

RETRO

ODA koordinerer digitalisering og transskribering af byråds- og sognerådsprotokoller

ODA koordinerer digitalisering og transskribering af kommunalbestyrelsernes protokoller. Projektets navn er RETRO (for retrodigitalisering).

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.

Formål

Det er ODA's ambition, at RETRO vil skabe et grundlag for, at kommunernes historie og arkiver får en central placering i fremtidens digitale historieskrivning, herunder Digital Humanities. Det skal både ske gennem tilvejebringelsen af data og gennem kompetenceløft.

RETRO skal medvirke til at koordinere og styrke arkivernes digitaliseringsindsats.

RETRO skal fremme brugerinddragelse i arkivernes digitale aktiviteter i form af crowdsourcing og lokale netværk.

RETRO skal fortælle lokale historier og fremdrage lokale beslutninger, der kan danne bagtæppe og inspiration for både den nationale og den lokalhistoriske formidling.

Danske arkiver ønsker gennem RETRO at bidrage til internationalt udviklingsarbejde inden for tekstgenkendelse, digital arkivformidling og crowdsourcing.

Landsdækkende projekt

Alle arkiver i ODA og Sammenlutningen af Lokalarkiver kan deltage i RETRO.

Hvordan foregår det?

Processen i dag er, at de enkelte arkiver selv uploader digitale billeder af deres protokoller til et program, Transkribus, der drives under EU-projektet READ under Horizon 2020. READ forvaltes af Universitat Innsbruck. Man transskriberer ogsa inde fra programmet. Programmet installeres pa de PC'er, som man arbejder med.

READ-projektet har ydermere forpligtet sig pa at levere en web-baseret udgave af programmet, som korer i en browser og derfor ikke kraver installation pa de enkelte PC'er. Dette program er dog ikke ferdigudviklet endnu.

Hvad er READ?

READ star for Recognition and Enrichment of Archival Documents og har til formal at udvikle en teknologi til at genkende handskrift i historiske dokumenter. Hertil kreves en stor mengde transskriberet tekst, der kan tjene som sammenligningsgrund.

Det bliver gjort via transskriberingsprogrammet Transkribus, som kan hentes gratis pa nettet. Programmet gor det muligt pa en enkelt og overskuelig made at transskribere hele protokoller, og det gor det fleksibelt for de personer, som deltager i transskriberingsprocessen, at tilga og overskue deres arbejde

Fordele ved at deltage

Som arkiv er det en stor fordel at have de primare kildesamlinger tilgaengeliggjort digitalt. Sogneradsprotokollerne er en vigtig kilde til at forsta sin lokalhistorie, og de beslutninger, der blev truffet. Ved at deltage i dette projekt og fa dem transskriberet (i Transkribus) tager man et stort skridt i retning af fa sine samlinger gjort sogbare og let tilgaengelige for arkivet og brugerne.

Aktuelt er transskriberet tekst dog kun sogbart i enten Transkribus-programmet eller gennem deres webservice, som kraver en front-end, for det kan presenteres for brugerne.

Hvad nu hvis det ikke virker med READ/Transkribus...?

Transkribus er forst og fremmest et almindelig og velfungerende crowdsourcingverktoj, der fungerer uafhangigt af den automatiske tekstgenkendelse. Transkribus arbejder ikke kun med handskrift, men kan ogsa genkende maskinskrift. Derfor har vi valgt at arbejde i Transkribus.

Som en bonus kan Transkribus ogsa bruges til at teste genkendelse af handskrift. Test viser, at Transkribus kan have en fejlmargen pa ca. 15 % ved den første automatiske tekstgenkendelse baseret pa en dokumentssamling, der ligner en typisk sogneradsprotokol med skiftende pennenforere og et relativt begranset antal sider. Hvis den samme hand skriver flere end ca. 10 sider, falder fejlraten allerede. Hvis vi ikke er tilfredse med resultaterne for genkendelse af handskrift, pavirker det ikke indskrivningsdelen.

Hvad skal man kunne for at være med i projektet?

Der stilles ikke store krav til arkiverne for at kunne deltage. Man skal levere dokumenter i digital form (jf. nedenfor). Man skal have en basal forståelse for it.

De frivillige eller medarbejdere, der skal transskribere, skal have evne til at læse forskellige typer af håndskrifter. Men kan man ikke læse de ældste protokoller fra 1840'erne, er det måske en bedre idé at tage fat i 1930'erne eller et moderne materiale.

Hvad skal man levere?

Man skal selv levere digitaliserede dokumenter. De kan være affotograferede eller skannede. De skal oprettes og navngives individuelt efter de retningslinjer, projektlederen giver. Skanningerne skal være tilstrækkeligt gode; de-skewing og god kontrast er et minimum.

Det enkelte arkiv skal selv rekruttere frivillige eller stille personale til rådighed til transskribering.

Formidling af data

Der er adskillige (gode) ideer til formidlingsdelen, men der er ikke taget så mange skridt i den retning endnu.

Data med de transskriberede og tekstgenkendte protokoller overføres via Transkribus-softwaren til READ. Det betyder, at data skal formidles fra READ/Transkribus. Det vil enten ske via READ's egen platform eller som eksport af data derfra til en brugergrænseflade tilpasset vores data og behov.

Hvis data skal stilles til rådighed for analyse, processering og fortolkning (big data), skal data kunne udtrækkes. Det kunne være til en relevant Åben Data-plattform.

Deltagerne får et datasæt med egne data.

Det er vores hensigt at udvikle en mere dynamisk hjemmeside, der giver indblik i fremdragne historier og erfaringer og skaber overblik over deltagere, progression, 'succesrater' fra READ osv.

Tilgængelighed

Relevante tilgængelighedsregler skal overholdes. Mange protokoller er f.eks. opdelt i åbne og lukkede sager. For lukkede sager anbefaler vi, at man som udgangspunkt anvender Arkivlovens regler (75 år). Åbne sager, anbefaler vi, at man kan offentliggøre, hvis data tidligere har været offentliggjort (jf. svar fra Kulturministeriet til Frederikssund Kommune, 7/11 2016).

Koster det noget?

Deltagelsen er gratis for medlemsarkiver, men ODA har finansieret projektet gennem bevillinger i 2015-2017, og Aarhus Stadsarkiv har bidraget med midler fra en udviklingspulje.

Deltagelsen i READ koster ikke noget.

Hvordan kommer man med?

Projektet bygger på erfaringer fra Aarhus Stadsarkiv, som derfor er projektholder.

Ønsker man at vide mere om projektet, og hvordan man deltager, kan man kontakte Aarhus Stadsarkiv på mail stadsarkiv@aarhus.dk.

Projektmedarbejder er arkivar Martin Philipsen Mølgaard, it-udvikler er Claus Juhl Knudsen og projektleder er Søren Bitsch Christensen, Aarhus Stadsarkiv.