



ARKIVVERKET
RIKSARKIVET

Samdok

samla samfunnsdokumentasjon

RAPPORT 2016



Delprosjekt Privatarkiv

Oppgave 10 Metodikk for høsting og langtidslagring av privatarkiv

Utarbeidet av
Arne-Kristian Groven / Ellen Røsjø

Rapportdato
18-03-2017



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 1 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

OPPGAVE	PRIV 2016_10 Metodikk for høsting og langtidslagring av privatarkiv (<i>videreføring av 2015/1</i>)
Ansvarlig delprosjekt	Privatarkiv
Arbeidsgruppens leder	Ellen Røsjø, Riksarkivet
Arbeidsgruppens medlemmer	Arne Kristian Groven (<i>Riksarkivet, fagansvarlig</i>), Martin Ellingsrud (<i>Arbark</i>), Pål Mjørland (<i>Fylkesarkivet i Oppland</i>), Stig Brunstad (<i>Bergen Byarkiv</i>), Jørgen Ø. Vik-Strandli (<i>Riksarkivet</i>), Per Kristian Ottersland (<i>Riksarkivet</i>), Torkel Thime (<i>Statsarkivet i Stavanger, Norsk olje- og gassarkiv</i>)
Målformulering	<ul style="list-style-type: none">• Verktøy og metode for høsting av digitalt skapt (privatarkiv)materiale er prøvd ut i minst 4 -5 forskjellige institusjoner.• Evalueringsrapport og anbefalinger foreligger.
Sammendrag	<ul style="list-style-type: none">• Denne SAMDOK-aktiviten har evaluert Archivemata og AtoM i totalt fire piloter for høsting av privatarkivmateriale. Programmene representerer begge «state-of-art» innenfor sitt fagfelt internasjonalt, i konkurranse med ikke alt for mange alternativer.• Gjennom pilot-studier i de fire organisasjonene har gruppen fremskaffet praktisk kunnskap om installasjon og bruk av verktøyene, og deres anvendelse på privatarkiv.• Verktøyene har vist seg hensiktsmessige på digitalt skapt privatarkivmateriale, men det er ingen (teknologiske) hindringer knyttet til anvendelse på digital informasjon fra statlig eller kommunal sektor.• Selv om prosjektdeltagerne har skaffet seg en del erfaringer, har vi ennå ikke nådd et samlet kompetansenivå som gjør oss i stand til å utnytte mulighetene som ligger i verktøyene til fulle, spesielt knyttet til optimalisering i forhold til skalérbarhet og automatisering av arbeidsprosessene. Dette medfører også at vi pr. i dag ikke med sikkerhet kan si om identifiserte forbedringspunkter er løsbare i verktøyene i dag, eller om det er endringsbehov utover riktig konfigurering av verktøyene.• Alt i alt har vi samlet mange positive erfaringer med verktøyenes funksjonalitet, og funnet noe vi bør undersøke nærmere.
Anbefalt oppfølging 2017	<p><u>Arkiverket:</u></p> <p>A. Arbeidet med høsting av privatarkiver:</p> <ul style="list-style-type: none">• Samordning av metodikk og verktøy med det som brukes for andre sektorer.• Proaktivt og systematisk arbeid basert på regionale bevaringsplaner.• En hensiktsmessig arbeidsdeling mellom lokalt, regionalt og nasjonalt nivå.• Digital kompetansebygging i institusjoner som jobber med privatarkiver. <p>B. Arbeidet med valg av hensiktsmessige verktøy/programvare:</p> <ul style="list-style-type: none">• Valg av hensiktsmessig programvare for håndtering av digitalt skapte arkiv. Revurdering av praksis med egenutvikling – fokus på bruk av verktøy og teknologi som er i bruk hos andre arkivinstusjoner – også internasjonalt.• Standardiserte "produksjonsløyper" for høsting/mottak/forvaltning av digitalt skapte arkiver, tilrettelagt for materiale fra alle sektorer.• Deling av kompetanse, metodikk og verktøy med alle deler av arkivsektoren. <p><u>SAMDOK 2017:</u></p> <ul style="list-style-type: none">• Delprosjekt PRIV gjennomfører en benchmark av AtoM vs Arkivportalen og Asta for sammenligning av funksjonalitet for arkivbeskrivelse og visning/gjenfinning av digitalt arkivmateriale.
Vedlegg	Rapporter fra pilotene: <ol style="list-style-type: none">1. Fylkesarkivet i Oppland2. Arbeiderbevegelsens Arkiv og Bibliotek (ArbArk)3. Bergen byarkiv4. Riksarkivet (Arkiverket)



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 2 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Innholdsfortegnelse

1. Innledning og bakgrunn	3
1.1 <i>Presisering og avgrensing av oppgaven</i>	3
2. Digital bevaring og tilgjengeliggjøring: Status i Norge	4
2.1 <i>Risiko forbundet med digital bevaring</i>	4
2.2 <i>Dagens digitale bevaringspraksis</i>	5
3. Prosjektgjennomføring	7
3.1 <i>Piloterings-deltakerne</i>	7
3.2 <i>Innhold i piloten</i>	7
3.3 <i>Prosjektets arbeidsform</i>	7
3.4 <i>Datasettene</i>	8
3.5 <i>Programvaren</i>	9
3.6 <i>Faser i arbeidsprosessen relatert til verktøy</i>	10
4. OAIS – grunnmuren i digital bevaring	11
4.1 <i>OAIS-standarden</i>	11
4.2 <i>OAIS sin funksjonsmodell</i>	11
4.3 <i>Informasjons-pakkene i OAIS</i>	12
4.4 <i>Praktiske utfordringer med OAIS</i>	13
4.5 <i>Andre standarder som kan øke bevaringsverdien</i>	13
4.6 <i>Faser i arbeidsprosessen relatert til OAIS-modellen</i>	14
5. Oppsummering av arbeid og funn	15
5.1 <i>Erfaringer med høsting av privatarkiver</i>	15
5.2 <i>Erfaringer vi har gjort oss med verktøyene</i>	15
6. Noen alternative veier videre	18
6.1 <i>Behov for videre utredning av Archivematica</i>	18
6.2 <i>Forbedringspotensiale integrasjon mellom Archivematica og AtoM</i>	18
6.3 <i>SiardSuite: Sikring av innhold for langtidsbevaring fra relasjonsdatabaser</i>	19
6.4 <i>Disk image: Sikring av innhold fra Mac-er og PC-er</i>	19
6.5 <i>Archivematica og AtoM i et framtidsscenario</i>	20
7. Konklusjon og anbefalinger	21
7.1 <i>Oppsummering</i>	21
7.2 <i>Potensialet i den piloterte programvaren</i>	22
7.3 <i>Anbefalinger om videre arbeid</i>	22
7.4 <i>Risiko forbundet med å vente</i>	24



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 3 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

1. Innledning og bakgrunn

I dag er det svært få arkivinstitusjoner som tar imot digitalt skapte privatarkiv. Dette til tross for at bedrifter, organisasjoner og personer allerede gjennom flere tiår i stor grad har sluttet å skape arkiv på papir. Vi risikerer store hull i vår kollektive hukommelse fordi (privat)arkiv som skapes i dag, ikke blir ivaretatt.

Generelle utfordringer på privatarkivfeltet er at bevaringen ikke er lovpålagt, men frivillig og avtalebasert. Det er ressurskrevende siden hvert enkelt tilfelle må håndteres spesielt. Samtidig er finansieringen av feltet underprioritert og ansvars- og oppgavefordelingen på feltet uklar. Digitale privatarkiv dør. Her fins mye materiale i form av fil- og mappestrukturer og databaser, det brukes ingen Noark-standard¹ - avleveringer må derfor planlegges tidlig. Vi kan ikke satse på å rakste i glørne etter konkurs/nedlegging. En mer planmessig bevaring må derfor skje med utvikling av bevaringsplaner for å kunne jobbe mer proaktivt for å utvikle en mer helhetlig samfunnsdokumentasjon. Blant de generelle tiltakene for å kunne få til dette er det at bevaringsinstitusjoner (i større grad enn i dag) må settes i stand til å ta imot og bevare digitale privatarkiv.

Langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale er generelt vanskelig, og det meste av bevaringsverdig informasjon er ikke tatt vare på, verken i Norge eller internasjonalt. Feltet er relativt umodent hva gjelder metodikk og teknologi, programvare-støtte (verktøystøtte). Men ser vi på det beste som finnes internasjonalt av teknologisk støtte for langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale, så finnes det i dag programvare-produkter på markedet som har vært i produktiv drift i snart ti år. Disse har nådd en modenhet og en kvalitet som det er verdt å se nærmere på.

Med dette som bakteppe ble det vedtatt å gjennomføre en oppgave i SAMDOK med tittelen «*Metodikk for høsting og langtidslagring av privatarkiv*» for å gjennomføre evaluering og utprøving av presumptivt gode programvare-kandidater hva gjelder langtidsbevaring og tilgjengeliggjøring av digitalt skapt materiale. Oppgaven ble lagt til delprosjekt Privatarkiv fordi bevaringsverdige digitale privatarkiv går tapt fordi institusjonene ikke er i stand til å ta imot og bevare dem. Dette er prekært for å oppnå en mer helhetlig samfunnsdokumentasjon. Men resultatene er like relevante også for materiale skapt i det offentlige.

1.1 Presisering og avgrensning av oppgaven

Målbeskrivelse:

- Valgte verktøy og metode for høsting er prøvd ut i minst 4 -5 forskjellige institusjoner
- Evalueringsrapport og anbefalinger foreligger.

Prosjektet har hatt to hovedfokusområder:

1. Å finne frem til en metodikk for høsting av privatarkiv som kan ivareta forskjellige informasjonstyper
2. Test av utvalgte verktøy som man tror kan være hensiktsmessige i dette arbeidet

I tillegg har det vært ønskelig å se metodikk og verktøy brukt på *privatarkiver* opp mot modeller og standarder for *generell håndtering av elektronisk skapt arkivmateriale*. I denne rapporten settes derfor arbeidet inn i en større sammenheng, illustrert ved OAIS-modellen (*se kapittel 4 nedenfor*).

¹ Med unntak av noen (større) bedrifter.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 4 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

2. Digital bevaring og tilgjengeliggjøring: Status i Norge

Det meste av bevaringsverdig digital informasjon er per i dag ikke tatt vare på. Privatarkivfeltet ligger nok etter både statlig og kommunal arkivsektor, men utfordringene er store på tvers av arkivsektoren.

I statlig sektor har Riksarkivet og noen Statsarkiv i flere tiår fokusert på Noark-uttrekk fra (generelle) Noark-godkjente sak-arkivsystem. Dette området har over tid stått overfor store kvalitative utfordringer, uten at dette har blitt skikkelig adressert før de senere årene:

- Artikkelen «*Uttrekkssammenligning- to uttrekk fra samme system*»², av Høiaas, Røros og Sødning fra 2016 tar for seg dette. En konsekvens av disse kvalitetsutfordringene er at unødvendig mye av Riksarkivets ressurser har blitt brukt på Noark-uttrekk de siste 15-20 årene.
- *En rapport utarbeidet av Riksarkivet i 2015*³, basert på svarene fra en statlig spørreundersøkelse hvor ca. 250 organer (store og små) innrapporterte, viser i overkant av 300 Noark-baserte elektroniske systemforekomster (IT-system/installasjoner i drift eller utfasede) og vel 2000 andre fagsystemforekomster (IT-system/installasjoner i drift eller utfasede). Det ble innrapportert over 1000 ulike(!) fagsystem. Dette er et betydelig antall flere fagsystemer enn hva man tidligere har antatt benyttes i statlig sektor. Disse tallene indikerer at svært mye bevaringsverdig digital informasjon fremdeles ikke er bevart også i statlig sektor.
- *Riksrevisjonens rapport*⁴ fra 2009 sier følgende: «Undersøkelsen viser manglende sikring og tilgjengeliggjøring av arkivmaterialet i kommunal sektor. Den manglende sikringen gir etter Riksrevisjonens vurdering økt risiko for at mye bevaringsverdig dokumentasjon går tapt. Undersøkelsen viser samtidig at Kulturdepartementets overordnede styring og oppfølging av sektoren er mangelfull.»
- SAMDOK-rapporten «*Digitalt skapt materiale i kommunal sektor 1985 til 2010 – kartlegging*»⁵, oppgave 2 fra 2014, forsøker å kvantifisere utfordringene i kommunal arkivsektor: «Gjennomsnittet for kommunene er beregnet til ca. 28,5 systemforekomster, totalt 12192 systemforekomster. Tallene må i høyeste grad betraktes som minimumstall.» Denne rapporten gir videre et estimat på tidsforbruk pr. IT-system med dagens metodikk, hvor det er snakk om ukeverk for å produsere kvalitativt god bevaring.

2.1 Risiko forbundet med digital bevaring

Det er flere typer risiko forbundet med bevaring og tilgjengeliggjøring av digitalt materiale, for eksempel:

- Risiko knyttet til at en har tatt vare på «bits and bytes» som etter hvert ikke kan leses og tolkes av programvare (lesbarhets-risiko)
- Risiko knyttet til at en ikke (lenger) har tillit til det bevarte og tilgjengeliggjorte digitale materialet, til at det er autentisk («tap av tillit»-risiko)

² <https://journals.hioa.no/index.php/arkiv/article/view/1670/1522>

³ <http://arkivverket.no/arkivverket/Arkivutvikling/Kartlegging-av-materiale-i-statleg-sektor/Rapport-for-kartleggingsprosjektet>

⁴ <https://www.riksrevisjonen.no/rapporter/sider/arkiv.aspx>

⁵ https://samdokdotcom.files.wordpress.com/2015/01/rapport-samdok-2014_2-kartlegging-av-digitalt-skapt-materiale-i-kommunal-sektor.pdf



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 5 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Lesbarhets-risiko knyttet eksempelvis til at:

- A. En konverterer til filformater som ikke overlever over tid: Feil valg gjøres med hensyn på konvertering (transformasjon) fra opprinnelig filformat til avledede filformater og representasjoner.
- B. Filformat-konvertering mislykkes og original-filer kastes, uten at dette oppdages i tide.
- C. Ved kryptering av filer, at krypterte filer ikke blir de-kryptert (gjort lesbare) forut for bevaring

Hva gjelder punkt A, så er det slik at noen filformater sannsynligvis forsvinner om noen få år, mens andre vil ha lengre levetid. En tommelfinger-regel er å forholde seg til formater som har mange brukere i dag.

Hva gjelder «tap av tillit»-risiko, så er tillit noe subjektivt. Noen av oss følger en ganske optimistisk tilnærming ved at vi basert på «magefølelsen» velger å tro på noe inntil det motsatte er bevist. Dette er på mange områder i livet en fornuftig strategi.

Med hensyn på langtidbevaring av digitalt arkivmateriale, så kan det vise seg å være en kostbar vei dersom «det motsatte bevises». I svært mange tilfeller vil det være umulig å gjenopprette tillit. Denne type optimistisk tillitsstrategi aksepterer vi på mange områder, men de fleste av oss vil ikke aksepteres dette i forhold til de transaksjonene vi gjør i nettbanken. Den bør ikke aksepteres i forhold til langtidbevaring og tilgjengeliggjøring da bevart digital informasjon noen ganger kan ende opp som bevis i rettsaker.

Men der hvor nettbankene kan skape tillit gjennom IT-krypto-verktøy (signering/bank-ID), kan ikke arkivene som driver langtidbevaring av digital informasjon gjøre det samme. Dette fordi de beste kryptoverktøy baserer seg på sertifikater med begrenset gyldighet.

Derfor må en ved langtidbevaring ta til takke med det nest-beste for å gjøre at flest mulig har tillit til arkivmateriale over lang tid. Dette inkluderer:

1. Bruk av sjekksummer, ved forflytning (i tid og rom) av digitale filer (et svakere krypto-verktøy)
2. I størst mulig grad å dokumentere filformat-konverteringer (endringer av filers format/representasjon), gjennom intelligente logger og dokumentasjon av programkoden som utfører filformat-konverteringer
3. Arkivbeskrivelser på alle nivåer, av bestand, opprinnelse, arkivskaper, arkivinstitusjon med mer
4. Automatisk registrering av tekniske metadata
5. Automatisk registrering av brukere som manipulerer materialet eller beskrivelsene.

Merk at sjekksummer, punkt 1, gjelder så lenge de digitale filene ikke gjennomgår en format-konvertering. Ved format-konvertering, punkt 2, må nye sjekksummer genereres og følge filene inntil neste konvertering, og så videre.

2.2 Dagens digitale bevaringspraksis

Generelt kan det slås fast at for få har (formell) kunnskap om digital bevaring og tilgjengeliggjøring i Norge i dag. Læresteder som Høgskolen i Oslo og Akershus har de siste årene begynt å utdanne bachelor-studenter på dette fagområdet, men kompetansen er jevnt over for lav.

Dette kan tilskrives også andre årsaker, utover at de fleste i dag ikke har formell fagbakgrunn på området, men mangel på kompetanse er et tungtveiende element. Et vel så viktig element er at arkiv-Norge i dag mangler gode verktøy, altså god programvare. Dessuten at en mangler god oversikt over de beste internasjonale programvare-verktøyene for langtidbevaring og tilgjengeliggjøring av digitalt



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 6 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

arkivmateriale, verktøy som har vært i produksjon opptil et tiår. Dette har igjen med manglende kompetanse å gjøre, og medfører at en kaster seg over egenutvikling av programvare uten å ha den tilstrekkelige innsikt og stor nok finansiering til at den resulterende programvaren noensinne kan måle seg med det beste som finnes internasjonalt.

Resultatet av disse forhold er at dagens bevaringspraksis virker både tidkrevende og tung og at en stadig sitter og venter på de egenutviklede verktøyene som ofte resulterer i skuffelser. Dagens praksis er heller ikke i stand til å sikre seg mot de risikoelement vi nettopp har berørt, på en god måte.

Ved digital bevaring er det mange deloppgaver som bør utføres, for eksempel:

- Kartlegging av verdifull digital informasjon (bevaringsplaner/bevarings- og kassasjonsvurderinger)
- Velge standarder for arkivbeskrivelser og metadata
- Registrering av arkivbeskrivelser, metadata og annen tilleggsinformasjon på ulike nivåer
- Uthenting av digital informasjon fra kilde-systemet, helst uten at autentisitet går tapt i første steg(!)
- Flytting av den digitale informasjonen, i tid og rom, til depot
- Sikre seg mot informasjonstap, f.eks. gjennom generering og validering av sjekksummer
- Virussjekk og karantene
- Identifisere og validere filformater
- Validere mottatt informasjons form og innhold, når dette er påkrevd, eksempelvis Noark-validering/testing
- Velge et velegnet (innpaknings-) format for langtidsbevaring, og et annet for visning
- Bevare opprinnelige filer i originale filformat
- Konvertere originale filer til presumptivt mer velegnede bevarings-filformater og visnings-filformater
- Generere unike identifikatorer, for å lette gjenfinning
- Osv.

Alle disse deloppgavene må/bør utføres og (automatisert) dokumenteres for å skape tilstrekkelig tillit til at den bevarte informasjonen er autentisk og for å sikre lesbarhet (visbarhet). Dagens praksis har mangler i forhold til flere av disse punktene.

Punktene illustrerer også behov for velegnet, helhetlig programvare. Da denne mangler i dag, er det derfor ikke noe rart at SAMDOK-kartleggingsrapporten fra 2014 sier: «Arbeidsmengden, tiden det tar å utføre arbeidsprosessen angitt over på en systemforekomst første gang vil for de mest kompetente medarbeiderne være ca. 1 månedsverk (= 4 ukeverk) med utgangspunkt i dagens metodikk.» Selv om de mest kompetente medarbeiderne gjør alle disse arbeids-stegene vil det allikevel kanskje ikke være nok, grunnet for dårlige verktøy/programvare.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 7 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

3. Prosjektgjennomføring

Prosjektet det her rapporteres fra, utgjør pilotfasen av Oppgave 10 i SAMDOK for 2016, *Metodikk for høsting og langtidslagring av privatarkiv*. Piloteringen er en videreføring av en aktivitet som startet i 2015, og arbeidet er avsluttet pr 31/12 2016.

3.1 Piloterings-deltakerne

Deltakere i piloten har være følgende arkivinstitusjoner/personer:

- Fylkesarkivet i Oppland v/ Pål Mjørlund
- Bergen byarkiv v/ Stig Brunstad
- Arbark (Arbeiderbevegelsens arkiv og bibliotek) v/ Martin Ellingsrud og Dag Juvkam
- Riksarkivet v/ Arne-Kristian Groven og Jørgen Østengen Vik-Strandli

Det er disse som lokalt har gjennomført uttesting gjennom å kjøre pilotstudier i egen institusjon.

Arne-Kristian Groven, Arkivverket, har vært faglig ansvarlig for arbeidet i gruppen, og Ellen Røsjø, Arkivverket, har vært arbeidsgruppens leder.

3.2 Innhold i piloten

Vi ønsket å se på:

- Depotets mottak av digitalt (skapt) arkivmateriale
- Hvordan sikre kvaliteten og ektheten til materialet som mottas
- Teste egnethet til utvalgt verktøystøtte/programvare
 - Kopieringsverktøy for fil- og mappestrukturer, e-post
 - SIARD som verktøy for bevaring av databaser
 - Konvertering av e-post med mer med Archivemata
- Ta uttrekk fra faste, etablerte systemer som finnes hos arkivskaper

3.3 Prosjektets arbeidsform

Prosjektet har hatt til sammen fire fellesmøter i tidsrommet juni 2015 til april 2016. Utover møter har kommunikasjonen i prosjektet stort sett foregått på e-post og telefon.

Prosjektet avholdt også en fagdag i Riksarkivet den 8. november, som ledd i arrangementet "*Digitalt tre dager til ende*", hvor 60-70 deltakere fra alle sektorer deltok. Dagen ble innledet av en representant fra *Artefactual Systems* som demonstrerte verktøyene *Archivemata* og *AtoM* (se kap 3.5 nedenfor), før deltakere i prosjektet hadde en gjennomgang av sine erfaringer med programvaren, som de har brukt i sitt arbeid. Arkivverket presenterte også noen av prosjektene det jobbes med. Det ble en bred diskusjon i salen om behov for funksjonalitet og løsninger i håndteringen av digitalt materiale, alt i alt svært nyttige innspill som prosjektet har tatt med videre i sitt arbeid.

Da noen av prosjektdeltakerne fra de ulike institusjonene følte behov for faglig påfyll forut for piloteringen fikk vi tatt et kurs i Ubuntu Linux server-administrasjon hos Redpill-Linpro AS i Oslo i mars 2016. I slutten av september 2016, på et tidspunkt hvor vi hadde opparbeidet en del egenkompetanse, dro vi også på



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 8 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

besøk hos Borthwick Institute for Archives i York, hvor dei flere år har benyttet seg av AtoM for håndtering av arkivbeskrivelser og var i ferd med å sette Archivemata i drift. De er også sentrale i Archivemata og AtoM fagforum/-nettverk i Storbritannia, nettverk som har europeisk deltagelse.

Programvaren Archivemata og AtoM ble installert lokalt hos Bergen byarkiv, Fylkesarkivet i Oppland og Riksarkivet. Arbeiderbevegelsens arkiv og bibliotek hadde ikke ved oppstart nok teknisk kompetanse til å installere i egen organisasjon. De valgte derfor å leie skytjenester fra Artefactual Systems, eierne av Archivemata og AtoM, og deres samarbeidspartnere i Tyskland. Denne erfaringen ble en berikelse for prosjektet, som da kunne skaffe seg erfaringer med IT-support i skyen.

Vi hadde da totalt fire ulike installasjoner, enten lokalt eller i skyen, og prosesserte de utvalgte datasettene her.

3.4 Datasettene

Vår utprøving/piloter ble gjennomført på privatarkiv-materiale i form av fil-/mappestrukturer i fire arkivinstitusjoner:

Fylkesarkivet i Oppland

- Diabetesforbundet avd. Oppland
 - 12 disketter som var samlet og lagret i et filhierarki der hver diskett fikk sin mappe. Det var ikke foretatt noe bearbeiding av innholdet, men produsert en fil som forklarte oppbyggingen av mappestrukturen.
- Hageselskapet avd. Oppland
 - Her ble det hentet en filstruktur bestående av mange mapper og dokumenter. Katalogen var i hovedsak administrert av hageselskapets sekretær, og hun hadde laget seg sitt eget system for å organisere dokumentene.
- Oppland arbeiderparti
 - Dette består av en filstruktur på 9.3 gigabyte fordelt på 9786 filer i 2267 mapper. Filstrukturen ble pakket i en 7zip fil for å beholde fil- og mappedatoer. Det ble i tillegg produsert en CRC sjekksum for hver fil.

ARBARK, Arbeiderbevegelsens arkiv og bibliotek

- Attac
 - Fil- og mappearkiv inneholdende nærmere 30 000 dokumenter (filer). Filene bestod av forskjellige formater, men i hovedsak var det doc/docx, PDF, power point, excel og JPEG-filer. Dataene ble hentet ut ved hjelp av kopiprogrammet Beyond Compare.

Bergen byarkiv

- Ramm Salbu Arkitekter
 - Arkivet inneholder ca. 40.000 filer fordelt på ca. 4300 mapper (etter den første runden med arkivbegrensning). Materialet er relativt omfattende, men har en relativt god orden (noe rot er det).



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 9 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Riksarkivet

- Sofieprisen (tidligere Sofiestiftelsen)
 - Dette er arkivet etter den nå nedlagte stiftelsen som har utdelt Sofieprisen i perioden 1998-2013. Dette består av en filstruktur på 2.1 gigabyte fordelt på 10.138 filer og 2.116 mapper.
- Union-bilder
 - Et utvalg av gamle bilder fra nå nedlagte Union brug i Skien, bestående av kun 32 bilder i en mappe. Størrelsen her var 157 megabyte.

3.5 Programvaren

Programvaren vi har prøvd ut er:

- *Archivematica*⁶, et (avansert) web-basert bevaringssystem som følger arbeidsflyten angitt i OAIS standarden⁷, ISO 14721. Programvaren er åpen kildekode og kan fritt lastes ned⁸.
- *AtoM*⁹ (Access to Memory), et web-basert verktøy for arkivbeskrivelser i hht. internasjonale standarder, visning og gjenfinning av digitalt innhold. Programvaren er åpen kildekode og kan fritt lastes ned¹⁰.

Både Archivematica og AtoM utvikles av samme programvare-leverandør, *Artefactual Systems*¹¹.

Vår opprinnelige ambisjon var også å teste ut verktøy for bevaring av relasjonsdatabaser, *SiardSuite* (Siard)¹². Dette har vi ikke fått gjort, men tar med litt i rapporten basert på prosjektpartnerenes erfaringer fra andre prosjekter.

Felles for Archivematica, Atom og SiardSuite er at de kan regnes å være «state-of-art» verktøy innen digital langtidsbevaring og tilgjengeliggjøring. Verktøyene og det teoretiske fundamentet har blitt presentert i internasjonale forskningspublikasjoner siden tidlig på 2000-tallet, og de får stadig flere brukere globalt sett. De har alle tre vært i produksjon i omtrent et tiår.

Et annet verktøy, *Preservica*, kan også regnes blant disse, men forskjellen er forretningsmodellen. Mens verktøyene vi har testet ut faller inn under kategorien åpen kildekode/fri programvare/åpne standarder, så tilhører Preservica den mer lukkede delen av IT-verdenen. Her er kunnskap om programmets indre struktur og datarepresentasjon ikke fritt tilgjengelig og lisenser for å få teste ut programvaren ville beløpe seg til flere hundretusen i vårt prosjekt. Dette var utenfor de økonomiske rammer vi hadde til rådighet.

⁶ <https://www.archivematica.org/en/>

⁷ http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

⁸ <https://www.archivematica.org/en/docs/archivematica-1.5/admin-manual/installation/installation/>

⁹ <https://www.accesstomemory.org/en/>

¹⁰ <https://www.accesstomemory.org/en/download/>

¹¹ <https://www.artefactual.com/>

¹² <https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 10 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

3.6 Faser i arbeidsprosessen relatert til verktøy

Nedenfor er de ulike fasene/arbeidsprosessene/hovedoppgavene som inngår i (langtids-) bevaring og tilgjengeliggjøring av digitalt innhold, satt opp mot eksisterende programvare som brukes i produksjon i ulike deler av arkiv- og museums-Norge, og programvaren vi har evaluert i dette SAMDOK-delprosjektet:

	FASE I ARBEIDSPROESSEN¹³	DAGENS PRAKSIS / VERKTØY	VERKTØY SOM ER EVALUERT
1	Mottak og preprosessering av digitalt innhold og assosiert tilleggsinformasjon	Ingen helhetlige system som enkelt kan automatisere og dokumentere delprosessene. Mye tidkrevende manuelt arbeid med (for) lite/dårlig programvarestøtte og for stor feilrisiko.	Archivematica
2	Bevaringsplanlegging (NB: OAIS terminologi ¹⁴)	Ingen helhetlige system	Archivematica
3	Depotstyring	ESSArch (Riksarkivet, KDRS, Bergen byarkiv), ellers direkte bruk av filsystem/databaser (uten spesifikke programoverbygg).	(A) Archivematicas depotstyringsmodul («storage management») med underliggende filsystem, eller (B) integrasjon med andre depotløsninger, eller (C) begge sammen.
4	Visnings-, gjenfinnings- og tilgjengelighetsstyring	Arkivportalen, Digitalarkivet, Digitalt museum, LoCloud. Lokalt utviklede innsynsløsninger.	AtoM (Access to Memory)
5	Vedlikehold av metadata og arkivbeskrivelser	Asta (hovedsakelig)	AtoM (Access to Memory)

¹³ Da har vi ikke tatt med forberedende arbeid som ligger forut for mottak av arkivverdig, digitalt materiale, herunder bevarings- og kassasjonsvurderinger, avtaleinngåelse, innsamling/«tapping»/transformasjon/uttrekk av innhold fra informasjonens opprinnelsessystem, test av innhold mot kravsett (eksempelvis Noark-standarden).

¹⁴ Her er det verdt å merke seg at OAIS snakker om ordet «bevaringsplanlegging» i en mer teknisk relatert kontekst knyttet til risikostyring for å unngå tap av lesbarhet/autentisitet. Det samme ordet har vanligvis en annen betydning i arkiv-Norge, hvor ordet «bevaringsplanlegging» knytter seg til å lage bevaringsplaner som kartlegger det som er verdifullt å ta vare på.

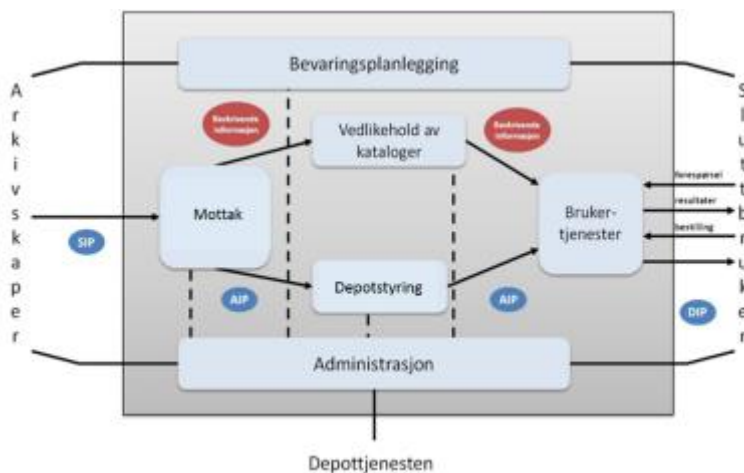


SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 11 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

4. OAIS – grunnmuren i digital bevaring

4.1 OAIS-standarden

Standarden ISO 14721 -OAIS, «Open archival information system», er den mest brukte standarden innenfor digital langtidsbevaring og aksess i dag. «Alle» bruker denne standarden.



Standarden består av to hoveddeler:

- Overordnede beskrivelser/kravsett knyttet til de ulike funksjoner/arbeidsprosesser.
- Overordnede beskrivelser av informasjons-«klumpenes» (pakkenes) struktur og innhold.

4.2 OAIS sin funksjonsmodell

Den overordnede funksjonsmodellen i OAIS gir oversikt over hvilke hovedfunksjoner og informasjonsflyt som må være på plass i et digitalt, langtidsbevarende arkiv.

Hovedaktivitetene er:

- Mottak av det digitale materialet og assosierte arkivbeskrivelser og metadata. («Ingest».)
- Bevaringsplanlegging¹⁵, risiko-styrt planlegging for å redusere informasjonstap og tillit. («Preservation Planning».)
- Vedlikehold av kataloger/arkivbeskrivelser. («Data Management».)
- Depotstyring. («Archival Storage».)
- Brukertjenester, det vil si brukertilgang, gjenfinning og tilgjengeligjøring. («Access».)

I tillegg er det i figuren over angitt noe som kalles SIP-er, DIP-er og AIP-er. Dette er angivelse av ulike informasjonstyper i OAIS og hvor i informasjonsflyten disse oppstår.

¹⁵ La oss nok en gang minne om at dette begrepet tradisjonelt brukes annerledes i arkiv-Norge.



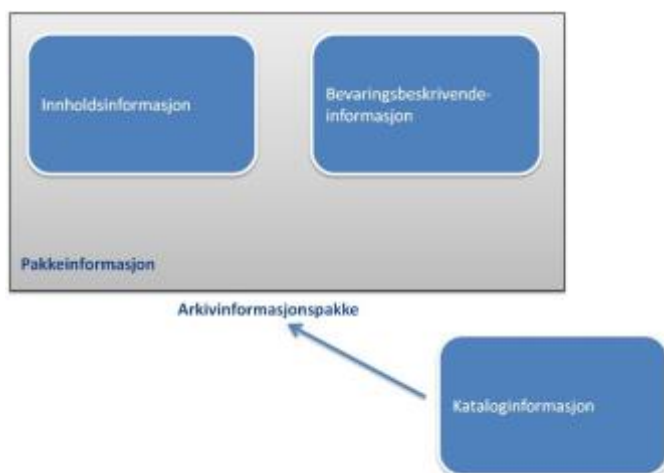
SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 12 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

4.3 Informasjons-pakkene i OAIS

SIP-er, DIP-er og AIP-er tre ulike typer såkalte informasjonspakker definert i OAIS, de kan forklares som følger:

- SIP, «Submission Information Package», eller såkalt **mottakspakke** på norsk. Denne kan enten mottas i sin helhet fra arkivskaper, organisasjonen som er opphav til informasjonen som ønskes bevart, eller skapes ved mottak i arkivinstusjonen, basert på mottak av ulike typer informasjon og arkivbeskrivelser.
- AIP, «Archival Information Package», eller såkalt **arkivpakke** på norsk. Denne skapes basert på SIP-en eller SIP-ene som mottas. Det kan være en-til-en, mange-til-en eller en-til-mange relasjon mellom SIP-er og avledede AIP-er. I AIP-er bevares informasjonen på de presumtively beste filformatene hva gjelder informasjonsrikdom og forventet levetid til filformatene. Floraen av arkivbeskrivelser og metadata (-standarder) er også rikere for AIP-er enn for DIP-er.
- DIP, «Dissemination Information Package», eller såkalt **visningspakke** (utleveringspakke) på norsk. Også her kan det være en-til-en, mange-til-en eller en-til-mange relasjon mellom AIP-er og (avledede) DIP-er. I DIP-er bevares informasjonen på de presumtively beste filformatene hva gjelder digital overføring og lesbarhet over internett. Filformater som kan komprimeres verdsettes da dette medfører raskere overføring til bruker. I praksis innenfor «state-of-art» løsninger genereres DIP-er enten fra AIP-er eller direkte fra SIP-er.

Selv om de tre ulike typer informasjonspakker i OAIS er forskjellige innholdsmessig, er de bygget på samme strukturelle inndeling¹⁶:



Informasjonen er strukturert i to hovedbolker:

- Innholdsinformasjon, som foruten data-ene inneholder informasjon som gjør det mulig å oppnå riktig tolkning av dataene.
- Bevaringsbeskrivende informasjon, for å sikre autentisitet og kontekstinformasjon

Metadata- og arkivbeskrivelser på ulike nivåer i en informasjonspakke blir også brukt til gjenfinning.

¹⁶ Kilde: Thomas Sørdring, undervisningsmaterieil, Høgskolen i Oslo og Akershus.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 13 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Ser vi på hvordan bevaringsbeskrivende informasjon («Preservation Description Information») er strukturert, så vil informasjonspakker i henhold til OAIS inneholde følgende typer informasjon (arkivbeskrivelser og metadata):

- Autentisitet- og integritets-sikrende informasjon («Fixity Information»)
- Proveniens-sikrende informasjon («Provenance Information»)
- Referanser («Reference Information»)
- Kontekst-sikrende informasjon («Context Information»)
- Aksess-rettighetsinformasjon («Access Rights Information»)

Ser vi på hvordan innholdsinformasjonen er strukturert i henhold til OAIS, så har vi følgende typer:

- Selve det digitale objektet, altså bits og bytes, sekvenser av nuller og enere. (Eller alternativt, et fysisk objekt, som finnes et eller annet fysisk sted!)
- Representasjonsinformasjon, beskrivelser av struktur og semantikk, som for digitale objekt er helt nødvendig for at bits og bytes skal kunne tolkes korrekt. (Representasjonsinformasjon inneholder informasjon om filformatet til det digitale objektet.)

4.4 Praktiske utfordringer med OAIS

Selve OAIS-standarden (OAIS-rammeverket) er et ca. 140 siders dokument som danner ryggraden i moderne digital (langtids-) bevaring. Standardisert informasjonsstruktur og arbeidsprosesser er bra, men som for alle standarder er det en avveining hva gjelder detaljeringsgrad: Har du en for detaljert standard, så vil den kunne realiseres av få. Skal den realiseres av mange, så vil detaljgraden gå ned.

Dette siste er tilfellet med OAIS. Det finnes derfor mange ulike realiseringer av OAIS. Om det er ulike realiseringer av arbeidsprosessene og informasjonsflyten, så har det mindre konsekvenser. Ulik programvares realisering av OAIS-informasjonspakkene gir større utfordringer.

4.5 Andre standarder som kan øke bevaringsverdien

Ulike standarder kan brukes på ulike nivåer i en OAIS informasjonspakke (SIP, AIP eller DIP) for å øke bevaringsverdien. Disse vil her bare kort bli nevnt, uten at listen på noen måte er utfyllende:

- PREMIS, «Preservation Metadata Maintenance Activity», standarden kan brukes for (strukturert) loggføring av manipulasjoner på informasjonspakker
- METS, «Metadata Encoding & Transmission Standard», kan brukes for å beskrive strukturen på digitalt innhold
- ISAD(G), «General International Standard Archival Description», for å beskrive arkivmaterialet
- ISAAR-CPF, «International Standard Archival Authority Records (Corporate bodies, persons, and families)»
- ISDF, «International Standard for Describing Functions»
- ISDIAH, «International Standard for Describing Institutions with Archival Holdings»
- DC, «Dublin Core»



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 14 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

4.6 Faser i arbeidsprosessen relatert til OAIS-modellen

Preprosessering (OAIS: «Ingest») og bevaringsplanlegging (OAIS: «Preservation Planning»)

Det er altså store hull i dagens praksis både hva gjelder programvarestøttet (1) mottak, og (2) (OAIS-) bevaringsplanlegging. Disse hullene vil kunne dekkes av Archivemata, noe som vil medføre:

- **Bedret kvalitet** på det som produseres for langtidbevaring og tilgjengeliggjøring da manuelle feilkilder elimineres og antall utførte kvalitets-økende tiltak blir langt høyere enn nåværende praksis.
- **Bedret sporbarhet** da vi kan snakke om en etterprøvbar bevis-kjede («chain of evidence») som dokumenterer hva som er gjort, framfor en optimistisk antagelse/tro på at noe er gjort riktig. (Noe som ofte ikke er tilfelle.)
- **Store tidsbesparelser** da tidsforbruket reduseres fra uker til timer/minutter.

Depotstyring (OAIS: «Storage Management»)

Her kan arkivinstusjonene deles inn i forskjellige kategorier:

- De som har valgt depotstyringsløsninger de seneste årene. Eksempelvis har Riksarkivet/Arkivverket valgt ESSArch.
- De som vurderer å velge depotstyringsløsninger
- De som ikke har avklart dette.

Benyttes Archivemata så kan Archivematas depotstyrings-grensesnitt benyttes dersom en ikke allerede har depotstyrings-verktøy i egen organisasjon. Dette gir blant annet aksess basert på tilgangskontroll. Det digitale innholdet er pakket inn i en informasjonsstruktur som er i henhold til OAIS-standarden, og det lagres i fil- og mappestrukturer.

Benyttes eksempelvis ESSArch som hos Arkivverket så kan Archivematas fil- og mappestruktur fungere som stedet hvor ESSArch henter det digitale innholdet fra. Dette krever visse tilpasninger, men muliggjør en sameksistens mellom Archivemata som «front-end» og ESSArch eller andre depotstyringsløsninger.

Internasjonalt finnes det flere gode og modne depotstyringsløsninger som kan benyttes. Flere av disse er tett integrert med Archivemata.

Visning og gjenfinning (OAIS: «Access»)

Internasjonalt benyttes AtoM, Access to Memory, av en rekke museer og arkivinstusjoner. AtoM tjener flere formål:

- Visning av alle mulige typer digital informasjon
- Effektiv gjenfinning av denne informasjonen, både via fulltekst-indeksering, geografiske referanser og via arkivbeskrivelser

AtoM dekker omtrent de samme behov som eksempelvis Arkivportalen eller Digitalarkivet. Vi ønsker i 2017 å gjennomføre en sammenligning av AtoM med disse teknologiene.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 15 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Vedlikehold av metadata og arkivbeskrivelser (OAIS: «Metadata management»)

Arkivbeskrivelsene i AtoM kan enkelt fylles ut av bruker, eller importeres inn/eksporteres ut av verktøyet. Disse er standardisert i henhold til kjente standarder som ISAD(G), ISAAR-CPF, ISDIAH, ISF med flere.

Det er også mulig å legge inn egne (eksempelvis nasjonale) standarder i AtoM. Da vil sammenfallende arkivbeskrivelser og metadata kunne nås via ulike standarder, uten at det krever dobbeltarbeid.

Hva gjelder arkivbeskrivelser, så har Asta-verktøyet mange brukere i Norge.

5. Oppsummering av arbeid og funn

5.1 Erfaringer med høsting av privatarkiver

Alle arkivene vi håndterte i prosjektet var basert på filstrukturer. Det var ingen databaseløsninger eller spesialiserte fag-/registreringssystemer. Dette tror vi er typisk for en stor gruppe privatarkiv, eksempelvis de som mottas fra ulike organisasjoner.

For private bedrifter kan nok situasjonen være en annen. I hvert fall dersom de er over en viss størrelse og befinner seg innenfor regulerte bransjer. Her vil det være større krav til organisert dokumentasjon og større krav til dokumentasjon av avtaler. Gjerne fulgt opp med registreringssystemer som bekrefter tidspunkter og håndtering av dokumentasjonen. Det samme vil nok også gjelde bedrifter som jobber med produktutvikling og trenger å dokumentere ideelle rettigheter.

Men generelt ser det ut som at (kvaliteten på) privatarkivene vi har håndtert vil være svært personavhengig. Institusjonene vi har hentet materiale fra har ikke hatt spesielle krav om dokumentasjon i forhold til eventuelle håndtering av klienter med mer. Det hadde nok vært ønskelig med bedre kontroll i forhold til ansatte eller samarbeidspartnere. Det lave ekstone kravet om dokumentasjon har gjort at dokumentasjon av saksbehandling har blitt håndtert uten en sterk styring eller kontroll. Struktur og navngiving av datafiler er basert på brukernes prioritering av viktighet og innhold. Datahåndteringen har derfor vært mindre regulert enn vi ser i offentlig sektor.

5.2 Erfaringer vi har gjort oss med verktøyene

Archivematica

Archivematica er en kompleks server-applikasjon (programvare) som må installeres på Ubuntu Linux server plattform, enten stabil versjon 14.04 LTS eller nyere 16.04 LTS. Vi (installerte og) kjørte førstnevnte versjon. Erfaringene vi gjorde var at Archivematica isolert sett var relativt grei å installere, men at det var vanskeligere enn vi trodde å koble Archivematica opp mot AtoM.

Funksjonaliteten følger strengt flyten i OAIS-standard, med menyvalg som har navnelikhet med (hoved-) funksjoner i OAIS-standard. Den første hovedfunksjonen som kjøres når digitalt materiale hentes kalles overføring, «Transfer», som igjen består av ulike delfunksjoner. Dette er ulike programmer som utfører isolerte deloppgaver. De fleste deloppgavene knyttet til overføring ble utført på en grei måte, slik prosjektgruppa ser det, med unntak av håndtering av såkalt «backlog» som er et oppsamlingsområde



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 16 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

dersom materiale kommer inn i biter og skal samles opp før videre prosessering. Her opplevde minst en prosjektpartner spesielle utfordringer. Generering av SIP fungerte også stort sett greit, med unntak av følgeproblem backlog-problematikken.

Depotstyringsmodulen til Archivemata, «Storage management»-modulen, har en ikke for avansert funksjonalitet. Men funksjonalitetene «storing/delete/search» AIP (arkivpakke) med mer fungerte greit. Likeledes fungerte lagring av DIP (visningspakke). Det er valgfritt om denne modulen brukes, avhengig av om en har andre systemer tilgjengelig for henholdsvis depotstyring av arkivpakker og visningspakker.

To problemstillinger dukket opp hos flere partnere. Den ene knyttet seg til henholdsvis filformat-identifikasjon og den andre til den regelstyrte konverteringen av filformater, fra opprinnelige filformater til henholdsvis presumptivt velegnede langtidbevarende formater eller presumptivt velegnede visningsformater. Eksempelvis vil TIFF være et velegnet langtidbevarende filformat, mens JPG er velegnet for visning. Et siste diskusjonspunkt dreier seg om koblingen mellom Archivemata og AtoM. Disse problemstillingene vil bli diskutert dypere senere i denne rapporten.

AtoM (Access to Memory)

AtoM er et attraktivt verktøy som dekker både arkivbeskrivelse og aktørbeskrivelse i henhold til de standarder som benyttes i stort omfang internasjonalt. Dette er en samstemt erfaring fra alle fire pilotene. Det er generelt enkelt å legge inn beskrivelser via brukergrensesnittet.

Import- og eksportfunksjonaliteten, eksempelvis på EAD-formatet, er noe som trekkes frem som spesielt bra blant prosjektpartnerne som har testet dette ut. Dette muliggjør relativt enkel deling av arkivbeskrivelser mellom flere installasjoner, samt å bidra til å gjøre registreringsarbeidet raskere. Denne delingen av beskrivelser kan følgelig også skje på tvers av institusjoner, men det ble ikke testet ut i prosjektet.

I AtoM får du tilgang til arkivbeskrivelser og digitale objekter på mange måter, via (avansert) søk, smålesings-funksjonalitet («browse»), navigasjonsfunksjonalitet, kontekstmeny med mer. Vår erfaring er at dette fungerer greit.

Det digitale materialet, objektene, det være seg (tradisjonelle) dokumenter av ulik filtype, excel-filer, ulike bilde-, lyd- eller videoformater vises enkelt i AtoM, i ulike kvaliteter. Objektene, i form av miniatyrer («thumbnails») og omsluttende arkivbeskrivelser befinner seg i samme visningsflate. Skal en aksessere objektet, trykker en bare på thumbnailen. For video-fremvisning er det eksempelvis en plugin som viser video i full eller redusert kvalitet/størrelse. Foruten video-plugins er det er andre innebygde plugins, eksempelvis for visning av PDF-dokumenter, ulike bilde- eller lydformater osv. Der det ikke finnes plugins, kan en allikevel trykke på miniatyren; og er riktig programvare installert på PC-en/plattformen, startes dette automatisk. Slik er det eksempelvis for MS Office-formatene og også for SIARD-filene vi sendte inn da vi hadde SiardSuite på PC-en. Dette siste muliggjør visning av databaseinnhold via AtoM.

AtoM skal kunne la seg installere på de ulike IT-plattformer, Linux, Mac og Windows. Erfaringene vi gjorde var at AtoM isolert sett var relativt grei å installere, men at det var relativt vanskelig å koble AtoM opp mot Archivemata. Dette overrasket oss litt. Både Archivemata og AtoM er komplekse server-applikasjoner, inneholdende web-server (Apache eller Nginx), Elasticsearch, Oracle Java, MySQL-server, PHP, Memcached



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 17 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

og Gearman job server. Installasjons-prosedyrene er veldokumenterte, med unntak av integrasjonen mellom Archivematica og AtoM. Her slet vi litt, men klarte det til slutt.

Forskjeller i kompleksitet mellom Archivematica og AtoM

Det er verdt å merke seg forskjellene på Archivematica og AtoM, på tross av at de begge har samme tjenerarkitektur (server-arkitektur).

Archivematica er en veldig prosesseringsintensiv, automatisert programvare hvor delkomponenter, programvare med ulikt opphav, gjenbrukes i Archivematica. Disse programkomponentene realiserer den store mengde delprosesser som til sammen utgjør en kvalitativ god langtidsbevaring.

Denne koblingen mellom programvare fra ulikt opphav er mulig fordi både Archivematica (og AtoM) program-overbygget og delkomponentene er utviklet under fri-programvare/åpen kildekode-lisenser. Disse lisensene er tilrettelagt for slik gjenbruk. Ettersom majoriteten av den beste langtidsbevarende programvaren som finnes internasjonalt i dag er under slike lisenser, vil Archivematica og AtoM kunne nyttegjøre seg det beste som finnes i svært stor grad.

Men det å administrere helheten som Archivematica gjør, er prosesseringsintensivt og komplekst selv om det kanskje ikke kan virke sånn fra et brukerperspektiv. Her har også prosjektdeltakerne en vei å gå da vi ikke kan påberope oss å ha forstått mulighetsrommet som ligger i de mange oppsett og konfigurasjonsmuligheter fullt ut ennå.

AtoM på den annen side er mye mindre kompleks. Her er det snakk om å ha god import/eksport-funksjonalitet basert på ulike internasjonale (eller nasjonale) standarder. Utover det skjer mye ved hjelp av brukerdialoger og datasettene gjennomgår ikke slik omfattende prosessering som for Archivematica.

Diskusjon

Før vi kort beskriver noen av de subjektivt opplevde utfordringene og svakhetene vi har støtt på, er det verdt å utdype hva vi mener med bruken av ord som svakhet og utfordring og videre hva denne opplevelsen kan skyldes.

- En opplevd svakhet eller utfordring er et avvik, nemlig avviket mellom hva verkøyet (enten Archivematica eller AtoM hver for seg, eller i samspill) presterer satt opp mot vår subjektive forståelse av idealsituasjonen.

Idealsituasjonen har i dette prosjektet stort sett vært en subjektiv opplevelse da vi er i en tidlig fase rent kompetansemessig. Uansett, dersom flere piloter opplever de samme svakheter og utfordringer, er det en indikasjon på at flere partners idealsituasjon ikke nås innenfor rammen av vår nåværende verktøybruk og verktøykompetanse. Årsakene til de opplevde svakheter og utfordringer knyttet til verktøystøtten kan følgelig ligge på flere nivå:

1. Vi har ikke oppdaget den mest velegnede funksjonaliteten eller de mest velegnede oppsettene (konfigurasjonene) ennå.
2. Vi har oppdaget noe reelt som vi ikke liker og henger oss opp i dette, men det kan være mer en detalj i den store sammenhengen.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 18 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

3. Vi er på sporet av noe substansielt, men er ikke på nåværende tidspunkt i stand til å vite helt sikkert om det kan tilskrives punkt 1 eller ikke.

Det gjenspeiler hvor langt vi har kommet i løpet av prosjektet. Vi har kommet langt, men videre arbeid gjenstår for å kunne yte programvaren full rettferdighet.

Men det vi kan si er at den overordnede opplevelsen er (veldig) positiv. Vi tror både Archivematica og AtoM representerer store positive bidrag til digital langtidsbevaring og tilgjengeliggjøring.

6. Noen alternative veier videre

6.1 Behov for videre utredning av Archivematica

Følgende tema må utredes videre:

- Automatiseringsgrad
- Progresjonsmonitorering og feilsituasjoner
- Feilsituasjoner knyttet til filformat-identifikasjon
- Regelstyrt konvertering, kalt normalisering, fra opprinnelig filformat til henholdsvis langtidslagringsformat og visningsformat

Vi har ennå ikke testet ut ulike automatiseringsgrad og konfigurasjoner knyttet til skalérbarhet. Dette bør sees nærmere på.

Når det gjelder monitorering av fremdrift/progresjon, så kunne en ønske seg mer visuelle tilbakemeldinger på om det virkelig går noe «under panseret» eller om det har stoppet opp. Dette er egentlig en mindre viktig ting, men det er forbedringspotensiale. Et annet og mer viktig tema knyttet til dette er hvordan ulike typer feil håndteres. Her har vi identifisert et forbedringspotensiale, slik vi ser det. Likeledes feilsituasjoner knyttet til filformat-identifisering.

Når det gjelder regelstyrte konverteringer (normaliseringer) så kunne vi ønske oss en funksjonalitet som var mer konfigurert og fleksibel. Videre at enkelte konverteringsprogram som ikke er tilgjengelig i dag kunne brukes. Det er imidlertid ikke sikkert at disse kan brukes grunnet programlisens-type.

6.2 Forbedringspotensiale integrasjon mellom Archivematica og AtoM

Eksport av ferdigprosessert innhold fra Archivematica til AtoM for visning av så vel arkivbeskrivelser som innhold er en av de store motivasjonene for at vi valgte å se på disse teknologiene.

Noen utfordringer knyttet til installasjon ble oppdaget av prosjektet, som slet litt med å få vertøyene samkjørte. Dette ble hos Riksarkivet løst ved hjelp av syv timer konsulentbistand, hos Fylkesarkivet i Oppland og Bergen byarkiv fikk de det til slutt til uten hjelp. Mye av problemene skyldtes fragmentert dokumentasjon, delvis kombinert med at det er relativt komplekse installasjoner og kompetanse. Her er det behov for å utvikle en detaljert veiledning for installasjon av AtoM/Archivematica, da med vekt på koblingen mellom de to. Dette arbeidet kan eksempelvis løses lokalt ved å utvikle en norsk veiledning.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 19 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Ved utprøving er det også identifisert enkelte delutfordringer knyttet til informasjonsutvekslingen fra Archivematica til AtOM, spesielt knyttet til store datasett. Dette vil måtte undersøkes nærmere.

6.3 SiardSuite: Sikring av innhold for langtidsbevaring fra relasjonsdatabaser

SiardSuite ble ikke direkte testet ut i prosjektet selv om det var opprinnelig intensjon. Da flere prosjektpartnere bruker dette verktøyet parallelt med SAMDOK-prosjektet vårt finner vi allikevel plass til en kort omtale om erfaringene vi har gjort oss.

SiardSuite er verktøyet som generer et kildenær representasjon av en relasjonsdatabase, for eksempel databasen som ligger i bunnen av et fagsystem eller et Noark-godkjent sak-arkivsystem. Kildenær fordi den i bunn og grunn speiler opprinnelig databasestruktur over i en XML-representasjon, pakket inn i en ZIP64-fil. Dette formatet har fått navnet SIARD. Kildesystemene SiardSuite håndterer er Oracle, MS Access, MS SQL Server, DB/2 og MySQL, med andre ord majoriteten av databasesystemene som finnes i dag. Speilingene (transformasjonene) fra disse databaseteknologiene er veldokumentert, med automatisert loggføring, noe som styrker tilliten til resultatet.

En SIARD-fil kan enten aksesseres via SiardSuite-brukergrensesnittet, via en standard nettleser eller det kan konverteres tilbake til en relasjonsdatabase dersom arkivinstitusjonen vil muliggjøre søk i database-materialet. Denne konverteringen skjer da til en standardisert versjon av SQL-databasespråket, kalt SQL:2008, noe som muliggjør tilgang på tvers av leverandører og plattformer (Oracle, MS Access, MS SQL Server, DB/2 og MySQL).

Det er verdt å merke seg at generering av databaseinnhold som SIARD-filer ligger forut for Archivematica prosesseringen, og SIARD-filer vil i Arhivematica og AtOM bli behandlet som en hvilken som helst annen fil, slik som docx-filer, PDF-filer osv.

SiardSuite testes ut eller brukes i produksjon i arkivinstitusjoner i over 50 ulike land. Verktøyet er fritt tilgjengelig ved henvendelse til Swiss Federal Archives¹⁷. Det finnes etter hvert også annen programvare både for generering og visning av SIARD-filer, for eksempel Database Preservation Toolkit¹⁸.

6.4 Disk image: Sikring av innhold fra Mac-er og PC-er

Vanlig kopiering fra en fil-mappestruktur til en annen har den ulempen at tids-/datostemplene på mapper endres til kopieringstidspunktet. Dette kan vanskeliggjøre analyse og forståelse av materialet.

En løsning kan eksempelvis være å lage en såkalt disk image. En disk image er en komplett kopi av harddisken, ned til bit-nivå. En disk image bevarer hver bit av data på harddisken på det tidspunktet bildet ble tatt.

Archivematica har en funksjonalitet for videre prosessering fra disk image.

¹⁷ <https://www.bar.admin.ch/bar/en/home/archiving/tools/siard-suite.html>

¹⁸ <http://www.database-preservation.com/>



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 20 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

6.5 Archivemata og AtoM i et framtidsscenario

Sameksistens med andre systemer

Archivemata og AtoM vil kunne sameksistere med eksisterende løsninger dersom arkivinstusjoner og museer skulle ønske det, eller de vil (presumptivt) kunne erstatte deler av eksisterende løsninger, samt fylle behov som pr. i dag ikke har nok programvare-støtte. Det er mange muligheter.

Det finnes også andre løsninger, som er sammenlignbare, men allikevel forskjellige. Et verktøy som er snakket om i Arkivverket heter Preservica. Dette er en totalløsning, hvor alle funksjoner nevnt over blir dekket i verktøyet. Dette til forskjell fra Archivemata og AtoM som kan dekke i ulik grad avhengig av krav til sameksistens med eksisterende programvare.

Behov for videre undersøkelser

På tross av alle disse positive funn er det allikevel nødvendig å avklare en del problemstillinger. Derfor anbefales oppfølgende SAMDOK-aktiviteter i 2017.

Et av punktene knytter seg til verktøyenes evne til å håndtere store samlinger av digitalt innhold. Dette er rett og slett et område vi ikke har rukket å få testet ut skikkelig.

Et annet punkt knytter seg til grensedragning mot andre typer bevarings- og tilgjengeliggjørings programvare. Programvare som arkivinstusjoner kanskje har kjøpt eller som er under utvikling. Grensedragninger mellom ulike bevarings- og tilgjengeliggjøringsverktøy vil måtte gå ned på hvordan eksempelvis METS og PREMIS er implementert i ulike verktøy.

Punktet over bør derimot ikke være en hindring for de arkivinstusjoner som ikke har denne type programvare!

Muligheter for videreutvikling av ønsket funksjonalitet

Videreutvikling av programvaren kan skje på tre måter. Enten(1) ved i dialog med Artefactual Systems å bestille funksjonalitet som kommer med i en av de neste standardutgivelsene, release-ene, av programvaren. Eller at man internt utvikler funksjonaliteten selv. Her er det to mulige veier å gå, enten (2) å ha en tett dialog med Artefactual Systems Inc. for å få funksjonaliteten inn i en framtidig standardutgivelse, eller (3) lage en lokal versjon, fork, av programvaren. Det siste alternativet er ikke å foretrekke da det krever lokal vedlikehold av programvaren.

Muligheter for brukerdiallog, support og skytjenester

Artefactual Systems¹⁹ Inc. som holder til i Vancouver i Canada har utviklet og er ansvarlige for vedlikehold og oppgraderinger av både Archivemata og AtoM, Access to Memory. Begge produktene ble påbegynt rett etter årtusen-skiftet basert på deltagelse og erfaringer opparbeidet i Interpares-prosjektet²⁰.

¹⁹ <https://www.artefactual.com/>

²⁰ <http://www.interpares.org/>



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 21 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Informasjon om produktene, inklusive brukerdokumentasjon og nedlasting av programvare kan finnes på hjemmesidene til henholdsvis Archivematica²¹ og AtoM²². I tillegg finnes det brukerforumsider²³ ²⁴for begge verktøyene hvor aktuelle tema bringes opp, slik som informasjon om bruker-nettverk og workshops, oversikt over videreutviklingsløpet for de to verktøyene, inklusive informasjon om hva slags informasjon som kommer i neste versjon og hvem som har sponset denne funksjonaliteten, med mer.

Vedlikeholds- og supporttjenester tilbys på tradisjonelt vis av Artefactual Systems. Vår erfaring er at det fungerte veldig bra på tross av plassering i Canada, spesielt da noen av konsulentene jobber fra Europa.

ARBARK leide skytjenester fra Artefactual i en firemåneders-periode. Løsningen innebar leie av lagringsplass, både for software og opplastet materiale, på en server lokalisert i Frankfurt. Teknisk støtte fra Artefactual var inkludert i perioden. Arbarks erfaring med skytjenester og support er gode. Artefactual har en god åpenhet. Men det koster penger.

Skytjenester på tvers av landegrensene er foreløpig lite aktuelt for offentlig arkivmateriale. Problemstillinger knyttet til norsk lovverk og krav om plassering av servere i Norge ble diskutert med Artefactual Systemes som ikke ser på dette som noe problem dersom det blir en norsk brukerbase.

7. Konklusjon og anbefalinger

7.1 Oppsummering

Denne SAMDOK-aktiviten, SAMDOK privat oppgave 10 for 2016, har evaluert Archivematica og AtoM (Access to Memory), to av de presumptivt beste verktøykandidatene innenfor digital langtidbevaring og tilgjengeliggjøring som finnes på markedet i dag. De representerer begge «state-of-art» innenfor sitt fagfelt internasjonalt, i konkurranse med ikke alt for mange alternativer.

Gjennom pilot-studier i fire organisasjoner, Fylkesarkivet i Oppland, Arbeiderbevegelsens Arkiv og Bibliotek (ARBARK), Bergen byarkiv og Riksarkivet (Arkivverket) har vi i 2016 skaffet oss praktisk kunnskap om installasjon og bruk av verktøyene. Vi har anvendt disse verktøyene på digitalt skapt privatarkivmateriale, men det er ingen (teknologiske) hindringer knyttet til å verktøy-anvendelse på digital informasjon fra statlig eller kommunal sektor.

Selv om prosjektdeltakerne har skaffet seg en del erfaringer, har vi ennå ikke nådd et samlet kompetansenivå som gjør oss i stand til å utnytte mulighetene som ligger i verktøyene til fulle, spesielt knyttet til optimalisering i forhold til skalérbarhet og automatisering av arbeidsprosessene. Dette medfører også at vi pr. i dag ikke med sikkerhet kan si om identifiserte forbedringspunkter (i forhold til relativt subjektive kriterier hos hver prosjektdeltaker) er løsbare i verktøyene i dag eller om det er endringsbehov utover riktig konfigurering av verktøyene.

²¹ <https://www.archivematica.org/en/>

²² <https://www.accesstomemory.org/en/>

²³ https://wiki.archivematica.org/Main_Page

²⁴ https://wiki.accesstomemory.org/Main_Page



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 22 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

Alt i alt har vi samlet mange positive erfaringer med verktøyenes funksjonalitet, og vi har funnet noe vi bør undersøke nærmere.

7.2 Potensialet i den piloterte programvaren

Vår konklusjon er at Archivemata og AtoM kan løse flere bevarings- og tilgjengeliggjøringsutfordringer, og arbeidsgruppen anbefaler at man går videre langs dette sporet for å få vurdert om disse verktøyene bør tas inn i «verktøykassen» for høsting av arkiver – i første omgang for privatarkiver, men muligens også for materiale fra andre sektorer.

Mange institusjoner tilknyttet SAMDOK har deler av en IT-infrastruktur for digital langtidbevaring og tilgjengeliggjøring på plass, andre har det ikke. Relativt til dette forskjelligartede landskap kan den uttestede programvaren passe inn, gjennom at den fleksibelt kan kobles mot annen programvare. I noen tilfeller vil det kreve en viss programmeringsjobb, men det er løsbart.

Archivemata og AtoM dekker også de ulike fasene i henhold til OAIS-standard på en måte som sikrer autentisitet og langsiktig tillit til den digitale informasjonen (*se tabell i kapittel 3.6*). Det vil gi langt høyere kvalitet enn dagens praksis i arkiv-Norge.

Vi skal ikke påstå at Archivemata og AtoM umiddelbart kan settes i produksjon da det (sannsynligvis) vil være behov for tilpasninger. Spesielt dersom sameksistens med andre verktøy er et krav.

Archivemata og AtoM har imidlertid en moden og åpen programkode som er nedlastbar uten kostnader, fordi programvaren (verktøyene) er utviklet under en fri- og åpen kildekode lisens. Programvaren har vært i drift i omkring ti år. Dette kombinert med at litt i underkant av hundre arkivinstitusjoner internasjonalt bruker verktøyene, og har ervervet seg teknisk kompetanse på dem, gjør at en kan ha tillit til at Archivemata og AtoM vil overleve i flere tiår til.

Forventet levetid på programvare er helt avhengig av faktorer som antall brukere, antall med teknisk kompetanse, god kommunikasjon med brukerne inklusive effektiv support og feil-/endringshåndtering. For Archivemata og AtoM kan supporttjenester kjøpes, slik Riksarkivet gjorde (i begrenset omfang) knyttet til installasjon i SAMDOK-prosjektet, eller slik ARBARK gjorde i forhold til å leie skytjenester (i Tyskland). Dette gjøres via Artefactual Systems, det kanadiske selskapet som eier og vedlikeholder programvaren.

7.3 Anbefalinger om videre arbeid

Arbeidet med høsting av privatarkiv

Teknisk sett skiller ikke håndtering av digitale private arkiv seg fra offentlige. En må være nødt til å håndtere de samme typer materiale både i statlig, kommunal og privat sektor – men selvsagt med variasjoner og ulikheter. Det vil være en fordel med større kompetansemiljøer i forhold til bevaring av elektroniske arkiver. Håndteringen krever sitt i forhold til arkivkompetanse, teknisk kompetanse og teknisk infrastruktur.

Dette peker i retning av tettere samarbeid på tvers av sektorer. Ettersom teknisk og arkivfaglig kompetanse på langtidbevaring er mangelfull innenfor alle sektorer i Norge i dag, vil det være ekstra viktig å samle kompetansen i levedyktige, langsiktige faglige nettverk og fora. Spesielt viktig er det å inkludere



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 23 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

privatarkivfeltet i størst mulig grad i denne helheten. SAMDOK er i så måte en fin arena som det bør bygges videre på.

Hvordan innsamlingsarbeidet skal prioriteres og koordineres bør ta utgangspunkt i regionale bevaringsplaner med og ha en arbeidsdeling der f. eks. en regional institusjon er akslet til å håndtere digitalt privatarkivmateriale. Her vil ulike lokale forhold være en kritisk suksessfaktor, eksempelvis dersom regional institusjon ikke har den rette kompetansen til å gjennomføre digital bevaring. Lokalt, regionalt og nasjonalt samspill vil uansett være viktig for å lykkes fullt ut, slik at kompetansen hentes på tvers av tilhørighet og de beste faglige miljøene kan benyttes. Hovedutfordringen er å få dette arbeidet høyere opp på prioritetslista, samt å få det finansiert. Hvis ikke vil mye verdifull samfunnsdokumentasjon gå tapt i løpet av få år!

Bevaring og tilgjengeliggjøring av digitale privatarkiv, enten det skjer i regi av kommunearkivinstitusjoner, Arkivverket eller ulike muséer eller private arkivinstitusjoner, løper den store risikoen at det blir alt for teknologisk fragmentert fordi de ulike løsninger som velges mangler standardisert informasjonsutveksling. Vi har sett store fordeler med standardisert informasjonsutveksling i AtoM, noe som illustrerer viktigheten av å nettopp dette. Dersom et system for visning av privatarkivmateriale kun kan aksesseres lokalt/isolert vil internetts mulighetsrom med å koble ulike informasjonskilder sammen forsvinne. Vi anbefaler derfor at standardisert informasjonsutveksling prioriteres.

Arbeidet med valg av hensiktsmessige verktøy/programvare

Det er stor risiko og store kostnader forbundet med å egenutvikle programvare, og ikke minst vedlikeholde den. Ofte vektlegges ikke dette sterkt nok når nye utviklingsprosjekt igangsettes. Skal en lykkes, så må programvaren få mange brukere; skal man få mange brukere, så må programvaren være av god kvalitet og brukerkommunikasjonen være tillitsvekkende. Forvaltning av programvare koster, Det koster mye mer enn de rene utviklingskostnadene. Kostnadene knytter seg til kunde-/brukeroppfølgning, samt videreutvikling som en konsekvens av brukerdialogen. Brukerforum og wikier må lages for å motta ris og ros, feilmeldinger og endringsønsker. I tillegg må slike henvendelser følges opp på daglig og ukentlig basis og plasseres inn i et periodisk regime av endringskontroll og nye releaser/programversjoner.

Programvare som støtter digital langtidbevaring og tilgjengeliggjøring vil være kompleks i struktur, men bør være brukervennlig. Den bør også støtte internasjonalt standardisert informasjonsutveksling. Bruken av informasjonsutveksling via internasjonale standarder er viktig. Hvis ikke mister arkiv-Norge muligheten til å anskaffe seg noe av den etter hvert gode og relativt modne programvaren som har blitt utviklet innenfor dette fagfeltet internasjonalt de siste 10-15 årene.

Vi er nå midt inne i diskusjoner omkring litt tilårskomne nasjonale IT-plattformer og programvare. Dette er programvare som ble utviklet for kanskje ti år siden eller mer, og som nå behøver oppgradering. Det snakkes om forberdringer og/eller nyutvikling. Det snakkes om Arkivportalen, om Digitalarkivet og det snakkes om Asta. Det finnes i dag innenfor fagfeltet internasjonalt utviklet programvare som er lagd under fri programvare/åpen kildekode-lisenser. Dette representerer en stabil kodebase som har vært i produksjon i flere år, eksempelvis Archivematica/AtoM. Denne kodebasen kan en (Arkivverket) velge å bygge videre på. Vårt råd er at en skynder seg langsomt i jakten på nye, gode løsninger. Kanskje AtoM kan bli den nye Arkivportalen? Da har man en solid kodebase å bygge videre på.



SAMDOK delprosjekt: Privat	Prioritert oppgave: 2016/10	SAK (ePhorte):	Dato: 10-03-2017	Side: 24 av 24
Forfatter: Arne-Kristian Groven	Tittel: Rapport 2016 – Metodikk for høsting og langtidslagring av privatarkiv			

I SAMDOK vil det i 2017 bli gjennomført en oppgave som er en direkte videreføring av verktøyperspektivet i dette prosjektet: PRIV 2017_1: *Benchmark AtoM vs Arkivportalen og Asta*, hvor målsettingen er å sammenligne AtoM (pluss Archivemata front-end) med henholdsvis Arkivportalen, Digitalarkivet og Asta, for arkivbeskrivelse og visning/gjenfinning av digitalt arkivmateriale. Identifiserte optimaliserings-/forbedringspunkter vil også bli fulgt videre opp.

Sammenligningen vil foregå fra to perspektiv:

- Et **kompletteringsperspektiv**, hvor vi ser på sameksistens mellom de nevnte teknologier og kravsetter utvekslingsformater.
- Et **erstatningsperspektiv**, hvor vi ser om AtoMs funksjonalitet kan erstatte en eller flere av de allerede eksisterende.

7.4 Risiko forbundet med å vente

Med alt som nå er nevnt kan det være en dårlig strategi å vente før en gjør noe med langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale i egen institusjon, arkiv eller museum.

Den optimale, altomfattende, garantert enkle løsningen på et utfordrende problem, «the silver bullet», finnes ikke når det gjelder langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale/informasjon.

Men ser vi på det beste som finnes internasjonalt av «state of art» verktøy for langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale, så finnes det i dag programvare-produkter på markedet som har vært i drift i snart ti år. Disse har nådd en modenhet og en kvalitet som det er verdt å se nærmere på. Dette er motivet for at vi her har sett nærmere på Archivemata, AtoM (og Siard/Siard Suite).

Håpet er at arkivinstitusjoner eller museer med ambisjoner om å forvalte digitalt materiale, finner rapporten nyttig og starter å bygge egen kompetanse rundt langtidsbevaring og tilgjengeliggjøring av digitalt arkivmateriale i egen institusjon.