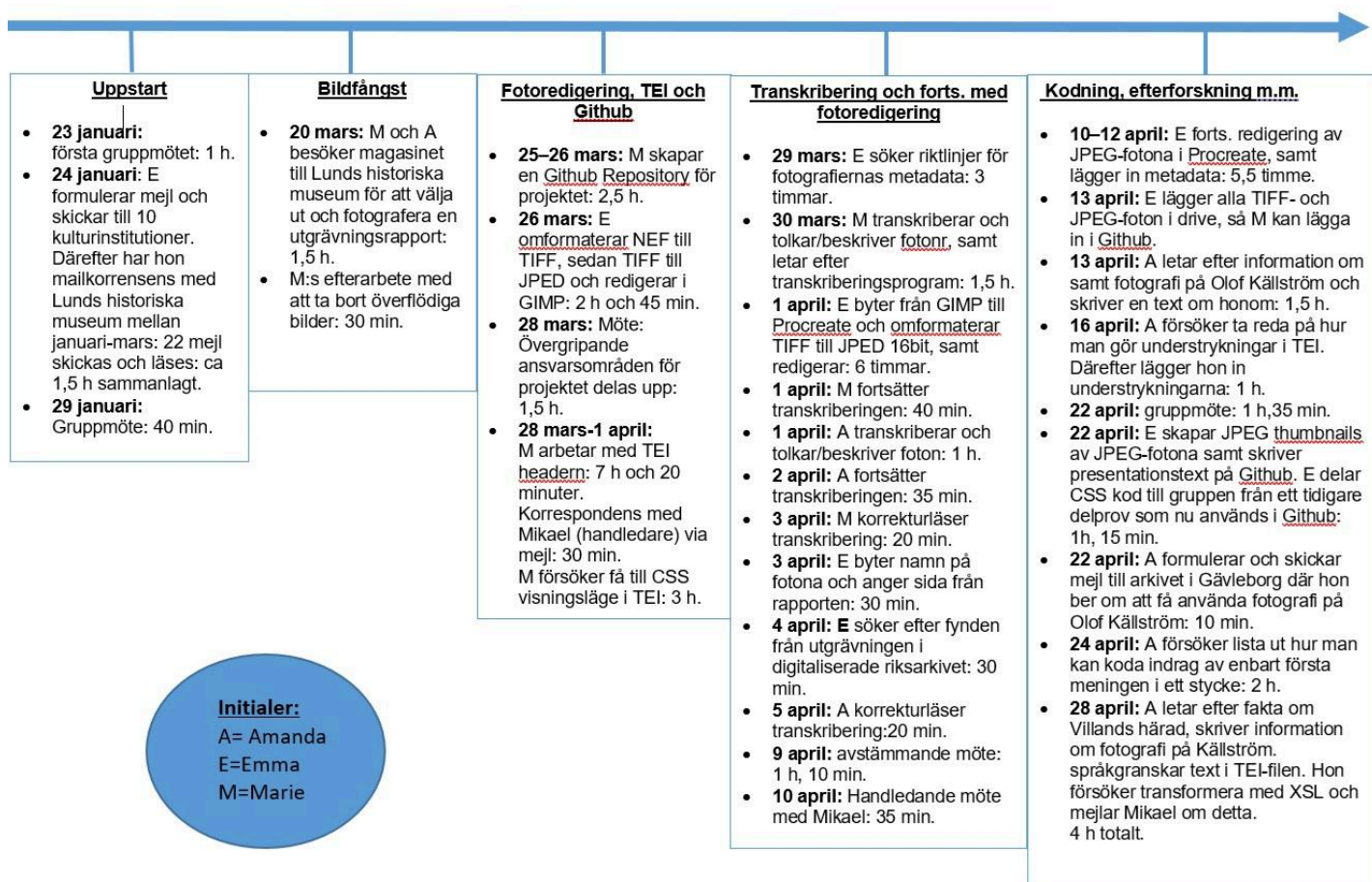
<https://doi.org/10.1108/OIR-06-2015-0193>

Villands härad (u.å). I *Nationalencyklopedin*. Hämtad 2025-04-28, från

<https://www.ne.se/uppslagsverk/encyklopedi/l%C3%A5ng/villands-h%C3%A4rad>

13. Bilagor

13.1 Bilaga 1: Tidslinje



(*fortsättning på nästa sida)

Visualliserings och rapportskrivning	Fortsättning på rapportskrivning, samt justeringar av webbsidan	Slutskedet av projektet
<ul style="list-style-type: none"> • 29 april: A börjar skriva på rapporten: 2 h. E tittar på tidigare projekt och skapar en skiss för visualisering av vårt projekt: 1 h. E sätter <u>waterstamp</u> på fotot från Arkiv Gävleborg: 10 min. E testar att redigera i <u>Github</u> CSS och XSL: 3 h. M <u>Lägger</u> in bakgrundsinformation om Olof Källström i XML-filen, jobbar med TEI <u>headern</u> och hemsidans layout. 3 h och 15 min. 3 h och 15 min. • 30 april: M kodar i XSL samt granskar och snyggar till XSL-dokumentet. Lägger även in en länk till <u>github</u> på webbsidan genom att <u>hårdkoda</u> i html 1 h, 20 min. • E ändrar alla JPEG-foton från färg till svartvitt: 30 min. • 2 maj: E <u>croppar</u> foton och redigerar i CSS-dokumentet: 1 h, 25 min. A försöker med hjälp av M lösa återkommande problem med transformering i Oxygen. A kan inte se alla delar av hemsidan i visningsläge i webbläsaren. A byter ut länkar till referenser som blivit fel, samt tar bort radbrytning (<u>Lb</u>) i XML-dokumentet för att se om det ser bättre ut: 1,5 h. E <u>croppar</u> foton: 45 min. E redigerar <u>css</u>: 40 min. • M snyggar till i XSL-dokumentet och justerar hur texten presenteras på hemsidans första sida. 40 min. • 5 maj: A skriver på rapporten: 40 min. A skapar en tidslinje för arbetsprocessen, i Microsoft Word och börjar fylla i den: 1 h, 10 min. E uppdaterar metadata i TEI: 40 min. E skriver på rapporten: 4 h. • 6 maj: A fortsätter med tidslinjen. 1 h, 40 min. A skriver på rapporten: 40 min. • 7 maj: A skriver på och redigerar i rapporten: 2 h. • 8 maj: A skriver på rapporten: 1 h. 	<ul style="list-style-type: none"> • 9 maj: A skrivit på rapporten: 1 h. A påbörjar dokument för sammanfattande tidsestimering: 10 min. M ändrar namn på <u>Github</u> och webbsidan. M korrekturläser TEI-dokumentet, kursiverar delar av transkriberingen och gör ett par mindre linebreak-justeringar i TEI: 45 min. M skriver på rapporten: 2 h, 50 min. M letar efter och hittar en passande copyright <u>license</u> och lägger in information om det i de berörda filerna. 55 min. • 10 maj: E skriver på rapporten: 1 h. • 11 maj: E skriver på rapporten: 2 h. • 12 maj: M gör små ändringar i TEI-dokumentet, samt fixar med valideringsfel i html-filerna: 1 h, 5 min. E skriver på rapporten: 10 min. • 13 maj: A fortsätter skriva ned och räkna ihop tid vi lagt på olika moment: 1 h, 20 min. • A tittar på hemsidan och funderat över ändringar: 10 min. M gör några mindre ändringar i html-filerna efter Mikael's förslag. M försöker byta namn på mappen för allt material och godkänner då ett kommando som dessvärre resulterar i att återställning sker till hur hemsidan såg ut för över en månad sedan. Genom en sparad uppdaterad kopia av mappen lyckas M skapa ett nytt <u>repository</u> med dessa filer: 2 h och 40 min. • 14 maj: A skriver på rapporten: 1,5 h. E skriver på rapporten: 10 min. • 15 maj: A skriver på rapporten: 50 min. A räknar på tidsestimering: 35 min. • 19 maj: A skriver på rapporten: 2 h. • 20 maj: A skriver på rapporten: 1,5 h. • M rättar till styckesindelningen för texten i hemfilen: 1 h 35 min. Samt skriver på rapporten: 1 h 35 min. 	<ul style="list-style-type: none"> • 21 maj: Gruppmöte för avstämning: 1 h, 30 min • M arbetar i samråd med E med att webbsidan ska se ut korrekt ut i webbläsare i mobiltelefon. M arbetar i TEI XML. 3 h, 50 min. M skriver på rapporten: 3 h och 15 min. E skriver på rapporten: 2 h. • 22 maj: E redigerar CSS: 1,5 h. • 23 maj: A skriver på rapporten: 2 h, 50 min. M lägger in info om copyright i <u>readme-filen</u>: 10 min. Samt i rapporten: 10 min. E skriver på rapporten: 5 min. • 24 maj: A skriver på rapporten: 1,5 h. M lägger in info om metadata på hemsidan i rapporten, och gör småändringar i TEI XML: 25 min. • 25 maj: A letar efter fakta och skriver en text om <u>Nosaby</u> kyrka: 35 min. E fotar <u>Nosabys</u> kyrka, redigerar, lägger in metadata och laddar upp fotot på <u>Github</u>: 25 min. A skriver på rapporten: 45 min. M lägger in fotografi som E tagit på <u>Nosaby</u> kyrka, samt A:s text om kyrkan i TEI-dokumentet. • 26 maj: A forts med tidslinjerna: 1 h, 15 min. A forts med tidsestimering: 55 min. A skriver på rapporten: 40 min. E redigerar bildformat och laddar upp på <u>Github</u>. E skriver på rapporten: 3 h. M ändrar bildformat på E:s foto så att webbläsaren kan läsa den: 30 min. M skriver på rapporten: 1 h. • 27 maj: A fixar det sista med tidslinjerna: 1 h. A räknar ihop det sista för tidsestimeringen: 45 min. A skriver och redigerar det sista på rapporten: 45 min • E skriver och redigerar det sista på rapporten: 45 min • M skriver och redigerar det sista på rapporten: 45 min

13.2 Bilaga 2: Tidsestimering

Gruppmöten:

1 h + 40 min + 1 h, 30 min + 1 h, 10 min + 35 min + 1 h, 35 min + 1h, 30 min

8 timmar x 3 = 24 timmar totalt

Efterforskning och skrivning av text till bakgrundsflik

1h 30 min + 2h 30min + 35 min

4 timmar och 35 min totalt

Sökning efter fynden i riksarkivets digitala samling:

30 min

Mailkorrespondens med Historiska museet vid Lunds universitet:

1 timme och 30 min

Bildfångst:

2 timme och 15 min (räknade också in efterarbetet här, samt fotografering vid kyrkan)

Fotoredigering och behandling av dess metadata, riktlinjer m.m:

3 h + 2 h, 45 min + 6 h + 30 min + 5 h, 30 min + 1 h, 15 min + 10 min + 30 min + 10 min +
10 min+30 min + 45 min +40 min

21 timmar och 55 minuter totalt

Transkribering:

1 h och 30 min + 40 min + 1 h och 35 min + 20 min + 20 min

4 timmar och 25 minuter totalt

Github och kodning i TEI, XSL och CSS:

2 h, 30 min + 7 h, 50 min + 3 timmar + 1 h + 2 h + 1 h + 3 h, 15 min + 1 h, 20 min + 1 h, 30
min + 40 min + 45 min + 55 min + 2 h, 40 min + 3 h + 40 min + 1 h, 30 min + 10 min + 25
min + 30 min (Inklusive mailkontakt med handledare)

34 timmar och 40 minuter totalt

Språkgranskning, allmän koll av hemsidan:

1 h + 10 min +10 min

1 timme och 20 min totalt

Rapportskrivning:

2 h + 40 min + 4 h + 40 min + 2 h + 1h + 1 h + 2h, 50 min + 1h + 2h + 10 min + 1h, 30 min +
10 min + 50 min + 2h+ 1h, 30 min +1h, 30 min + 3h, 15 min + 2 h, 50 min + 5 min + 1h, 30
min + 25 min + 45 min + 40 min + 3 h + 1 h + 45 min + 45 min + 45 min

40 timmar och 35 minuter totalt

Tidslinjen:

1 h, 10 min + 1 h, 40 min + 1h, 15 min + 1 h

5 timmar och 5 min totalt

Dokument för tidsestimering:

10 min + 1h, 20 min + 35 min + 15 min + 45 min

3 timmar och 5 minuter totalt

Total tid för hela digitaliseringsprojektet:

*141 timmar och 55 minuter totalt**

*Utöver denna tid har vi haft kontinuerlig kontakt om projektet i Messenger, men detta är svårt att tidsbedömma och står därför inte med. Läsning av kurslitteratur står inte heller med. En del extra läsning har behövts för att vi ska lära oss moment till projektet.

För att kunna slutföra tidslinjer och tidsestimering har vi behövt sluta räkna i tid. Därför har en del arbete som gjorts sista dagen inte kunnat skrivas in.