

Prosjektrapport

**Farvel DBS. Utredning av metoder for datahøsting fra private
databasesystemer**

Arbeiderbevegelsens arkiv og bibliotek

2019

Innhold

Innhold	1
Begreper	4
1.0. Innledning.....	7
1.1. Prosjektets formål og måloppnåelse.....	7
1.2. Organisering	8
1.3. Bakgrunn for prosjektet.....	8
1.4. Elektroniske arkiver og arkivuttrekk	10
2.0. Kartlegging - innhenting av informasjon fra LOs fagforbund	12
3.0. Serveroppgradering og etablering av digitalt sikringsmagasin	15
4.0. Informasjonsinnhenting fra andre arkivinstitusjoner og aktører	16
5.0 Om håndtering av el-arkiver i digitalt depot	18
5.1 OAIS-modellen	18
5.2 Konvertering og anbefalte arkivformater	18
5.3 Arkivmaterialets integritet og autentisitet	19
5.4 Potensielle mottakssystemer	20
5.5 Mer om kunnskapsbehov i digitalt depot	21
6.0. Aktuelle metoder	23
6.1. Noark 5-uttrekk	23
6.2. SIARD/Decom	24
6.3. Leverandøruttrekk	27
7.0. Arkade 5	28
8.0 Asta Mapper	31
9.0 Testuttrekk basert på aktuelle metoder.....	32
9.1. Acos Avlevering.....	32
9.2. SIARD 2.1.60.....	33
9.3. Full Convert Spectral Core.....	33
10.0 To caser: systemer med og uten Noark-godkjenning.....	34
10.1. Fellesforbundet.....	34
10.2. Norsk arbeidsmandsforbund	36
10.2.1. Leverandøruttrekk	36
10.2.1.1. Eksporten fra systemet	37
10.2.1.2. Sikring av SIP.....	39

10.2.1.3. Produksjon av AIP: begrensninger i antall tegn.....	41
10.2.1.4. Produksjon av AIP: konvertering av dokumenter	42
10.2.1.5. Produksjon av AIP: aktør- og arkivbeskrivelse.....	44
10.2.1.6. Produksjon av AIP: ferdigstilling.....	47
10.2.1.7. Søkemuligheter og gjenfinning	48
10.2.1.8. Vurdering av testuttrekk og tanker om digitalt depot	48
10.2.2. Forsøk på SIARD-uttrekk	50
11.0 Vurdering av metoder.....	53
11.1. Noark 5-uttrekk	53
11.2. Uttrekk fra systemer uten Noark-godkjenning.....	54
11.3. SIARD-uttrekk	54
12.0 Generelt om krav	56
13.0 Datasikkerhet.....	57
13.1 Digitalt materiale vs. fysisk materiale.....	57
13.2 Mulige situasjoner som kan forårsake skade på eller tap av digitalt materiale.....	58
13.3 Løsninger og konkrete tiltak hos Arbark for å oppnå ønsket grad av sikring og beskyttelse	58
13.3.1 Oppsett av brannmur	59
13.3.2 Antivirusløsning	59
13.3.3 Tilgangsstyring, både for vanlige brukere på administrasjonsnettet, og for administratorkontoer	60
13.3.4 Sikring av data ved innsamling av arkiv	60
13.3.5 Lagring av flere duplikater og generasjoner av materialet, og sikkerhetskopiering til flere lokasjoner og i flere generasjoner.....	61
13.3.6 Inndeling av nettverket i separate soner med tilgangsstyring	63
13.3.7 Systemovervåkning	65
13.3.8 Opplæring av ansatte i grunnleggende datasikkerhet, og bevisstgjøring om potensielle trusler	65
13.3.9 Gode og gjennomarbeidede rutiner for bruk av arkivmateriale, tildeling og fjerning av tilganger og avvikshåndtering.....	66
13.3.10 Risikovurderinger og kriseplaner	66
13.3.11 Revisjonsplan	67
14.0 Datasikkerhet og arkivinnhenting	68
14.1 Alternativ 1: normal løsning	68
14.2 Alternativ 2: utvidet løsning.....	68
14.3 Alternativ 3: ekstra sikker løsning	69

15.0 Veiledninger	70
15.1. Bevaring og kassasjon	70
15.2. Personvern	71
16.0 Seminarer i regi av Arbark og deling av kompetanse	73
17.0 Tilgjengeliggjøring	74
18.0 Avslutning	77
18.1 Utrekksmetodikk	77
18.2 Digitalt magasin og sikkerhet	77
18.3 Tilgjengeliggjøring	78
Programvare som det refereres til i rapporten	79
Litteratur	80
Lover og forskrifter	80
Standarder	81
Andre kilder	81

Begreper

Acos Websak: Saksbehandlingssystem som brukes av Arbark. Leveres av Acos.

ADDML (Archival Data Description Markup Language): Arkivverkets standard for teknisk beskrivelse av datasett. Har form som en fil som dokumenterer innholdselementer og deres relasjoner.

AIP (Archival Information Package): Brukes i OAIS-modellen om digitale arkivpakker (inneholdende filer og nødvendig metadata) som skal langtidsbevares. Utgangspunktet for en AIP er SIP.

Arbark: Arbeiderbevegelsens arkiv og bibliotek.

Compendia Sak: Saksbehandlingssystem som brukes av Norsk arbeidsmandsforbund. Driftet av leverandøren Compendia.

CSV (Comma-separated values): Filformat med rene tekstfiler som skiller informasjon ved hjelp av kommategn.

DIAS (Digital arkivpakkestruktur): Norsk modell som angir hvordan foreliggende standarder skal anvendes i en arkivpakke.

DIP (Dissemination Information Package): Brukes i OAIS om visningspakken (basert på AIP) som gjøres tilgjengelig for brukere.

EAC-CPF (Encoded Archival Context – Corporate bodies, Persons and Families): Standard for å beskrive arkivskapere og andre aktører i OAIS-baserte arkivpakker.

EAD (Encoded Archival Description). Standard for logisk beskrivelse av arkivmateriale i OAIS-baserte arkivpakker.

Evry: Leverandør av Noark-systemene ESA og Elements.

FTP (File Transfer Protocol): Verktøy som kan brukes til å overføre filer mellom et nettverk og en klient/PC.

GDPR (General Data Protection Regulation): EUs personvernforordning, innført i Norge i juli 2018.

IBM Notes: Gruppevareløsning med dokumentdatabase i bunn, tidligere Lotus.

IKA (Interkommunale arkiv): Selvstendige interkommunale selskaper som blant annet yter depottjenester for sine eierkommuner.

ISAD(G) (General International Standard Archival Description): Internasjonal standard for arkivbeskrivelse.

ISAAR(CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons and Families): Internasjonal standard for beskrivelse av arkivskapere og aktører.

KDRS (Kommunearkivinstusjonenes digitale ressurscenter): Nasjonalt el-arkiv-samarbeid bestående hovedsakelig av kommunearkivinstusjoner.

LO (Landsorganisasjonen i Norge): Norges største hovedsammenslutning av arbeidstakere, bestående av 25 fagforbund.

MODARK (Modernisering av arkivoverføring): Prosjekt i Arkivverket som utvikler tiltak for overføring av digitale arkiver fra offentlig forvaltning til arkivdepot.

NOARK (Norsk arkivstandard): Standard for journalføring og arkivering, utviklet av Riksarkivaren.

OAIS (Reference Model for an Open Archival Information System): Modell med prinsipper for bevaring av digitale objekter.

ODBC (Open Database Connectivity): Standard for tilkobling til database.

PEANUTS (Private elektroniske arkiver i Norge – uttrekk, testing, sikring): Prosjekt gjennomført av Arbark i 2015-2016, med støtte fra Arkivverkets arkivutviklingsmidler.

SAMDOK: Riksarkivarens program for arbeid med samlet samfunnsdokumentasjon. Prosjektet ble påbegynt i 2013 og avsluttet i 2017.

SHA (Secure Hash Algorithm): Betegnelse på en samling sjekksumfunksjoner.

SIARD (Software Independent Archiving of Relational Databases): Åpent format for bevaring av databaser.

SIP (Submission Information Package): Brukes i OAIS om den digitale aksesjonen depot mottar fra arkivskaper. SIP skal bevares uendret.

TDR (Trusted Digital Repository): ISO-standard (ISO 16363) som angir krav til et digitalt arkivdepot. En videreføring av TRAC.

Tieto: Leverandør av Noark-systemet Public 360.

TRAC (Trustworthy Repositories Audit & Certification): Internasjonal standard som angir krav til et digitalt arkivdepot.

XML (Extensible Markup Language): Markeringspråk og filformat som kan organisere data i en hierarkisk struktur.

1.0. Innledning

1.1. Prosjektets formål og måloppnåelse

I 2016 søkte Arbeiderbevegelsens arkiv og bibliotek (Arbark) om støtte fra Riksarkivarens utviklingsmidler for arkivsektoren til å gjennomføre det treårige prosjektet «Farvel DBS. Utredning av metoder for datahøsting fra private databasesystemer». Prosjektet fikk innvilget støtte høsten 2016, og etter egen søknad i 2017 og 2018 ble det til sammen bevilget kr 1 425 000 til prosjektet.

Prosjektet startet i september 2016, og ble slutført i august 2019. Det hadde tre separate prosjektfaser: fase I (2016-2017), fase II (2017-2018) og fase III (2018-2019).

I den opprinnelige søknaden fra 2016 om utviklingsmidler ble målet for prosjektet spesifisert som: å utrede forutsetningene for at depotinstitusjoner skal kunne høste, langtidslagre og gjøre tilgjengelig informasjon fra databasesystemer (DBS) skapt av private arkivskapere. Den generelle tilnærmingen gjorde at prosjektet ville «ha betydelig overføringsverdi for andre arkivinstitusjoner som arbeider med privatarkiv [...] og også museer og biblioteker med depotfunksjon for privatarkiver». Prosjektet skulle munne ut i «en grundig rapport som vil fungere som veileder for private arkivskapere på den ene siden, og for private og offentlige depotinstitusjoner for privatarkiver på den andre siden, med hensyn til kompetansebygging, etablering av rutiner og vurdering av investeringer som må til for å sikre den framtidige høstingen, langtidslagringen og formidlingen av informasjon fra databasesystemer». Slik skulle prosjektet «være et viktig bidrag til kompetansebyggingen hos private arkivskapere og i depotinstitusjoner for privatarkiver».

Med den foreliggende prosjektrapporten og de veiledere og andre dokumenter som prosjektet har produsert, og som er vedlagt prosjektrapporten, kan det konkluderes med at prosjektets overordnede målsetting er oppfylt. Den foreliggende prosjektrapporten med vedlegg utgjør nettopp et viktig bidrag til kompetansebygging innenfor det norske arkivfeltet knyttet til håndtering av databasebaserte private arkiver.

Prosjektsøknaden for prosjektfase I (2016-2017) satte opp følgende generelle delmål: 1) Kvalitetssikring av kartleggingen [av LO-forbundenes arkivsystemer]; 2) Testing av funksjonsanalyse som metode; 3) Personsensitivt materiale; 4) Kompetanse i arkivdanningen og depotinstitusjonen; 5) Tilgjengeliggjøring av arkiver; 6) Digitalt depot; 7) Resultat [prosjektrapportering].

Med unntak av delmål 5) Tilgjengeliggjøring av arkiver, er alle delmålene oppnådd i prosjektperioden. Tilgjengeliggjøring av databasebaserte digitalarkiver har vist seg å være et komplisert og lite utforsket område i en slik grad at prosjektgruppen ikke hadde kapasitet til å utrede dette spørsmålet i noen særlig grad innenfor prosjektets rammer.

Prosjektsøknaden for prosjektfase II (2017-2018) presiserte et utvidet mål bilde, der 1) gjennomføring av uttrekk fra databasesystemer samt 2) utvikling av en teknisk infrastruktur som ville gjøre Arbark i stand til å bevare digitalt arkivmateriale på mellomlang sikt, ble integrert som målsettinger i prosjektet. Begge disse målsettingene er nådd i løpet av prosjektperioden.

I tillegg inkluderte prosjektsøknaden for prosjektfase II nye delmål: 1) Kompetansebygging i arkivdanningen: Utarbeiding av minstestandard og veileder; 2) Kompetansebygging i Arbark: Testuttrekk og funksjonsanalyse, herunder uttrekk fra systemer både med og uten Noark-godkjenning; 3) Etablering av midlertidig digitalt depot; 4) Samarbeid og deling av kompetanse. Når det gjelder delmål 1) ønsket arkivskaperne at Arbark laget en bevaring og kassasjonsveiledning (se vedlegg). Noen minstestandard var ikke etterspurt. Som det ellers vil framgå av prosjektrapporten er de andre delmålene oppnådd i løpet av prosjektperioden.

1.2. Organisering

Prosjektansvarlig har vært daglig leder ved Arbark, Frank Meyer (2016-2018), Ole Martin Rønning (2018-2019). Arbeidet har vært ledet av en styringsgruppe bestående av Ole Martin Rønning, Tor Are Johansen (førstearkivar og koordinator for arkiv, Arbark) og Max Tønnesen (IT-rådgiver, Arbark). Arkivar Dag Juvkam ved Arbark har vært prosjektleder. Prosjektets referansegruppe har, i tillegg til medarbeidere fra Arbark, bestått av arkivpersonale fra LO, Fagforbundet, Fellesforbundet, Forbundet for ledelse og teknikk og Norsk arbeidsmandsforbund.

1.3. Bakgrunn for prosjektet

I offentlig sektor har man lov- og regelverk som skal sikre at arkivverdig materiale blir bevart. Slikt regelverk finnes ikke i privat sektor, noe som har bidratt til en skjevhet i samfunnsdokumentasjonen. Arbark er blant annet depotinstitusjon for LO og de 25 fagforbundene som er tilsluttet LO. LO er Norges største hovedsammenslutning for

arbeidstakere og dermed en viktig aktør i det norske samfunnet. Bevaring av LO-organisasjonenes arkiver er viktig sett ut i fra myndighetenes uttalte mål om å sikre en mest mulig helhetlig samfunnsdokumentasjon. Men bevaring er også viktig for å sikre LO-medlemmenes rettigheter. Arbark opplever at både forbundene selv og LOs juridiske avdeling ofte etterspør gamle saker. Det er særlig snakk om yrkesskadesaker, oppsigelser eller dokumentasjon fra tarifforhandlinger, og dokumentene kan være fra mange tiår tilbake.

I dag bruker flere av LOs organisasjoner elektroniske saksbehandlingssystemer. Fram til nå har likevel de fleste også arkivert på papir. Imidlertid har Arbark de siste årene fått signaler om at flere av disse organisasjonene vurderer en overgang til fullelektronisk arkivering. Arbark må derfor bygge opp kompetanse på mottak av denne typen arkivmateriale. Det finnes ikke styrende nasjonale rutiner for mottak av elektroniske privatarkiver. Dermed er det ingen enkel oppskrift vi kan adoptere. Ved prosjektstart i 2016 hadde Arbark ikke noen digital depotløsning og hadde ennå ikke mottatt en eneste avlevering fra et elektronisk saksbehandlingssystem. Det var lite kunnskap om det som foregikk på el-arkiv-området rundt om i depot-Norge, eller hvilke verktøy som ble brukt av depotinstitusjoner for å håndtere digital arkivinformasjon. Kompetanse om uttrekksprosesser var også lav ute hos arkivskaperne.

Til sammen utgjorde dette en kritisk situasjon ettersom viktig informasjon kunne gå tapt.

Behovet for kompetansebygging og mangelen på etablerte rutiner utgjorde bakgrunnen for prosjektet. Naturlig nok ble LO-forbundene valgt som case-studie i prosjektet, gitt Arbarks ansvar for bevaring av arkiv fra denne samfunnssektoren. Samtidig er arkivdanningen i de forskjellige forbundene interessant på grunn av den store variasjonen: Noen forbund er store i kraft av høye medlemstall, mens andre er mindre og har færre ressurser til rådighet. Størrelsen på organisasjonene har åpenbart noe å si for antall ansatte som jobber med arkiv. Men organisasjonens ressurser får også konsekvenser for hvilke sak-/arkivsystemer som brukes (Noark vs. ikke-Noark). Det er rimelig å anta at slike variasjoner speiler den generelle situasjonen i privat sektor, noe som øker prosjektets overføringsverdi for andre institusjoner som jobber med lignende problemstillinger.

Selv om Arbark før dette prosjektet ikke hadde erfaring med databaser, har vi mottatt fil- og mappearkiver. Dette er arkiver som typisk benyttes av organisasjoner i privat sektor, hvor arkivinformasjon lagres i mappestrukturer i Windows. Arbark fikk i 2015 støtte fra Riksarkivarens arkivutviklingsmidler til å gjennomføre prosjektet «Private elektroniske

arkiver i Norge – uttrekk, testing, sikring» (PEANUTS). Her tok vi imot mappearkivet til organisasjonen Attac, og testet ut systemene *Archivematica* og *Access to Memory* for å se hvor egnet disse var til å håndtere slikt materiale. Vi kunne dermed ta med oss kompetanse fra PEANUTS-prosjektet inn i arbeidet med de teknologisk sett mer krevende database-arkivene.

Ved prosjektstart hadde Arbark som nevnt ikke en tilfredsstillende lagringsløsning for digitalt materiale. Utviklingen og oppbyggingen av et digitalt depot var integrert i prosjektet, inkludert vurderinger angående datasikkerhet. Prosjektrapporten beskriver nærmere dette arbeidet.

Også tilgjengeliggjøring av database-basert arkivmateriale har vært en del av prosjektet. Det finnes foreløpig ikke nasjonale løsninger på dette området. Vi vil i rapporten redegjøre for status når det gjelder Arbarks muligheter til å tilgjengeliggjøre digitalt skapt arkivmateriale.

1.4. Elektroniske arkiver og arkivuttrekk

Prosjektet har vært rettet mot elektroniske saksbehandlingssystemer med en underliggende database. Saksbehandlingssystemet, altså det logiske laget hvor saksbehandlerne jobber, er bygget over en database som i realiteten er organisasjonens arkiv. Derfor kaller vi i denne rapporten saksbehandlingssystemene for «sak-/arkivsystemer».

Når man snakker om databaser menes ofte klassiske relasjonsdatabaser. En relasjonsdatabase består av separate tabeller med informasjon, hvor data er koblet sammen ved hjelp av primær- og fremmednøkler. En primærnøkkel identifiserer en unik rad i en tabell bestående av flere rader, mens en fremmednøkkel er et felt i en tabell som peker til et felt – vanligvis primærnøkkelen – i en annen tabell.

Relasjonsdatabaser omtales ofte som *strukturerte data*. Databaser kan også ha en svært enkel oppbygning. I dokumentorienterte baser er dokument og metadata lagret sammen i basen enkeltvis og ikke spredt over flere tabeller. Et eksempel på denne typen *semi-strukturerte data* er de Notes-baserte systemene vi kommer nærmere tilbake til i rapporten.

Databaser kan altså spenne mellom en nærmest ustrukturert samling av data, til svært komplekse relasjonsdatabaser hvor innhold og metadata er spredt over mange hundre tabeller.

De norske arkivinstitusjonene som har erfaring med elektroniske arkiver (Arkivverket, de kommunale arkivinstitusjonene og de store byarkivene) har valgt en strategi for å bevare

digitalt skapt arkivmateriale som baserer seg på *arkivuttrekk*. Dette innebærer å gjøre arkivverdig materiale uavhengig av produksjonssystem, altså å frigjøre dokumenter og metadata fra proprietære, låste systemer og lagre dem i langtidsformater. Et arkivuttrekk kan sies å være en konvertering av basen til et avleveringsformat (for eksempel et Noark 5-uttrekk).

Denne strategien er lite kjent utenfor arkivkretser. Noen arkivskapere vurderer å arkivere i selve sak-/arkivsystemet, uten å lage avleveringsuttrekk eller skrive dokumentene ut på papir. Dette er uheldig. Saksbehandlingssystemer er utviklet for å legge til rette for effektiv saksbehandling, ikke for å langtidsbevare informasjon. Med fullelektronisk arkivering er det en risiko for informasjonstap på grunn av teknologiforgjengelighet; med teknologisk utvikling fornyes og skiftes både systemer, programvare og fysisk hardware ut, og for at informasjonen skal fortsette å være lesbar over tid, må man foreta stadige konverteringer til nye formater og plattformer.¹ På lang sikt er det heller ikke gunstig å være avhengig av en kommersiell aktør for å bevare informasjon. For å sikre at digitalt skapt informasjon blir sikret og bevart for ettertiden, må informasjonen dermed hentes ut og lagres i systemuavhengige formater.

Vi understreker at det er dataene i sak-/arkivsystemene som er interessante for Arbark, og at det er disse prosjektet har vært rettet mot. Hos LOs organisasjoner er de spesialiserte fagsystemene som inneholder potensielt arkivverdig informasjon i hovedsak medlemssystemer. Medlemsinformasjonen ønsker de færreste av organisasjonene å avlevere til Arbark, og Arbark har fram til nå heller ikke vært interessert i å ta imot denne. Basert på kontakt med arkivskaperne, vurderer Arbark det slik at arkivinformasjonen som har verdi for ettertiden befinner seg i sak-/arkivløsningene. Dette understøttes av at medlemssystemspesifikk informasjon i liten grad etterspørres av brukere.

¹ Et illustrerende eksempel på hyppige teknologiske endringer: vi dro ut til en arkivskaper for å hente et fil- og mappearkiv som lå på en gammel PC. Planen var å kopiere arkivet over til en disk vi hadde med, men operativsystemet på PCen, Windows XP, klarte ikke å lese vår nye disk.

2.0. Kartlegging - innhenting av informasjon fra LOs fagforbund

Prosjektet har brukt LO og LOs fagforbund som case-studie. Høsten 2016 innhentet Arbark systematisk informasjon fra LO og fagforbundene om deres elektroniske sak-/arkivsystemer. Informasjonsinnhenting foregikk gjennom møter med personale i organisasjonene. Kartleggingen var en oppfølging og oppdatering av en tidligere informasjonsinnhenting fra 2013.

Kartleggingen viste at flere ikke bruker sak-/arkivsystemer, men i stedet lagrer materiale i fil- og mappestrukturer i Windows. Dette gjelder mindre forbund som ikke har behov for, eventuelt ikke ser seg råd til, et avansert elektronisk system.

Foruten LO selv, bruker følgende forbund sak-/arkivsystemer:

- Creo
- El & IT
- Fagforbundet
- Fellesforbundet
- Fellesorganisasjonen
- Forbundet for ledelse og teknikk
- Handel og kontor
- Industri Energi
- Norsk arbeidsmandsforbund
- Norsk nærings- og nytelsesmiddelarbeiderforbund
- Norsk offisersforbund
- Norsk sjømannsforbund
- Norsk tjenestemannslag

Systemene i LOs organisasjoner er ganske varierte. Kartleggingen viser at seks av organisasjonene bruker Noark-godkjente systemer: Acos Websak, Public 360 og Elements. Det er i hovedsak de store organisasjonene som bruker disse systemene. Noark, Norsk arkivstandard, er utviklet av Arkivverket og spiller en sentral rolle for digital langtidsbevaring i offentlig sektor. Standarden stiller strenge krav til metadata i arkivdanningen, den har krav

om uttrekksfunksjonalitet, og den angir et uttrekksformat. Dette skal gi en garanti for at arkivinformasjonen kan hentes ut av et system og leveres til depot i et bestemt format.

I privat sektor er man ikke pålagt å bruke Noark-godkjente systemer. En organisasjon kan fritt velge etter hva som passer best for økonomi og organisasjonens størrelse. Åtte av LO-forbundene bruker systemer uten Noark-godkjenning, noe som i utgangspunktet gjør dem mer krevende å håndtere, særlig med tanke på at det ikke finnes et standardisert avleveringsformat. Enkelte av disse ikke-Noark-systemene er relasjonsdatabaser (noe som innebærer at de skal være mulig å «redde», for eksempel ved hjelp av SIARD, se nedenfor), mens andre har dokumentorientert database-teknologi i bunn.

Under den forrige kartleggingen i 2013 hadde flere forbund relativt nylig tatt i bruk sak-/arkivsystemer, og var fortsatt i en innkjøringsfase. Flere av dem hadde høsten 2016 gjort seg opp en mening om systemene. Informasjonsinnhentingene tegner et bilde som kanskje er symptomatisk for privat sektor: Selv om de fleste forbundene bruker de samme systemene og hevder de kommer til å fortsette med disse, har flere i løpet av den 3-årige prosjektperioden (2016-2019) skiftet system, og ytterligere flere vurderer å bytte. Noen av de litt mindre forbundene ga uttrykk for at systemet de bruker er for komplisert, og at de derfor ønsker enklere løsninger. Hvis funksjonene og grensesnittet i systemet er lite brukervennlig kan det bidra til dårlige arkivrutiner. Risikoen for dårlige arkivrutiner øker ytterligere i organisasjoner som ikke har egen arkivtjeneste. Men også noen av de større forbundene med egen arkivtjeneste har skiftet system. Fellestrekket ved misnøyen er at systemene ikke er godt nok tilpasset arkivskapers behov. En årsak til byttene er altså misnøye, en annen er at systemleverandør slutter å drifte systemet eller ikke etterkommer arkivskapers ønske om videreutvikling og ny funksjonalitet.

Problemstillinger knyttet til rutiner i arkivdanningen ble diskutert underveis i prosjektet, både internt på Arbark og i prosjektets referansegruppe. Konklusjonen ble at de omfattende variasjonene innad i LO-systemet, både når det gjelder systemer, organisasjonsstørrelse og arkivkompetanse, setter grenser for hva Arbark kan gjøre på et generelt nivå for å bidra til bedre elektronisk arkivdanning. Arbark kan gi råd om bevaring og kassasjon og skape en grunnleggende bevissthet hos arkivskaperne om verdien av gode arkivdanningsrutiner. Men selve den praktiske opplæringen må utføres av systemleverandørene, etter initiativ fra arkivskaperne selv.

Under kartleggingen antydte flere forbund at det kan bli aktuelt med en elektronisk avlevering i løpet av få år. Samtidig kom det fram at noen av de litt mindre forbundene virket skeptiske til i det hele tatt å avlevere til depot. Årsaken var at de ikke helt ser poenget med å avlevere, så lenge de har et system som de lettvisst kan hente informasjon fra. Papirarkiver er én ting, de tar mye plass og må på et tidspunkt gjøres noe med. Elektroniske arkiver derimot tar ikke fysisk plass. Samtidig er arkivuttrekk komplisert og forutsetter høy grad av teknisk kompetanse. Arkivskaper må som regel bruke systemleverandør for å produsere vellykkede uttrekk. Dette arbeidet koster penger. Sett fra de mindre forbundenes ståsted er skepsisen til avlevering forståelig; private arkivskapere er jo ikke pålagt å avlevere arkiv. Å argumentere blant arkivskaperne for sikker langtidsbevaring av digitale arkiver blir derfor en pedagogisk oppgave for Arbark.

3.0. Serveroppgradering og etablering av digitalt sikringsmagasin

Arbarks tekniske infrastruktur var ikke på noen måte tilfredsstillende ved prosjektstart. Særlig var det et behov for å oppdatere gammel maskinvare og utdaterte programvareversjoner. Som en del av prosjektet ble det i 2017 foretatt omfattende investeringer i Arbarks IT-infrastruktur for å bygge opp et digitalt depot. I påvente av en felles nasjonal lagringsløsning er det et mål å sikre at Arbark har tilstrekkelig kapasitet, maskin- og programvare til å kunne håndtere mottak og bevaring av elektroniske arkiver på mellomlang sikt. En del av oppgraderingen bestod i å opprette en dedikert Windows-basert lagringsløsning på Arbarks servernettverk. Det ble lagt vekt på å konfigurere systemet med en god lagringskapasitet som lett skal kunne utvides i takt med økende behov. I motsetning til de store aktørene i arkivsektoren har ikke Arbark et digitalt depotforvaltningssystem. De elektroniske arkivene vi mottar vil foreløpig lagres i det digitale magasinet i vanlige mappestrukturer. Magasinet ligger på internt nett bak en brannmur. Se mer om dette under punkt 13, Datasikkerhet.

Konfigurasjonen av oppgraderingen er vedlagt rapporten.

4.0. Informasjonsinnhenting fra andre arkivinstusjoner og aktører

Ettersom Arbark ved prosjektstart ennå ikke hadde mottatt arkivuttrekk, har lærdom fra andre depotinstusjoner vært helt nødvendig. I norsk arkivsektor er det, bortsett fra noen enkeltinstusjoner, Arkivverket og Kommunearkivinstusjonenes digitale ressurscenter (KDRS) som har kommet lengst i arbeidet med elektroniske arkiver. KDRS er et nasjonalt el-arkiv-samarbeid bestående hovedsakelig av kommunale arkivinstusjoner. KDRS jobber med å utvikle og effektivisere langtidslagringen av digitale arkiver.²

I begynnelsen av prosjektet besøkte vi flere interkommunale arkivinstusjoner (IKAer) for å lære mer om deres metoder. Gjennom prosjektperioden har vi også deltatt på samlinger i regi av KDRS og hatt møter med Arkivverkets MODARK-prosjekt.³ Vi følger også med på Arkivverkets forvaltningsforum, som hadde oppstartsmøte i november 2018.

Forvaltningsforumet er en arena for kommunikasjon og samhandling i sektoren, og har som funksjon å sikre sammenheng mellom standarder og verktøy over tid.

Denne informasjonsinnhenting og samarbeidet med andre aktører har vært avgjørende for Arbarks kompetansebygging i løpet av prosjektperioden.

Bakgrunnen for besøkene og kontakten med andre aktører var altså behovet for å lære om hvilke metoder som brukes for å langtidsbevare elektroniske arkiver. Konkret gikk dette ut på hvordan man henter materiale ut av systemene og håndterer informasjonen i etterkant. Og ikke minst, hvordan gjøres informasjonen tilgjengelig for sluttbruker?

Metodikken for langtidslagring er kort sagt at filer og metadata hentes ut av fra de proprietære produksjonssystemene som er i bruk ute hos arkivskaperne, og at filene konverteres fra forgjengelige formater og over til arkivformat.

Det er mange utfordringer knyttet til håndtering av elektroniske arkiver, også hos de depotinstusjonene som har kommet lengst på området. Dette gjelder særlig tilgjengeliggjøring: Fram til nå ikke har det ikke eksistert noen nasjonal strategi for tilgjengeliggjøring og innsyn i database-baserte arkiver i Norge.⁴ Vi merket oss at både

² kdrs.no

³ Modernisering av arkivoverføring. MODARK er en videreføring av MAVOD-prosjektet og jobber med tiltak for overføring av digitale arkiver fra offentlig forvaltning til arkivdepot.

⁴ KDRS har imidlertid fått utviklet en innsynsløsning for Noark 5-baserte arkiver (KDRS Innsyn). Det finnes sannsynligvis hjemmesnekrede løsninger rundt om, men altså foreløpig ingen nasjonale løsninger. På sikt kan kanskje Digitalarkivet bli en portal for alle slags digitale arkiver.

Riksarkivet og IKAene har fokusert på å «redde» arkivene, hovedsakelig på grunn av manglende ressurser og det at feltet – på tross av utvikling – fortsatt er umodent. De siste årene har det blitt utviklet flere nye verktøy som skal effektivisere og forenkle arbeidet med elektroniske arkiver, og Arbark ønsker så langt det er mulig og hensiktsmessig å følge de offentlige depotenes metoder. Vi kommer tilbake til noen av disse verktøyene.

Å følge med på standardisering og utvikling er viktig for Arbark. Men det er noen forskjeller mellom Arbarks arkivskapere i privat sektor og depotene i kommunal sektor. I kommunal sektor er mange av utfordringene knyttet til fagsystemer, altså systemer som håndterer helt spesifikke oppgaver eller funksjoner. Slike fagsystemer kan etter hvert gå ut av bruk, og derfor kan det haste med å få høstet arkivinformasjonen. For Arbarks del vil det være snakk om uttrekk fra aktive saksbehandlingssystemer. Det betyr at systemet driftes av en leverandør som vil kunne bistå ved elektronisk avlevering.

5.0 Om håndtering av el-arkiver i digitalt depot

5.1 OAIS-modellen

Reference Model for an Open Archival Information System (OAIS) er en overordnet internasjonal standard med prinsipper for håndtering av digitale objekter. OAIS-modellen er selve bibelen når det kommer til langtidsbevaring av digitalt arkivmateriale. Den er først og fremst en konseptuell modell, og den sier lite konkret om hvordan et arkivdepot kan implementere modellen i praksis. I Norge er OAIS konkretisert gjennom DIAS-standarden,⁵ som ligger til grunn for Arkivverkets og KDRS' digitale depotforvaltningssystem, ESSArch.

I OAIS inngår en informasjonsmodell hvor tre typer arkivpakker er sentrale: *Submission Information Package* (SIP), *Archival Information Package* (AIP) og *Dissemination Information Package* (DIP). Digital arkivinformasjon skal leveres til depot samlet som en avgrenset logisk enhet, en SIP. Aksesjonen/uttrekket depotet mottar fra arkivskaper inngår altså i SIP. Etter mottak skal SIP-pakken gjennom en bearbeidingsprosess som resulterer i en AIP, som langtidsbevares i depot. AIP skal være selvdokumenterende og inneholde nødvendige logiske og tekniske metadata. Kjernen i AIP er ofte den originale avleveringspakken supplert med depotets operasjoner ved mottak. DIP er pakken som gjøres tilgjengelig for eventuelle brukere. DIP-pakken vil kunne være et helt arkiv, eller et *subset* med utvalgt informasjon og metadata som er relevant for bruker.

5.2 Konvertering og anbefalte arkivformater

Når et arkivdepot mottar arkivmateriale påtar depotet seg samtidig å forvalte og tilgjengeliggjøre materialet. Dette gjelder uavhengig av om det er en papirbasert eller digital aksesjon. I papirverdenen er det slik at dersom man plasserer aksesjonen i syrefrie bokser og mapper og setter boksene i et passe temperert magasin, er mye av jobben gjort. Siden papirmateriale er svært bestandig trenger man ikke drive aktivt vedlikehold. Etter at arkivet er tilgjengeliggjort i et registreringssystem (for eksempel Asta/Arkivportalen) kan man dermed si at arkivet er bevart for ettertiden. I en digital verden er det litt annerledes.

⁵ <https://www.arkivverket.no/forvaltning-og-utvikling/regelverk-og-standarder/dias-prosjektet-digital-arkivpakkestruktur>

Digitale data er forgjengelige – i takt med den teknologiske utviklingen skiftes formater ut og erstattes av nye. For at arkivinformatjonen skal overleve må depot derfor flytte data over til nye formater og plattformer. Slik vedlikeholdes arkivmateriale gjennom *migrering*. Dette er en jobb som egentlig aldri vil ta slutt.

Arkivering av digitalt materiale (bitstrømmer) skiller seg fra tradisjonell arkivering på en grunnleggende måte: For å kunne omskape de arkiverte dataene til informasjon som er lesbar for mennesker, må man ha en «fremviser» i form av programvare. For et arkivdepot knytter det seg noen problemstillinger til dette: lisenser (både på programvare og proprietære dataformater), endring i formater over tid, nødvendig plattform som den aktuelle programvaren skal kunne kjøres på etc. Dette er bakgrunnen for valget om å konvertere innholdet til åpne formater som har bred støtte i arkivverdenen, slik som PDF/A, som i dag er foretrukket arkivformat for rene tekstdokumenter. PDF/A regnes som mer bestandig og varig enn andre formater nettopp fordi det er åpent og plattformuavhengig, og fordi det er utviklet spesielt for arkivering.⁶

Et annet utbredt arkivformat er XML, *Extensible Markup Language*. Selve filformatet .xml er et tekstformat som organiserer data i en hierarkisk struktur der merker (<>) gir informasjon om innhold (se 6.1, Noark 5-uttrekk). Formatet er hensiktsmessig for lagring av metadata ved arkivuttrekk, og det er åpent og plattformuavhengig.⁷

Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver (også kjent som Riksarkivarens forskrift) § 5-17 angir andre godkjente filformater.⁸ Arbark forsøker å følge disse bestemmelsene, men ekskluderer likevel ikke andre formater. Dersom materiale som avleveres ikke er konvertert til et arkivformat gjør Arbark denne jobben selv.

5.3 Arkivmaterialets integritet og autentisitet

I tillegg til dokumentkonvertering er sjekksum-verifisering sentralt i et digitalt depot. Bruk av sjekksummer er den vanligste metoden for å undersøke om digitale filer har blitt tuklet med. Dette handler egentlig om arkivinformatjonens autentisitet og integritet – brukere må kunne

⁶ Standarden ISO 19005 setter visse krav til PDF/A som skal sikre dokumentets informasjonsinnhold over tid, blant annet ved å forby avansert funksjonalitet som lyd og video (som kan sette grenser for hvor lenge dokumentet lar seg lese).

⁷ ICTU 2002, XML and digital preservation, s. 9.

⁸ https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286#KAPITTEL_5

stole på at arkivinformasjonen er hva den utgir seg for å være, og at det ikke er skjedd uautoriserte endringer i materialet.

En sjekksum er en lang, unik rekke med bokstaver og tall som beregnes på grunnlag av det samlede bit-innholdet i en fil. I et arkivdepot kan summen brukes som grunnlag for identitetskontroll: Ved mottak kan depottjenesten ved hjelp av programvare regne ut sjekksummer for filene, og dersom man på et senere tidspunkt vil kontrollere at en fil er uendret, kan man beregne en ny sjekksum og foreta en sammenligning med den opprinnelige. Dersom de to sjekksummene er like, er filens integritet bekreftet. Hvis det har skjedd endringer i dokumentet, vil sjekksummen ikke være den samme. (Merk at verifisering ved sjekksum ikke forteller *hva* som er endret, *hvem* som har gjort endringen og hvorfor, eller når endringen har skjedd. Sjekksummene sier kun *at* noe har endret seg.)

Samtidig er det slik at når et dokument konverteres fra ett format til et annet, endres bit-strømmen, selv om informasjonsinnholdet i dokumentet er det samme. Derfor må nye generasjoner av arkivpakken (inneholdende nye genererte sjekksummer) bevares i framtiden, slik at depotet kan påvise at innholdet er uendret.

I forhold til analogt materiale innebærer bevaring av el-arkiv som vi ser noen utfordringer. Men selv om teknologien har endret seg, er de arkivfaglige prinsippene de samme. Som med papirarkiv må det også lages arkivbeskrivelser i et digitalt depot for å dokumentere informasjonens opphavs- og brukssammenheng. Arkivbeskrivelsene bidrar sammen med sjekksummene til å sikre materialets autentisitet.

5.4 Potensielle mottakssystemer

Arbark har foreløpig ikke et digitalt depotforvaltningssystem. Selv om systemet ikke er i aktiv bruk, har vi installert standardversjonen av Archivemata.⁹ Vi vurderer det slik at standardversjonen først og fremst er et mottaksprogram for fil-/mappearbarkiver og bilder.

Det er mye positivt å si om Archivemata. Systemet er kompatibelt med OAIS-modellen og basert på automatiserte mikrotjenester, slik som fil-identifisering, konvertering og sjekksum-generering. Det kan altså gjøre mange av operasjonene beskrevet ovenfor. Men systemet har også visse mangler. For eksempel har ikke Archivemata noen tilfredsstillende katalog eller lagringsløsning. Pakkene som lastes opp og kjøres gjennom Archivemata havner i *Archival*

⁹ <https://www.archivemata.org/en/>

Storage som enkeltstående AIPer. Herfra kan pakkene inneholdende filer og metadata lastes ned lokalt. Men dette gjør ikke nødvendigvis materialet håndterlig og gjenfinnbar.

Arbark er usikker på hvor egnet programvaren er for store og hierarkisk svært sammensatte arkiver, noe som er det vanligste for Arbark når det gjelder fil- og mappearkiver. Hvis man ønsker en mer helhetlig depotløsning, kan Archivemata integreres med andre programmer, siden det har åpen kildekode. For eksempel kan systemet brukes sammen med Access to Memory, som driftes av samme selskap som har utviklet Archivemata, Artefactual Systems fra Canada.¹⁰ Access to Memory er først og fremst et visningsverktøy, et slags Arkivportalen for filarkiver, men er også godt egnet til å lage arkivbeskrivelser.

En annen mulighet er integrering mellom Archivemata og RODA (Repository of Authentic Digital Objects). RODA har også åpen kildekode og er som Archivemata en digital depotløsning med funksjonalitet basert på OAIS.¹¹ Men nettopp fordi systemene har overlappende funksjonalitet og dermed forskjellig realisering av arkivpakker, vil det i dag være vanskelig å integrere dem. Imidlertid pågår det et EU-finansiert prosjekt som skal spesifisere pakkestruktur på tvers av teknologier og systemer.¹² På sikt kan det altså bli mulig å importere og eksportere mellom RODA og Archivemata. Men det gjenstår å se. Arbark vil uansett jobbe mer med disse problemstillingene framover.

5.5 Mer om kunnskapsbehov i digitalt depot

Operasjoner og vedlikehold i digitalt depot er nødvendig for at informasjonen skal fortsette å være tilgjengelig, og for å opprettholde autentisitet og integritet. Dette mottak- og vedlikeholdsarbeidet innebærer håndtering etter *best practices* og ikke minst OAIS-standarden, som vi allerede har vært inne på.

En annen standard Arbark har hentet inspirasjon fra er *Trusted Digital Repository* (TDR).¹³ TDR erstattet i 2012 TRAC-standarden (*Trustworthy Repositories Audit & Certification*) og ble samtidig ISO-godkjent (ISO 16363). TDR er kompatibel med OAIS og fastsetter et sett av krav som bør oppfylles av virksomheter som skal bevare digital informasjon, for at disse virksomhetene skal kunne anses som pålitelige. Standarden tar for seg organisatorisk

¹⁰ <https://www.accessmemory.org/en/>

¹¹ <https://rodain.roda-community.org/>

¹² <https://www.eark-project.com/>

¹³ <https://public.ccsds.org/pubs/652x0m1.pdf>

infrastruktur, selve forvaltningen av den digitale informasjonen, og teknisk infrastruktur og sikkerhet. Institusjoner kan søke og formelt få status som TDR-sertifiserte, men et arkivdepot kan også bruke standarden som et internt verktøy for å oppnå størst mulig grad av profesjonalitet. For Arbark har TDR særlig vært viktig med tanke på integritets- og autentisitetssikring, og i arbeidet med å designe teknisk infrastruktur for digitalt depot.

I tillegg til standarder er selvsagt kunnskap om lovverk viktig i et digitalt depot, også i privat sektor. I privat sektor har lovverket åpenbart en viktig veiledende funksjon som kan bidra til sikring av arkivene. For et arkivdepot er særlig Riksarkivarens forskrift sentral. For å generere sjekksummer og konvertere til arkivformat trengs også kjennskap til programvare og verktøy. Når det gjelder programvare finnes det mange muligheter, både lisensiert og fri programvare.¹⁴

Institusjoner som skal ta imot mer kompliserte el-arkiver vil være avhengig av en del IT-kompetanse. Samtidig er det viktig få på plass rutiner for mottak og praktisk håndtering av el-arkiv, etter hvert som man begynner å motta slike, og her må arkivarene fra starten ha en sentral rolle. Vår vurdering er at kompetanse på dette best bygges ved å sette av ressurser til testing og praktisk utprøving.

¹⁴ Et godt sted å få oversikt er i boka *Archival Arrangement and Description. Analog to Digital* av Lois Hamill. En annen publikasjon Arbark har hatt nytte av er rapporten *Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house*, utgitt av Online Computer Library Center (OCLC). Her går man gjennom stegene ved mottak og kommer med anbefalinger til programvare som kan brukes i de forskjellige stegene.

6.0. Aktuelle metoder

På bakgrunn av våre funn i prosjektet vurderer vi det slik at det er tre metoder som er aktuelle for Arbark når det gjelder database-uttrekk og bevaring: 1) Noark 5-uttrekk; 2) SIARD/Decom-metoden; og 3) leverandøruttrekk.

6.1. Noark 5-uttrekk

Ettersom offentlige arkivskapere er pålagt å bruke Noark-godkjente systemer, har flere norske depotinstitusjoner erfaring med dette. Her kan altså Arbark utnytte kompetanse som allerede finnes.

Noark 5-standarden angir hva et arkivuttrekk skal inneholde, hvordan uttrekket skal være organisert og hvilket format dataene skal være i. Graden av standardisering gjør Noark 5-uttrekk godt egnet for langtidsbevaring. I tillegg skal Noark 5-systemer ha en modul som skal kunne gjøre selve uttrekksjobben. Det er imidlertid ikke gitt at uttrekksmodulen fungerer uten bistand fra systemleverandør, noe vi kommer tilbake til.

Selv om Arbark ennå ikke har mottatt et Noark 5-uttrekk fra våre arkivskapere, har Arbark i prosjektperioden (og gjennom videreutdanning for en arkivar) opparbeidet erfaring med formatet. Utrekksformatet ser slik ut:

Navn	Endringsdato	Type
dokumenter	30.04.2018 11:52	Filmapp
addml.xsd	30.04.2018 11:50	XSD-fil
arkivstruktur.xml	30.04.2018 11:50	XML-fil
arkivstruktur.xsd	30.04.2018 11:50	XSD-fil
arkivuttrekk.xml	30.04.2018 11:50	XML-fil
endringslogg.xml	30.04.2018 11:50	XML-fil
endringslogg.xsd	30.04.2018 11:50	XSD-fil
loependeJournal.xml	30.04.2018 11:50	XML-fil
loependeJournal.xsd	30.04.2018 11:50	XSD-fil
metadatakatalog.xsd	30.04.2018 11:50	XSD-fil
offentligJournal.xml	30.04.2018 11:50	XML-fil
offentligJournal.xsd	30.04.2018 11:50	XSD-fil

I Noark 5 er filen «arkivstruktur.xml» sentral. Her beskrives arkivets logiske oppbygging fra øverste nivå og ned til dokumentfil. For eksempel:

```
<arkiv>
  <arkivdel>
    <klassifikasjonssystem>
```



```

<klasse>
  <mappe>
    <registrering>
      <dokumentbeskrivelse>
        <dokumentobjekt>

```

De ulike nivåene vil ha tilknyttet metadata – altså informasjon som beskriver dokumentasjonen. Metadataene er essensielle i arkivsammenheng fordi de binder dokumentene til konteksten de er skapt i og sikrer dokumentenes bevisverdi.¹⁵ I arkivstruktur.xml-filen kan «dokumentbeskrivelse» og «dokumentobjekt» for eksempel se slik ut:

```

67 | <dokumentbeskrivelse>
68 |   <systemID>8baddba8-616b-4bb6-92e2-a96873b14aa2</systemID>
69 |   <dokumenttype>Brev</dokumenttype>
70 |   <dokumentstatus>Dokumentet er ferdigstilt</dokumentstatus>
71 |   <tittel>Forespørsel – Anker skipsverfts bedriftsarkiv – Deponeringsvilkår (klausul) og inngåelse-
72 |   <opprettetDato>2015-01-02T10:18:14</opprettetDato>
73 |   <opprettetAv>Jonas Larsen</opprettetAv>
74 |   <tilknyttetRegistreringSom>Hoveddokument</tilknyttetRegistreringSom>
75 |   <dokumentnummer>1</dokumentnummer>
76 |   <tilknyttetDato>2015-01-02T10:18:14</tilknyttetDato>
77 |   <tilknyttetAv>Jonas Larsen</tilknyttetAv>
78 |   <dokumentobjekt>
79 |     <versjonsnummer>1</versjonsnummer>
80 |     <variantformat>Arkivformat</variantformat>
81 |     <format>pdf</format>
82 |     <opprettetDato>2015-01-02T10:18:14</opprettetDato>
83 |     <opprettetAv>Jonas Larsen</opprettetAv>
84 |     <referanseDokumentfil>dokumenter/Deponeringsvilkår og avtale.pdf</referanseDokumentfil>
85 |     <sjekksum>84415768828BB1CF85B72FB670DC23F9A3BAF8F8D497B46FEB6805DCDAFEA01
86 |     <sjekksumAlgoritme>SHA256</sjekksumAlgoritme>
87 |     <filstoerrelse>182987</filstoerrelse>
88 |   </dokumentobjekt>
89 | </dokumentbeskrivelse>

```

Dersom man er ute etter et spesifikt dokument kan man gjøre søk i «arkivstruktur.xml». Under nivået «dokumentobjekt» finner man «referanseDokumentfil». Her ser man hvor i filkatalogen («dokumenter»-mappen) man finner dokumentet.

6.2. SIARD/Decom

Riksrevisjonen fastslo i 2010 at store mengder digitalt skapt materiale fra kommunal sektor kan gå tapt som resultat av manglende rutiner og metodikk for bevaring av elektroniske arkiver. Etter dette har kommunale arkivinstitusjoner satt i gang tiltak for å sikre arkivene. Kommunearkivinstitusjonenes digitale ressurscenter (KDRS) har de siste årene utviklet en

¹⁵ Riksarkivet 2018, s.18.

produksjonslinje for å tilrettelegge for langtidsbevaring og innsyn. Grovt sagt er stegene i metodikken at man:

- henter dataene ut fra databasen, enten ved SIARD-uttrekk eller Noark-uttrekk
- med programvaren Arkade 5 (se kapittel 7.0) kan man så opprette en mottakspakke, en såkalt SIP, som kan sendes til depot for behandling
- i depot kan SIARD-uttrekk (SIP) berikes med informasjon, ved hjelp av «det digitale ordningsbordet» Decom. Man har da det OAIS betegner som en AIP, en arkivpakke
- AIPen kan så legges inn i KDRS' digitale depotsystem for langtidslagring
- for Noark 5-uttrekk kan man se på materialet ved hjelp av innsynsløsningen «KDRS Innsyn»

Pakken som blir tilgjengelig for innsyn kalles en DIP, og er oftest en bearbeiding av AIP. KDRS jobber også med å lage DIP-pakker fra SIARD-uttrekk.¹⁶

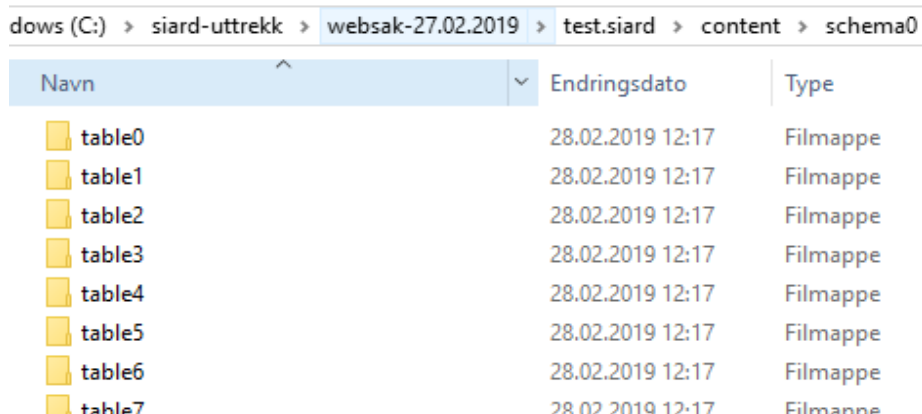
SIARD (Software Independent Archiving of Relational Databases) er et langtidsformat (og et uttrekksverktøy) utviklet av det sveitsiske nasjonalarkivet,¹⁷ som de senere årene har fått utbredelse i Europa. Etter Riksarkivarens forskrift § 5-12e er SIARD godkjent som uttrekksformat. IKAene bruker SIARD for å «redde» de elektroniske arkivene (primært fra gamle fagsystemer) som står i fare for å forsvinne ute i kommunene. SIARD er, som Noark 5-formatet, XML-basert, men minner mer om rådata. Litt forenklet kan man si at Noark 5-uttrekk er ordnede arkiver, mens SIARD er uordnet: alt kommer med i uttrekket, og samtidig forsvinner det logiske laget når informasjonen trekkes ut av produksjonssystemet. Det er imidlertid mulig å laste en SIARD-fil opp i et database-system for slik å gjenskape informasjonen. SIARD konverterer kun fra relasjonsdatabaser.

I løpet av prosjektperioden gjorde vi flere testuttrekk fra våre egne baser, blant annet fra Acos Websak. SIARD-uttrekkene ble gjort mot en kopi av Websak-basen. (Som vi skal komme tilbake til finnes det flere måter å produsere SIARD-uttrekk på.) I forkant av uttrekket gikk vi inn i selve basen og så nærmere på tabeller og basens oppbygning. Websak-basen inneholder over 400 tabeller. Blant annet som resultat av uforståelige navn er det ofte vanskelig å vite

¹⁶ «Produksjonslinje – automatisert formidling av fagsystem».

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

hva de ulike tabellene inneholder. Mappene i SIARD-uttrekket som speiler tabellene i basen, ligger i mappen «schema0»:



Navn	Endringsdato	Type
table0	28.02.2019 12:17	Filmappe
table1	28.02.2019 12:17	Filmappe
table2	28.02.2019 12:17	Filmappe
table3	28.02.2019 12:17	Filmappe
table4	28.02.2019 12:17	Filmappe
table5	28.02.2019 12:17	Filmappe
table6	28.02.2019 12:17	Filmappe
table7	28.02.2019 12:17	Filmappe

Det er en utfordring at det logiske laget forsvinner ved SIARD-uttrekk. For eksempel må man lete i uttrekket etter hvor selve dokumentfilene befinner seg. Dokumentene er nemlig ikke lagret i en egen mappe, slik de er i Noark 5-uttrekk. For å finne dokumentene sjekket vi størrelsen på et stort antall mapper av gangen, og kunne slik snevre ned slik at vi til slutt klarte å finne fram til mappene med dokumentfiler. Filene var ikke lagret som tekstfiler, men som binærfiler, såkalte *BLOB*er.¹⁸ For å identifisere og stadfeste filformater og metadata må man derfor bruke et fil-identifiseringsprogram, for eksempel DROID (Digital Record Object Identification),¹⁹ som kan fastslå filtype, størrelse, alder, med mer. Ved SIARD-uttrekk kan man altså «redde» digitale arkiver, men hvordan gjør man uttrekkene mer tilgjengelige for bruk?

For å løse dette har KDRS fått utviklet verktøyet Decom, som omtales som et «digitalt ordningsbord». Decom er del av den nevnte produksjonsløypa til KDRS hvor man først foretar et SIARD-uttrekk fra et bestemt system, og deretter legger SIARD-fila inn i Decom, hvor man kan tilføre uttrekket beskrivelsesdata via et malsett for det aktuelle systemet.²⁰ Gjennom beskrivelse av tabeller, kolonner og felter skal programvaren bidra til å gjøre arkivinformasjonen lettere å forstå. Decom gir også mulighet til å gjenbruke malsett for like eller lignende systemer. På denne måten effektiviseres arbeidet med de elektroniske arkivene i kommunesektoren, hvor det finnes svært mange forskjellige systemer som må håndteres. I tillegg har programmet en dokumentkonverteringsmodul som konverterer til arkivformat.

¹⁸ *Binary Large Object* – en samling binære data lagret som en enkelt enhet i et system. BLOBer brukes gjerne til lagring av fildata direkte inn i databaser.

¹⁹ DROID er et verktøy for fil-identifisering som blant annet brukes av The National Archives i Storbritannia.

²⁰ Espeland 2019.

Arbark har testet ut en demoversjon av Decom i KDRS-regi. Brukergrensnettet er enkelt og intuitivt, men arbeidet med å lage malsett og beskrivelser er krevende. Man er avhengig av tilgang til fagsystemets system- og brukerdokumentasjon, samt skjermdumper. Etter at arkivaren har beskrevet den opprinnelige SIARD-filen, kan den berikede SIARD-filen eksporteres fra Decom. Beskrivelsene som er gjort i Decom vil være synlige i filen «metadata.xml». Samtidig vil de oppdaterte malsettene lagres i Decom, slik at andre vil kunne dra nytte av dem.

En eventuell investering i Decom må for Arbarks del vurderes opp mot hvor mye bruk vi faktisk har for programmet. Decom er som sagt utviklet for å dekke et behov i kommunedepotene med tanke på fagsystemer, som vanskelig kan «mappes» inn i en hierarkisk Noark-struktur. Ettersom Arbark møter på litt andre utfordringer regner vi det (foreløpig) som lite sannsynlig at vi begynner å bruke Decom i nær framtid.

6.3. Leverandøruttrekk

Dette er betegnelsen vi bruker på uttrekk fra systemer uten Noark 5-godkjenning, hvor man ikke bruker SIARD, men får systemleverandør til å gjøre jobben. Dette kan være arkiv fra både klassiske relasjonsdatabaser og dokumentorienterte baser, eller andre databasestrukturer.

Arbark ønsker i hovedsak å følge nasjonale standarder og metoder. I privat sektor er dette likevel ikke alltid mulig, for eksempel i tilfeller hvor arkivskapere har dokumentdatabase-systemer uten Noark-godkjenning. At vi ikke får arkivinformasjonen i et standardisert format, innebærer at Arbark først må gjøre en vurdering av hva vi ønsker å trekke ut fra systemet. Deretter må arkivskaper bestille produksjon av uttrekk hos leverandøren. Et problem med slike uttrekk er at metadataene som følger med uttrekket kan være svært mangelfulle. Systemleverandøren forstår ikke nødvendigvis viktigheten av å få med all metadata. I forkant av uttrekksproduksjon må Arbark derfor samarbeide tett med både leverandør og arkivskaper, slik at vi sikrer at arkivinformasjonen er mest mulig komplett og håndterbar. Metoden forutsetter i alle tilfeller at arkivskaper er villig til å betale systemleverandøren for arbeidet denne utfører (noe som for øvrig også gjelder for produksjon av Noark-uttrekk).

Leverandøruttrekk er så vidt vi vet foreløpig eneste mulighet for uttrekk fra IBM Notes-baserte systemer, som brukes av flere LO-forbund. Disse er ikke relasjonsdatabaser, og dermed kan (sannsynligvis) heller ikke SIARD brukes (se punkt 10.2.2.).

7.0. Arkade 5

For at arkivuttrekk skal kunne leses og ha verdi for ettertiden må den tekniske kvaliteten på uttrekket testes. Tidligere har denne testingen foregått i depot. Dette har fungert dårlig. Lite tilfredsstillende verktøy har medført at arkivskaper ikke har hatt mulighet til å teste selv, og depotet har på sin side ikke hatt tilgang til datagrunnlaget. Når arkivdepotet ikke har godkjent uttrekket på grunn av dårlig datakvalitet, har arkivskaper måtte gjøre et nytt uttrekk. Dermed har deponeringsprosessen blitt kostbar og lite effektiv.

Et av verktøyene som har kommet de siste årene er Arkade 5. Dette er et gratis, brukervennlig program som forenkler test- og overføringsprosessen. Programmet er bestilt av og utviklet for Arkivverket. Det er meningen at arkivskaper selv skal kunne kjøre uttrekket gjennom Arkade 5, før det avleveres til et depot. Arkade 5 tester den tekniske kvaliteten på uttrekk, produserer testrapporter og pakker inn data. Testingen baserer seg på Arkivverkets standard for teknisk beskrivelse av datasett, ADDML.²¹ Standarden har form som en fil som dokumenterer innholdselementer og deres relasjoner.

Prosjektet har testet ut Arkade 5 med gode resultater, og Arbark er svært positiv til programvaren. Det er nettopp slike brukervennlige, standard-baserte programmer vi trenger for å kunne bevare digitalt skapt materiale.

For Arbarks del er Arkade 5 kun aktuelt for uttrekk fra Noark 5-systemer (siden Arkade forutsetter ADDML-filer i bunn, og Arbark ikke har programmer til å generere ADDML-filer).

Når man laster inn katalogen og kjører et uttrekk gjennom Arkade får man en testrapport med feil og avvik:

²¹ *Archival Data Description Markup Language*: <https://www.arkivverket.no/forvaltning-og-utvikling/regelverk-og-standarder/andre-arkivstandarder/addml-archival-data-description-markup-language>

Arkade 5

Testing av uttrekk

Filsti: C:\Users\DAJ\Desktop\Dags uttrekk
UUID: a5516a41-4ec7-45c6-94ac-5f9c479e26f9
Arkivtype: Noark5
Fil som blir prosessert nå: arkivstruktur.xml
Antall XML-elementer prosessert: 22

Start testing

Vis rapport

Opprett pakke

Ny kjøring

Meldinger:

Validering av XML i henhold til skjema

arkivstruktur.xml - Linje 25: Elementet <http://www.arkivverket.no/standarder/noark5/arkivstruktur:arkivperiodeStartDato> er ugyldig - Verdien 2008-01-02T12:15:00 er ugyldig i henhold til datatypen <http://www.arkivverket.no/standarder/noark5/metadatakatalog:arkivperiodeStartDato> - Streng 2008-01-02T12:15:00 er ikke en gyldig Date verdi.

arkivstruktur.xml - Linje 26: Elementet <http://www.arkivverket.no/standarder/noark5/arkivstruktur:arkivperiodeSluttDato> er ugyldig - Verdien 2010-12-30T12:14:01 er ugyldig i henhold til datatypen <http://www.arkivverket.no/standarder/noark5/metadatakatalog:arkivperiodeSluttDato> - Streng 2010-12-30T12:14:01 er ikke en gyldig Date verdi.

Basert på rapporten kan arkivskaper selv rette opp feilene i uttrekket, for eksempel i XML-filene. Man skiller mellom feil som er gjort i produksjonsfasen (dokumentproduksjon) og feil i uttrekksprosessen. Feil i førstnevnte bør bevares fordi det ivaretar autentisiteten (det er autentisk at feil er gjort). Arkivskaper kan da gi utfyllende forklaringer. Feil i uttrekksprosessen bør derimot rettes opp. Uttrekket skal speile produksjonssystemet og ikke inneholde feil som er gjort av uttrekksprodusenten.

Etter at rettingene er utført kan man kjøre uttrekket gjennom Arkade på nytt, og så opprette en arkivpakke. Dette gjøres ved først å registrere metadata (hvem er arkivskaper, hvilket system, hvem har laget uttrekk, hvilken depotinstitusjon osv.).

Når man har fylt inn disse dataene oppretter Arkade arkivpakken som etter OAIS-terminologi kalles SIP. Det er denne pakken som skal leveres til depot:

Legg til	Pakk ut	Prøv	Kopier	Flytt	Slett	Egenskaper
F:\Dag\Arkadepakke-9e3c734d-e1b5-458d-ac31-aff74ab20dc7\9e3c734d-e1b5-458d-ac31-aff74ab20dc7						
Navn	Størrelse	Endret	Opprettet			
administrative_metadata		2019-01-29 12:33	2019-01-29 12:35			
content		2019-01-29 12:26	2019-01-29 12:35			
descriptive_metadata		2019-01-29 12:33	2019-01-29 12:35			
dias-mets.xml	11 174	2019-01-29 12:33	2019-01-29 12:35			
dias-mets.xsd	155 707	2019-01-29 12:33	2019-01-29 12:35			
log.xml	1 011	2019-01-29 12:33	2019-01-29 12:35			

Selve Noark-uttrekket (dokument-mappen og XML-filene) ligger inne i mappen «content». Arkade-pakken er basert på Arkivverkets DIAS-arkivpakkestruktur.

Vi presiserer at det ikke er depotinstitusjonen som skal bruke Arkade 5 i avleveringsprosessen. Arkade er utviklet nettopp for å gjøre avleveringsprosessen mer effektiv ved å forhindre at uttrekk ikke godkjennes av depot. Arkade-testingen og opprettelse av arkivpakke gjøres altså hos arkivskaper, eventuelt med bistand fra leverandør, etter at leverandør har laget uttrekket.

Bruk av Arkade 5 er gratis. Programmet kan lastes ned på Arkivverkets hjemmesider.

8.0 Asta Mapper

Et annet, nytt verktøy er Asta Mapper, utviklet av Stiftelsen Asta. Det kan brukes til å gjøre uttrekk fra både Noark-systemer og fagsystemer.

Asta Mapper fungerer på følgende måte: Først gjøres det uttrekk fra databasen. Basert på en testrapport analyseres så arkivets innhold i forhold til kravene som ligger i Noark 5.²² Asta Mapper inneholder mange retteprogrammer som maskinelt korrigerer avvik, mangler og feil i dataene. Korrigeringen dokumenteres i aktuelle poster (og i filen INFO.txt). Deretter skjer det en automatisert mapping av dataene til Noark 5-formatet. Filene som skal avleveres genereres maskinelt.

Asta Mapper er uttrekksmetodikk som leveres av Stiftelsen Asta. Det er med andre ord stiftelsen som foretar uttrekk og tar betalt for det. Slik sett er Asta Mapper et alternativ til å la systemleverandørene gjøre jobben.

De foreløpige erfaringene med Asta Mapper er gode: en omfattende deponering fra Justis- og beredskapsdepartementet ble godkjent av Arkivverket på første forsøk, noe som er oppsiktsvekkende bra med tanke på alle problemene man vanligvis har hatt med elektroniske deponeringer.²³ Selv om de foreløpige erfaringene er gode, er de samtidig få. Arbark vurderer likevel Asta Mapper som en høyst aktuell metode for Noark-uttrekk. Prisen på tjenesten avhenger av antall arbeidstimer, og arbeidsmengden vil variere fra uttrekk til uttrekk. Analysen tidlig i prosessen brukes til å beregne hvor mye arbeidet vil koste.²⁴

Basert på de foreløpige erfaringene med Asta Mapper kommer Arbark til å anbefale at våre arkivskapere henter inn pristilbud fra Asta før de bestiller uttrekk fra systemleverandør.

²² Evensen, Negaard 2017.

²³ <https://www.stiftelsenasta.no/media/1351/erfaringsrapport-asta-mapper.pdf>

²⁴ Stiftelsen Asta, udatert, s. 15.

9.0 Testuttrekk basert på aktuelle metoder

Kompetansebygging har stått sentralt i prosjektet. For å få praktisk erfaring har vi gjennomført testuttrekk. Slike tester gir reell erfaring med de ulike stegene i uttrekksprosessen, og mer innsikt i hva som kreves av depotinstitusjon og arkivskaper, samtidig som vi har kunnet utarbeide rutinebeskrivelser.

For å prøve ut metodene gjorde vi flere testuttrekk fra vårt eget Acos Websak-system.

Vi prøvde ut tre forskjellige typer uttrekksteknologi: Acos Avlevering, SIARD 2.1.60 og Full Convert Spectral Core.

9.1. Acos Avlevering

Acos Avlevering er uttrekksmodulen til det Noark 5-godkjente systemet Websak. Arbark tenkte i utgangspunktet at vi kunne hjelpe arkivskaperne med å produsere uttrekk, dersom vi selv fikk dette til. Da ville kostnadene ved uttrekk bli mindre. Kanskje kunne vi også hjelpe de organisasjonene som har andre Noark 5-systemer, ettersom uttrekksmodulene fungerer på samme måte. Vi var imidlertid klar over at dette var vel optimistisk – vår kontakt med andre depotinstitusjoner tilsa nemlig at uttrekksmodulene ikke fungerer slik de var tenkt.

Første steg var å lage en kopi av Websak-basen. Deretter gikk vi inn i denne og journalførte og avsluttet saker.

Så prøvde vi uttrekksmodulen Acos Avlevering. Selve modulen er i og for seg grei. Man velger først avleveringsformat og kildebase. Deretter angir man databaseinnstillinger og kobler til basen. Etter å ha valgt hvor uttrekket skal lagres defineres tidsperiode. Vi valgte en snever tidsperiode for å være sikre på at vi hadde gjort alt som trengtes for å få et vellykket uttrekk. Deretter startet vi konverteringen.

Under denne fikk vi stadige feilmeldinger, eksempelvis: «Exception of type 'Acos.Websak.Avlevering.Noark5Avlevering.DataTilgang.EkskludertMappeException' was thrown».

Delvis i samarbeid med Acos klarte vi å identifisere en del feil, som vi fikk rettet opp slik at vi kunne prøve på nytt. Vi kom da videre. Men det dukket hele tiden opp nye feil, og dersom vi hadde fortsatt ville vi måttet betale Acos utover vanlig brukerstøtte, noe vi ikke ønsket. Følgelig klarte vi ikke å få gjennomført Websak-uttrekk på egen hånd.

Vår erfaring med Acos Avlevering er at modulen fungerer dårlig. Det kan være mange grunner til dette. Mye skal være gjort riktig for at man selv skal klare å produsere et vellykket uttrekk. Slik modulen er i dag vil det uansett ikke være aktuelt for Arbark å produsere Noark-uttrekk selv. Altså må arkivskapere med Noark 5-systemer samarbeide med leverandør – eventuelt Stiftelsen Asta – for å produsere uttrekk. Dette er for øvrig i tråd med et allment prinsipp i det norske arkivmiljøet; at det skal skilles mellom den som produserer uttrekket og den som tester det i etterkant.

9.2. SIARD 2.1.60

Versjon 2 av SIARD kom i april 2018. Vi prøvde SIARDs uttrekksprogramvare i mai og juni samme år. Programvaren er tilgjengelig på det sveitsiske nasjonalarkivets nettsider. Bruken er gratis, men det kreves en lisens.

Programmet fungerer, men da vi testet var programvaren såpass fersk at det sannsynligvis var en del «bugs». Selv om vi fikk gjennomført noen uttrekk, ble resultatene derfor ikke helt vellykkede. Programvaren er også lite intuitiv og litt vanskelig å bruke. Arbark vil neppe bruke SIARDs egen programvare for uttrekk, siden vi trenger enklere verktøy.

9.3. Full Convert Spectral Core

Full Convert Spectral Core er et brukervennlig program som produserer SIARD-filer og som kan hente fra mange forskjellige plattformer.²⁵ Vi ble tipset om programvaren av kollegaer i andre institusjoner, og prøvde derfor ut en demoversjon. Programmet støtter versjon 2 av SIARD.

Ved bruk av Spectral Core trenger man nettilgang og tilgang til aktuell databaseserver. Fremgangsmåten er slik at man velger databaseserver (for eksempel SQL for Websak), og logger inn. Deretter ser man hvilke baser som er tilgjengelig på serveren. Man velger så port, og konverterer databasen til SIARD. Programmet er lisensiert, men prisen er overkommelig: i september 2018 betalte vi \$ 699 for en ettårig lisens, altså i overkant av 6000 kr. Ved SIARD-uttrekk kommer Arbark til å bruke Full Convert Spectral Core framfor SIARDs egen programvare.

²⁵ <https://www.spectralcore.com/fullconvert>

10.0 To caser: systemer med og uten Noark-godkjenning

Da vi tidlig i prosjektet hadde møter med LOs organisasjoner, forhørte vi oss om hvilke forbund som kunne være interessert i gjøre testuttrekk fra sine systemer. Vi mente begge parter hadde interesse av dette. For Arbarks del var det viktig å gå gjennom prosessen steg for steg, for å få praktisk erfaring og dermed kunne etablere rutiner for hvordan reelle elektroniske avleveringer skal gjennomføres. Også forbundene ville ha nytte av å gjennomføre uttrekk for læringens del. Ikke bare det aktuelle forbundet som gjør uttrekket – men også de andre forbundene, siden erfaringene fra prosessen blir delt med de andre.

Det var kun to forbund som mente at dette kunne være aktuelt: Fellesforbundet og Norsk arbeidsmandsforbund. Begge er ressurssterke forbund. Samtidig ønsker de å bli fullelektroniske. Fellesforbundet bruker det Noark-godkjente systemet Acos Websak. Norsk arbeidsmandsforbund bruker Compendia Sak, som ikke har Noark-godkjenning og som heller ikke har relasjonsdatabase i bunn.

Det skal sies at dette var prosesser som ikke gikk helt knirkefritt, da arbeidet stadig ble satt på vent, mye på grunn av kostnader. Fellesforbundet ønsket heller ikke å bruke penger på et testuttrekk, og dermed ble det i stedet snakk om en faktisk avlevering. Arbark hadde helst ønsket et testuttrekk. Fellesforbundets uttrekk lot seg dessverre ikke gjennomføre i løpet av prosjektperioden. Men vi fikk likevel med oss noen nyttige erfaringer fra de tidlige stegene i en avleveringsprosess.

10.1. Fellesforbundet

Det første steget mot avlevering var at Arbark foretok en bevaringsvurdering av Fellesforbundets elektroniske arkiv. En slik vurdering av systemers informasjonsinnhold vil være første steg i de fleste elektroniske avleveringer til Arbark.

Vurderingen baserte seg på en funksjonsanalyse av organisasjonen, forbundets arkivnøkkel og tett kontakt med forbundets arkivtjeneste. Vi ønsket å foreta en funksjonsanalyse for å undersøke om en slik analyse egner seg for våre organisasjoner. Det skal sies at vi ikke gikk i dybden i saksbehandlingsprosesser, men vi mener likevel analysen var nyttig for å vurdere dokumentasjonen som skapes. Siden fagforbund har relativt klart definerte funksjoner, kan analysen av Fellesforbundet også fungere som utgangspunkt for vurdering av andre forbunds dokumentasjon.

En funksjonsanalyse er kort sagt en undersøkelse av dokumentasjonsprosessene i en organisasjon. Ved å klargjøre hvilken dokumentasjon som skapes, og hvorfor, kan man vurdere hva som er viktig å bevare for ettertiden. Funksjoner sier noe om hva en organisasjon gjør: hvorfor eksisterer organisasjonen, hvilke oppgaver utfører organisasjonen for å oppfylle sin funksjon, og hva slags dokumenter resulterer saksbehandlingsprosessen i? I følge Fellesforbundets vedtekter er forbundets hovedformål å fremme sine medlemmers lønns- og arbeidsvilkår. For å oppfylle denne funksjonen forhandler eksempelvis forbundsledelsen med motparten i arbeidslivet, og forhandlingene resulterer i overenskomster. Forbundet skal også forsvare sine medlemmers interesser i nedbemanningsprosesser og lignende. Dokumenter som produseres i disse prosessene må følgelig bevares – Arbark mener at dette materialet har en objektiv historisk verdi, men det er også slik at forbundene selv etterspør materialet lenge etter at det er gått ut av administrativ bruk.

I arbeidet med bevaringsvurderingen så vi også på forbundets arkivnøkkel. Generelt sett anbefaler vi bevaring av hele arkivserier. Målet er å skille ut materiale av et visst omfang. Vurdering på et detaljert nivå vil ofte resultere i at bare små mengder kasseres, og det vil da være lite hensiktsmessig å bruke mye ressurser på en slik gjennomgåelse. Forbundet ønsket imidlertid en vurdering av klassene i arkivnøkkelen, og dette ble derfor også gjort. En detaljert gjennomgang av arkivnøkkelen ble dermed sendt til forbundet sammen med funksjonsanalysen. Bevaringsvurderingen er vedlagt denne rapporten.

Vi presiserer at denne vurderingen gjaldt dokumentasjonen som skapes. Når det gjelder metadataene som produseres av systemet og av saksbehandlere/arkivarer, er disse i stor grad gitt av strukturen i Websak (arkiv – arkivdel – klassifikasjonssystem – klasse – (saks)mappe – journalpost – dokument).

Deretter begynte Fellesforbundet arbeidet med å gjøre klart til avlevering gjennom å rydde i systemet og avslutte saker.

I februar 2019 meldte forbundet at det ønsket å vente med gjennomføring av uttrekket til etter en oppdatering er foretatt, sannsynligvis i 2020. Uttrekket kommer dermed til å lages etter at prosjektperioden er over.

10.2. Norsk arbeidsmandsforbund

En viktig del av prosjektet har vært å få erfaring med uttrekk fra sak-/arkivsystemer uten Noark-godkjenning. Disse systemene er særlig interessante for Arbark, siden flere av arkivskaperne våre bruker dem. Resultatene har samtidig stor overføringsverdi til andre virksomheter innenfor privat sektor. Systemer uten Noark-godkjenning innebærer en enda større utfordring enn Noark-systemene, siden det følgelig ikke eksisterer noen standard for hvordan materialet skal være organisert når det hentes ut.

Under kartleggingen i LO-forbundene kom det fram at flere bruker systemer basert på Lotus/IBM Notes. Notes er en såkalt gruppevareløsning, med en dokumentdatabase i bunn i stedet for en klassisk relasjonsdatabase. Et Notes-uttrekk kan på mange måter heller sammenlignes med et fil- og mappeuttrekk enn uttrekk fra en relasjonsdatabase (men dette avhenger av hva slags løsning for dokumenthåndtering den enkelte organisasjon har valgt å bygge i Notes). Det er vår antagelse at mengden metadata som kommer med i et uttrekk sannsynligvis vil være mer mangelfull enn komplett.

Norsk arbeidsmandsforbund er ett av flere forbund som bruker et Notes-system. Forbundet er også av dem som ønsker å bli fullelektronisk relativt raskt. Av den grunn ville vi gjerne komme i gang med testuttrekk så tidlig som mulig i prosjektperioden. Dette viste seg imidlertid å være vanskelig å få til, og IT- og arkivtjenesten i Arbeidsmandsforbundet fikk ikke klarsignal til å sette i gang med forberedelser til uttrekk før senhøsten 2018.

Planen var å gjøre to typer uttrekk: ett uttrekk utført av systemleverandør, for å få erfaring med eksportrutinen og vurdere resultatet, og et SIARD-uttrekk, for å se om det var mulig å få ut en SIARD-fil fra et Notes-system. Det understrekes at leverandøruttrekket kun var et testuttrekk. Forbundet har også skrevet ut sakene for de aktuelle årene og arkivert disse på papir.

10.2.1. Leverandøruttrekk

Forbundet ønsket at Arbark skulle ta seg av kontakten med leverandøren. Det viste seg at leverandøren hadde en eksportrutine fra Notes-basen som, slik vi forstod det, ikke var prøvd ut tidligere. Eksportrutinen skulle resultere i mapper med saksnummer som navn, inneholdende dokumenter og metadata. Uttrekket skulle gjøres fra historisk base, altså en arkivdel som kun inneholder avsluttede saker.

Utfordringen med ikke-Noark-godkjente systemer er at de i liten grad er standardiserte når det gjelder uttreksformat. Dokumentene i Arbeidsmandsforbundets system Compendia Sak konverteres heller ikke til arkivformat når saken avsluttes. Etter dialog med leverandøren antok vi likevel at vi kunne få et brukbart resultat, og ikke minst ville et uttrekk være verdifullt med tanke på kompetansebygging og etablering av rutiner.

I det følgende gir vi en beskrivelse av en elektronisk avlevering, fra uttrekk til ferdig arkivpakke.

10.2.1.1. Eksporten fra systemet

For å sikre dataintegritet ønsker Arbark primært at leverandør produserer uttrekk og sender det til oss. Dette gir den beste grenseoppgangen mellom roller, og vi unngår samtidig en lite ønsket tilgang rett inn i andre organisasjoners datasystemer. I dette konkrete tilfellet ble det likevel slik at vi fikk tilgang til systemet på en lokal PC hos Arbark. Etter at tilgang var gitt skulle utvikler/leverandør legge til rette for at vi selv på en enkel måte kunne foreta eksporten. Leverandøren anbefalte metoden og forbundet var informert. Vi hadde et minimum av rettigheter inne i systemet, og vi anså dermed løsningen som trygg nok.

Tilgang til systemet forutsetter programvare og brukertilgang. Programvaren IBM Notes ble lastet ned fra en såkalt FTP-server,²⁶ hostet av leverandør. Vi installerte programvaren lokalt på en PC hos Arbark, slik at vi hadde en Notes-klient hos oss. Deretter hjalp Compendia oss med å opprette en bruker, og dermed hadde vi tilgang.

Leverandøren laget så en «eksportér-knapp» inne i systemet, slik at vi enkelt kunne foreta eksporten. Vi lastet ned uttrekket lokalt til C-disk. Derfra kunne vi senere overføre uttrekket til Arbarks digitale magasin.

Vi måtte prøve flere ganger før eksporten gikk gjennom. Første eksportforsøk stoppet på 25 prosent, da det viste seg at to forskjellige saker hadde samme saksnummer. Nedlastningen feilet da eksporten prøvde å opprette en fil som allerede var opprettet. Leverandør rettet dette ved å skrive om eksportkoden slik at eventuelle saker med duplikat saksnummer fikk en teller etter seg (første med saksnummer 10-00657 får mappenavn 10-00657, påfølgende sak får mappenavn 10-00657_1).

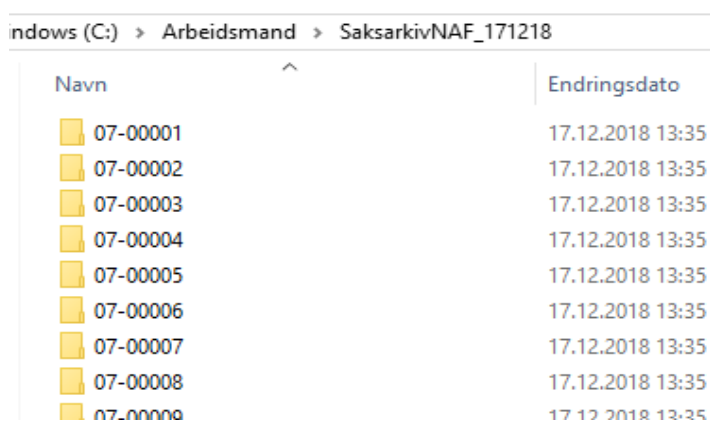
²⁶ En *File Transfer Protocol* er et verktøy som kan brukes til å overføre filer mellom et nettverk og en klient/PC. FTP-overføring innebærer typisk innlogging på en nettbasert server og nedlasting av filer.

Andre eksportforsøk stanset på 59 prosent. Denne gangen var det noen spesielle tegn i et dokument som førte til at dokumentet ikke kunne lagres på disk.

Tredje forsøk stoppet på 60 prosent. Dette skyldtes en annen variasjon av dokument med rart filnavn (sannsynligvis også problemer med spesielle norske tegn).

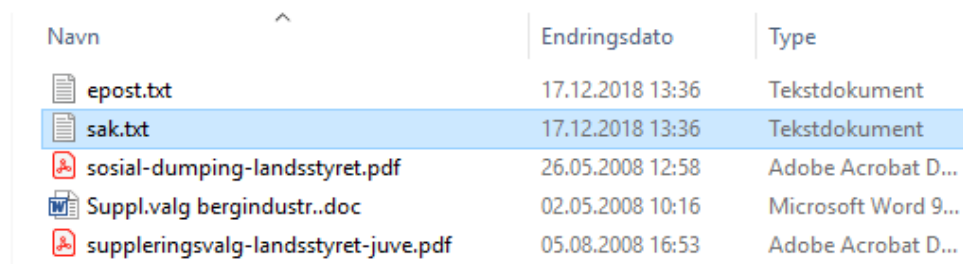
Fjerde forsøk lyktes. Selve eksportjobben tok ca. 45 minutter. Uttrekket hadde 23 254 filer og var på 9,94 GB.

Uttrekket var organisert på samme måte som saksoversikten inne i Compendia Sak, altså etter år (fra 2007) og løpenummer:



Navn	Endringsdato
07-00001	17.12.2018 13:35
07-00002	17.12.2018 13:35
07-00003	17.12.2018 13:35
07-00004	17.12.2018 13:35
07-00005	17.12.2018 13:35
07-00006	17.12.2018 13:35
07-00007	17.12.2018 13:35
07-00008	17.12.2018 13:35
07-00009	17.12.2018 13:35

Inne i saksmappen lå dokumentene:



Navn	Endringsdato	Type
epost.txt	17.12.2018 13:36	Tekstdokument
sak.txt	17.12.2018 13:36	Tekstdokument
sosial-dumping-landsstyret.pdf	26.05.2008 12:58	Adobe Acrobat D...
Suppl.valg bergindustr..doc	02.05.2008 10:16	Microsoft Word 9...
suppleringsvalg-landsstyret-juve.pdf	05.08.2008 16:53	Adobe Acrobat D...

Metadata for dokumentfilene (opprettelsestidspunkt osv.) var med i eksporten. Metadata for hver enkelt sak var lagret inne i saksmappene, i en semikolonseparert fil kalt «sak.txt» (CSV-format).²⁷ I filen lå følgende metadata: saksnummer, sakstittel, opprettelsestidspunkt, status, siste aktivitet, møtedato, saksbehandler, alternativ saksbehandler, arkivnøkkel, type og lesetilgang. Et problem med denne løsningen er at sakstittel ikke er lett tilgjengelig. Sakstittel burde ideelt sett vært på mappenivå. Dette lot seg likevel løse siden vi fikk saksoversikten, inneholdende saksnummer og sakstittel, av forbundet. Saksoversikten har form av et Excel-

²⁷ *Comma-separated values*: Filformat hvor komma brukes til å separere informasjon.

dokument. Med denne har vi en inngang til arkivet, noe som er en forutsetning for effektiv framfinning.

I tillegg til «sak.txt» lå det en «epost.txt»-fil i mange av mappene, med følgende metadata: saksnummer, journalnummer, type, fra e-post, kopi til e-post, blindkopi til e-post, emne, innhold, status, saksbehandler, alternativ saksbehandler, vedlegg.

Metadata for «brev.txt» var: saksnummer, journaltype, type, fra, til, til adresse, emne, innhold, status, saksbehandler, alternativ saksbehandler, vedlegg.

«Sak.txt» lå i alle saksmapper, mens «epost.txt» og «brev.txt» logisk nok kun lå i mappen dersom saken inneholdt e-post eller brev.

Selv om vi alt i alt var godt fornøyde med resultatet av eksporten var det likevel noen åpenbare problemer. For det første var ikke filene konvertert til arkivformat (se 10.2.1.4, Konvertering av dokumenter). For det andre hadde ikke alle filbetegnelsene meningsbærende navn. For eksempel hadde dokumenter som var scannet inn i sak-/arkivsystemet fått løpenummer. Slike mangler når det gjelder titler antar vi er ganske vanlig.

Det var også mange hull i saksnummerrekkefølgen. Neste saksnummer etter sak 15-00355 kunne for eksempel være 15-00360. Slike hull var gjennomgående i hele arkivet. Det manglet med andre ord mange saker. Arkivtjenesten i forbundet forklarte dette med at uttrekket vi fikk kun besto av avsluttede saker. Uttrekket ble gjort fra en historisk base i systemet som leverandøren hadde opprettet fordi den aktive basen var for stor. Det vil si at det finnes mange gamle saker (fra 2007 og framover) som fortsatt er aktive hos arkivskaper. Dersom forbundet blir fullelektronisk tilsier dette at det vil komme en framtidig avlevering som inkluderer disse sakene.

Hull i saksnummerrekkefølgen kan forekomme. Depotinstitusjonen må dokumentere dette, slik at det finnes informasjon som forklarer hullene i den ferdige arkivpakken.

10.2.1.2. Sikring av SIP

Etter at vi hadde fått eksporten fra systemet, foretok vi en virussjekk. Deretter sjekket vi om filene var lesbare. Samtidig hadde vi underveis i prosessen lesetilgang i produksjonssystemet. Dermed kunne vi foreta stikkprøver og sjekke om mapper og filer i vår eksport stemte overens med informasjonen i systemet. Vi kunne også sjekke feilene som leverandøren påpekte ved uttrekksproduksjonen. Vi fant aldri noen avvik under disse stikkprøvene.

Etter mottakskontroll ble uttrekket kopiert fra lokal C-disk til Arbarks digitale magasin. Kun to av Arbarks arkivarer har tilgang til lagringsområdet. Det originale uttrekket er etter OAIS-terminologi en SIP. Etter OAIS skal opprinnelig SIP bevares uendret og integritetssikret. SIP skal fungere som en back-up som det ikke skal gjøres endringer på. Dette gjør det mulig å kunne spore depotets operasjoner.

Her er begrepene autentisitet og integritet viktige. En arkivinstitusjon skal kun utføre operasjoner som er nødvendige for at arkivmaterialet skal vedlikeholdes på best mulig måte. Underveis, helt fra mottak, er det avgjørende at arkivinformatjonens autentisitet og integritet er opprettholdt. I arkivfaglig sammenheng betyr dette at brukere skal kunne stole på at et dokument er hva det utgir seg for å være, og at det ikke skjer uautoriserte endringer i dokumentene.

Sjekksum-verifisering er som nevnt en vanlig metode for å bevise at det ikke har skjedd endringer i digitale filer. Sjekksum-generering var ikke del av eksportfunksjonen fra Compendia. Vi måtte derfor regne ut sjekksummer for dokumentene selv. Til dette brukte vi programvaren Fixity AVP.²⁸ Programvaren er gratis. Fixity kan overvåke digitalt arkivmateriale over tid, ved å foreta jevnlig skanninger. Først gjør Fixity en førstegangsskanning av materialet. Alle filene i uttrekket får da utregnet en sjekksum basert på bit-innhold. Det lages to rapporter fra hver skanning: en oversiktsrapport («reports») og en historikk-rapport («history»). Slik får man dokumentasjon bakover i tid. «Reports»-filen oppsummerer resultatet av skanningen:

```
fixity_2019-01-16-123602_Test.tsv - Notisblokk
Fil Rediger Format Vis Hjelp
Fixity report
Project name      Test
Algorithm used    sha256
Date              2019-01-16
Time Elapsed      0 hrs 0 min 1 seconds
Total Files       1
Confirmed Files   1
Moved or Renamed Files 0
New Files         0
Changed Files     0
Removed Files     0
Confirmed File:   H:\Testmappe\Test.docx
```

«History»-filen inneholder på sin side selve sjekksummene:

²⁸ <https://www.weareavp.com/products/fixity/>

```
Test_2019-01-16-123602.tsv - Notisblokk
Fil Rediger Format Vis Hjelp
H:\Testmappe;
99 12:36:00 99 99
2019-01-16 12:36:02
.db||-||1
sha256
69b3a4f8aacdbb46bcd29cad091143d9262996a223c78f52db2f0785da4be268
```

Arkivaren definerer hvor ofte man ønsker at materialet skal sjekkes for endringer (for eksempel hver måned). Ved neste skanning skal de eksisterende sjekksummene bekreftes. Dersom det har skjedd noe kan arkivaren få beskjed via e-post, eller sjekke manuelt i rapportene. I Fixity-programmet kan man også sette på filter for visse typer filer, for eksempel skjulte systemfiler, slik at disse ikke skannes. Fixity har SHA256-algoritme som standardinnstilling.²⁹

Vi laget også sjekksum for den samlede SIP-pakken. I Windows 10 har 7-Zip funksjonalitet for å regne ut sjekksum for mapper. Vi brukte 7-Zip til å lage sjekksum for SIP, la deretter sjekksummen inn i en PDF-fil og plasserte denne på samme mappenivå som SIP-pakken.

10.2.1.3. Produksjon av AIP: begrensninger i antall tegn

Vi anså nå originaluttrekket som sikret. Vi måtte imidlertid også lage en pakke som i størst mulig grad er selvforklarende for ettertiden. Det første vi gjorde i den forbindelse var å lage en arbeidskopi. Det var denne som skulle bearbeides og lages om til en ferdig arkivpakke (AIP).

Kopieringen ble gjennomført fra lokal C-disk til Arbarks digitale magasin. Da vi kopierte originaluttrekket for å lage arbeidskopien fikk vi feilmelding på 11 dokumentfiler: «Finner ikke dette elementet». Det er litt usikkert hva årsaken var, men trolig skyldtes feilen (svært) lange fil- og mappenavn, ettersom Windows har visse begrensninger på antall tegn i filstien til et dokument.

Begrensningene i antall tegn må man være oppmerksom på. Filsystemet i Windows (vi bruker Windows 10 Pro, versjon 1803) tillater maksimalt 255 tegn *i hele filstien*, fra rotfil ned til filnavn. Da vi etter kopieringen sammenlignet arbeidskopien med originaleksporten på C-

²⁹ Det finnes flere mulige sjekksum-algoritmer, blant annet CRC, SHA-1, SHA-384 og SHA-512. Arbark foretrekker SHA-256 da det er krav om denne i Riksarkivarens forskrift § 5-31(9).

disk, oppdaget vi at verken arkivstørrelse (GB) eller antall filer stemte overens. For at mappenavn skulle være mest mulig meningsbærende hadde vi først nokså omstendelige navn på mappene. Samtidig ble det flere mapper innover til dokumentene fra rotfilen:

[aab-fsrv](#) (Q:) > [ARK-1646, Norsk arbeidsmandsforbund](#) > [Saksarkiv 2007-2015 - originale filer](#) > [SaksarkivNAF_171218](#)

Dersom et dokument ligger i en filsti der det totale antall tegn overskrider 255, kommer det opp en feilmelding når dokumentet skal åpnes: «Beklager, vi finner ikke [filstien]».

Sannsynligvis telles heller ikke disse dokumentene når man går inn på Egenskaper og ser antall filer (og arkivstørrelse).

De 11 filene vi fikk feilmelding på ved kopiering måtte overføres manuelt: vi noterte saksnummer og dokumenttittel fra feilmeldingen, fant så disse igjen i originaleksport på C-disk, og overførte filene en og en. Vi kortet også ned mappenavnene. Deretter gjorde vi en ny sjekk, og da stemte antall filer og arkivstørrelse.

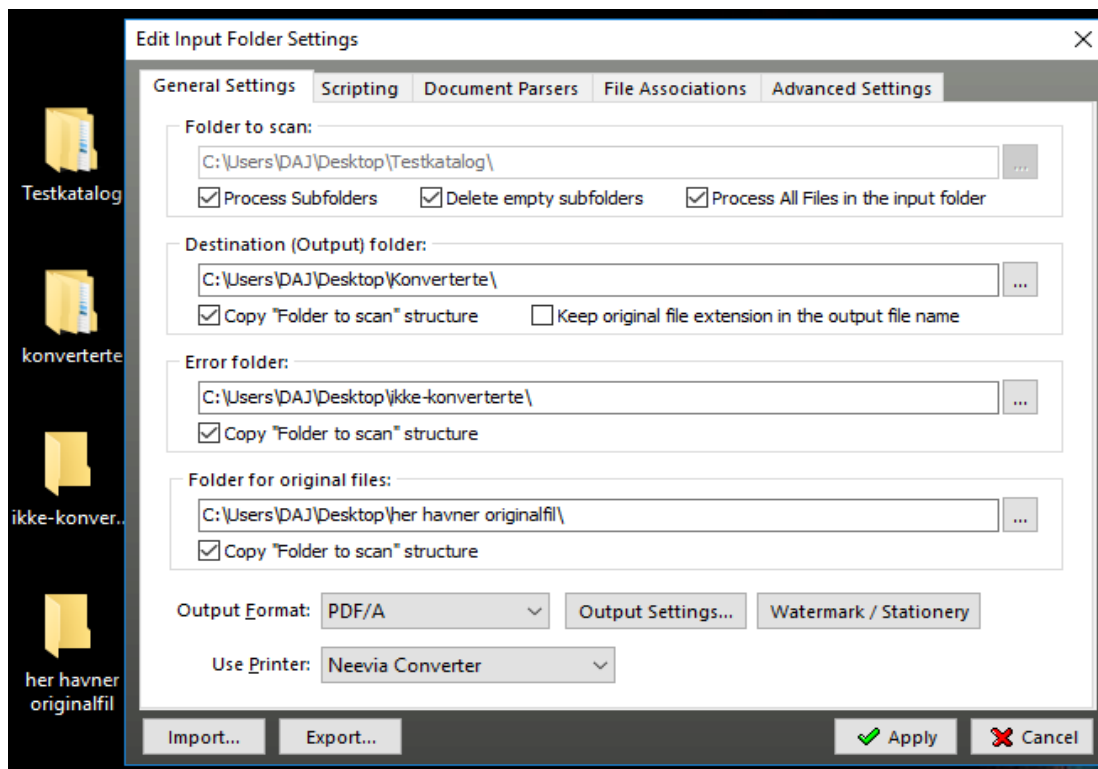
Denne kopieringen ble først gjort fra lokal C-disk til Arbarks digitale magasin. Men av hensyn til ytelse og hastighet fant vi ut at vi heller ville jobbe med arbeidskopien på C. Vi slettet derfor arbeidskopien på sikringsområdet som hadde møtt på tegnproblemene, og kortet samtidig ned på mappenavn. Da vi senere overførte den ferdige AIPen/arbeidskopien fra C-disk til digitalt magasin, gikk det knirkefritt.

Det er likevel noen usikkerhetsmomenter knyttet til tegnbegrensningene som vi må forsøke å få klarhet i. Sannsynligvis kan andre depotinstitusjoner komme med innspill. Det kan også hende at det finnes egnet programvare som løser dette.

10.2.1.4. Produksjon av AIP: konvertering av dokumenter

Dokumentene i uttrekket var ikke konvertert til arkivformat. Dette måtte vi derfor gjøre selv. Hoveddelen av dokumentene var enten Word-filer eller vanlig PDF-format, altså ikke PDF/A. Konvertering inngikk derfor i arbeidet med AIP. Til å konvertere filer i mappehierarkier og legge dem tilbake på riktig plass, brukte vi konverteringsprogrammet Neevia Document Converter Pro.³⁰ I programmet velger man hvilke mapper som de konverterte dokumentene skal plasseres i:

³⁰ <https://neevia.com/convert-to-pdf/>



Vi ønsket kun å konvertere Word-filer og vanlige PDF-filer til PDF/A. Vi konverterte ikke Excel-filer, da disse kan bli vanskelige å lese ettersom deres opprinnelige form ikke samsvarer godt med PDF/A. I denne omgang konverterte vi heller ikke metadata-filene (.txt), delvis av samme grunn. Samtidig er rene tekstfiler også godkjent arkivformat.³¹ Siden txt-filene ikke ble konvertert til PDF/A ble de ikke låst for endringer. Vi kan imidlertid sikre oss mot endringer ved å skrivebeskytte mappene i digitalt magasin.

For å unngå å konvertere disse formatene gikk vi inn i «File Associations» og valgte «PASSTHROUGH» for de aktuelle formatene.

Deretter ba vi programmet om å legge de konverterte filene i samme mappe som de ikke-konverterte filene. Vi huket også av for at den opprinnelige mappestrukturen skulle opprettholdes. Dokumentene fra *Folder to scan* havner da i samme mappe, uavhengig av om dokumenter konverteres eller ikke.

På grunn av mengden dokumenter stod Neevia og jobbet i flere dager. Konverteringen ble gjort på en normalt kraftig, bærbar PC (12 GB ram).

³¹ Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver (Riksarkivarens forskrift) § 5-17 1a.

En del filer ble ikke konvertert. Dette var hovedsakelig filer med ukjent format. Disse måtte overføres manuelt.

Opprettelsesdato, opprinnelig filtype og størrelse forsvant med PDF/A-konverteringen. På tross av at det i Neevias brukermanual står at programmet har funksjonalitet til å få med slike metadata, klarte ikke vår versjon (6.9) dette. Dette er selvfølgelig ikke ideelt, men heller ikke kritisk siden sentrale metadata ofte finnes inne i dokumentene (og selvfølgelig i SIP-pakken). Bevaring av opprettelsesdato og filtype er etter vår mening særlig viktig ved konvertering av dokumenter i et fil/mappearkiv, og vi fortsetter å lete etter programmer som klarer å få med opprinnelige metadata ved kopiering.

10.2.1.5. Produksjon av AIP: aktør- og arkivbeskrivelse

Arbark ønsker i størst mulig grad å følge nasjonale og internasjonale standarder. Vi har tidligere deltatt i to SAMDOK-delprosjekter ledet av Arkivverket.³² I sammenheng med det første av disse fikk vi støtte fra Arkivverket til det tidligere nevnte PEANUTS-prosjektet, hvor vi testet ut tilgjengeliggjøringsprogrammet *Access to Memory* (AtoM). AtoM er nå installert på Arbark.

AtoM er standard-basert og kan blant annet brukes til å lage aktør- og arkivbeskrivelser. At arkivinformasjonen blir beskrevet er avgjørende, siden man ved å sette materialet inn i en kontekst gjør ettertiden i stand til å forstå materialet. Arbark har lang erfaring med å beskrive LO-forbundenes arkiver og hva de inneholder, og det finnes allerede arkiv- og aktørbeskrivelser av Norsk Arbeidsmandsforbund i Asta. Men vi ville likevel ha med beskrivelser som del av arkivpakken slik at den blir selvbeskrivende og man unngår å måtte gå et annet sted for å finne informasjon. Beskrivelsene i AtoM kan eksporteres som XML-filer, og bevares altså som del av arkivpakken.

AtoM er nett-basert. For å lage aktørbeskrivelse skriver man adressen i nettleseren, logger på AtoM og oppretter ny *Authority record*. På norsk er dette en aktørbeskrivelse, i vårt tilfelle en beskrivelse av arkivskaper. Aktørbeskrivelsesfeltene i AtoM er basert på standarden International Standard Archival Authority Record for Corporate Bodies, Persons and

³² *Delprosjekt privatarkiv 2016, Metodikk for høsting og langtidslagring av privatarkiv og Delprosjekt privatarkiv 2017, Sammenligning av programvare for arkivbeskrivelse og visning/gjenfinning av digitalt arkivmateriale.*

Families, også kjent som ISAAR (CPF). Dette er en faglig, ikke-teknisk standard utviklet av International Council on Archives.

Vår opplevelse er at AtoM er meget brukervennlig og enkel å bruke til beskrivelser:

The screenshot shows the 'Edit Authority record - ISAAR' interface for 'Norsk arbeidsmandsforbund'. The interface is divided into several sections:

- Identity area**: A section for identifying the authority.
- Description area**: A section for providing a description of the authority.
- Dates of existence ***: A field containing '1895 -'.
- History**: A text area containing the text: 'Stiftet 13. april 1895 som Det norske vei- og jernbanearbeiderforbund. Skiftet navn til Det norske sten- jord- og bergarbeiderforbund 1900. Innmeldt i LO 1. januar 1901.'
- Places**: A text area containing the text: 'Forbundet er landsomfattende. Forbundskontoret ligger i Oslo, men forbundet har avdelinger rundt om i hele Norge.'

Etter å ha fylt inn de feltene man ønsker, lagres og eksporteres beskrivelsen.

Aktørbeskrivelsen eksporteres da som en XML-basert variant av ISAAR, nemlig EAC-CPF (Encoded Archival Context – Corporate bodies, Persons and Families). Denne EAC-filen lagres som del av arkivpakken.

Arkivbeskrivelsen ble laget på samme måte: vi opprettet ny *Archival description* i AtoM. Her er beskrivelsesfeltene basert på General International Standard Archival Description (ISAD), også kjent som ISAD(G).

Fonds ARK-1646 - Norsk arbeidsmandsforbund (Draft)

Identity area		»
Reference code	ARK-1646	
Title	Norsk arbeidsmandsforbund	
Date(s)	• 2007 - 2015 (Creation)	
Level of description	Fonds	
Extent and medium	23 254 filer (9,94 GB) lagret i filsystem i Windows	
Context area		»
Name of creator	Norsk arbeidsmandsforbund (1895 -) Administrative history: Stiftet 13. april 1895 som Det norske vei- og jernbanearbeiderforbund. Skiftet navn til Det norske sten- jord- og bergarbeiderforbund i 1898. Tok navnet Norsk arbeidsmandsforbund i 1900. Innmeldt i LO 1. januar 1901.	
Repository	Arbeiderbevegelsens arkiv og bibliotek	
Archival history	Uttrekk gjort fra Arbeidsmandsforbundets elektroniske sak-/arkivsystem, Compendia Sak. Eksport utført 17. desember 2018. Testing og sikring av materiale pågår. Filkonvertering er gjennomført.	
Immediate source of acquisition or transfer	Eksport fra Compendia utført 17. desember 2018	
Content and structure area		»
Scope and content	Arbeidsmandsforbundets saksarkiv fra årene 2007 til 2015. Arkivet inneholder blant annet saker fra styrende organer, medlemssaker og	

Clipboard

[Add](#)

Explore

[Reports](#)

[Browse as list](#)

[Browse digital](#)

Import

[XML](#)

[CSV](#)

Export

[Dublin Core 1.1](#)

[EAD 2002 XML](#)

Related people and organizations

[Norsk arbeidsmandsforbund \(Creator\)](#)

Den ferdige beskrivelsen eksporteres som en EAD.xml-fil. På samme måte som EAC-CPF samsvarer med ISAAR(CPF), er EAD (Encoded Archival Description) en implementering av ISAD(G).

Felter i EAD/ISAD(G) og EAC-CPF/ISAAR(CPF) overlapper. Derfor skal man helst ikke fylle ut aktørrelevant informasjon i selve arkivbeskrivelsen (EAD).

Også EAD-filen ble lagt ved AIP.

Arkivbeskrivelsen i AtoM ble relativt overordnet. Vi anser dette som holdbart siden informasjon lenger ned i arkivhierarkiet finnes i sak.txt-filene og i selve saksmappene. Men vi ser likevel et potensiale i det å få metadata fra slike uttrekk inn i en hierarkisk struktur ala Noark. Metadata for hver enkelt sak lå som vi har sett inne i filen «sak.txt». Dette er filer i såkalt CSV-format, hvor kommategn skiller informasjonen fra hverandre. Ved å få metadataene fra CSV-filene over i en xml-fil som ligner Noark 5-filen «arkivstruktur.xml» kan vi få en mer enhetlig arkivkatalog som kan utveksles med andre systemer. I forhold til CSV er XML langt mer hensiktsmessig som utvekslingsformat.³³ Og interoperabilitet er viktig, for eksempel ved eksport til Arkivportalen eller Digitalarkivet.

³³ Sødning 2017.

Selv om det kan være fordelaktig å mappe strukturen i Compendia-uttrekket mot Noark 5, har vi foreløpig ikke gjort dette.

10.2.1.6. Produksjon av AIP: ferdigstilling

Arbeidet med AIP foregikk på lokal C-disk. Etter at filer var konvertert og beskrivelsesfiler laget, overførte vi AIP til Arbarks digitale sikringsområde. Her ble den lagt i samme overordnede mappe som originaluttrekket (SIP), slik at man har to versjoner av «Norsk arbeidsmandsforbunds saksarkiv 2007-2015»: en uendret original og en bearbeidet kopi. AIPen fikk følgende mappestruktur:

```
«ferdig_arkivpakke»
  «administrative_metadata»
    «repository_operations»
      «Fixity_reports»
    «content»
      <selve uttrekket/saksmapper med filer>
    «descriptive_metadata»
      <arkivbeskrivelse- og aktørbeskrivelsesfiler (EAC og
EAD)>
```

Mappestrukturen er et forsøk på tilpasning til Arkivverkets DIAS-standard.³⁴

Etter overføringen genererte vi sjekksummer for dokumentfilene ved hjelp av Fixity, på samme måte som vi hadde gjort med SIP-pakken. Sjekksummene for filene ble lagt i «Fixity_reports». Vi brukte også 7-Zip til å lage sjekksum for «ferdig_arkivpakke». Samlet sjekksum ble lagt på samme nivå i hierarkiet som den ferdige pakken, i filen «ferdig_arkivpakke_checksum_sha256.pdf».

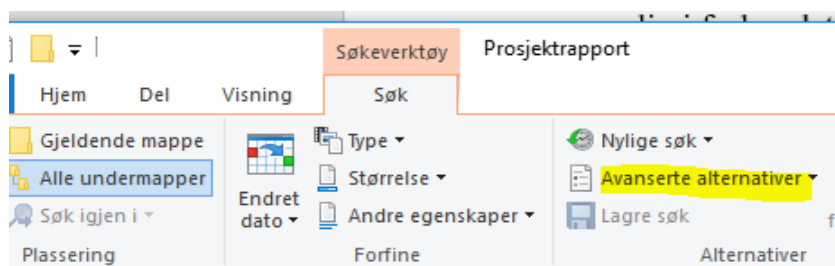
Som tidligere nevnt var vi avhengig av saksoversikten fra systemet for å kunne søke i materialet. Saksoversikten inneholder både saksnummer og tittel. Denne ble tilsendt og lagt i «content» sammen med brukermanualen til Compendia Sak.

En manuell logg med operasjonene vi utførte underveis ble lagt til i «repository_operations». Det samme gjelder et notat som beskriver konverteringen.

³⁴ Riksarkivaren 2012, s. 22.

10.2.1.7. Søkemuligheter og gjenfinning

Når det gjelder søkemulighetene er disse ganske gode i Windows 10. Ettersom vi fikk saksoversikt av arkivskaper, har vi kontroll på hvilket saksnummer som hører til hvilken sak. I tillegg kan man i Windows' søkeverktøy gå inn i Avanserte alternativer og velge «Filinnhold».



Man kan da gjøre søk på innhold i dokumentfiler. Når vi i tillegg har saksoversikt bidrar dette til (relativt) effektiv framfinning for saksbehandlere.

10.2.1.8. Vurdering av testuttrekk og tanker om digitalt depot

Som sagt er erfaringen fra dette uttrekket svært nyttig for vår kompetansebygging. Per mai 2019 er det to forbund som bruker Compendia daglig. Samtidig er det flere forbund som har lignende systemer som bygger på samme programvareplattform, IBM Notes. Dette betyr: 1) at vi kan bruke samme metode på nytt, altså at vi laster ned programvaren og eksporterer fra systemet; 2) at vi nå vet hvordan Notes' uttrekksformat ser ut. Formatet har en svært enkel hierarkisk oppbygning.

Arbarks test- og mottaksrutiner vil være enklere enn rutinene i offentlige depotinstitusjoner. Arkivverkets testing har bestått i å utføre prosedyrer for å verifisere at innholdet i en SIP er teknisk konsistent, korrekt og komplett, om det oppfyller definerte struktur- og formatkrav, og om det er tilknyttet de tekniske og logiske metadata som kreves for å bevare informasjonen med opprettholdt lesbarhet og autentisitet.³⁵ Selv om det nå finnes verktøy som forenkler prosessen (for eksempel Arkade 5), har ikke Arbark tilstrekkelig med ressurser, kompetanse og erfaring til å kunne håndtere digitale arkiver på samme måte som Arkivverket og IKAene.

³⁵ Riksarkivaren 2012, s. 38.

Arbark er avhengig av brukervennlige arbeidsverktøy. Programmene vi brukte ved mottak av Arbeidsmandsforbundets arkiv var Fixity, Neevia og AtoM. Disse ble valgt mye på grunn av deres brukervennlighet. En av barrierene mot å ta i bruk nye verktøy er ofte at man ikke skjønner akkurat hva programmene gjør og hvorfor. Derfor var gode brukermanualer viktig for oss. Etter at vi hadde mottatt uttrekket måtte vi likevel bruke en del tid på testing av programmene, for å få dem til å gjøre akkurat det vi ønsket. Særlig gjaldt dette Neevia, selv om vi har brukt dette tidligere (men på litt andre typer oppgaver). Neevia har ganske mange funksjoner. Derfor er det en fordel å ha noe erfaring med programmet før man går i gang med en stor konverteringsjobb (særlig for uttrekk med mange mapper hvor filer skal plasseres riktig).

Vår vurdering er at alle disse programmene har lav brukerterskel og fungerer godt. Arbark vil inntil videre fortsette å bruke dem, men vi utelukker ikke mer leting etter andre verktøy. Særlig er det ønskelig med et konverteringsverktøy som tar med seg metadata fra opprinnelig fil.

Ved mottaket av Arbeidsmandsforbundets arkiv kunne flere ting vært gjort annerledes. For eksempel burde generering av sjekksummer vært en del av eksportfunksjonen fra systemet. I stedet måtte vi altså selv lage sjekksummer. PDF/A-konvertering burde også vært del av eksportrutinen.

Vi burde også ha brukt fil-identifiseringsprogrammet Digital Record Object Identification (DROID). Denne gangen gjorde vi fil-identifiseringen manuelt ved å ta stikkprøver, noe som ikke er nok. Muligens burde vi også brukt sammenligningsverktøyet Beyond Compare.³⁶ Alt dette er erfaringer vi tar med oss videre.

Depottjenesten skal ha kontroll med hva som blir gjort i mottaksprosessen og i det videre arbeidet med å lage arkivpakker. Logging er derfor viktig. Siden Arbark ikke har noe digitalt depotforvaltningssystem opprettet vi en loggfil der vi manuelt skrev ned hva vi gjorde underveis. Denne loggfilen ble ikke så detaljert og grundig som den burde ha vært. En lærdom av dette er at informasjon knyttet til mottak bør standardiseres. Arbark har derfor laget en spesifikasjon – et ferdigdefinert skjema – som skal fylles ut når vi mottar en digital aksesjon. Her inngår dateringer, informasjon om hvordan avleveringen er gjennomført, hvem som har foretatt forskjellige operasjoner osv. Hele prosessen skal dokumenteres. Noe av denne informasjonen er dekket gjennom bruk av Arkade 5 (for Noark-uttrekk) og aksesjonsføring i

³⁶ <https://www.scootersoftware.com/>

Asta, men bare delvis. Spesifikasjonen er vedlagt prosjektrapporten.³⁷ Dette er en første versjon som med tiden vil forbedres.

Vi har også, basert på våre erfaringer, utarbeidet en rutinebeskrivelse for de ulike stegene i prosessen med å lage arkivpakker ved mottak fra ikke-Noark-systemer. Samme rutinebeskrivelse kan også komme til nytte ved andre typer avleveringer (både Noark-systemer og fil/mappearkiver), samtidig som den har overføringsverdi til andre depotinstitusjoner. Også rutinebeskrivelsen er vedlagt rapporten.³⁸

Oppsummert gir erfaringene fra prosjektet oss grunn til å tro at vi skal kunne ta imot og håndtere materiale fra lignende systemer som Compendia. Dette forutsetter imidlertid at arkivskapere prioriterer å lage eksport fra sine systemer.

10.2.2. Forsøk på SIARD-uttrekk

I tillegg til leverandøreksporten ønsket vi et SIARD-uttrekk. Det kan virke noe eksperimentelt å konvertere et uttrekk fra dokument-database til SIARD (relasjonsdatabase), ettersom det er en viss risiko ved slike transformasjoner. Men når vi først hadde muligheten ønsket vi likevel å prøve ut om vi kunne få en SIARD-fil fra Notes som vi kunne teste.

Full Convert Spectral Core henter ikke direkte fra Notes-baser, men det skal likevel være mulig å få uttrekk fra slike systemer og deretter viderekonvertere til SIARD-formatet. Denne migreringen gjøres via en såkalt ODBC-driver – et grensesnitt mot Notes som fungerer som oversetter.³⁹ I følge systemleverandør ville eksporten komme i Excel-format. Spectral Core har funksjonalitet for å konvertere fra Excel til SIARD.⁴⁰ Driveren er utviklet av IBM, som i dag eier Notes.

Vi ønsket at leverandøren skulle utføre eksporten og sende den til oss, så kunne vi bruke Spectral Core, som vi har lisens på. Leverandøren ville imidlertid ikke bruke tid både på egen eksport og vår metode. Som vi allerede har vært inne på hadde vi tilgang til selve systemet, og i tillegg fikk vi en brukerveiledning for hvordan ODBC-driveren skulle kunne brukes. I

³⁷ «Skjema for mottak av elektronisk arkiv».

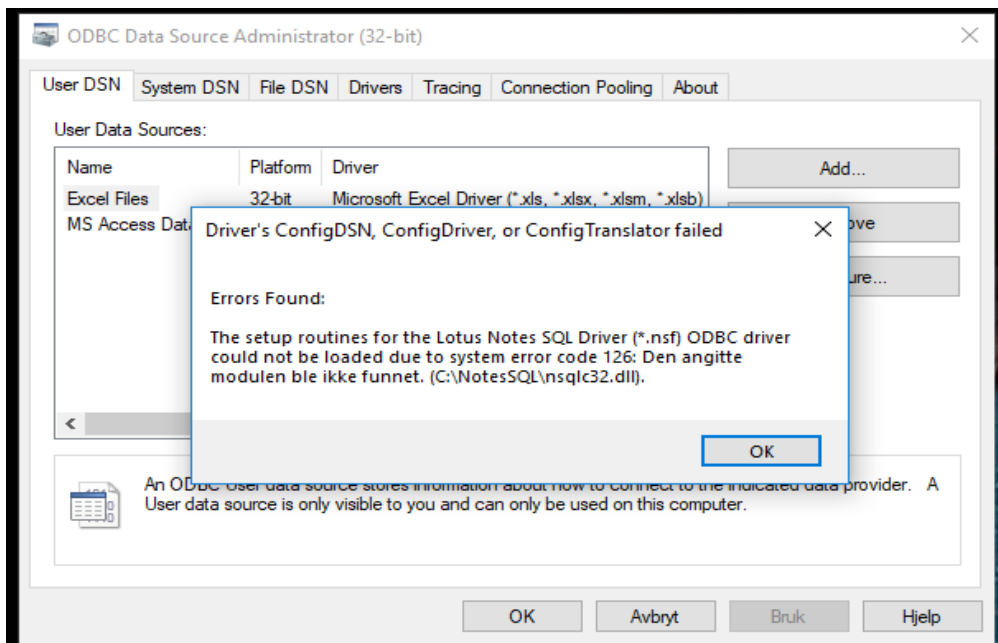
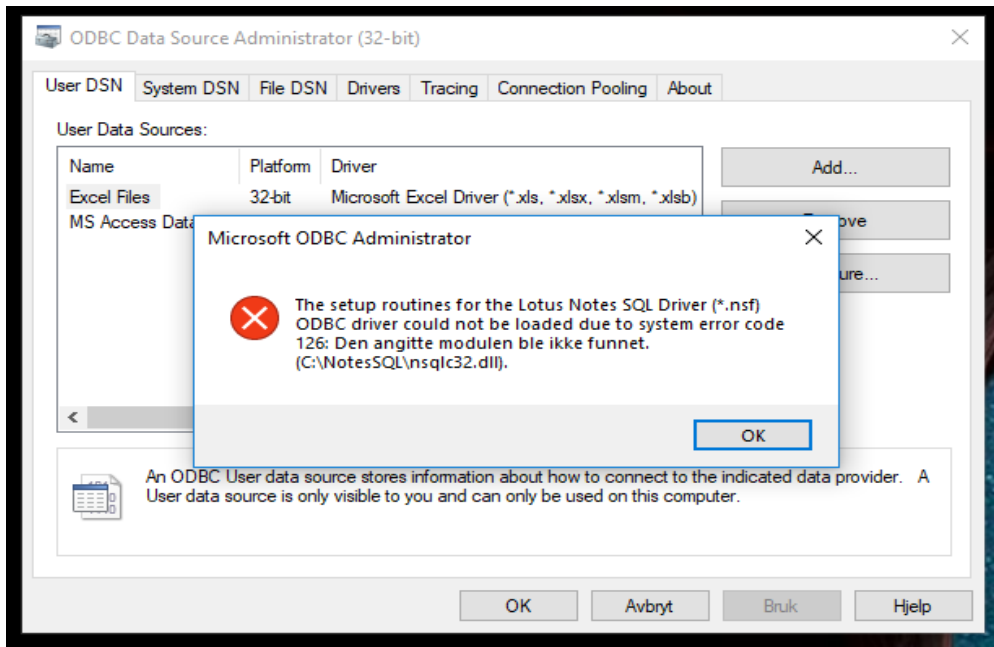
³⁸ «Rutiner ved mottak av elektroniske arkiver».

³⁹ ODBC: Open Database Connectivity, en standard for tilkobling til de aller fleste typer databaser. Tilkobling forutsetter at databasesystemprodusenten (Microsoft, IBM, etc.) har laget en ODBC-driver.

⁴⁰ <https://www.fullconvert.com/databases/siard>

utgangspunktet ønsker vi at arkivskaper/leverandør gjør uttrekksjobben, men samtidig var dette en ekstra utprøving som passet godt inn i prosjektprofilen og ga oss flere erfaringer.

Via en FTP-server fikk vi lastet ned IBMs ODBC-driver. Denne fikk vi imidlertid ikke til å fungere. Vi fikk disse feilmeldingene:



Arbarks personale klarte ikke å løse problemet. Vi sendte også feilmeldingen til Compendia, men de hadde heller ikke erfaring med dette.

Dette tilsier at det er vanskelig å få til SIARD-uttrekk fra Notes-systemer. Selv om vi er åpne for å ta imot SIARD-filer fra relasjonsdatabaser, er vi skeptiske til å benytte metoden som er beskrevet her, særlig når resultatet ble såpass dårlig. Konklusjonen er at leverandørene må lage uttrekk fra slike systemer.

11.0 Vurdering av metoder

11.1. Noark 5-uttrekk

Flere av arkivskaperne Arbark har ansvar for bruker Noark-systemer. Dette er i stor grad de store og ressurssterke organisasjonene som vi samarbeider mye med, og som har tradisjon for å avlevere til oss. At det finnes et uttrekksformat som er håndterbart, er derfor viktig for at arkivinformasjon kan bevares.

Selv om Noark 5-systemer har en uttrekksmodul, fungerer ikke den nødvendigvis hensiktsmessig uten bistand fra leverandør. I dette prosjektet har vi prøvd ut avleveringsprogrammet til Acos, med dårlige resultater. For å få vellykkede uttrekk har arkivskaperne derfor to alternativer: 1) å få leverandørene til å produsere uttrekk; 2) bruke Stiftelsen Asta og programvaren Asta Mapper.

At arkivskaper blir avhengig av leverandørens hjelp for å produsere uttrekk, og må betale for det, kan være uheldig. Gitt at det ikke er avleveringsplikt for private, ser man her en reell hindring for elektroniske avleveringer. Samtidig er det i hovedsak de store organisasjonene med mest betalingsevne som har gått til innkjøp av Noark-systemer. Disse har ofte behov for gammelt arkivmateriale, samtidig som de gjerne er bevisste betydningen av å ta vare på historisk materiale. Dermed har de stor egeninteresse av å avlevere arkiv. Vi har da også fått tydelige signaler fra flere arkivskapere om at det i løpet av få år vil bli avlevert Noark-arkiv.

At det skal skilles mellom hvem som produserer uttrekket og hvem som tester i etterkant, er i tråd med vanlig praksis. Når uttrekket er produsert bruker leverandør/arkivskaper Arkade 5 til å lage en arkivpakke. Arkivpakken overføres så til Arbark. Leverandørene som leverer Noark 5-godkjente systemer til LO-forbundene, er Acos, Evry og Tieto. Alt tyder på at disse vil ha erfaring med bruk av Arkade 5.

Ved mottak av Noark 5-uttrekk vil vi følge rutinebeskrivelsen for depotoperasjoner som ble utformet under prosjektperioden i forbindelse med uttrekket fra Norsk arbeidsmandsforbund. Selv om denne beskrivelsen er laget med utgangspunkt i en ikke-Noark-avlevering, kan den likevel fungere veiledende. Etter hvert som vi får erfaring med forskjellige typer avleveringer vil det utarbeides flere rutinebeskrivelser.

11.2. Uttrekk fra systemer uten Noark-godkjenning

Arbark ønsker i hovedsak å følge nasjonale standarder og metoder. I privat sektor er dette imidlertid ikke alltid mulig, for eksempel når arkivskapere bruker dokumentorienterte databasesystemer uten Noark-godkjenning. Leverandøruttrekk vil i slike tilfeller være en aktuell metode. Det samme gjelder for relasjonsdatabase-systemer som ikke følger Noark.

Uttrekket vi mottok fra Arbeidsmandsforbundet viser at vi kan motta ikke-Noark-baserte arkivuttrekk, også i tilfeller hvor dokumentene ikke er konvertert til arkivformat. Det er et håndterbart antall organisasjoner som i dag har systemer uten Noark-godkjenning. Det vil si at for hver avlevering kan vi gå i dialog med forbund og leverandør og høre hva leverandøren kan «tilby». Her handler det om å sikre at avleveringen i størst mulig grad tilfredsstiller kravene våre. Særlig viktig er det å få med metadata fra systemet.

Vi ser at mottak av digitale privatarkiver ikke vil følge en like strømlinjeformet modell som for offentlige arkiver. Ikke-Noark-uttrekk må medføre en viss grad av fleksibilitet fra vår side.

I slike prosesser er det i alle tilfeller nødvendig med tett kontakt både med leverandør og arkivskaper. Derfor er det helt klart en fordel at Arbark samarbeider jevnlig og relativt tett med arkivpersonalet i ulike fagforbund.

11.3. SIARD-uttrekk

SIARD-uttrekksformatet er per i dag noe krevende å håndtere for Arbark. Men andre depotinstitusjoner får stadig mer erfaring med formatet, og KDRS er i ferd med å utvikle en løype hvor SIARD har en sentral rolle. Derfor er det grunn til å tro at SIARD vil utvikle seg framover, også med tanke på innsyn.

Det skal også sies at selv om det er mange likheter, er det også noen forskjeller mellom Arbarks oppgaver og de utfordringene IKAene møter på. IKAene må eksempelvis redde arkivinformasjon fra fagsystemer som holder på å dø ute i kommunene. Arbark er generelt sett ikke interessert å ta imot fra fagforbundenes medlemssystemer, og fagforbundene ønsker som oftest heller ikke å avlevere dette. For Arbark sin del er database-uttrekk aktuelt fra *aktive* saksbehandlingssystemer som er i bruk i relativt store organisasjoner.

Derfor vil Arbark foretrekke at leverandør produserer ferdige uttrekk, framfor bruk av SIARD. SIARD kan likevel være aktuelt, for eksempel for å sikre at all informasjon kommer med i et Noark 5-uttrekk. SIARD som uttrekksmetode ble særlig aktuelt etter publisering av

en artikkel i Tidsskriftet Arkiv i 2016. Funnene i artikkelen er bekymringsfulle: Ved å sammenligne to forskjellige uttrekk fra samme Noark-system, viser forfatterne hvordan det i dette tilfellet manglet tusenvis av dokumenter.⁴¹ SIARD henter alt fra produksjonssystemet og forhindrer slik mulige dokumenttap.

Samtidig er SIARD interessant å følge med på siden forbundenes systemflora er i stadig utvikling. I tilfeller hvor det er aktuelt med SIARD-uttrekk kommer Arbark til å foreta uttrekket selv ved bruk av Full Convert Spectral Core.

Programvaren Decom har som vi tidligere har vært inne på, en sentral rolle i produksjonsløypa til KDRS. En eventuell investering her vil være helt avhengig av hvor mye bruk vi faktisk vil få for programmet, og det er fremdeles usikkert. Selv om Arbark foreløpig ikke har planer om å gå til innkjøp av Decom, følger vi likevel med på utviklingen.

⁴¹ Høiaas, Hansen Rørås, Sødning 2016.

12.0 Generelt om krav

Som depotinstitusjon er Arbark nødt til å stille visse krav til hva vi kan imot. Et minstekrav er at vi får tilstrekkelig data og metadata til at vi kan lage en selvbeskrivende AIP (altså en arkivpakke hvor man ikke må lete etter informasjon andre steder for å forstå materialet). Men utover det må vi samtidig være realistiske og akseptere at mer spesifikke krav ikke alltid blir fulgt. I privat sektor er det ingen automatikk i dette. Samtidig vil avvik medføre en del ekstra arbeid for Arbarks arkivarer, med å få avvikende uttrekk opp til et visst nivå. På et tidspunkt må Arbark vurdere om vi skal ta betalt for denne jobben. Dette vil innebære at vi tar betalt for «avstanden» det er mellom våre krav og den aktuelle avleveringen.

Ved avlevering er det en fordel om sjekksummer følger med i uttrekket. Det er åpenbart også best om filene er konvertert til arkivformat. Vi har vurdert om dette er krav vi bør stille til arkivskaper og leverandør ved avlevering. Etter vår oppfatning er det ikke hensiktsmessig å stille slike absolutte krav, siden vi ønsker å gjøre terskelen for avlevering så lav som mulig. I tilfeller hvor sjekksum og konvertering ikke er del av uttrekket, kan Arbark gjøre disse operasjonene selv. Vi vil likevel selvfølgelig påpeke at det er ønskelig med slike eksportfunksjoner i vår kontakt med leverandør og arkivskaper.

13.0 Datasikkerhet

For å kunne ivareta funksjonen som arkivdepot for digitalt materiale, må vi ta særskilte hensyn rundt beskyttelsen av de digitale dataene vi samler inn og oppbevarer.

Dette gjelder alle prosessene som berører det digitale materialet, både ved innsamling, bearbeiding, oppbevaring, vedlikehold og tilgjengeliggjøring.

I det følgende tegner vi et bilde av de truslene og uønskede hendelsene vi per i dag kan komme ut for og følgelig må ta forholdsregler mot, og vi spesifiserer tekniske tiltak som bør iverksettes for å sikre depotet best mulig mot disse hendelsene.

13.1 Digitalt materiale vs. fysisk materiale

Fysisk og digitalt materiale har noen grunnleggende forskjeller som har betydning for både hvilke risikoer de er utsatt for, og hvilke tiltak som er mulige for beskyttelse og langtidsbevaring:

Duplisering av fysisk materiale for lagring på flere lokasjoner er både tid- og plasskrevende, og i noen tilfeller umulig – 100% identiske kopier kan ikke lages, men kopier som bevarer *informasjonen* tilnærmet identisk er mulig å lage. Duplisering av digitalt materiale over flere lokasjoner kan på sin side være både tidkrevende og kreve høy lagringskapasitet, men er enkelt å sette opp og gjennomføre. Dette gjør at risiko for tap av materiale ved hendelser som vannlekkasje/oversvømmelse, jordskjelv, brann eller annet er mindre for digitalt materiale enn for fysisk materiale.

Uautorisert uthenting/fjerning/kopiering av fysisk materiale er en langsom prosess, som kan gi synlige indikasjoner på aktivitetene. Tilsvarende for digitalt materiale er en «usynlig» prosess, der datamengdene kopieres i stedet for å flyttes, og der dataoverføring kan settes opp slik at den blir så og si usynlig for sanntidsovervåkning.⁴²

Fysisk materiale vil brytes ned over tid. Digitalt materiale vil også nedbrytes over tid, i den forstand at de fysiske mediene det digitale materialet lagres på vil forringes. Men innholdet vil kunne dupliseres med jevne intervaller uten at kvaliteten svekkes, og slik motvirkes nedbrytingen. Flere kopier av digitalt materiale vil også kunne oppbevares på separate medier og lokasjoner, benyttes til å utføre en automatisk gjensidig kontroll av innholdet

⁴² Dette kan til en viss grad unngås ved utvidet overvåkning og analyse av nettverkstrafikk.

(sjekksumberegning og –verifisering), og varsle om og korrigere endringer som oppstår, slik at materialet bevares og forholdene som forårsaker endringene kan korrigeres.

Digitalt materiale består egentlig av to deler: Selve dataene/innholdet, og programvaren som viser innholdet på en egnet leser. Det kan være rene dokumenter som lagres som enkeltstående filer, men det kan også være større databaser og/eller datamengder som krever spesielle applikasjoner (og/eller spesielle versjoner av disse) for å kunne vises i sin opprinnelige tilstand. Langtidsbevaring av digitalt materiale innebærer enten å bevare en instans av (programvare-)plattformen for visning av innholdet, eller å sørge for regelmessig konvertering av det lagrede materialet når nødvendig, slik at materialet kan vises på de til enhver tid tilgjengelige fremvisningsapplikasjonene. Det innebærer også å lagre informasjon om dataformatene og datastrukturene sammen med selve dataene, slik at et eventuelt visningsprogram kan gjenskapes hvis ingen andre muligheter finnes.⁴³

Endringer i fysisk materiale vil som oftest kunne spores, mens endringer i digitalt materiale i utgangspunktet ikke vil kunne påvises. Dette kan avhjelpes med sjekksumberegninger og nøyaktig loggføring av alle endringer som skjer med materialet.

13.2 Mulige situasjoner som kan forårsake skade på eller tap av digitalt materiale

I hovedsak kan vi gruppere uønskede hendelser i disse hovedkategoriene:

1) Tap av materiale; 2) Uønsket endring av materiale (herunder også kryptovirus); 3) Uønsket tilgang til, og/eller spredning av materiale; og 4) Tap av informasjon om formater og/eller metadata slik at materialet blir «dødt».

13.3 Løsninger og konkrete tiltak hos Arbark for å oppnå ønsket grad av sikring og beskyttelse

Hvordan kan man så beskytte seg mot uønskede hendelser? Vi regner følgende tiltak som viktige:

⁴³ Her er det forskjell på strukturert og ustrukturert materiale: for fil-/mappearkiver er det selve enkeltfilenes dataformat som er viktig, mens for sak-/arkivsystemer må også databasestruktur og den logiske strukturen i applikasjonen bevares.

13.3.1 Oppsett av brannmur

En brannmur har flere oppgaver, hvorav de viktigste er:

- Kontrollere trafikk inn og ut av nettverket, og stoppe alle uønskede tilkoblingsforsøk
- Analysere nettverkstrafikk, og stoppe trafikk som ser mistenkelig ut eller som ikke følger protokollene
- Tilby VPN-tilkoblinger (virtuelle, private nettverk) mellom lokasjoner, og for brukere på hjemmekontor/ute av huset

Arbark har infrastruktur på to lokasjoner per i dag, og benytter brannmurer fra SonicWall både for individuell beskyttelse på lokasjonene, og for å knytte lokasjonene sammen i et VPN-nett. Innenfor vår nåværende situasjon og planlagte endringer er disse tilstrekkelige, men av kapasitets- og sikkerhetshensyn bør disse etter hvert vurderes å skiftes.

13.3.2 Antivirusløsning

En god antivirusløsning beskytter enhetene i nettverket mot skadelig programvare, både ved å regelmessig kontrollere alle lagrede filer og ved å kontrollere alle filer hver gang de brukes (sanntidsbeskyttelse). Noen kriterier må være på plass for en tilfredsstillende virusbeskyttelse:

- Løsningen må kjøre på alle enheter der det skjer brukerpålogging og der det er mulighet for å kjøre nettleser med internett-tilgang
- Løsningen må ikke være så ressurskrevende at det i vesentlig grad går ut over ytelsen til den enkelte enheten
- Leverandøren av løsningen må sørge for regelmessige oppdateringer av antivirusmønstrene som benyttes for å gjenkjenne skadelig programvare
- Løsningen bør, hvis organisasjonen er over en viss størrelse, ha en sentral administrasjonskonsoll som kan overvåke og fjernadministrere klientene

Arbark benytter antivirus fra Trend Micro. Denne er installert på alle aktive arbeidsstasjoner.

13.3.3 Tilgangsstyring, både for vanlige brukere på administrasjonsnett, og for administratorkontoer

En del tiltak er viktige for å oppnå ønskelig grad av kontroll med hvem som har tilgang til hvilket materiale:

- Brukerkontoer gis så begrenset tilgang som mulig innenfor de nødvendige rammene gitt av den enkelte brukers oppgaver
- Rettigheter på lokal PC begrenses (her vil det i praksis i et Windows-miljø være nødvendig at det åpnes for at bruker kan tildeles midlertidige forhøyede rettigheter for å utføre enkelte vedlikeholdsoppgaver og/eller tilpasninger på egen PC)
- Rettigheter til delte ressurser (filområder) tildeles grupper, og personer meldes inn og ut av de respektive gruppene for å få tilgang til nødvendige ressurser eller applikasjoner. Disse gruppene kan defineres etter avdeling, funksjon, lokasjon, eller etter andre kriterier. Ved å benytte grupper i tilgangsstyringen unngår man «lekkasjer» der tilgangsstyringen blir uoversiktlig fordi enkeltpersoner har blitt gitt tilganger til spesifikke områder/mapper i lagringshierarkiet
- For sensitivt materiale kan logging av tilganger vurderes. Filserveren vil da logge forespørsler om tilgang til de aktuelle delte områdene, og denne loggen kan analyseres i etterkant både for å kartlegge bruk av tilganger, og tilganger som er nektet

Arbark regulerer dette ganske bra per i dag, men det bør jevnlig tas en gjennomgang og revisjon av brukertilganger, grupped medlemsskap, og rettighetstildelinger på delte ressurser.

13.3.4 Sikring av data ved innsamling av arkiv

Ved innsamling av digitale arkiver er det viktig å sørge for så stor grad av isolasjon som mulig, for å forhindre at eventuell skadelig programvare «smitter» enten fra våre systemer/nettverk og ut til arkivgivere, eller fra deres systemer/nettverk og inn til oss. Det er også viktig allerede i innsamlingsfasen å gjøre nødvendige tiltak for å sikre autentisiteten til det innsamlede materialet, og følge opp disse tiltakene i alle ledd inntil arkivdataene er ferdig bearbeidet, pakket og lagret i digitalt depot.

Sikringstiltak for disse prosessene kan være:

- Bruk av egen aksesjons-PC, som renses (re-images) hver gang før bruk, slik at det ikke overføres skadelig programvare mellom arkiv og arkivskapers nett eller systemer
- Bruk av egen ekstern harddisk med kryptering for lagring av arkivmaterialet i transitt. Om denne skulle komme på avveie vil ingen kunne få tilgang til materialet
- Etablering av karantene for innsamlet materiale, der materialet oppbevares i en periode etter innsamling (tre uker er anbefalt), slik at eventuelle nulldagssårbarheter kan oppdages og skadelige filer fjernes før materialet frigis til arkivnettet
- Etablering av sluse-PC som foretar viruskanning av innsamlet materiale, og verifisering av metadata, før materialet overføres til karantene
- Utarbeiding av sjekksummer for hver enkelt innsamlet fil og for hele innsamlingspakken allerede på innsamlingstidspunktet, og verifisering av disse sjekksummene i alle ledd

Ettersom vi sannsynligvis må ut og hente digitale arkiver er særlig dedikert aksesjons-PC og sluse-PC viktige tiltak. Aksesjons-PCen må ha en kryptert ekstern harddisk som brukes ved innhenting. Maskinen bør være tilstrekkelig kraftig, slik at den har kapasitet til operasjonene som skal utføres under innsamling (spesielt sjekksummergenerering og kryptering). Sluse-PCen vil stå som en port mellom administrasjonsnettet og arkivnettet for overføring av innhentet materiale.

Det må også bygges opp scripts og funksjoner for å understøtte oppgavene til disse to maskinene, og rutiner for bruk av disse.

13.3.5 Lagring av flere duplikater og generasjoner av materialet, og sikkerhetskopiering til flere lokasjoner og i flere generasjoner

Der det er teknisk mulig, vurderer vi å sette opp lagringsløsninger som kjører automatisk duplisering av innhold mellom flere enheter/lokasjoner. En lagringsløsning kan også gjerne ha innbygd versjonskontroll og/eller snapshot-funksjonalitet, slik at endrede filer ikke lagres over gamle filer, men lagres som en nyere versjon i tillegg til gammel fil, eller at det tas periodiske

snapshots av filsystemet, slik at man kan hente tilbake en tidligere versjon av en fil dersom det skulle være nødvendig.

«3-2-1-regelen» er en gylden regel innenfor sikkerhetskopiering: Minst tre kopier, minst to forskjellige fysiske medier, og minst en kopi på en annen lokasjon. Den kan med fordel utvides til «3-2-1-offline» - der man i tillegg til ovenstående alltid har minst en kopi som ikke er elektronisk tilkoblet systemet.

Ved oppretting av digitalt depot og back-up av dette, vil datamengdene som skal ivaretas etter hvert bli ganske store. Behovet for trygghet og stabilitet er derfor stort. Maskinvaren som løsningen bygges på må vurderes nøye, og vurderingen må basere seg på flere kriterier, der pris ikke nødvendigvis er det viktigste. Siden vi jobber i et langtidsperspektiv må levetiden til både back-up-løsning og medier brukt til back-up overveies, slik at vi ved intervaller som er kortere enn offline-mediens forventede levetid kan revidere løsningen og erstatte mediene.⁴⁴

Arbark benytter i dag servere og lagringsenheter fra Hewlett Packard, og virtualiseringsløsningen VMWare som underliggende fysisk plattform. Dette gir en stabil og fleksibel plattform for oppbygging av infrastrukturen. Alle servere kjører som virtuelle maskiner i VMWare, og back-up foretas av systemet Veeam back-up, som bruker funksjonalitet i VMWares løsning til å ta back-up av serverne uten at de berøres av selve back-up-prosessen.

Med utgangspunkt i dette systemet vil en 3-2-1-offline-back-up-strategi kunne bestå av:

- Selve de aktive dataene som ligger i produksjonssystemet (1. kopi)
- En online back-up til diskbasert løsning som kjøres med hyppige intervaller til magasinet i Folkets Hus, med flere generasjoner back-up (2. kopi)
- En online back-up til et separat disksystem i Arbarks fjernmagasin som tar vare på flere generasjoner back-up (3. kopi)
- En tapebasert back-up som kjøres med sjeldnere intervaller, der tapene fysisk fjernes fra tapestasjon ved fullført back-up og eventuelt oppbevares på en tredje lokasjon (offline-kopi)

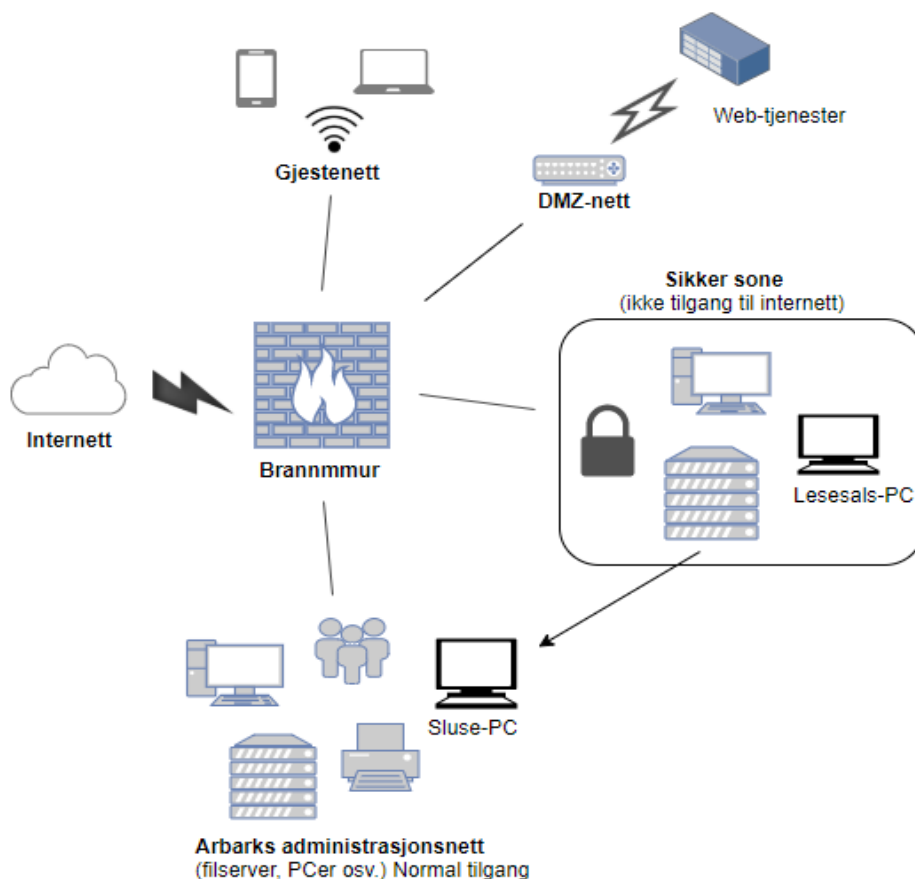
⁴⁴ Skybaserte løsninger kan være en mulighet, men det finnes ikke så mange aktuelle alternativer. I tillegg kan dette bli en kostbar løsning når datamengdene etter hvert øker opp mot og over tresifret antall terabytes.

Denne løsningen kan settes opp i Veeam back-up, men krever innkjøp av en ekstra lagringsenhet for plassering i magasin i Folkets Hus, og et tapebasert lagringssystem for offline-back-up.

13.3.6 Inndeling av nettverket i separate soner med tilgangsstyring

Lagringssikkerhet er en forutsetning for at arkivinformatjon holdes intakt. Som depotinstitusjon må Arbak minimere risikoen for at noe uforutsett skal kunne skje med de digitale arkivene. Et naturlig tiltak for å øke sikkerheten vil være å dele systemet i flere logisk separerte nettverk etter funksjon.

En sannsynlig inndeling for Arbak ser slik ut:



- Administrasjonsnett => Her eksisterer alle «vanlige» systemer og alle ansattes arbeidsstasjoner. Dette nettet har normal tilgang til internett, og brukere har en viss tilgang til å gjøre endringer i egen plattform, innenfor gitte grenser.

- Arkivnett (sikker sone) => Her lagres og bearbeides alt digitalt materiale som mottas. Digitalt magasin ligger altså på arkivnettet. Enheter på dette nettet har ikke tilgang til internett, og heller ikke tilgang til ressurser på andre nettverk, og det er satt opp særskilt streng tilgangsstyring og trafikkovervåkning. Brukere som skal ha tilgang hit blir tildelt egne brukerkontoer separat fra kontoer på administrasjonsnett. Det settes opp en egen «nedlåst» lesesalsterminal der besøkende til bibliotek/lesesal kan gis tilgang til digitalt materiale på forespørsel, uten mulighet til å kopiere ut fra terminalen.
- Gjestenett => Her kobles «usikre» enheter til, slik som mobiltelefoner, PCer fra besøkende til lesesalen, etc. Dette har ingen tilgang til andre nettverk internt.
- DMZ-nett (Demilitarized Zone) => Nettverk som benyttes for maskiner som skal tilby tjenester ut mot internett og samtidig tjenester inn mot administrasjonsnett/andre nett.
- Testnett => Et separat nett som benyttes til testing og klargjøring av utstyr, og som er isolert fra de andre nettene, men der tilganger kan justeres ved behov.

Ved å gjennomføre en slik inndeling vil man få isolert brukernes arbeidsområder og lagring av administrative data fra arkivdata og digitalt depot. Ved å begrense tilgangen mellom de ulike nettverkene, slik at det ikke er tilgang fra arkivnettet mot internett, ikke tilgang fra de andre nettene mot arkivnettet, ikke tilgang fra gjestenett mot noen interne ressurser, og så videre, minsker man risikoen for uønskede hendelser betraktelig. Ved et eventuelt datainnbrudd eller passordlekkasje vil en inntrenger kunne nå administrasjonsnettet, men ikke automatisk komme videre til arkivnettet.

Ved å aktivere loggføring på detaljnivå på delte ressurser på arkivnettet, kombinert med automatisk sjekksverifisering av filer, og eventuelt med versjonering av alle lagrede ressurser, vil man både kunne spore hvilke dokumenter som har blitt åpnet, hvem som har åpnet dem, og hvilke endringer som eventuelt er gjort.

Andre tiltak som bør tas stilling til er:

- Fysisk sikring av lokalene der systemene befinner seg. Kun nødvendig personale bør ha fysisk tilgang til serverpark og nettverksinfrastruktur, inkludert back-up-lagre

- Loggføring av all filtilgang på lagringsplattformen

Arbark har per i dag kapasitet til å dekke det nødvendige behovet.

13.3.7 Systemovervåkning

Ved å sørge for automatisk overvåkning og varsling av alle forhold som kan påvirke driftsmiljøet blir man i stand til å håndtere små driftsavvik før de får utviklet seg til å ramme flere deler av systemet, eller før de får vokst seg større.

Det finnes tilgjengelige løsninger for dette i dag. Flere av disse er open-source eller community-utgaver som kan lastes ned og installeres uten kostnad, og som gir mer enn tilfredsstillende funksjonalitet og pålitelighet.

Andre tiltak er sentralisert logginnsamling og –analyse, og rutinemessig penetrasjonstesting for å avdekke eventuelle svakheter ved oppsettet, som kan muliggjøre uautorisert tilgang utenfra.

Dette vil ikke kreve noen investeringer.

13.3.8 Opplæring av ansatte i grunnleggende datasikkerhet, og bevisstgjøring om potensielle trusler

De fleste vellykkede dataangrep i dag er rettet mot de ansatte, og i hovedsak via e-post. Ved å sørge for tilstrekkelig bevisstgjøring kan man redusere risikoen for at slike angrep lykkes til en brøkdel.

Hvis man i tillegg etablerer rutiner der den enkelte ansatte vet hvordan oppståtte situasjoner skal håndteres, vil man kunne redusere konsekvensene ved et eventuelt vellykket angrep.

Følgende temaer kan det lages enkle opplæringsjesjoner rundt:

- Epost-svindel
 - «CEO-svindel»
 - Ondsinnede vedlegg
 - Phishing
 - «Trusselmail»

- «Social Engineering»
- Virus, malware, løsepengevirus
- Passordsikkerhet
- Datalagring/back-up av egne arbeidsområder og filer
- GDPR og tankegang rundt personvern
- Datasikkerhet på sosiale media
- Generelt nettvett

13.3.9 Gode og gjennomarbeidede rutiner for bruk av arkivmateriale, tildeling og fjerning av tilganger og avvikshåndtering

Det finnes allerede rutiner for innsamling, bearbeiding, oppbevaring og tilgjengeliggjøring av papirbasert arkivmateriale. Prinsippene herfra kan også brukes for digitalt materiale, men i tillegg må systemer og rutiner tilpasses «flyktigheten» i materialet, og vår moderne «alltid på nett»-verden.

Det må også på plass administrasjonsrutiner for håndtering av brukerkontoer og tilganger, oppdagelse og håndtering av avvik, inklusive rapportering.

13.3.10 Risikovurderinger og kriseplaner

Det bør også utarbeides risikovurderinger og kriseplaner både for administrasjonsnett og for arkivnett.

En risikovurdering skal inneholde en oversikt over de forskjellige typene hendelser som kan oppstå, vurdering av sannsynligheten for at den enkelte hendelsen skal oppstå, og vurdering av konsekvenser dersom hendelsen inntreffer.

En kriseplan skal inneholde tiltak for de enkelte hendelsene i risikovurderingen, inkludert rutiner for varsling og eskalering, dersom det er nødvendig.

Her finnes det materiale og *best practices* vi kan lene oss mot for å få drahjelp til å utarbeide vårt eget rutinebibliotek.

13.3.11 Revisjonsplan

Alle løsninger, rutiner og planer må kvalitetssikres og oppdateres med jevne mellomrom.

Vi må lage en revisjonsplan som inkluderer alle elementene som kan kreve endringer og oppdateringer, inklusive revisjonsplanen selv, og rutiner for å gjennomføre, kvalitetssikre og godkjenne internrevisjonen.

14.0 Datasikkerhet og arkivinnhenting

Ved høsting av el-arkiv vil Arbark i mange tilfeller dra ut til arkivskaper og foreta uttrekk. Dette gjelder først og fremst for fil- og mappearkiver, men også SIARD-uttrekk kan bli aktuelt. SIARD-uttrekk vil gjøres ved hjelp av Full Convert Spectral Core, som vi har lisens på.

Når vi skal ut til organisasjonene blir datasikkerhet viktig. Vi har gjort oss noen tanker om hvordan arkivinnhenting kan gjøres i praksis. I det følgende bruker vi SIARD-uttrekk som eksempel.

14.1 Alternativ 1: normal løsning

Ved SIARD-uttrekk tar vi med en av arkivarenes PCer hvor Full Convert Spectral Core er installert. Uttrekk foretas med Spectral Core, og uttrekket legges på medbrakt ekstern harddisk. Denne harddisken skal støtte hardware-kryptering: man installerer et program på maskinene disken skal tilkobles, og deretter må disken «låses opp» med et passord når den skal brukes. Dataene blir dermed krypterte og passordbeskyttet, selv om både PC/ekstern disk skulle komme i hendene på uvedkommende.

Når vi kommer tilbake til Arbark begynner arbeidet med depotoperasjoner (konvertering, sjekksum-generering osv.) fra arkivarens PC. Når arbeidet er utført, overføres arkivet til Arbarks digitale sikringsmagasin.

14.2 Alternativ 2: utvidet løsning

Alternativ 2 innebærer at vi bruker en egen dedikert PC for arkivhenting/uttrekk.

«Innsamlings-PCen» har installert Spectral Core, som brukes til å produsere SIARD-uttrekk, som igjen legges på kryptert harddisk. PCens tilgang til en organisasjons elektroniske arkiv må settes opp spesielt i hvert enkelt tilfelle, og maskinen vil ikke ha tilgang til noen ressurser på hverken arkivskaperorganisasjonens eller Arbarks IT-systemer med mindre denne tilgangen gis og settes opp eksplisitt. PCen har også installert antivirus- og brannmur-programvare, og «rens» etter hver gang den er brukt. Rensing vil i hovedsak bestå av følgende operasjoner:

- PC benyttes til innhenting av data ute hos arkivskaper

- PC skrur av når innhenting er fullført

PC tas inn på «rensestasjon» hos Arbark

- PC startes opp fra en dedikert USB-minnepinne med kloneprogram
- Et forhåndsgenerert *image* som inneholder Windows, databasedrivere, Spectral Core, antivirus og andre nødvendige komponenter lastes inn på PC
- PC startes opp fra det nyinstallerte imaget – alle eventuelle innstillinger og tilpasninger som er gjort, og alle data som eventuelt har blitt lagret lokalt på PC blir fjernet.

Ved å gjennomføre et slikt regime vil vi minimere risiko for eventuell overføring av uønskede elementer fra Arbarks IT-miljø ut til den aktuelle organisasjonens infrastruktur, og *vice versa*.

Etter at SIARD-uttrekket er hentet hos arkivskaper, overføres SIARD-filen til sikker sone/arbeidsområde hos Arbark. Derfra kan vi jobbe med uttrekket. Når endelig arkivpakke er ferdig overføres den til sikringsmagasinet.

14.3 Alternativ 3: ekstra sikker løsning

Tredje alternativ innebærer samme metode som alternativ 2, bare med en ekstra sikkerhetsmekanisme. Etter at arkivpakken er «ordnet» overføres den til sikringsmagasinet. Inn til dette området er det som vi ser av figuren under punkt 13.3.6 en sluseordning, og det vil ikke være tilgang ut. Dette minimerer risikoen for at noe uforutsett skal skje. Slusingen består i at disken kobles til en dedikert PC på administrasjonsnettet til Arbark. Denne maskinen låser opp disken, og overfører dataene til et eget overføringsområde på lokal disk. Server i sikker sone vil skanne etter data på dette området, og starte overføring så snart det finnes data her. Overførte data blir automatisk fjernet fra transittområdet.

Når data er overført til sluse-PC, kjøres et renseprogram som først fyller overføringsdisken med data slik at eventuelle innsamlede data blir overskrevet. Deretter formateres disken, slik at eventuelle skjulte elementer blir fjernet.

Av disse tre løsningene anser vi alternativ 3 som det beste.

15.0 Veiledninger

15.1. Bevaring og kassasjon

På våre besøk ute i LOs organisasjoner ble det raskt klart at det var mye usikkerhet om bevaring og kassasjon, ikke bare blant saksbehandlere, men også i arkivpersonalet. Flere etterlyste veiledning i hva Arbark ønsker å motta. Etter diskusjon i prosjektets referansegruppe ble vi enige om at Arbark skulle utarbeide en bevarings- og kassasjonsveiledning for LO-forbundene.

Arbark har lang erfaring med å motta arkiv fra LOs fagforbund, og vi vet følgelig mye om hva som befinner seg i arkivene. Samtidig har vi kunnskap om hva som ofte blir etterspurt, både fra historikere og forbundene selv. Dermed kan vi gi gode råd om hva som bør bevares for ettertiden.

På grunn av fagforbundenes organisasjonsmessige likhet, og fordi deres hovedfunksjon er den samme,⁴⁵ er dokumentasjonen som skapes i ulike forbund relativt likeartet: utover materiale fra styrende organer vil et forbundsarkiv typisk inneholde saker som er ført på vegne av medlemmer, samt opprettelser av lokale tariffavtaler og krav overfor arbeidsgiverorganisasjonene i tariffoppgjørene.

Selv om det som produseres i en organisasjon som oftest har verdi på kort sikt, betyr ikke det nødvendigvis at alt er like interessant i et historisk perspektiv. Veiledningen fokuserer på materiale som vi regner med vil ha interesse for ettertiden, og vi luker for eksempel vekk materiale som jevnlig mottas fra andre organisasjoner og som ikke resulterer i saksbehandling. På den annen side ønsker Arbark i størst mulig grad å bevare arkiver slik de ble skapt i samtiden, og det går en grense for hvor detaljerte vi både vil og kan være i våre anbefalinger. En bevarings- og kassasjonsveiledning som skal gjelde mange arkivskapere må nødvendigvis bli av generell art. Derfor må det alltid foretas en bevaringsvurdering i forkant av en avlevering.

Kassasjonsveiledningen vil bli lagt ut på Arbarks nettsider, og den er vedlagt denne rapporten («Bevaring og kassasjonsveiledning»).

⁴⁵ Innholdsmessig vil formålsparagrafen i vedtektene til to ulike fagforbund i stor grad samsvare.

15.2. Personvern

I fagforbundenes sak-/arkivsystemer finnes det naturligvis personopplysninger. Spørsmålet om hvordan man skal håndtere personopplysninger ble aktuelt med innføringen av EUs personvernforordning – *General Data Protection Regulation* – i juli 2018. Forordningen angir «særlige kategorier av personopplysninger» - informasjon av sensitiv karakter, for eksempel knyttet til helse, religiøs overbevisning og lignende. Disse særlige kategoriene er spesielt viktige for oss, da forbundenes medlemssaker (særlig oppsigelser eller yrkesskader) kan inneholde dokumentasjon som omfattes av disse kategoriene.

Fram til nå har fagforbundene avlevert på papir. Ulike forbund har hatt ulike avleveringspraksis når det gjelder personsensitivt materiale. Noen avleverer alle typer materiale på en gang og overlater håndteringen til oss, mens enkelte avleverer sensitivt materiale atskilt fra det regulære saksarkivet. Når vi ordner dette materialet blir det en egen arkivserie. Andre igjen ønsker ikke å avlevere personsensitivt materiale overhodet.

Ved avlevering inngår Arbark en avtale med arkivskaper om bruk av arkivet. Siden Arbark er en privat depotinstitusjon må *alle* som ønsker tilgang til et av våre arkiver søke om innsyn. Arbark har en standardavtale hvor det står at det er vi som avgjør innsynsforespørsler og bruk av materialet. Dette er det ideelle for oss, og også vanlig praksis. Men noen arkivskapere ønsker mer kontroll, slik at brukere må søke arkivskaper om innsyn. Både hva som faktisk blir avlevert og grad av klausulering kan altså variere fra arkivskaper til arkivskaper. Disse varierende praksisene vil sannsynligvis fortsette.

I forkant av innføringen av personvernforordningen sommeren 2018 fikk Arbark mange forespørsler fra fagforbundene om hva forordningen ville bety i arkivsammenheng. Flere vurderte å begynne med sletting, noe vi klart frarådet. Forbundsarkivarene i prosjektets referansegruppe var også veldig opptatt av dette spørsmålet. I juni 2018 laget vi derfor en veiledende uttalelse som ble lagt ut på Arbark sine nettsider. I denne tar vi for oss det vi mener er forordningens mest sentrale punkter når det gjelder arkivfunksjonen. Uttalelsen er vedlagt projektrapporten.

Arbark er også representert i Samarbeidsforumet for personvern i LO-forbundene. Forumet har representasjon fra de fleste av LOs organisasjoner, og holder jevnlig møter hvor man diskuterer ulike problemstillinger knyttet til personvernforordningen.

Det er fortsatt ting som er uklare når det gjelder implementeringen av forordningen i det norske lovverket. Bruk av materiale fra forbundenes sak-/arkivsystemer vil derfor foreløpig

kun være tilgjengelig for avleverende forbund. For mer om dette, se punkt 17.0, Tilgjengeliggjøring.

Det har blitt laget flere veiledninger i dette prosjektet, både for arkivskaperne og depot. En av disse gir veiledning i hvordan arkivskapere, dvs. LOs organisasjoner, skal gå fram ved avlevering av elektronisk arkiv. Veiledningen er vedlagt denne rapporten («Veiledning for avlevering»).

16.0 Seminarer i regi av Arbark og deling av kompetanse

I prosjektperioden har det blitt arrangert to seminarer knyttet til sentrale problemstillinger. Det første hadde periodisering og uttrekk som tema, og det andre omhandlet personvernforordningen. Begge seminarene ble holdt for personale fra LOs organisasjoner. Slik kontakt ut mot arkivskaperne er del av en strategi for å øke kompetansen om elektroniske arkiver og relaterte problemstillinger, både for oss på Arbark og for arkivfaglig personale i organisasjonene.

Det første seminaret vi arrangerte tok for seg temaet periodisering og uttrekk. Det ble arrangert i september 2017. Her inviterte vi to foredragsholdere som var knyttet til Norsk Arkivråd.

I november 2018 holdt vi et nytt seminar, denne gangen om personvernforordningen. Som tidligere nevnt fikk vi før innføringen av forordningen mange forespørsler fra LO-forbundene om hva forordningen får å si for arkivering og bevaring av informasjon. Vi inviterte derfor Janet Birkedal Martin fra Seksjon for bevaringsvurdering og privatarkiv i Arkivverket, og Herbjørn Andresen, førsteamanuensis ved OsloMet, til å snakke om temaet.

Vi fikk gode tilbakemeldinger på begge seminarene.

Vi ser på deling av kompetanse som sentralt i arbeidet med elektroniske arkiver. Arkivmiljøet er lite, og alle tjener på størst mulig grad av samarbeid. Utover egne seminarer har vi derfor også delt erfaringer fra prosjektet i forum hvor dette har vært naturlig. For eksempel la vi fram foreløpige erfaringer under et arrangement med Arbarks nordiske søsterinstitusjoner i 2017 og på Privatarkivkonferansen i 2018. I mai 2019 deltok vi også på erfaringsseminar om utviklingsmidler for privatarkiv, hvor alle som fikk støtte fra Arkivverket i 2016-2017 var invitert til å snakke.

17.0 Tilgjengeliggjøring

Tilgjengeliggjøring har vært en del prosjektet. Da vi skrev søknaden var vi usikre på hva som fantes av systemer for søk og innsyn i database-arkiver i Norge. Selv om man rundt om har laget spesialiserte innsynsløsninger for forskjellige systemer, foreligger det ikke nasjonale løsninger av typen Arkivportalen. Vi har sett nærmere på muligheter for Arbark til å tilgjengeliggjøre el-arkiv.

For Arbarks del er tilgjengeliggjøring to (litt forskjellige) ting:

- Arkivskapere har ofte bruk for avlevert materiale etter en viss tid. Slik sett er Arbark arkivskapernes forlengede arm.
- Arbark har også som oppdrag å gjøre materiale tilgjengelig for forskning.


Begge deler handler grunnleggende sett om det samme: hensikten med å bevare arkivinformatjon er at det skal kunne brukes.

LO-organisasjonenes arkiver inneholder personsensitive opplysninger, og det er selvsagt ikke aktuelt å gjøre saksarkivene åpent tilgjengelige. Men siden fagforbundene selv har behov for å bruke materialet etter at det er avlevert til Arbark, trenger vi effektive innsynsløsninger. Underveis i prosjektet ble vi oppmerksom på at KDRS har etablert en innsynsløsning for Noark 5-uttrekk: KDRS Innsyn. Løsningen er utviklet av selskapet Piql og ligger åpent tilgjengelig på github.⁴⁶ Alle som ønsker kan laste ned og benytte løsningen. Programmet har to hovedfunksjoner: «søk» og «eksporter».

Det kan gjøres søk i uttrekk som er lastet opp til programmet. Søk skjer mot innhold i filen «arkivstruktur.xml». Man kan gjøre fritekstsøk på alle nivåene (arkiv, arkivdel, mappe, dokumentbeskrivelse osv.) og metadatainnførsler (tittel, forfatter osv.). Dersom man for eksempel får treff på dokument-tittel kan man huke av «Dokumentobjekt» i arkivstrukturen, og trykke «Vis» for å se dokumentet:

⁴⁶ <https://github.com/piql/insight>

KDRS Innsyn



Søk: deponerings

Søkevalg

Vedlegg

Skill mellom små og store bokstaver

Inkluder Noder

Klasse Import

Arkiv Arkivskaper

Arkivdel Korrespondansepart

Registrering Mappe

Dokumentobjekt Klassifikasjonssystem

Dokumentbeskrivelse Avskrivning

Informasjon

Versjonsnummer 1

Variantformat Arkivformat

Format pdf

Opprettet Dato 2015-01-02T10:18:14

Opprettet Av Jonas Larsen

Referanse Dokumentfil dokumenter/Deponeringsvilkår og

Lokasjon	Søketreff
Registrering	System ID 27e6dd85-118b-43dc-a9c3-209c86e31899 Opprettet Dato 2015-01-02T10:18:14 Opprettet Av Jonas Larsen Arkivert Dato 2015-01-07T14:00:15 Arkivert Av Ole Jan Hansen Registrerings ID 2015/1-1 Tittel Forespørsel - Anker skipsverfts bedriftsarkiv - Deponeringsvilkår (klausul) og inngåelse deponerings- og ordningsavtale Offentlig Tittel Forespørsel - Anker skipsverfts bedriftsarkiv - Deponeringsvilkår (klausul) og inngåelse deponerings- og ordningsavtale Journalår 2015 Journalsekvensnummer 1 Journalpostnummer 1 Journalposttype Inngående dokument Journalstatus Arkivert Journaldato 2015-01-07 Dokumentets Dato 2017-01-02
Dokumentbeskrivelse	System ID 8baddba8-616b-4bb6-92e2-a96873b14aa2 Dokumenttype Brev Dokumentstatus Dokumentet er ferdigstilt Tittel Forespørsel - Anker skipsverfts bedriftsarkiv - Deponeringsvilkår (klausul) og inngåelse deponerings- og ordningsavtale Opprettet Dato 2015-01-02T10:18:14 Opprettet Av Jonas Larsen Tilknyttet Registrering Som Hoveddokument Dokumentnummer 1 Tilknyttet Dato 2015-01-02T10:18:14 Tilknyttet Av Jonas Larsen
Dokumentobjekt	Versjonsnummer 1 Variantformat Arkivformat Format pdf Opprettet Dato 2015-01-02T10:18:14 Opprettet Av Jonas Larsen Referanse Dokumentfil dokumenter/Deponeringsvilkår og avtale.pdf Sjekksum 8441576882881CF85B72FB670DC23F9A3BAF8F80497B46FEB6805DCDAFEA01B Sjekksum Algoritme SHA256 Filstørrelse 182987
Registrering	System ID 90d2ccb7-47c2-46e7-be16-4f93bc1b841a Opprettet Dato 2015-01-05T08:11:16 Opprettet Av Jonas Larsen Arkivert Dato 2015-01-07T14:00:15 Arkivert Av Ole Jan Hansen Registrerings ID 2015/1-1 Tittel Forespørsel - Anker skipsverfts bedriftsarkiv - Deponeringsvilkår (klausul) og inngåelse deponerings- og ordningsavtale Offentlig Tittel Forespørsel - Anker skipsverfts bedriftsarkiv - Deponeringsvilkår (klausul) og inngåelse deponerings- og ordningsavtale Journalår 2015 Journalsekvensnummer 1 Journalpostnummer 1 Journalposttype Inngående dokument Journalstatus Arkivert Journaldato 2015-01-07 Dokumentets Dato 2017-01-02

Programvaren lager også rapporter: Dersom et dokument skal sendes til en saksbehandler kan relevante nivåer i arkivstrukturen merkes og eksporteres, slik at saksbehandler får med nødvendig metadata.

Som vi ser er løsningen veldig «arkivarsentert», og den forutsetter kunnskap om Noark 5. For Arbark vil KDRS Innsyn være nyttig når fagforbundene henvender seg til oss og spør etter materiale fra Noark-systemer.

En annen mulighet, som er aktuell for alle typer uttrekk, er den enkle løsningen: når forbundene etterspør materiale kan vi gjøre søk i filutforskeren eller i selve dokumentene. I Noark-uttrekk kan man også søke i arkivstruktur.xml-filen.

Disse løsningene er først og fremst aktuelle med tanke på forbundenes egne behov for dokumentasjon. Når det gjelder tilgjengeliggjøring i forskningssammenheng har prosjektet vurdert det slik at den foreløpig beste løsningen er bruk av låst lesesals-PC i Arbarks lokaler. En slik løsning samsvarer med opplegget vi skisserer i punkt 13.3.6. I et arkivuttrekk vil en del, men ikke all informasjon være personsensitiv. På en lesesals-PC kan vi tilgjengeliggjøre etterspurt materiale ved å overføre fra vårt sikringsområde. Denne PCen må ha brukerbegrensninger og den må være nedlåst. Det vil si at den er uten nett-tilgang, at utskriftsmulighetene er låst, og at USB-portene er inaktive. Dokumentasjon som ikke er personsensitiv kan for eksempel tilgjengeliggjøres over nett.

Dette er løsningene vi foreløpig ser for oss når det gjelder tilgjengeliggjøring. Men dette området er i kontinuerlig utvikling. For eksempel initierte KDRS våren 2019 et nytt tilgjengeliggjøringsprosjekt, «Produksjonslinje – automatisert formidling av fagsystem», der Arbark deltar i prosjektgruppen. Prosjektet skal komplettere KDRS' produksjonsløype for el-arkiver og jobber med å finne måter for å effektivisere tilgjengeliggjøring av fagsystem. Målet er å utvikle en løsning som på bakgrunn av SIARD-uttrekk kan produsere en innsynsfil, altså en DIP. Dersom dette realiseres vil SIARD-uttrekk bli enda mer aktuelt for Arbark.

18.0 Avslutning

18.1 Uttreksmetodikk

Informasjonsinnhenting i prosjektet viser at 14 av LOs organisasjoner har sak-/arkivsystem. Seks arkivskapere har Noark-godkjent system, åtte bruker system uten slik godkjenning. Per i dag har LOs fagforbund relativt god kontroll på arkivene sine. De fleste holder ennå fast på papirarkivering, men på sikt regner vi med at man også i denne sektoren begynner med fullelektronisk arkivering.

Det er tre aktuelle måter å hente arkivinformasjonen ut av systemene på: 1) Dersom systemet er Noark-godkjent kan det produseres et Noark-uttrekk; 2) dersom det er et relasjonsdatabase-system kan man lage SIARD-uttrekk med Spectral Core; eller 3) systemleverandør lager uttrekk fra system uten Noark-godkjenning (dette kan være relasjonsdatabaser eller andre databasestrukturer).

Avlevering fra Noark-systemer kan gjøres ved at systemleverandør produserer uttrekk, og kjører det gjennom Arkade 5 før avlevering. Alternativt kan Stiftelsen Asta produsere uttrekket.

SIARD-uttrekk kan gjøres fra relasjonsdatabase-systemer med eller uten Noark-godkjenning. For Arbark er SIARD foreløpig kun aktuelt som tilleggsuttrekk fra Noark-systemer, for å sikre at all informasjon kommer med.

Vi foretrekker at leverandør produserer uttrekk. Dokumentorienterte databaser har vist seg å kunne transformeres til en klassisk hierarkisk arkivrepresentasjon, men denne typen enklere databaseteknologi betinger mindre etablerte metoder. Når leverandør skal lage uttrekk fra et ikke-Noark-system, må depot, arkivskaper og systemleverandør derfor samarbeide for at uttrekksformatet skal bli mest mulig tilfredsstillende.

18.2 Digitalt magasin og sikkerhet

Som del av prosjektet har Arbark oppgradert serverparken. I sammenheng med oppgraderingen ble det opprettet et digitalt depot med streng tilgangsstyring. Sikkerhet er avgjørende for at Arbark skal kunne ivareta funksjonen som digitalt arkivdepot. Vi jobber nå med å dele nettverket i flere logisk separerte soner etter funksjon: På administrasjonsnett

finner man alle vanlige systemer og de ansattes arbeidsstasjoner. Her er det normal internett-tilgang. I sikker sone, også kalt arkivnett, lagres digitalt materiale som mottas. Enheter på dette nettet har ikke tilgang til internett eller til ressurser på andre nettverk. Brukere får tildelt egne brukerkontoer separat fra kontoer på administrasjonsnett.

Ved å skille arkivnettet fra det vanlige nettverket vil man få isolert brukernes arbeidsområder og lagring av administrative data fra arkivdata og digitalt depot. Og ved å begrense tilgangen mellom de ulike nettverkene minsker man risikoen for uønskede hendelser.

18.3 Tilgjengeligjoring

I påvente av nasjonale løsninger kommer Arbark til å sette opp en innsynsløsning på lesesalen. Denne vil bestå av en «låst» lesesals-PC hvor etterspurt materiale overføres fra digitalt magasin. Når forbundene etterspør materiale fra sine Noark-systemer vil vi benytte KDRS Innsyn, eventuelt søke i mappestrukturen og xml-filene i uttrekket.

Ved starten av prosjektet så vi ikke for oss at det ville gi svar på alle de utfordringene Arbark møter i arbeidet med databasebaserte arkiver. Med prosjektet har vi likevel kommet i gang, det er bygget opp en lagringsløsning slik at vi faktisk kan ta imot denne typen elektronisk materiale, og vi har fått praktisk erfaring med å håndtere det. Men det skjer mye på dette området, og man må hele tiden holde seg oppdatert. For å sikre de elektroniske arkivene Arbark har ansvar for, og slik bidra til en mer helhetlig samfunnsdokumentasjon, vil arbeid med el-arkiver også framover være en prioritert oppgave i vår institusjon.

Programvare som det refereres til i rapporten

7-Zip: program som kan pakke og komprimere mapper og innhold.

Access to Memory: tilgjengeliggjøringsystem for digitale filer, med funksjonalitet til å lage arkivbeskrivelser. Utviklet av Artefactual Systems Inc.

Arkade 5: Arkivverkets testverktøy for arkivuttrekk.

Archivemata: system for mottak og bevaring av digitale filer. Utviklet av Artefactual Systems Inc.

Asta Mapper: verktøy/metodikk for uttrekk av elektroniske arkiver, levert av Stiftelsen Asta.

Documaster Decom: verktøy som kan beskrive SIARD-filer på tabell- og feltnivå.

DROID – Digital Record Object Identification: programvare som identifiserer filformater.

ESSArch: digitalt depotforvaltningssystem som brukes av flere aktører i arkiv-Norge, deriblant Arkivverket og KDRS.

Fixity AVP: program som regner ut sjekksummer for dokumentfiler.

Full Convert Spectral Core: lisensiert programvare som produserer SIARD-filer.

KDRS Innsyn: innsynsløsning for Noark 5-uttrekk.

Neevia Document Converter Pro: lisensiert programvare som konverterer dokumenter til arkivformat (PDF/A).

RODA: Repository of Authentic Digital Objects. System for mottak og bevaring av digitale filer.

SIARD: Software Independent Archiving of Relational Databases. Standardisert format for databasebevaring og programvare som kan lage filer i SIARD-format.

Litteratur

ABM Skrift # 43 (2007). Minnehåndtering. Metode for digital langtidslagring i kommunal sektor.

Andresen, Herbjørn (2019). Fremfinning og bruk av digitalt skapt materiale i arkivdepotene. I *Norsk Arkivforum* 25. Arkivarforeningen, Oslo.

ICTU (2002). XML and digital preservation. Digital preservation testbed white paper. Hentet 20.06.2019 fra

<https://pdfs.semanticscholar.org/048d/b1a85e5a687f5a778c8342b9168708e0c172.pdf>.

Erway, Ricky and Barrera-Gomez, Julianna (2013). Walk This Way: Detailed Steps for Transferring Born-Digital Content from Media You Can Read In-house. Dublin, Ohio, Online Computer Library Center, Inc. (OCLC).

Hamill, Lois (2017). *Archival Arrangement and Description. Analog to Digital*. Rowman & Littlefield Publishing Group, Inc.

Høiaas, P. B., Hansen Rørås, C., Sødring, T. (2016). Uttrekkssammenligning – to uttrekk fra samme system. *Tidsskriftet Arkiv*, 7. Hentet 21.06.2019 fra <https://doi.org/10.7577/ta.1670>

Marthinsen, Jørgen Helge (2012). Arkivteori – en innføring. Riksarkivaren skriftserie 36.

Riksarkivaren (2012). Vel bevart! Rapport fra samarbeidsprosjektet DIAS – Digital arkivpakkestruktur.

Riksarkivet (2010). Digitalt og autentisk. Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale. Prosjektrapport.

Riksarkivet (2010). Digitalt og autentisk. Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale. Kortversjon av prosjektrapport.

Sødring, Thomas (2017). Introduction to XML. Hentet 20.06.2019 fra

<http://edu.oslomet.no/ark2200/h17/slides/week1/Introduction%20to%20XML.odp>

Lover og forskrifter

Riksarkivarens forskrift (2017). Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver (FOR-2017-12-19-2286). Hentet

10.07.2019 fra: <https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286?q=riksarkivarens%20forskrift>

Standarder

Audit and Certification of Trustworthy Digital Repositories (TDR) (2011). CCSDS. Consultative Committee for Space Data Systems.

Reference Model for an Open Archival Information System (OAIS) (2012). CCSDS. Consultative Committee for Space Data Systems.

Riksarkivet (2018). Noark 5. Standard for elektronisk arkiv. Versjon 5.0.

Andre kilder

Espeland, Sigve (2019). «KDRS produksjonslinje». Presentasjon på KDRS-workshop, 7. mai 2019.

Evensen, M., Negaard, T. (2017). «Presentasjon og demonstrasjon av Asta Mapper». Presentasjon på Riksarkivet, 12. oktober 2017.

Stiftelsen Asta (udatert). Erfaringsrapport. Justis- og beredskapsdepartementet. Deponeringsprosjektet. Hentet 21. august 2019:

<https://www.stiftelsenasta.no/media/1351/erfaringsrapport-asta-mapper.pdf>

Riksarkivaren (2009). *Godkjente dokumentformater ved avlevering og deponering. Særlig om overgangen til PDF/A*. Hentet 25. mai 2019:

https://www.arkivverket.no/sok/_/attachment/inline/eb3f91e6-10cf-4eb6-a855-90baadcad7bc:99d66752ac201be40573cd8a5e90134f7fc178a1/Rundskriv%20om%20PDF-A%20av%2022.1.2009.pdf