



19. september 2024

Side 1 af 7

Afrapportering for Retrodigitaliseringsprojektet 2017-2024

Baggrund

Kommunalbestyrelsen (byråd og sogneforstanderskab/sogneråd) er omdrejningspunktet i det kommunale selvstyre. Dens beslutningsdokumenter er centrale for det kommunale arkiv og for historieforskning lokalt og nationalt. De indeholder fortællingen om den lokale udmøntning af statslige politikker og kilder til det politiske liv fra græsrodder til organisationer og partier. Kommunalbestyrelsen er også arbejdsgiver og udfører af stort og småt. Dens beslutningsprotokoller er dermed en indgang til både konkrete menneskeskæbner og politiske beslutninger, det strategiske niveau og til forvaltningsniveauet.

I 2020 kunne Kommunernes Landsforening fejre 50-året for kommunalreformen i 1970. Kommunalreformen betød, at landets kommuner blev reduceret fra 1098 til 277, og sondringen mellem sogne- og købstadskommuner blev formelt ophævet. Det ville Organisationen Danske Arkiver (ODA) gerne være med til at markere. Frem mod 2020 blev Retrodigitaliseringsprojektet derfor igangsat. Dette skete ved henvendelser til landets arkiver og involvering af frivillige i retrodigitaliseringen.

De danske arkiver står gennem ODA fælles i prioriteringen af kommunalbestyrelsernes forhandlingsprotokoller som en af de vigtigste historiske kilder til lokalsamfundenes liv og beslutningsprocesser.

Retrodigitaliseringen af kilderne er også vigtig for bevaringen af protokollerne og giver mulighed for at nå nye målgrupper i historieformidlingen. RETRO (for retrodigitalisering) blev igangsat med lige netop disse hensyn for øje.

ODA begyndte allerede i 2017 koordineringen af en landsdækkende skanning og transskribering af kommunalbestyrelsernes protokoller. Digitaliseringen af forhandlingsprotokollerne gør dem mere læselige, søgbare på fuldtekst og tilgængelige for hvermand igennem projektets hjemmeside.

RETRO vil betyde, at et væld af nye studier bliver mulige for forskere, studerende og borgere, når spørgsmål om lokale såvel som nationale forhold stilles til Danmarks historie.

Kultur og Borgerservice

Borgerservice og Biblioteker
Aarhus Kommune

Aarhus Stadsarkiv

Dokk1, Hack Kampmanns Plads 2
8000 Aarhus C

Direkte telefon: 41 85 65 45

Sagsbehandler:

Mia Gulvad Jørgensen



Projektets formål

19. september 2024
Side 2 af 7

- Udvikling af tekstgenkendelsesmodeller, der kan anvendes på ældre historisk håndskrift indenfor det danske sprogområde og automatisere transskriberingen af samme gennem maskinel tekstgenkendelse.
- Opbygning af et up to date-crowdsourcing-tilbud, der ikke kræver egen vedligeholdelse eller udvikling. Dette gøres gennem samarbejder mellem danske arkiver, universiteter og andre institutioner i både lokale, nationale og internationale rammer.
- Omdannelsen af analoge historiske data til regulære datasæt, f.eks. til forskning og brug for digitale humanistiske metoder.
- Gennemførelse af maskinlæsning af mindst 375.000 dokument sider afhængig af dokumenttype.
- Etablering af en platform og grundlag for at videreføre projektet på frivillig basis efter projektophør.

Under covid-pandemien fik projektet til formål at udvikle et digitalt tilbud og at skabe et digitalt fællesskab for frivillige, særligt ældre, som var i højrisiko for at blive smittede og derfor isolerede i eget hjem.

Kunstig intelligens trænes til håndskriftsgenkendelse

Retrodigitaliseringsprojektet er et digitaliserings- og transskriberingsprojekt, der har til formål at gøre danske sogne- og byrådsprotokoller tilgængelige online, læsbare og fuldtekstsøgbare for alle.

Projektet har anvendt softwaren Transkribus, som gør det muligt at træne kunstig intelligens til automatisk at transskribere håndskrevne dokumenter. Når et tilstrækkeligt antal dokumenter er trænet, kan man på baggrund af materialet skabe en tekstgenkendelsesmodel, som kan bruges på andre lignende materialer.

Disse tekstgenkendelsesmodeller kan derfor bruges til at transskribere projektets protokoller, der er skrevet med både gotisk trykskrift og håndskrift, skråskrift og formskrift, men kan også bruges til transskribering af et væld af andre kilder – håndskrevne såvel som trykte.

I dette projekt er der arbejdet med kommunalbestyrelses (by- og sogneråds) forhandlingsprotokoller fra perioden 1841-1970.



19. september 2024
Side 3 af 7

Projektets målsætninger blev nået

På Aarhus Stadsarkiv er der blevet transskriberet mere end 55.000 sider, hvoraf 18.000 sider er korrekturlæst. Det betyder, at Stadsarkivet med hjælp fra frivillige og kunstig intelligens har nået at transskribere stort set alle by- og sognerådsprotokoller inden for Aarhus Kommunes grænser for perioden 1841-1940. Fokus har været på denne periode, da nyere materiale ikke kan offentliggøres pga. GDPR.

På landsplan har mere end 25 arkiver deltaget i projektet. Deltagelsen er blevet sikret ved Aarhus Stadsarkivs mellemkomst. De deltagende arkiver har været engageret på forskellige vis. Nogle anvender frivillig arbejdskraft, mens andre er professionelt drevne. Ved projektafslutningen i september 2024 er lige under 500.000 sider blevet transskriberet. Resultaterne fra projektet udstilles på hjemmesiden <https://app.transkribus.org/sites/danske-kommunearkiver>.

Projektet har vist sig så bæredygtigt, at transskriberingsarbejdet fortsætter ved frivillig arbejdskraft efter projektets afslutning, og sitet vil blive opdateret løbende i takt med, at der kommer mere materiale til. Der er truffet aftale med Dansk Center for Byhistorie om, at centret vil varetage den videre drift af sitet og publiceringen af dataene fra Retrodigitaliseringsprojektet.

Retrodigitaliseringsprojektet har skabt – eller været medskabere af – fire tekstgenkendelsesmodeller; en til gotisk håndskrift fra 1700-tallet, en til gotisk håndskrift fra 1800-tallet, en til gotisk trykskrift og en til skråskrift og formskrift.

Projektet har sikret en demokratiseret adgang til vores lokale historie, idet den information, arkivalierne indeholder, ikke længere kun er forbeholdt de eksperter, som kan læse ældre dansk håndskrift, men nu bliver tilgængelig i læsbar form og fuldtekstsøgbar for alle interesserede.

Desuden er der gennem projektet blevet skrevet flere danske vejledninger til Transkribus, som kan være en hjælp for brugere, der har svært ved at klare sig med de engelske vejledninger på READ-COOPs hjemmeside. Vejledningerne kan findes på projektets Transkribus Site, Danske Kommunearkiver, som også indeholder information om projektet og dets udarbejdede tekstgenkendelsesmodeller samt et bibliotek med alle de transskriberede protokoller. Sitet rummer desuden et bagkatalog af blog-indlæg, der blev skrevet og udgivet undervejs i projektet.



Projektets datasæt er blevet offentliggjort på dataportalen på opendata.dk.

19. september 2024

Side 4 af 7

Projektet har givet Aarhus Stadsarkiv (og andre arkiver) en ny og kompetent gruppe af frivillige, som har tilegnet sig kompetencer til transskriberings- og korrekturlæsningsarbejdet gennem Transkribus. Gruppen af frivillige vil efter projektets afslutning transskribere andre samlinger af arkivalier, som stadsarkivet vil kunne offentliggøre på AarhusArkivet.dk. Her vil sognerådsprotokollerne fra Aarhus og omegn også komme til at ligge.

Gennem Retrodigitaliseringsprojektet er kendskabet til Transkribus blevet udbredt i Danmark, og også andre interessenter bruger det i deres arbejde.

Undervejs blev der søgt fondsmidler til at udvide retrodigitaliseringen til andre nordiske lande. Et netværk blev skabt og et samarbejde etableret med arkiver og kulturinstitutioner rundt i norden, men denne udvidelse faldt grundet manglende fondsbevilling.

Donatorer

Projektet er blevet gennemført med støtte fra Augustinus Fonden, Aage og Johanne Louis-Hansens Fond, Velux Fonden, Region Midtjyllands kulturpulje for 2019, Organisationen Danske Arkiver, Arkivudvalget for Aarhus Kommune og de danske arkivers almindelige virke og budgetter. Desuden har Rigsarkivet stillet skanningsudstyr til rådighed for underprojektet "Demokratiets Rødder".

Organisationen Danske Arkiver (ODA) har medfinansieret Retrodigitaliseringsprojektet, mens sekretariatet for projektet har været placeret organisatorisk ved Aarhus Stadsarkiv.

Samarbejdspartnere

Projektet er udført i regi af ODA i samarbejde med en række danske stads- og lokalarkiver. De enkelte arkiver og deres frivillige har koordineret den lokale arbejdsindsats. Samarbejdet har sikret, at projektet geografisk er udbredt over det meste af Danmark.

Projektets følgegruppe bestod af Nina Javette Koefoed, professor i historie ved Institut for Kultur og Samfund, Aarhus Universitet, Charlotte Voss, arkivar og formand for ODA, Niels Elers Koch, direktør for Trap Danmark og Lex.dk, Helle Strandgaard Jensen, lektor i historie ved Institut for Kultur og Samfund, Aarhus Universitet, Allan Vestergaard, enhedschef for Brugerservice ved Rigsarkivet og Max Odsbjerg Pedersen, informationsspecialist ved Det Kgl. Bibliotek. Følgegruppen har ydet sparring



og været ambassadør for projektet samt været med til at udvikle metoder for, hvordan Transkribus kan bruges i praksis.

19. september 2024
Side 5 af 7

En tak skal særskilt lyde til Nina Javette Koefoed, professor ved Institut for Kultur og Samfund ved Aarhus Universitet, og Johan Heinsen, professor ved Institut for Politik og Samfund ved Aalborg Universitet, som har bidraget med sparring og data til opbygning af projektets tekstgenkendelsesmodeller.

Aarhus Stadsarkiv og ODA er medlemmer af READ-COOP. READ-COOP ejer og udvikler Transkribus, som er den software, projektet har anvendt til tekstgenkendelse. READ-COOP er et internationalt kooperativ af forskere og institutioner, som via Transkribus arbejder for at udvikle tekstgenkendelsesmodeller inden for forskellige sprogområder.

Milepæle

- Projektet blev indledt i 2017 i forbindelse med, at Aarhus Stadsarkiv og Organisationen Danske Arkiver (ODA) frem mod 2020 ønskede at markere 50-året for Kommunalreformen af 1970.
- I 2017 blev projektsekretariatet etableret ved Aarhus Stadsarkiv. Det var finansieret af ODA. Transskriberingen blev igangsat samme år via Transkribus.
- Projektet fik sin egen hjemmeside – retrodigitalisering.dk.
- I 2020 blev den første danske tekstgenkendelsesmodel til håndskrift bygget med kunstig intelligens i Transkribus. Modellen bestod af materiale fra Aarhus, Faxe, Gentofte og Næstved, som på det tidspunkt udgjorde de deltagende arkiver i projektet, og modellen kunne læse skråskrift fra 1870'erne og frem.
- I 2021 modtog projektet donationer fra Augustinus Fonden, Aage og Johanne Louis-Hansens Fond samt Velux Fonden til videreudvikling af computermodeller, der kunne læse historiske, håndskrevne dokumenter.
- I 2022 blev syv workshops afholdt i forskellige landsdele for interesserede arkiver. Derigennem sluttede nye deltagere sig til projektet, som siden fortsatte med at vokse støt.
- I april 2023 blev den første tekstgenkendelsesmodel til 1800-talsgotisk håndskrift offentliggjort. Modellen blev hjulpet på vej af lektor i historie ved Aalborg Universitet, Johan Heinsen, der stillede godt og



vel 300.000 transskriberede ord fra slutningen af 1700-tallet og starten af 1800-tallet til rådighed.

19. september 2024
Side 6 af 7

- I juni 2024 afsluttedes projektet, men transskriberingsarbejdet fortsætter lokalt ved de deltagende arkiver. Resultaterne vil løbende blive lagt op på projektets nye Transkribus Site "Danske Kommunearkiver".
- I Aarhus er mere end 55.000 sider blevet manuelt eller maskinelt transskriberet, og af dem er 18.000 sider blevet korrekturlæst. Siden projektets begyndelse har mere end 60 frivillige arbejdet med protokollerne fra Aarhus Kommune.
- På landsplan har mere end 25 arkiver deltaget i projektet. I alt er næsten 500.000 sider blev transskriberet ved projektets afslutning.

Resultaterne peger fremad

Projektets forløb og resultater har et enormt potentiale til at bidrage til forskning på tværs af fagfelter i det danske sprogområde.

Konkret åbner de onlinetilgængelige, letlæselige og fuldtekstsøgbare by- og sognerådsprotokollerne for nye undersøgelser om nærdemokratiet og den helt nære danmarkshistorie for et langt bredere publikum.

Tekstgenkendelsesmodellerne gør det muligt at bearbejde andre håndskrevne dokumenter hurtigt og effektivt, så de let kan undersøges i dybden på samme måde som by- og sognerådsprotokollerne. Herved kan kilde-materialet anvendes af alle – og ikke kun de ganske få, der kan læse håndskriften. Modellerne gør også, at de mange ressourcer, der tidligere blev brugt på transskriberingen, kan bruges andre steder.

Retrodigitaliseringsprojektet skaber også potentiale for internationale samarbejder, f.eks. på tværs af Norden. Et dansk-tysk samarbejde med fokus på grænseområdet kunne også være en mulighed.

I kølvandet på Retrodigitaliseringsprojektet er der allerede opstået adskillige andre digitaliserings- og transskriberingsprojekter, som Aarhus Stadsarkiv har sparrat med og vejledt. Det er derfor forventeligt, at flere nye tekstgenkendelsesmodeller vil blive udarbejdet i fremtiden. Det tyder i det hele taget på, at mulighederne med Transkribus har startet en bølge inden for digitalisering og transskribering af arkivalier, og at det fremadrettet vil føre til tilgængeliggørelsen af større og større mængder materialer.



Tekstgenkendelsesmodellerne hjælper allerede forskellige institutioner og virksomheder med at læse og søge i gamle, håndskrevne sager og papirer, som kan være afgørende for deres arbejde og beslutninger i dag. Et eksempel herpå er Tinglysningen, der bruger modellerne, når de skal læse gamle skøder.

19. september 2024
Side 7 af 7

Organisationen Danske Arkiver (ODA) og Aarhus Stadsarkiv vil takke alle samarbejdspartnere for et lærerigt og særdeles perspektivrigt projekt – og alle bidragydere for at gøre det muligt.

Projektet har sat en retning for, hvor langt branchen kan nå, når professionelle og frivillige kræfter forenes, og hvor løfterige perspektiverne er for brugen af kunstig intelligens til at bringe historien til live ved at 'oversætte' gammel og ofte ulæselig tekst til søgbar data.