Brukermøte Produktkategori Verktøy

Produktmøte 2020.05.27 kl. 13.00 - 15.00

Referat fra møtet:

*1. Velkommen og litt informasjon v/Siri Faye-Lund*

* Kort intro og presentasjon om Arkivverkets satsninger Innebygd arkiv og Digitalarkivet

*2. Arkade – ny funksjonalitet og kommende planlagt funksjonalitet v/Erik Aaberg*

* Erik holdt presentasjon og fortalte om ny og planlagt ny funksjonalitet i Arkade. Ny versjon kommer før sommeren i år. Presentasjonen vil bli tilgjengelig i etterkant.
* Det ble informert om en del mindre endringer og et par større (“velge hvilke tester
man vil kjøre” og deteksjon av filformat).
* Det ble også informert om bedring av ytelse (spesielt Disk I/O) og omlegging til nyere versjoner av programvare (.Net Core og biblioteker). Relatert til dette er en oppgradering av kommandolinje grensesnittet, som kan kjøre “kryss-plattform” og som gjør automatisering enklere.
* Erik informerte også om at det ikke er noe nytt om Arkadukt siden produktet er satt på sidelinjen og vi ikke har fått budsjett for å utvikle det noe videre.

#### 3. Tilbakemeldinger og forslag

* Om deteksjon av filformater. Arkade finner hvilken filtype en fil er i henhold til “Pronom” kategorisering[[1]](#footnote-32158). Men hva gjør vi med det (se også presentasjonen)? Skal det testes mot om filformater matcher forskriften? Skal det kreves i uttrekk? Skal det testes på filendelse?
* DROID og Siegfried: Innspill fra Sigve med spørsmål om Arkade bruker DROID. Erik informerer om at Arkade bruker Sigefried[[2]](#footnote-259), og det ble en liten diskusjon med sammenligning av de to som i utgangspunktet benyttes til samme formål.
* Bruk av Pronom-koder: Pål understreket at Pronom-kodene er viktig å få per fil heller enn aggregert. På spørsmål fra Erik svarte Pål at de ikke er flinke nok til å bruke Pronom-kodene foreløpig, men at planen er å benytte disse til vedlikehold i langtidsbevaring.
Det ble konkludert med at Arkade **finner** filtype, **lagrer** filtype pr. fil, men **viser** antall av hver type aggregert i rapporten. I første versjon vil ikke Arkade bruke identifisert filtype til å sjekke mot Riksarkivarens forskrift eller lignende.
* Diskusjon om versjoner av PDF-formatet og testing / validering:
Man var innom jungelen av PDF-formater, hvorav (max) 8 er tillatt i forskriften. Man ble enige om at Arkade via Siegfried identifiserer hvilken PDF “subtype” fila tilhører. Dette lagres i XML-rapportene, men brukes ikke for testformål (foreløpig). Det vil imidlertid være nyttig info i ettertid.
* ADDML og fagsystemer: Spørsmål fra Erik om det lages ADDML for fagsystemer og om det da brukes Arkade. Sigve Forteller at når de jobber med fagsystemer så bruker de utelukkende SIARD, og kjører da via Decom en DROID-analyse på filer i SIARD uttrekket. Videre når det gjelder mulighetene for innsyn så er det først og fremst i retning JSON vi tenker i arkivmiljøet i kommunal sektor. Sigve anså ADDML som overflødig - i det minste for fagsystemer
* ADDML og Arkadukt: Geir kommenterer at det er enighet i forumet om at ADDML for fagsystem er en tung jobb, og at Arkadukt var ment å skulle brukes til å lage en slik beskrivelse. Nå som det er stilt på sidelinjen og ikke vedlikeholdes eller utvikles, blir det en for tung jobb å lage ADDML til fagsystemer.
* Arkade og testing på gyldighet (“compliance”) av dokumenter i forhold til standardene: Pål stilte spørsmål om det ville være mulig å få til å teste filformater for dokumenter tilhørende fagsystemer der disse ligger utenfor selve fagsystemet. Diskusjon rundt dette. Erik nevner at det finnes andre og bedre eksterne verktøy for dette og at det derfor ikke virker hensiktsmessig å inkludere en slik funksjon i Arkade. Det vil imidlertid være mulig i fremtiden å “koble på” ekstern programvare, slik at den kalles fra Arkade, og der Arkade mottar resultatene. Det er slik det er gjort med Siegfried.
* Arkade og SIARD-testing: Pål spør videre om det vil være aktuelt å lage en funksjon i Arkade for å teste SIARD-filer. Diskusjon rundt dette. Erik svarer at det har vært diskutert, og at de har sett på et mulig samarbeid med det danske verktøyet ADA[[3]](#footnote-9296), men at det ikke er gått videre med dette foreløpig.
* Arkade og størrelse på uttrekk som testes: Nooshin stiller spørsmål om det er noen størrelsesbegrensning på uttrekkene en tester i Arkade. Litt diskusjon rundt dette også med tanke på at dette skal pakkes og at det da nå er en en grense på 1 Tb for de arkiv vi mottar. Erik og Jørgen T. informerer om at det på et tidligere tidspunkt eksisterte en slik begrensning, men at denne nå er fjernet. Det er også en kontroll av diskstørrelse nå i forkant av selve testingen slik at en får vite om det er nok plass før en setter i gang en større testing (heller enn å få vite dette når disken er full etter at halve testen er utført) noe som er en stor fordel ved testing av større uttrekk.
* Pål nevnte at det ble uoversiktlig i rapporten når det var svært mange arkivdeler, og der ikke alle ble listet. Dette blir annerledes i ny versjon, der alle arkivdeler grupperes og listes.
Relatert til dette: Arkivdelens navn vil nå vises i rapporten i tillegg til at system id angis.
* Generell tanke rundt mange forekomster av samme type i rapporten. Det må gjøres et kompromiss mellom kompletthet og lesbarhet. Det ble nevnt som en mulighet av testpunktene kunne lenke til mer komplette undersider med lister, i stedet for at alt tas med i hovedrapporten. Team Arkade tar dette med som en tanke videre.
* Arkade og SIARD: Spørsmål om Arkade kan lage SIARD-filer. Erik informerer om at det ikke er mulig. Arkade kan “pakke inn” i SIARD filer i henhold til vår DIAS pakkestandard, men ikke produsere.
* Arkade, Noark og SIARD: Sigve spiller inn at Arkade er veldig Noark-tung og at i kommunene er omtrent 90 % av det de bevarer fagsystemer, som bevares i SIARD-formatet. Han mener derfor at Arkade bør få litt flere ben å stå på, særlig nå som skjebnen til Noark virker noe uviss. Litt diskusjon rundt dette. Erik påpeker at denne typen valg av retning i utviklingen er på strategisk nivå, men det er åpenbart at SIARD brukes stadig mer. Erik sier han likevel fortsatt er i tvil om Arkade burde teste SIARD. Litt diskusjon rundt ulike muligheter for å inkludere SIARD-testing på noe vis i Arkade, Lars J foreslår at en kan bake inn en SIARD-validator i Arkade. Dette krever litt nærmere spesifikasjon. SIARD blir stadig mer viktig, men det finnes andre verktøy som kan gjøre mye (Spectral Full Convert[[4]](#footnote-10210), Database Preservation Toolkit[[5]](#footnote-16629), SIARD Suite[[6]](#footnote-31263)), og som nevnt danske ADA. Arkade utvikling vil ta kontakt med Sigve/KDRS/Lars for å tenke litt på hvordan man evt. kan integrere SIARD bedre med Arkade.
* Ytelse: Av “historiske” grunner leser Arkade nå filer fra disk flere ganger ved testing. Dette kan optimaliseres, og vil bli bedre i neste versjon. Dette kan gi vesentlige forbedringer ved testing av arkiver som inneholder svært mange filer.
1. <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx> [↑](#footnote-ref-32158)
2. https://www.itforarchivists.com/siegfried/ [↑](#footnote-ref-259)
3. https://www.sa.dk/da/offentlig-forvaltning/it-leverandoerer/ada-testprogram-arkiveringsversioner/ [↑](#footnote-ref-9296)
4. https://www.fullconvert.com/ [↑](#footnote-ref-10210)
5. https://database-preservation.com/ [↑](#footnote-ref-16629)
6. https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html [↑](#footnote-ref-31263)