

Arkivverket
Postboks 4013 Ullevål stadion
0806 OSLO

Sted:
Sandefjord

Deres referanse:
2021/15483

Vår referanse:
2022/78

Dato:
18.10.2024

Sluttrapport

Bakgrunn

Prosjektet «Ordning og bevaring av digitalt skapt næringslivsarkiv – Scanrope» ble avsluttet i juli 2023. Prosjektet ble søkt om og tildelt midler i 2021. Målet var å ordne og bevare de digitalt skapte arkivene etter Tønsbergbedriften Scanrope, som hadde et omfang på 372 GB, hvorav ca. 25 % var i arkivformat. Et av delmålene var å fokusere på konvertering av tekniske tegninger til arkivformat. Arkivet skulle bevares i sikringsdepot hos KDRS og gjøres tilgjengelig for publikum.

Oppstarten av prosjektet ble utsatt til høsten 2022 på grunn av forsinkelser i ansettelsen av IT-arkivarer og tilhørende opplæring. Opplæringsuttrekket var også vesentlig mer komplekst på grunn av en tilvekst av e-poster med flere ukurante formater, som økte volumet betydelig.

Delmål i prosjektet

Lage AIP-0 – Arkivpakke for langtidsbevaring (uten konvertering av formater)

- Identifisere og slette søppelfiler (temporære filer og tomme filer)
- Analysere filer og formater. Finne filer med feil endelse og endre disse til korrekt type for å gjøre dem lettere lesbare og mulige å konvertere
- Pakke originalene med Arkade 5
- Grovt dokumentere innholdet i Asta og publisere på arkivportalen
- Overføre til Digitalt depot

Lage AIU-0 – Tilleggspakke med arkivformater

- Skille ut alt som ikke var i arkivformater

- Konvertere filer til arkivformater med flere ulike tilnærminger
 - Trinnvis redusere antall formater som ikke lot seg konvertere
 - Bruke filformatanalysen til å håndtere ulike formater
 - Prioritere tekniske tegninger og Microsoft Office-filer først
 - Deretter ta grupper av filtyper basert på antall gjenværende filtyper
- Pakke de konverterte filene i arkivformat i en tilleggspakke AIU. Filer som ikke kunne konverteres, forblir i AIP-0 for eventuell senere konvertering
- Overføre til Digitalt depot

Gjennomføring

En egen virtuell maskin (VM) ble satt opp med ulike konverteringsprogrammer. Vi anskaffet lisenser på følgende programmer/pakker:

- Neevia Document Converter Pro
- Total Outlook Converter Pro

Før vi startet konverteringen, ryddet vi bort det vi anså som søppelfiler.

Filformatanalysen dannet noe av grunnlaget for dette. Analysen viste oss tomme filer, og en rekke temporære Office-filer ble slettet. Pronomenanalysen i Arkade 5 krasjet, så vi brukte Siegfried til å utføre filanalysene. Filer med feil endelser ble rettet med batch-script.

Erfaringer fra test case brukt i opplæring før Scanrope prosjektet tilsa at virus problematikk var utbredt i gamle eposter. Ved å isolere den virtuelle maskinen fra internett/nettverk forøvrig minimerte vi risiko for spredning. Siden den virtuelle maskinen har et rent image å starte ifra er det lett å starte forfra i et nytt prosjekt men rene ark uten virus fra tidligere kjøring.

Neevia Converter Pro

Det ble også installert flere programpakker som opprinnelig skapte filene. Hvis Neevia ikke hadde innebygd konvertering, ble det forsøkt å bruke runtime-versjonen av det proprietære programmet med utskrift til PDF. Noen leverandører har laget gratis runtime-versjoner for at brukere kan lese, men ikke redigere filer. I noen tilfeller ble fullt funksjonelle programpakker med 30 dagers prøveperiode installert på VM-en.

Spesielt tekniske tegninger fikk nytte av disse runtime. Gjennom tidene vara det utvekslet og lagret tegninger fra flere proprietære system.

Basert på erfaring fra forprosjektet fant vi ut at tiden det tar å konvertere var langt høyere enn først antatt. Ikke alle formater lot seg konvertere til arkivformater.

Første fase: Vi brukte Neevia Converter. Denne programvaren gjør det mulig å grovsortere filer i flere hauger basert på om konverteringen var vellykket eller ikke.

Konverterte filer ble plassert i en mappe, og originalfilene havnet i en suksessmappe. Filer som ikke kunne konverteres, havnet i en egen "feil-mappe".

Fra det tidligere prosjektet visste vi at flertrådet prosess med å ta flere filer samtidig kunne henge seg opp og dytte konverterbare filer til feil-mappen. Ved å flytte kjente konverterbare filformater fra feil-mappen tilbake til "start-mappen" og kjøre Neevia i flere iterasjoner, oppnådde vi en akseptabel mengde feil.

Forbedringspunkter

Å flate ut 3 dimensjonale tegninger i 2D med PDF kan diskuteres om er en vellykket strategi. Noen blir bort når en dimensjon blir borte. Vi håper at informasjonen er godt nok i et fremtidig bruk. Målet vårt har ikke vært å gjenskape et produksjonsbehov, men et dokumentasjonsbehov. Vår mening er at pdf'ene dekker et slikt behov. Å bevare i 3D vil være bedre, 2D er godt nok i mange tilfeller. Originaler før konvertering er også bevart slik at det er mulig på senere tidspunkt å gjøre ny konvertering.

Mye gjøres for å standardisere uttrekk rundt Siard. Filarkiv er ikke en database i tradisjonell forstand, så produksjonsløypen for databaser med Siard passer ikke helt. Windows og andre operativsystemer har ulike former for metadata for filer og kataloger, som navn, sti, opprettet, endret, rettigheter, filstørrelse, attributter osv. I fremtiden vil jeg anbefale at slike metadata bevares i en enkel database slik at man kan lage Siard-uttrekk. Et enkelt verktøy for å finne disse metadataene er Power Query i Excel kan avlese metadata på mapper og filer

Dette kan være en database med bare filnavn og sti, eller man kan velge alle relevante metadata som filsystemet har. Dette er etter vår mening en tilnærming som gjør at filsystemet passer bedre inn i produksjonslinje-konseptet utviklet i KAI-miljøet. Man vil også kunne formidle og tilgjengeliggjøre data i «innsynsløsninger» optimalisert for Siard.

Feil som kunne vært håndtert bedre

Passordbeskyttelse eller rene malfiler utgjorde stor del av feilsituasjonene. Prosessen med re-konvertering kunne vært gjort smartere ved å bruke feilloggen til Neevia for å unngå å legge tilbake filer som åpenbart ikke ville konverteres.

Dette er et klart forbedringspunkt. En slik analyse av hvilke feil som oppstod i Neevia kunne ha blitt brukt til å hindre rekonvertering av filer med kjente feil, i stedet for å kjøre iterasjon etter iterasjon på de samme filene til vi følte at det ikke hjalp mer å kjøre på nytt.

Oppsummering av resultater og gevinster

Det ble laget to arkivpakker: en arkivpakke AIP-0 med filene slik de var før konvertering, og en tilleggspakke AIU-0 med konverterte filer.

AIP-0:

503 542 filer hvor av 160988 ca 32% var i arkivformater 68% var i andre formater (338.140 filer)

AIU-0:

165 402 nye filer i arkivformater. Som utgjør grovt regnet 50% færre filer som ikke var i arkivformater. Totalt sett. 170 tusen filer i AIP-0 lot seg ikke konvertere, det betyr ikke at disse filene er uleselige, men kan bli fortere forgjengelige.

Prosjektregnskap:**Ordning og bevaring av digitalt skapte privatarkiv**

<p>Prosjektets mål er å ordne og bevare de digitalt skapte arkivene etter Scanrope. Omfanget er på 372 Gb, ca. 25% er på arkivformat.</p> <p>Prosjektet inneholder en nyskapende del med konvertering av tekniske tegninger til arkivformat.</p> <p>Arkivet skal bevares i sikringsdepot i KDRS og formidles ut til publikum.</p> <p>Det er også bevart eldre papirarkiver fra Scanrope.</p>	<p>Søknadssum: 1750 000</p> <p>Budsjett: Lønnskostnader 245000 Lisenser 45000 Kontor/administrasjon 5000 Konsulent 55000</p>
--	--

Regnskap

Inntekter	2022	2023	sum
Øremerket tilskudd	175 000		175 000
Utgifter			
Lønnsmidler	102 010	72 991	175 001

Det ble ikke behov for å leie inn konsulent, men det ble brukt mer intern kompetanse både i opplæring og prosjektledelse. Administrasjon og programvare er innekket i egeninnsatsen på kr 175.000

Vennlig hilsen,
Rolf B. Holte
ITarkivar