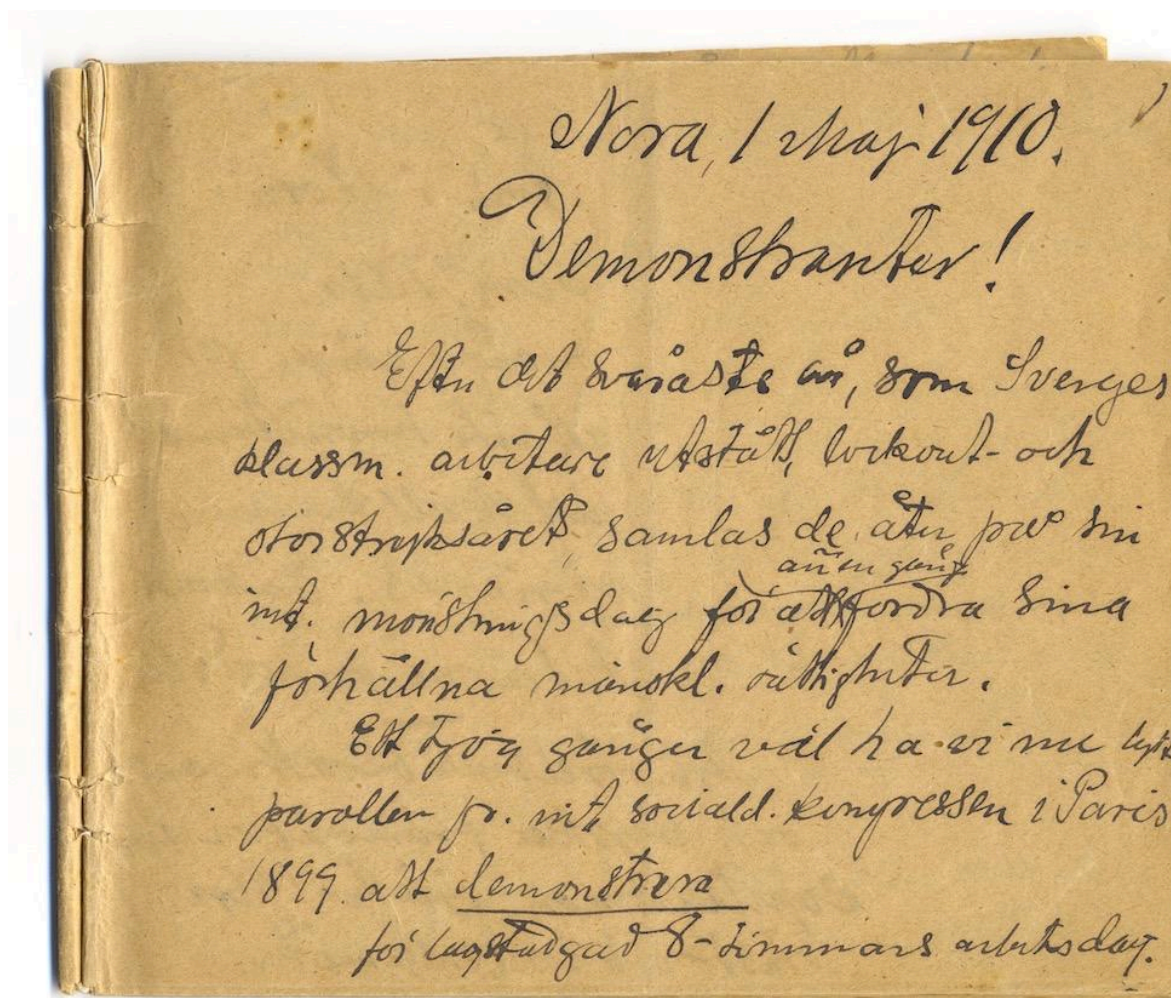


”Demonstrera alla dagar!”

Digitalisering av Elma Danielssons manuskript



Projektrapport Digitalisering för bevarande och tillgänglighet

Högskolan i Borås VT 2025

Rebecca Lager

Emilia Nilsson

Peeter Sällström Randsalu

1. Inledning

Inom ramen för detta projekt har vi, i samarbete med Arbetarrörelsens arkiv i Skåne, digitaliserat två av Elma Danielssons handskrivna manuskript: ”1:a majtal i Nora 1910” och ”Tal i Askersund från den 8 maj 1910”.

Vi har skannat och transkriberat dessa handskrifter, och med hjälp av TEI skapat en textkritisk, digital utgåva av manuskripten.

Manuskripten finns nu bevarade både i sin tidigare fysiska form, och i sin nya digitala form. Tillsammans med Arbetarrörelsens arkiv i Skåne kommer vi att se till att de digitala bilderna och dess metadata arkiveras på deras servrar på ett sätt som innebär att långsiktigt bevarande säkras.

Vi kommer även, tillsammans med arkivet, se till att den digitala utgåvan av manuskripten finns tillgänglig via arkivets egen hemsida, men just nu finns den tillgänglig på vår egen Github (<https://p33t3r.github.io/digitalisering/>). Här finns inskannade bilder av manuskripten, diplomatarisk transkription av texten, och en moderniserad normaliserad transkription. Kopplat till detta finns kommentarer, referenser och länkar. Där finns också texter som ger en bild av Elma Danielssons (politiska) liv, som vi har skrivit genom att gå till andra källor.¹

1.1 Arbetarrörelsens arkiv

Arbetarrörelsens arkiv i Skåne är en ideell förening, som i sina lokaler i Malmö hyser en av södra Sveriges ”mest omfattande samlingar av [a]rbetarrörelsehistoria”, och genom att medlemsorganisationer fortsätter att deponera material, så fortsätter samlingarna att växa. Forskare såväl som allmänhet är välkomna till deras lokaler, för att titta på materialet i samlingarna. (Arbetarrörelsens arkiv i Skåne, u.å)

En av de mer välbesökta samlingarna är Axel och Elma Danielssons personarkiv. Tidigare var det framför allt det material som gällde Axel Danielsson, en av den tidiga arbetarrörelsens förgrundsgestalter, som efterfrågades i arkivet, men efter att en staty av Elma Danielsson tog plats i Rörsjöparken i Malmö i maj 2023 – vilket rönt en del uppmärksamhet i media – riktades intresset mer och mer mot Elma Danielsson, och idag är hennes manuskript det mest efterfrågade materialet i arkivet.

Endast en mycket liten del av arkivets material finns i digitaliserad form, då resurser för ett mer omfattande digitaliseringsprojekt saknats.

1.2 Elma Danielsson

Elma Danielsson var en pionjär inom den svenska arbetarrörelsen. Genom sitt politiska och journalistiska arbete var hon en stark röst för socialismen och för kvinnors rättigheter. Hon var med och grundade flera kvinnliga fackföreningar och kvinnoklubbar, till exempel grundade

¹ Framför allt Anna-Maria Lindgren och Marika Lindgren Åsbrinks bok *Systrar, kamrater! Arbetarrörelsens kvinnliga pionjärer* (2007).

hon 1888 arbetarrörelsens första politiska kvinnoförening: Malmö Kvinnliga Arbetareförbund, där hon även var ordförande. På socialdemokratiska partikongressen 1905 och 1908 var Elma Danielsson ombud, liksom på kvinnokongressen 1908 (Lindgren & Lindgren Åsbrink, 2007).

Elma Danielsson som en skicklig talare. Under 1900-talets början anlätades hon ofta för att hålla föredrag hos partiorganisationer och fackföreningar, inte bara i Skåne utan på många ställen i landet, och hon ”lockade många åhörare, med föredrag som täckte ett brett register av frågor och ämnesområden” (Svenskt kvinnobiografiskt lexikon, 2018).

2. Dokumentation av tillvägagångssätt

2.1 Val av material

När vi fick kontakt med Arbetarrörelsens arkiv i Skåne, var de mycket intresserade av vad ett samarbete kunde innebära. Endast en mycket liten del av arkivets material finns i digitaliserad form, då resurser för ett mer omfattande digitaliseringsprojekt saknats, och de hade flera förslag på möjliga material.

Eftersom syftet med projektarbetet var att genomföra ett småskaligt digitaliseringsarbete, med fokus på djup och kvalitet snarare än kvantitet, behövde vi avgränsa materialet.

Cowick (2018) menar att en av de första sakerna man behöver göra är att bestämma vilka urvalskriterier man utgår från när man väljer material. Det bästa sättet att göra det, menar hon, är att utvärdera möjliga material utifrån parametrarna *värde, efterfrågan, skick, rättsliga och etiska överväganden* och *tillgänglighet* (Cowick, 2018, s. 12).

Utifrån dessa parametrar fanns det framför allt två material som var aktuella: Axel Danielssons dagböcker och Elma Danielssons manuskript. Dessa material uppfyllde samtliga kriterier: de är unika objekt med stort historiskt värde, de efterfrågas ofta av forskare och allmänhet, de är i skört skick som riskerar att försämrast av ytterligare användning, de skyddas inte av upphovsrätt, och slutligen finns inte dessa objekt digitalt tillgängliga sedan tidigare.

Axel Danielssons dagböcker var dock i ett så pass skört skick att vi bedömde att det skulle behövas mer avancerad teknisk utrustning än den vi hade tillgång till, varför valet föll på Elma Danielssons manuskript.

Dessa manuskript består av fyra separata handlingar, i den fysiska formen av enkla rektangulära block/häften på ca 16 gånger 19 centimeter. Blocken är av papper och är sköra framför allt i bindningen. Sidorna har veck och nötta kanter, som smular sig vid oförsiktig hantering.

Texten är skriven för hand med svart bläck och med blyerts.

Efter övervägande tillsammans med handledare beslöt vi att välja endast ett av talen, bland annat för att ordentligt kunna utforska möjligheterna att skapa mervärde kring materialet med hjälp av TEI.

Vi valde ut det första av Elma Danielssons manuskript (manuskript A1) ”1:a majtal i Nora 1910”, dels för att det var det av manuskripten som textmässigt kändes mest sammanhållet, dels för att det kändes mindre skört än de övriga.

Senare i processen upptäckte vi innehållsliga och formmässiga kopplingar med ett av de andra manuskripten, och valde därför att digitalisera även ”Tal i Askersund från den 8 maj 1910” (manuskript A2).

2.2 Val av teknisk utrustning

I enlighet med Cowicks (2018) rekommendationer ville vi säkerställa att vi hade den hårdvara och mjukvara som behövdes för arbetet med det material som skulle digitaliseras (Cowick, 2018, s. 32).

Vi beslutade att genomföra bildfångsten med arkivets egen skanner, en Epson Perfection 4490. Skannern blev rengjord med trasa och glasputs för att få bort smuts och fettfläckar från glaset. I övrigt krävdes inga åtgärder.

Datorn som vi använde, och som hade den skannerprogramvara och de drivrutiner som behövdes, tillhörde en av oss i gruppen. Vi säkerställde, i enlighet med Cowick (2018), att vi hade tillräckligt utrymme på hårddisken för att arkivera filerna, samt tillräckligt med digitalt lagringsutrymme för att spara kopior på filerna (Cowick, 2018, s. 32).

2.3 Pilotförsök

För att få en bild av vilka utmaningar som bildfångst och textfångst skulle innebära valde vi att genomföra en provskanning och testade sedan att använda automatisk textigenkänning, OCR, för att göra texten maskinläsbar.

2.3.1 Provskanning och beslut kring bildfångst

Vi hade gärna skannat blocken i hela uppslag, men då dessa inte fick plats i skannern, beslutade vi att skanna blocken sida för sida i stället. Vi bestämde dock att bindningen skulle finnas med på varje bild, för att i någon mån bevara känslan av ett block, samt möjliggöra för framtida digitalt montage av sidorna.²

För att testa olika inställningar för bildfångsten valde vi ut sidor med olika kvalitet från manuskript A1: en sida med skrift i bläck, en med skrift i blyerts.

Cowick (2018, s. 80) hävdar att sällsynta böcker bör skannas med minst 300 dpi, men helst mer. Även Deutsche Forschungsgemeinschaft (2013, s. 6) rekommenderar minst 300 dpi. Vi provskannade sidorna med 1 200 dpi, 800 dpi, 600 dpi och 300 dpi. Värdet av hög bildkvalitet vägdes mot storleken på bildfilerna.³

² Ett sätt att kompensera något för förlusten av originalets materialitet. Björk (2015) skriver: ”The materiality of a document is arguably the dimension that is most obviously affected when the document is transferred to digital format.” (Björk, 2015, s. 203)

³ Vi bedömde det inte som troligt, men vi ville även optimera möjligheten att kunna använda OCR.

| Upplösning | Filstorlek på testfilerna | Anmärkningar |
|------------|---------------------------|---|
| 1 200 dpi | 157 MB | För stor fil för att arbeta med |
| 800 dpi | 73 MB | Orimlig filstorlek i förhållande till bildkvaliteten |
| 600 dpi | 39 MB | Optimal avvägning bildkvalitet/filstorlek |
| 300 dpi | 10 MB | Ej rimlig kvalitetssänkning för långsiktigt bevarande |

Tabell 1: Skanningsinställningar

Bilder skannade med 800 dpi hade varit hanterbara för lagring av det utvalda materialet, men med tanke på skalbarhet valde vi att skanna med 600 dpi, vilket genererade bilder som upptog nästan hälften så stort lagringsutrymme. Vid 600 dpi var dock bildkvaliteten mycket god, så att inte bara texten utan även papprets kvaliteter tydligt framträdde.

Vi provade även skanning med olika färgrymder, där vi eftersträvade likhet med originalet,⁴ och valde den ”neutrala” färgbalans som hörde till skannerns automatiska inställningar.

2.3.2 Försök till textfångst med OCR

Enligt Tanner (2004) är det, i de flesta fall, inte kostnadseffektivt att använda OCR (Optical Character Recognition) på handskriven text, då handskrift bara låter sig fångas under “very controlled circumstances and not usually as an automated activity” (Tanner, 2004, s. 9).⁵

Eftersom det har skett stora tekniska framsteg på området sedan 2004, ville vi ändå undersöka möjligheten att använda någon typ av handskriftsigenkänning (HTR) för att underlätta arbetet. Vi provade Riksarkivets verktyg HTRflow (Riksarkivet, 2025), både med en lokal installation samt den demo som finns tillgänglig på https://huggingface.co/spaces/Riksarkivet/htr_demo.

Vi provade med flera olika versioner av de skannade filerna, med olika upplösningar och olika färgrymder, men det visade sig att ingen av dessa modeller producerade text som gick att använda utan mycket omfattande och helt genomgripande efterbearbetning, varpå vi beslöt att göra transkriptionerna manuellt. Se även bilaga 3.

2.4 Bildfångst

Skanningen pågick under två förmiddagar. En person ansvarade för materialet, och en person för datorn och inställningar för skannern, att namnge och spara filer och göra backuper. Då vi

⁴ Även här ville vi optimera möjligheterna för OCR.

⁵ Enligt Tanner (2004) kan man överväga att använda OCR även om ord- eller teckenkorrekthet är relativt låg, om syftet till exempel är *indexering* eller *full text retrieval* (Tanner, 2004, s. 3f). Detta var dock inte aktuellt för oss.

i detta skede inte var helt på det klara med hur texterna i blocken hängde ihop, gav vi bildfilerna provisoriska namn och sidnummer.⁶

Manuskript A1 och A2 skannades in enligt valda inställningar på skannern.

Blocken skannades i mitten av skannern för att tillåta beskärning av bildfilen.

Bilderna behövde inte efterredigeras. Vi skannade med ett par millimeters ”ram” omkring, och använde samma storlek för alla bilder.

Vi kontrollerade att varje skannad bild blev godtagbar. I några fall godtog vi sidor med viss lutning, då det inte var möjligt att trycka ned bindningen utan risk att skada materialet.

Längre fram i processen undersökte vi möjligheten att skanna in sammanlagt tre sidor på nytt. En sida hade ett vikt hörn, som visade sig dölja en bokstav. Denna sida skannade vi in på nytt, och lät bildfilen ersätta den gamla. Men en annan sida med ett vikt hörn, och en sida med ett litet hål i mitten med en flik vikt mot baksidan, lät sig inte skannas om, då pappret visade sig vara alltför känsligt för hantering. Vi kunde dock säkerställa att dessa vikningar inte dolde någon skrift.

2.4.1 Angående långsiktigt bevarande av bildfilerna

Cowick (2018) betonar vikten av att välja ett stabilt, utbrett filformat som stödjer inkludering av metadata (Cowick, 2018, s. 43). Bilderna sparades därför som TIFF-filer. Enligt Cowick ska man ska välja klara och tydliga filnamn, som innehåller tillräckligt mycket information för att det ska vara möjligt att förstå vad filen innehåller (Cowick, 2018, s. 37). Vi valde därför att namnge bilderna utifrån de beteckningar som används hos Riksarkivet, där materialet finns katalogiserat under referenskod SE/MSA/AA001964/F:V (Riksarkivet, u.å.). Sålunda fick bildfilerna namn enligt följande mall:

AA001964_F_V_Elma_Danielsson_A1_Tal_Nora_1910_bild01.tif

Då de skannade filerna var för stora för att lägga upp på Github,⁷ valde vi att ladda upp TIFF-filerna i originalstorlek till Internet Archive (2025) och länka till dem från hemsidan.⁸ Tanken är dock att allt material ska kunna lagras långsiktigt hos Arbetarrörelsens arkiv i Skåne.

Bilderna sparades även som betydligt mindre JPEG-filer⁹ på Github. Storleken var en avvägning för att bilderna skulle kunna ses på hemsidan med tillräckligt hög upplösning för att urskilja detaljer, men också vara små nog för att göra hemsidan tillgänglig.

⁶ Manuskript A1 är vikt så att talets första sida utgör blockets första sida. Själva blockets första sida är i själva verket en annan, men vi beslöt att i sidnumreringen utgå från talet, eftersom vikningen antyder att det är så blocket använts, både av Elma Danielsson själv och av dem som bevarat och använt blocket.

⁷ Ca 40 MB per fil för drygt 100 filer.

⁸ Internet Archive säger själva om sitt uppdrag att: ”The Internet Archive [...] is building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, people with print disabilities, and the general public. Our mission is to provide Universal Access to All Knowledge.” (Internet Archive, 2025).

⁹ 1 000 x 840 pixlar (ca 400 kB)

2.5 Textfångst

2.5.1 Transkription

Transkriptionen gjordes i ett delat kalkylark, där en person gjorde den primära transkriptionen, och en person korrekturläste/transkriberade parallellt och lämnade kommentarer i samma dokument. I ett sista skede korrekturlästes texten av den tredje av gruppens medlemmar.

Frågor som uppstod rörde tillvägagångssättet vid och nivå på transkriptionen (se diskussion i analysdel).

För att åstadkomma en diplomatarisk transkription ville vi fånga så många som möjligt av originaltextens egenskaper.

Handskriften återgavs med de grafem som använts (bokstäver, skiljetecken, siffror), den stavning som använts (inklusive stavfel och förkortningar), och med de ändringar, överstrykningar och tillägg som gjorts av författaren. Transkriptionen återgav även styckeindelningar och radbyten.

Tolkningsförfarandet var att gå från del till helhet, och från helhet till del, i en tolkningsspiral: att förstå innehållet genom kontexten, och kontexten genom innehållet; att förstå meningar genom stycken, att förstå ord genom meningar, och att förstå bokstäver genom ord, och tvärtom. I en handskrift kan de enskilda bokstäverna, tagna ur sitt sammanhang, vara obegripliga. Men genom att utläsa ordet, kan bokstäverna identifieras.

Om ett ord eller en bokstav var otydlig eller oläslig, så markerades detta.

2.5.2 Kodning med TEI

Efter transkription och korrekturläsning överfördes texten till TEI-format enligt Text Encoding Initiatives (Text Encoding Initiative Consortium, 2025) riktlinjer för textkodning.

TEI är en internationell och interdisciplinär standard för textkodning som används av bibliotek, museer, förlag och forskare världen över för att representera olika slags textmaterial i digital form (Text Encoding Initiative Consortium, 2025). TEI-formatet är ett transparent XML-format, och det rekommenderas som det bästa valet för långsiktig arkivering av digital text (Deutsche Forschungsgemeinschaft, 2013, s. 33).

När det gäller digitala objekt, så är det inte stabilitet, utan rörlighet, som är en grundförutsättning för ett långsiktigt bevarande (Björk, 2015, s. 76). Att använda standardiserade format som TEI för textkodning och TIFF för bildmaterial gör att materialet blir tillgängligt för lång tid framöver, och materialet kan enkelt omvandlas till andra format om man så skulle vilja, utan att för den skull gå miste om metadata eller annan avgörande information.

Våra TEI-filer utgick från Wout Dillens mall (Dillen, 2024) med en TEI-fil per tal/block, och i varje fil använde vi oss av huvudelementen <header>, <facsimile> och <text>. (För en redogörelse för samtliga använda element, se bilaga 1.)

I huvudelementet <header> lade vi in de metadata som var relevanta för vårt projekt: information om det fysiska materialet, det genomförda projektet, de institutioner som är kopplade till materialet och projektet samt ämnesord¹⁰ som är relevanta för materialet. (Text Encoding Initiative Consortium, 2025).

Elementet <facsimile> beskriver de bilder som ligger till grund för arbetet, och inom elementet <text> ligger transkriptionen av de olika sidorna. TEI-element användes för att koppla ihop sidorna med respektive faksimil, och för att märka upp textrader och stycken.

Utöver den diplomatariska transkriptionen skapades även en moderniserad, normaliserad text. För att kunna göra detta användes TEI-taggar för att märka upp förkortningar och deras respektive utskrivna former, originalformer och moderniserade/normaliserade former, Elma Danielssons ändringar i texten, tillägg i texten där det är uppenbart att ord eller skiljetecken fattas i originalet och noter till texten med kommentarer och hänvisningar.

3. Resultat

Projektarbetet har resulterat i att Elma Danielssons manuskript A1 ”1:a majtal i Nora 1910” och A2 ”Tal i Askersund från den 8 maj 1910” nu finns bevarade både i sin tidigare fysiska form, och i sin nya digitala form.

Tillsammans med Arbetarrörelsens arkiv i Skåne kommer vi att se till att de digitala bilderna och deras metadata arkiveras på deras servrar på ett sätt så att långsiktigt bevarande säkras.

Från arkivets sida uttrycker man en lättnad över att de sköra originalmanuskripten kan skyddas från hantering, genom att de nu finns tillgängliga digitalt (O. Wallengren, personlig kommunikation, 14 maj 2025).

Projektarbetet har resulterat i att en textkritisk digital utgåva av manuskripten finns tillgänglig genom Github, och så småningom genom Arbetarrörelsens arkiv i Skånes egen hemsida.

Hemsidan utgår i grunden från Wout Dillens mallar (Dillen, 2024). Då hemsidan ska bli en del av Arbetarrörelsens arkiv i Skåne har vi använt de typsnitt och stilval som används på arkivets egen hemsida (Arbetarrörelsens arkiv i Skåne, 2025).

Startsidan beskriver projektet och ger en kortfattad introduktion till Elma Danielsson. De två digitaliserade talen presenteras på var sin sida, med en sida för den diplomatariska transkriptionen, en sida för den moderniserade texten, med länkar tillbaka till den diplomatariska transkriptionen, och en sida för enbart bilderna. I den diplomatariska vyn är bilderna klickbara för att öppna dem i full storlek i en ny flik.

¹⁰ Lämpliga ämnesord valdes ut ur KvinnSams kontrollerade vokabulär (Göteborgs Universitetsbibliotek, u.å.) och Svenskt kvinnobiografiskt lexikons samling av nyckelord (Svenskt kvinnobiografiskt lexikon, u.å.)

Ytterligare ett resultat av projektet är den dokumenterade arbetsprocessen. Eftersom projektet är genomfört med skalbarhet i åtanke, kan den dokumenterade arbetsprocessen förhoppningsvis vara användbar för arkivet, i ett fortsatt digitaliseringsarbete.

Från arkivets sida uttrycks ett intresse för att förstå arbetsprocessen, samt för hur TEI kan användas både för att bevara och tillgängliggöra fler delar av arkivets material.

4. Projektets relevans för kulturarvsbevarande i allmänhet

Mats Dahlström (2011) skriver om hur tidigare studenter digitaliseringsprojekt, trots sin småskalighet och sina tekniska begränsningar, uppvisar både möjligheter och utmaningar som bibliotek och andra kulturbeskrivande institutioner kan möta när de påbörjar ett digitaliseringsarbete av äldre skriftligt material. (Dahlström, 2011, s. 93) Detta kan sägas även om vårt projekt.

Vi har velat skapa kvalitet och djup i digitaliseringen av Elma Danielssons manuskript, och utforska var som är möjligt att göra med hjälp av TEI. Vi har själva inspirerats av andra digitaliseringsprojekt, så som "The Beckett Digital Manuscript Project" (<https://www.beckettarchive.org/>), och projektet "Paper Bullets: The Civil War Letters of John and Phebe Miller" (<https://scholarlyediting.org/issues/41/paper-bullets/>), och vi kan hoppas att vårt eget projekt också på något sätt kan bidra med inspiration kring hur man kan göra en småskalig, kvalitativ, kritisk digital edition av ett äldre skriftmaterial. Särskilt eftersom vi har använt oss av teknik som är tillgänglig även för dem med begränsade resurser.

Vi tror att Elma Danielssons manuskript har relevans på flera sätt, och kan studeras ur olika perspektiv: politiska, kulturella, språkliga m.fl.

Olof Wallengren, som är föreståndare för arkivet, menar att Elma Danielssons handlingar och manuskript tillhör arkivets mest värdefulla resurser. Han menar att dessa handlingar gör det möjligt för människor (forskare, studenter, allmänheten) att ta del av "en politisk idévärld som låg i framkant i sekelskiftets arbetar- och kvinnorörelse". Framför allt har hennes tal inspirerat många, eftersom de ger uttryck för "en skarpsynt feministisk analys av samhället" med ett budskap som är "lika relevant idag" (O. Wallengren, personlig kommunikation, 26 maj 2025).

Conway (2010) menar att kulturarvssektorn nu utmanas av "large-scale information providers" som Google, när det handlar om att definiera vad "bevarande" innebär för något. Han skriver: "In the age of Google, nondigital content does not exist, and digital content with no impact is unlikely to survive." (Conway, 2010, s. 64) Projektet med att digitalisera Elma Danielssons manuskript är i detta avseende ett sätt att hålla dem levande.

Manuskripten kan i sin digitala form ytterligare bidra till inte bara ett bevarande, utan ett levandegörande av kulturarvet. Genom Elma Danielssons många historiska, politiska och kulturella referenser, till händelser, personer och texter, så pekar manuskripten hela tiden vidare mot andra texter, andra dokument.

5. Kritisk analys

Tillvägagångssätt vid bildfångst och textfångst påverkas i grunden av hur vi ser på förhållandet mellan källmaterial och reproduktion/representation/tolkning, eller vilken benämning man nu vill använda för de digitala objekt som skapas för att representera källmaterialet.

I projektet med Elma Danielssons manuskript arbetade vi med både faksimiler och transkriptioner i två versioner, och följaktligen rörde det sig om flera tolkningar av manuskripten.¹¹ Dahlström (2019) använder kartan som en metafor för att beskriva förhållandet mellan representation och källmaterial. Han skriver:

A representation, both a text transcription and a facsimile image, is something other than the original it represents. It is akin to the principle of a map. A map is a representation of a landscape, not the landscape itself. (Dahlström, 2019, s. 201)

Att använda TEI för att presentera transkriptioner av texten i källdokumentet, innebar stora möjligheter att fånga källdokumentets egenskaper.

Pierazzo (2011, s. 463) diskuterar på vilket sätt ”digital editions” skiljer sig från ”printed editions” när det kommer till att presentera material så som transkriptioner och manuskript, och hon menar att en ”digital diplomatic edition” är något helt annat än en tryckt motsvarighet.

En diplomatarisk transkription¹² reproducerar så många som möjligt av källdokumentets egenskaper, med de tecken som står till buds. Detta kan inkludera egenskaper som till exempel rader och radbrytningar, förkortningar och olika former på bokstäver (Pierazzo, 2011, s. 463f).

Ingen transkription kan fånga ”hela” källdokumentet, utan transkriptionen är med nödvändighet en tolkning (Pierazzo, 2011, s. 464–465).¹³ Men medan tekniken, i en tryckt version av en diplomatarisk transkription, *begränsar* möjligheterna att fånga källdokumentets egenskaper, så är det i en digital version tvärtom tekniken som gör möjligheterna näst intill *obegränsade* (Pierazzo, 2011, s. 466).

Under arbetet med transkriptionerna var vi tvungna att förhålla oss till de möjligheter som TEI innebar, och själva begränsa detaljnivån på återgivningen. Vi ställde oss samma fråga som Pierazzo (2011, s. 466): ”Which features of the primary source are we to reproduce in order to be sure that we are following ’best practice’?”

Björk (2015) lyfter i sin avhandling ett exempel från en intervju, där en respondent resonerar kring digitalisering, utifrån de detaljerade protokollen för TEI-kodning:

¹¹ Att vi var tre personer som arbetade med texterna, gjorde det också tydligt att det inte alltid var självklart hur texten skulle tolkas.

¹² ”Transcription” och ”edition” används ibland synonymt, och kommer i följande resonemang användas på det viset. Pierazzo (2011) menar dock att medan en ”transcription” kan vara både en process, en produkt och en relation (mellan två dokument), så är en ”edition” en presentation av en transkription. (Pierazzo, 2011, s. 464)

¹³ Vissa forskare vill beskriva tolkningarna i termer av översättning (translation), från ett semiotiskt system till ett annat. (jfr Pierazzo, 2011, s. 465)

[T]here is a break-even point where it is no longer meaningful to try to convey detailed characteristics of the document in code and where an image simply would prove more efficient (jokingly referring to the option of encoding coffee stains in authors' manuscripts): "... am I encoding the text or the document?" (Björk, 2015, s. 214)

Inte bara möjligheterna med TEI fick oss att reflektera kring vad det egentligen var vi digitaliserade. En annan diskussion handlade om hur abstrakt verk och fysiskt dokument flöt in i varandra.

Vi upptäckte till exempel att textinnehållet i blocken ibland sträcker sig över de fysiska blocken. I manuskript A1 finns anteckningar till det tal som finns i manuskript A2, och vissa bearbetningar av texten i A1 tycks ha gjorts i samband med att texten i A2 skrivits.

Under arbetet talade vi om "talen", "manuskripten", "blocken", "texten" och menade ibland samma saker, ibland olika saker. Vi frågade oss om det var dokumentet vi digitaliserade eller texten, innehållet, verket, informationen?

Björk (2015) menar att det finns skillnader i hur arkiv, bibliotek och museum ser på digitalisering, och han menar att det kan ha att göra med skillnader i synen på vad som utgör själva materialet. Museiföremålet har sitt betydelse i sin status som artefakt, sin form och mening. Biblioteks-föremålet, dokumentet, har sin betydelse i sin mening, det vill säga i sitt textinnehåll (Björk, 2015, s. 13).

Elma Danielssons manuskript är unika objekt, vilket gör dem intressanta både som texter och som artefakter. Begreppet "manuskript" kan i sig inrymma både det abstrakta verket och det fysiska objektet.

6. Slutsatser, rekommendationer

Vi ser stora fördelar med småskalig, kritisk digitalisering.

Inom ramen för detta projekt hade vi inte möjlighet att gå vidare med de kopplingar som vi fann mellan texterna i manuskript A1 och A2. Vi ser stora möjligheter att i en fortsatt digitalisering koppla resurserna i Axel och Elma Danielssons personarkiv till varandra.

Elma Danielssons manuskript skulle med fördel kunna kopplas till brev, intyg, program och förteckningar av olika slag. Säkerligen finns intressanta kopplingar att göra i förhållande till Axel Danielssons dagböcker och till andra delar av samlingen, och till ytterligare andra dokument i samlingar hos andra institutioner.¹⁴ Här finns trådar som sträcker sig genom tid och rum, och med hjälp av TEI (eller annan teknik) finns alla möjligheter att låta dokumenten tala med varandra.

Digitalisering med fokus på kvalitet och djup är tidskrävande. Dahlström (2009) skriver:

¹⁴ Samtliga årgångar av *Arbetet* har till exempel redan digitaliserats av KB (tidningar.kb.se), så där handlar det snarare om att länka de urklipp med Elma Danielssons texter som finns bevarade hos arkivet till motsvarande redan digitaliserade tidningssidor.

Kritisk digitalisering är mycket kostsam relativt antalet digitala utprodukter. Den är långsam och får begränsas till i sammanhanget ytterst få objekt. Den förutsätter textvetenskaplig kompetens. [...] Den kan dessutom ofta rikta sig till en begränsad mängd slutanvändare. Resultatet kan riskera [att] ”glömmas bort” efter digitaliseringen, och inte heller spridas genom marknadsföring och annan uppmärksamhet. (Dahlström, 2009, s. 181)

Frågan är huruvida ett sådant här projekt verkligen leder vidare till fortsatt digitalisering av viktiga kulturarvsdokument, eller om de kulturarvsbevarande institutionerna ser arbetet som för tidsödande och kostsamt, trots den stora nyttan inte minst när det gäller unikt och skört material.

Vår förhoppning är att någon annan kan ta upp stafettpinnen och fortsätta arbetet med att digitalisera det unika material som finns hos Arbetarrörelsens arkiv i Skåne, och att de kan använda detta projekt som grund.

Referenser

- Arbetarrörelsens arkiv i Skåne. (u.å.) *Arbetarrörelsens arkiv i Skåne*. <https://www.aarkiv.se>
- Arbetarrörelsens arkiv i Skåne (u.å.) *Vår verksamhet*. Hämtad den 26 maj 2025 från <https://www.aarkiv.se/about-me/>
- Björk, L. (2015) *How reproductive is a reproduction? Digital transmission of text-based documents*. University of Borås.
- Cowick, C. (2018) *Digital curation projects made easy : a step-by-step guide for libraries, archives, and museums*. Rowman & Littlefield.
- Conway, P. (2010). Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas. *The Library Quarterly: Information, Community, Policy*, 80(1), 61–79. <https://doi.org/10.1086/648463>
- Dahlström, M. (2009). Kritisk digitalisering. I N. D. Lund, J. Andersen, H. D. Christensen, C. G. Johannsen, & L. Skouvig (Red.), *Digital formidling af kulturarv: Fra samling til sampling*. Multivers.
- 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.
- Dahlström, M. (2019). Copies and facsimiles. *International Journal of Digital Humanities*, 1(2), 195–208. <https://doi.org/10.1007/s42803-019-00017-5>
- Deutsche Forschungsgemeinschaft (2013). *Practical Guidelines on Digitisation*. https://www.dfg.de/resource/blob/176110/76abec10bdc30b41f18695145003d6db/12-151-v12_16-en-data.pdf
- Dillen, W. (2024) *A template for publishing digitisation projects on GitHub Pages*. SSLIS/DHCM-template. <https://github.com/sslis/DCHM-template>
- Göteborgs Universitetsbibliotek. (u.å.). *Ämnesord för databasen KVINNSAM*. <https://www2.ub.gu.se/kvinn/kvinnsam/listor/amnesord.html>
- Internet Archive. (2025) *About the Internet Archive*. <https://archive.org/about>
- Lindgren, A. & Lindgren Åsbrink, M. (2007). *Systrar, kamrater!: arbetarrörelsens kvinnliga pionjärer*. Stockholm: Idé och tendens.
- Pierazzo, E. (2011). A rationale of digital documentary editions. *Literary and Linguistic Computing*, 26(4), 463–477. <https://doi.org/10.1093/lilc/fqr033>
- Riksarkivet. (u.å.) *Manuskript och handlingar gällande Elma Danielsson*. <https://sok.riksarkivet.se/arkiv/WglOKowoWKIAzy8FmBfCwE>
- Riksarkivet. (2025) *HTRflow*. <https://ai-riksarkivet.github.io/htrflow/latest/index.html>
- Svenskt kvinnobiografiskt lexikon. (u.å.) *skbl.se – Nyckelord*. <https://skbl.se/sv/nyckelord>

Svenskt kvinnobiografiskt lexikon. (2018, 8 mars). *Elma Charlotta Danielsson*. I skbl.se (artikel av Christina Carlsson Wetterberg). Hämtad den 21 maj 2025 från <https://www.skbl.se/sv/artikel/ElmaDanielsson>

Tanner, S. (2004). *Deciding whether Optical Character Recognition is feasible*. King's Digital Consultancy Service.

https://www.researchgate.net/profile/Simon-Tanner/publication/265748433_Deciding_whether_Optical_Character_Recognition_is_feasible/links/5bd7475e92851c6b27972809/Decidingwhether-Optical-Character-Recognition-is-feasible.pdf

Text Encoding Initiative Consortium. (2025). *The TEI Guidelines: P5 Version 4.9.0*. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

Bilaga 1: Tidslinje

Under hela arbetets gång har gruppen haft löpande kontakt med hjälp av en Whatsapp-grupp, samt träffats över Zoom vid behov.

| | |
|-----------|--|
| v. 3 – 11 | Uppstart – försök att få kontakt med någon intresserad institution Läsning av kurslitteratur Inläsning av riktlinjer för TEI och anvisningar för Github |
| v. 12 | Möte med Arbetarrörelsens arkiv i Skåne Uppstartsmöte med handledare |
| v. 13 | Första kontakten med materialet Urval, provskanning och beslut kring bildfångst |
| v. 14 | Skanning av manuskript och påbörjande av transkription samt research kring företeelser som nämns i manuskripten – människor, platser, citat osv. |
| v. 15 | Fortsatt transkription och research Påbörjande av korrekturläsning och textkodning Påbörjande av XSL- och CSS-kodning Möte med handledare |
| v. 16 | Fortsatt transkription och research Fortsatt korrekturläsning och textkodning Fortsatt arbete med XSL- och CSS-kodning Inläsning av material om Elma Danielsson |
| v. 17 | Avslutning av transkription Fortsatt korrekturläsning och textkodning Fortsatt arbete med XSL och CSS Påbörjande av projektrapport |
| v. 18 | Fortsatt korrekturläsning och textkodning Fortsatt arbete med XSL och CSS Fortsatt arbete med projektrapport 1 maj |
| v. 19 | Korrektur av textkodningen Färdigställande av XSL och CSS Fortsatt arbete med projektrapport |
| v. 20 | Möte med Arbetarrörelsens arkiv för uppdatering och omskanning av vissa sidor Fortsatt arbete med projektrapport |
| v. 21 | Fortsatt arbete med projektrapport Färdigställande av introduktionstext till hemsidan |
| v. 22 | Färdigställande av projektrapport Publicering av slutgiltig version av hemsidan på Github |

Bilaga 2: Textkodning

Då förhoppningen är att arbetet med digitalisering av Arbetarrörelsens arkivs material ska kunna fortgå även efter detta projekts slut följer här en mer genomgripande översikt av vårt arbete med TEI-kodningen av texterna, och en förteckning över de TEI-element som vi har använt.

Efter transkription och korrekturläsning överfördes texten till TEI-format enligt Text Encoding Initiatives (Text Encoding Initiative Consortium, 2025) riktlinjer för textkodning. Våra TEI-filer utgick från Wout Dillens mall (Dillen, 2024), och vi använde oss av huvudelementen <header>, <facsimile> och <text>.

I elementet <header> har vi använt elementen <titleStmt>, <publicationStmt>, <sourceDesc>, <profileDesc> och <encodingDesc>. Dessa element inkluderar metadata om det fysiska materialet, det genomförda projektet, de institutioner som är kopplade till materialet och projektet samt ämnesord som är relevanta för materialet (Text Encoding Initiative Consortium, 2025). Lämpliga ämnesord valdes ut ur KvinnSams kontrollerade vokabulär (Göteborgs Universitetsbibliotek, u.å) och Svenskt kvinnobiografiskt lexikons samling av nyckelord (Svenskt kvinnobiografiskt lexikon, u.å.)

Elementet <facsimile> beskriver de bilder som ligger till grund för arbetet, och inom elementet <text> ligger transkriptionen av de olika sidorna.

Varje enskild sida gavs TEI-taggen <div>, med attributen @type="page" för att särskilja dem från andra <div>-taggar i dokumentet, @n för sidnummer och @facs för länkar till respektive faksimil. Varje sida inleddes också med taggen <pb> för "page beginning" och varje textrad inleddes med <lb> för "line beginning". Inom varje sida inneslötts varje stycke av taggen <p>.

Då vi ville producera både en diplomatarisk transkription och en moderniserad, normaliserad text använde vi oss av TEI-taggar <abbr> och <expn> för förkortningar och deras respektive utskrivna former, <orig> och <reg> för originalformer och moderniserade former av ord, <sic> och <corr> för sådant som förekom i originaltexten men som av olika anledningar har justerats i den moderniserade texten, samt <supplied> för tillägg i texten där det är uppenbart att ett ord eller ett skiljetecken fattas i originalet. Dessa element användes med elementet <choice> för att under XSL-transformeringen kunna välja vilka textformer som skulle visas. Vi använde oss även av taggen <note> för att lägga till noter till texten, där vi fann att så behövdes.

För Elma Danielssons egna strykningar och ändringar använde vi TEI-taggar för strykningar och <add> för tillägg, samt <delSpan> för strykningar som gällde mer än ett ord eller en mening. Vi använde TEI-taggen <handShift> för att visa var Elma Danielsson bytte från bläckpenna till blyertspenna och tvärtom, samt <emph> med attributen @rend="double" eller @rend="triple" för att markera understrykningar med ett, två eller tre streck. Vid de tillfällen när Elma Danielsson citerar en dikt eller gör en uppräkningslista på flera rader använde vi elementet <lg>, "line group" med elementet <l> för att beteckna varje rad. Slutligen använde vi taggen <unclear> för att markera ord eller tecken som var svårtydda.

Vi använde även elementet <name> med attributen @type="person" och @type="place" för att identifiera personer, även fiktiva sådana, och platser i manuskripten. Vi har inte arbetat vidare med dessa element, men man kan tänka sig att de kan användas framöver vid till

exempel automatisk analys av texterna, för att särskilja den fiktiva personen Fredag i Daniel Defoes *Robinson Kruse* från veckodagen fredag, eller staden Nora från personnamnet Nora.

Ett problem som uppstod under TEI-kodningen var en begränsning hos XML-formatet, nämligen att det till sin natur har en hierarkisk struktur, där olika element inte kan "gå omlott". Innan man kan stänga en `<div>`-tagg måste alla andra element inom `<div>`-elementet först ha stängts i sin tur, vilket gör att det inte går att på ett enkelt sätt hantera stycken (`<p>`) som sträcker sig över flera sidor. Detta löstes genom att för `<p>`-element som sträckte sig över flera sidor använda attributet `@part`, med värdena "I" för "Initial", "M" för "Medial" och "F" för "Final" (Text Encoding Initiative Consortium, 2025, §17.3). Därmed kunde dessa uppdelade stycken med hjälp av XSL-transformering sättas samman till ett stycke för den moderniserade transkriptionen.

Vi använde också elementet `<pc>` ("punctuation character") för avstavningar vid rad- eller sidbrytningar tillsammans med attributet `@break="no"` vid påföljande `<lb>`-tagg, för att markera att dessa element skulle tas bort och orden sammanskrivas i den moderniserade transkriptionen.

För att beskriva textens struktur användes följande TEI-element:

| TEI-element | Används för |
|---------------------------------------|--------------------|
| <code><div @type="page"></code> | varje enskild sida |
| <code><pb></code> | sadbrytning |
| <code><lb></code> | radbrytning |
| <code><p></code> | stycke |

För att beskriva Elma Danielssons ursprungliga text användes följande TEI-element:

| TEI-element | Används för |
|---|---|
| <code></code> | strykningar |
| <code><add></code> | tillägg |
| <code><delSpan></code> | strykningar som gäller mer än ett ord eller en mening |
| <code><emph></code> , även med attributen <code>@rend="double"</code> eller <code>@rend="triple"</code> | understrykningar med ett, två eller tre streck |

| | |
|--|--|
| <code><lg></code> och <code><l></code> | dikter eller uppräknningar med flera rader |
| <code><handShift></code> med attributen <code>@medium="pencil"</code> och <code>@medium="ink"</code> | byte mellan bläckpenna och blyertspenna |

För våra egna ändringar och tillägg användes följande TEI-element:

| TEI-element | Används för |
|--|---|
| <code><abbr></code> och <code><expan></code> | förkortningar och deras respektive utskrivna former |
| <code><orig></code> och <code><reg></code> | originalformer och moderniserade former av ord |
| <code><sic></code> och <code><corr></code> | sådant som förekom i originaltexten men som av olika anledningar har justerats i den moderniserade texten |
| <code><supplied></code> | tillägg i texten där det är uppenbart att ett ord eller ett skiljetecken fattas i originalet. |
| <code><unclear></code> | svårtydda eller oläsliga ord eller tecken |
| <code><note></code> | noter och anteckningar |

Bilaga 3: HTR-resultat

Här följer ett par exempel på hur den automatiska textigenkänningen fungerade (eller inte).

HTRflow på sida 1 i manuskript A1 skannad med en upplösning på 300 dpi:

Nora, 1 Maj 1910,
Domonshanter;
Efter och uplästes år, som Sverges
klassen arbetare utstått, sockert och
storstryksäret samlas de åter på sin
an en gauf
mk. mönstrings dag för att fordra Sina
förhållna mantal. rättigheter.
Estt hög järhu väl ha vi nu tydt
parollen /3. intz Socild. konpiesen i Paris
1899. att demonstra
för lagtadgad 8- timmars arbetsdot-

HTRflow på sida 1 i manuskript A1 skannad med en upplösning på 600 dpi:

Nora, 1 Maj 1910.
Domonsharter,
Eftu och årådts är, som Sverijes
dussin. arbetare utstått, sockert och
stortrejkåret samlas de åter på ny-
antu plup
mk. mönstrings dag för attfordra Sina
förhållna marskt. öäktskuter.
Ett tjog gånger väl ha vi nu tyt
parollen p. intz Socild. konpiesen i Paris
1899. att demonstre
för lagstadgad 8-timmars arbetsdag.

Vår manuella transkription:

Nora, 1 Maj 1910.

Demonstranter!

Efter det svåraste år, som Sverges
klassm. arbetare utstått, lockout- och
storstrejkåret, samlas de åter på sin
int. mönstringsdag för att än en gång fordra sina
förhållna mänsk. rättigheter.

Ett tjog gånger väl ha vi nu lytt
parollen fr. int. sociald. kongressen i Paris
1899 att demonstrera
för lagstadgad 8-timmars arbetsdag.

Det hade förmodligen gått att få bättre resultat om modellerna hade tränats på just Elma Danielssons texter, och att skapa sådana modeller skulle kunna utgöra en vidareutveckling av det här projektet, då det nu finns både bild och text att träna på.