Prosjektrapport

Fra grasrota til Stortinget – å dokumentere politiske partier i en digital tid

Arbeiderbevegelsens arkiv og bibliotek, Museene i Akershus og Norsk Industriarbeidermuseum

2023

Innhold

Begreper og forkortelser ... 4
1.0. Innledning ... 5
1.1. Formål og måloppnåelse ... 5
1.2. Prosjektets samarbeidspartnere ... 6
1.3. Organisering ... 6
1.4. Bakgrunn for prosjektet ... 7
1.5. De politiske partiene ... 9
2.0. De politiske partienes digitale arkiver ... 11
2.1. Situasjonen i sentralorganisasjonene ... 11
2.1.1. Organisasjon, roller og arkivrutiner ... 11
2.1.2. Arkivløsningene ... 11
2.1.3. Kommunikasjon og sosiale medier ... 12
2.1.4. Audiovisuelt materiale ... 13
2.2. Oppsummering av funn på regionalt (fylkeskommunalt) og lokalt (kommunalt) nivå ... 13
2.2.1. Organisasjon, roller og arkivrutiner ... 14
2.2.2. Arkivløsningene ... 16
2.2.3. Kommunikasjon (e-post) og sosiale medier ... 17
2.2.4. Audiovisuelt materiale ... 19
2.2.5. Fokus på sikkerhet ... 19
2.2.6. Avlevering av arkiver ... 20
2.2.7. Oppsummering regionalt og lokalt nivå ... 20
2.3. Bevaring og kassasjon ... 21
3.0. Avleveringsmetoder og periodisering ... 22
3.1. Avlevering på portabel disk ... 22
3.2. Avlevering fra skytjeneste ... 22
3.3. Utkopiering hos arkivskaper ... 23
3.4. Periodisering ... 23
4.0. Mottak av digitale arkiver ... 26
4.1. Sentrale utfordringer knyttet til langtidsbevaring ... 26
4.2. OAIS – Open Archival Information System ... 27
4.3. Andre standarder ... 30
4.4. Filformater og strategier ... 31
4.5. Særlige utfordringer med fil-/mappearkiver ... 33
4.5.1. Lange filstier ... 33
4.5.2. Bevaring av opprinnelige metadata ... 35

4.6. Beskrivelse av «manuell» mottaksrutine hos Arbark før prosjektstart 36
5.0. Archivematica og RODA 39
5.1. RODA 39
5.2. Archivematica 40
5.3. Sammenligning av systemene 42
5.4. Archivematica som mottakssystem på Arbark 42
5.5. Tilleggsoperasjoner 43
6.0. E-post 45
6.1. E-post – the good and the bad 45
6.2. E-post i partiene - situasjonsbeskrivelse 46
6.3. Bevaring av e-post – det pragmatiske valget 46
6.4. Bevaring og e-post 48
6.4.1. Capstone 48
6.4.2. Programmer 51
6.4.3. Bevaringsverdig e-post i mindre partiorganisasjoner – en anbefaling 53
7.0. Sosiale medier 54
7.1. Nedlasting fra nettleser 54
7.2. Automatisk netthøsting – *Webrecorder* 57
7.3. DigiFacket-prosjektet 58
8.0. Digitalarkivet 60
8.1. Langtidsbevaring – Archivematicas begrensninger 60
8.2. Testing av arkivopplasting til Digitalarkivet 61
8.3. Opplasting av reelle data til bevaring i Digitalarkivet 62
9.0. Avslutning 64
9.1. Vilje til avlevering 64
9.2. Metoder for mottak 64
9.3. Bevaring og kassasjon 65
9.4. E-post 65
9.5. Sosiale medier 66
9.6. Samarbeid og framdrift 66
9.7. Formidling 67
Litteratur 68

Begreper og forkortelser

AIP: *Archival Information Package*. Brukes i OAIS-modellen om digitale arkivpakker som skal langtidsbevares. Pakkene inneholder både filer og nødvendige metadata. Utgangspunktet for en AIP er SIP, se under.

BagIt: Pakkeformat for digitale bevaringspakker.

Capstone: Metodikk for bevaring av e-post hvor nøkkelroller i en gitt organisasjon prioriteres.

DIAS: *Digital arkivpakkestruktur*. Arkivverkets pakkeformat for digitale bevaringspakker.

DIP: *Dissemination Information Package*. Brukes i OAIS om visningspakken som gjøres tilgjengelig for arkivbrukere. DIP er basert på AIP.

EAC-CPF: *Encoded Archival Context – Corporate bodies, Persons and Families*. Standard for beskrivelse av arkivskapere og andre aktører i OAIS-baserte arkivpakker.

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

E-ARK: Pakkeformat for digitale bevaringspakker.

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

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

METS: *Metadata Encoding and Transmission Standard*. Standard for beskrivende, administrativ og strukturell metadata vedrørende digitale objekter.

OAIS: *Open Archival Information System*. Modell med prinsipper for bevaring av digitale objekter.

PREMIS: *Preservation Metadata: Implementation Strategies*. Metadatastandard for langtidsbevaring av digitale objekter.

SIP: *Submission Information Package*. Brukes i OAIS om den digitale aksesjonen depotet mottar fra arkivskaper. SIP er utgangspunktet for AIP og DIP.

1.0. Innledning

1.1. Formål og måloppnåelse

Høsten 2019 søkte Arbeiderbevegelsens arkiv og bibliotek (Arbark), Museene i Akershus (MiA) og Norsk Industriarbeidermuseum (NIA) om støtte fra Arkivverkets utviklingsmidler på arkivfeltet til å gjennomføre det treårige prosjektet «Fra grasrota til Stortinget – å dokumentere politiske partier i en digital tid». Prosjektet tok utgangspunkt i de politiske partiene som medlemsbaserte politiske organisasjoner. Partivirksomheten på sentralt, regionalt og lokalt plan resulterer i en omfattende dokumentasjon som i all hovedsak er digitalt skapt. Både ut ifra et perspektiv om en samlet samfunnsdokumentasjon og som en dokumentasjon på de demokratiske prosessene i Norge, har det avgjørende betydning at partienes funksjon som grasrotpolitiske verksteder og sentrale politikkutformende aktører blir dokumentert og sikret for ettertiden. Prosjektet var designet for å finne fram til en generell ramme for høsting, sikring og tilgjengeliggjøring av digital dokumentasjon fra medlemsbaserte politiske organisasjoner. Videre hadde prosjektet følgende delmål: 1) å utrede utvalgte partiorganisasjoners digitalt baserte virksomhet og tilhørende elektroniske arkivdanning, 2) foreta bevaringsvurderinger og en prioritert utvelgelse av dokumentasjon for arkivering, 3) vurdere konkrete handlingsmåter for å kunne høste og sikre prioritert digital dokumentasjon, og 4) vurdere metodikk for tilgjengeliggjøring av høstet digital informasjon.

Prosjektet fikk innvilget støtte høsten 2019 og ble igangsatt januar 2020. Det ble sendt nye søknader i 2020 og 2021 som også ble innvilget. Prosjektet ble sluttført i desember 2022. Denne rapporten og medfølgende vedlegg beskriver løsningene prosjektet har kommet fram til.

På tross av koronasituasjonen som oppstod i mars 2020, klarte prosjektet å kartlegge og utrede partiorganisasjonenes digitale arkivdanning innenfor gitt tidsramme. Koronanedstengningen var en utfordring tidlig i prosjektet, men med Teams som erstatning for fysiske møter var det mulig å gjennomføre kartleggingen. Bevaringsvurderinger, dvs. skriftlige bevarings- og kassasjonsveiledninger, ble laget etter at kartleggingen var fullført. Bevaringsvurderingene ligger som vedlegg til denne rapporten. Flere konkrete handlingsmåter for høsting og sikring har blitt vurdert og testet av prosjektets arbeidsgruppe. Det eneste målet som ikke er nådd er delmål 4. Tilgjengeliggjøringsløsningen som prosjektets deltakere vil benytte – Digitalarkivet

– er ikke ferdig utviklet. Det var derfor ikke mulig å prøve denne ut og gjøre vurderinger innenfor prosjektperioden.

1.2. Prosjektets samarbeidspartnere

Arbark er Norges største private arkivinstitusjon, med hovedoppgave å bevare arbeiderbevegelsens kulturarv og formidle den videre til nye generasjoner. Institusjonens sentrale bevaringsområde er fagbevegelsen og de politiske partiene på venstresiden. Arbark har de senere årene fått tildelt støtte fra Arkivverkets utviklingsmidler til flere prosjekter.

MiA er Vikens største kulturinstitusjon og har forvaltningsansvar for 20 museer, 30 besøkssteder og har flere kompetansesentre. I tillegg eier MiA tre samlinger. MiA er fotobevaringsinstitusjon og forvalter rollen som fylkeskoordinerende ledd for privatarkiv i Akershus.

NIA er et konsolidert museum i østre del av Telemark med sju besøkssteder og nesten 100 000 besøkende årlig. I tillegg til besøksarenaene har NIA betydelig virksomhet knyttet til forvaltning av mer enn 110 kulturhistoriske bygg, to store fartøy, flere kaianlegg, en stor kunstsamling, arkiver, foto og samlinger innenfor våre fagfelt. NIA er ett av tre museer i Telemark som til sammen utgjør Telemarkarkivet, et faglig nettverk for dem som arbeider med privatarkiv i museene i fylket. NIA har dessuten en avtale om drift av Arbeiderbevegelsens arkiv i Telemark (AAT). Avtalen innebærer at NIAs arkivar bevarer og tilgjengeliggjør AATs arkiver for medlemmer og brukere. AAT er en selvstendig institusjon med et valgt styre, der medlemmene i form av partilag og fagforeninger innenfor arbeiderbevegelsen selv eier arkivene.

1.3. Organisering

Prosjektansvarlig har vært daglig leder ved Arbark, Ole Martin Rønning. Arbeidet har vært ledet av en styringsgruppe bestående av Katja Nicolaysen (MiA), Tor Are Johansen (Arbark), Dag Juvkam (Arbark) og Dan Velle/Guro Nordby (NIA). Dan Velle var NIAs representant i styringsgruppa fra januar 2020 til desember 2021, da Guro Nordby tok over.

Arbeidsoppgavene har blitt utført av en arbeidsgruppe med arkivpersonale fra institusjonene: Kim Pham og Alvilde Mejlænder fra MiA; Kine Kristoffersen (fra januar 2020 til mars 2021), Dan Velle (fra mars 2021 til og med desember 2021) og Stian Elvsås (2022) fra NIA; Martin Ellingsrud, Dag Juvkam og Odd-Rune Hansen fra Arbark. Dag Juvkam har vært prosjektleder.

Prosjektet har hatt en referansegruppe med representanter fra ulike deler av arkivsektoren. Referansegruppa har blitt holdt orientert om prosjektets utvikling og har kommet med innspill underveis. Den har bestått av Tone Hemli (Arbeiderpartiet), Rolf Holte (Vestfoldarkivet), Arnulf Melgård (Skien SV), Jorunn B. Olafsdottir (Arkivverket) og Anneli Sundqvist (OsloMet).

1.4. Bakgrunn for prosjektet

Mens lovverket bidrar til å sikre bevaring av offentlige arkiver, er arkiv fra privat sektor i liten grad lovregulert. De statlige arkivinstitusjonene kan stille krav til arkivskaperne om at arkivene må være velordnet eller i godkjente formater før de deponeres. Slike virkemidler har ikke private arkivinstitusjoner. En kanskje større utfordring er endringene den teknologiske utviklingen har medført for selve arkivdanningen og arbeidsmetodene: I dag er det mange organisasjoner i privat sektor som ikke lenger arkiverer på papir, men i forskjellige digitale løsninger. Arkivinstitusjonene har ikke klart å holde tritt med denne utviklingen, da mange fortsatt har mye jobb med papirarkivene. Til sammen kan konsekvensen bli at ettertiden vil ha tilgang til kommunalt og statlig arkivmateriale, mens det blir mer tilfeldig hva historikere og andre får tilgang til fra privat sektor. Av disse grunnene er det viktig at arkivinstitusjonene bygger kompetanse på digitalt skapte arkiver.

Mange av Arbarks arkivskapere har de senere årene gått over til fullelektronisk arkivdanning. Arbark har derfor jobbet med å heve kompetansen på digitale arkiver. Tidligere har Arbark fått prosjektmidler fra Arkivverket til å gjennomføre prosjektene *PEANUTS* («Private elektroniske arkiver i Norge – uttrekk, testing, sikring», 2015-2016) og *Farvel DBS. Utredning av metoder for datahøsting fra private databasesystemer* (2016-2019). I PEANUTS-prosjektet mottok Arbark det digitale mappearkivet til organisasjonen Attac Norge. Formålet var å teste ut hvorvidt depotprogrammet Archivematica og tilgjengeliggjøringsprogrammet Access to Memory kunne brukes ved mottak av fil- og

mappearkiver, altså arkiver som lagres i mappestrukturer i Windows. Siden den gang har Archivematica fått ny funksjonalitet, og systemet har blitt utforsket videre i dette prosjektet. I Farvel DBS ble det jobbet med metoder for høsting og langtidslagring av arkivmateriale skapt i elektroniske saksbehandlingssystemer. LO og LOs fagforbund ble benyttet som *case*. Disse prosjektene har bidratt til betydelig kompetanseheving på el-arkiv-området for Arbarks del, men en fullt utviklet mottakspraksis har ikke blitt etablert.

MiA forvalter som nevnt rollen som fylkeskoordinerende ledd for privatarkiv i Akershus. I bevarings- og samhandlingsplan for privatarkiv Akershus 2019-2025, effektmål 2.5 Forvaltning og fornyelse, legges det vekt på tiltak som kan øke kompetanse innenfor feltet digitalisering. Bevaringsplanen har også et eget kapittel om privatarkiv i Viken fylkeskommune der det påpekes mangel på spisskompetanse på elektroniske arkiv. For MiA har derfor deltakelse i prosjektet vært svært relevant og lærerikt. Metodikken som har blitt utprøvd vil kunne benyttes langt bredere enn bare for politiske partier. Prosjektet har gitt et løft og styrket kompetanse på digitalt skapte privatarkiv for MiA, noe som også gir arkivarene bedre grunnlag for å sikre digitalt skapte arkiv i fremtiden.

Som del av Telemarksarkivet har NIA bevaringsansvar for privatarkiver, som i økende grad blir fullelektroniske. I tillegg har NIAs arkivar gjennom den nevnte avtalen med AAT et særskilt ansvar for bevaring og tilgjengeliggjøring av arbeiderbevegelsens arkiver. Bevaringsplan for privatarkiv i Telemark konkluderer med at det er lite digitalt skapt materiale som er bevart i Telemark, og at innsamlingen av slikt materiale bør prioriteres. Prosjektet har styrket det faglige grunnlaget for innsamling av digitalt skapte arkiver i framtida.

Det er ikke mange institusjoner i Norge som har jobbet dedikert med elektroniske arkiver fra privat sektor. I arbeid med bevaring av el-arkiv har oppmerksomheten ofte vært rettet mot sak-/arkivsystemer. Fokus i dette prosjektet har derimot vært på såkalt ustrukturert arkivmateriale; arkivinformasjon som ikke befinner seg i databaser med en gitt struktur, men i forskjellige mer fleksible løsninger, slik som mappestrukturer i Windows, e-postsystemer og sosiale medier. Fil- og mappearkiver er utbredt hos arkivskaperne som leverer til Arbark, MiA og NIA. Trolig utgjør de en stor del av de digitale arkivene i privat sektor.

I organisasjoner med sak-/arkivsystem overfører saksbehandlerne e-post som inngår i en sak fra e-postsystem til sak-/arkivsystem. Slik bevares e-post typisk i offentlig sektor. Men organisasjonene som omfattes av vårt prosjekt har ikke elektronisk sak-/arkivsystem. E-post

blir dermed liggende i e-postsystemet. Altså blir ikke informasjonen sikret og langtidsbevart. At mange også benytter private e-postkontoer, kompliserer bevaringen ytterligere.

Sosiale medier inngår også i prosjektet. Arkivverket har i samarbeid med Retriever utviklet en løsning for automatisk høsting fra Facebook, Twitter og Instagram.[1] Innlegg fra de offentlige profilene til partiledere, stortingspolitikere og sentrale partiorganisasjoner høstes regelmessig. Det er kun åpningsinnlegg som bevares, ikke påfølgende kommentarer.[2] Det er svært positivt at Arkivverket har påtatt seg denne jobben. Likevel dekker dette bare en del av den utadrettede politiske virksomheten i Norge. For å bidra til en mest mulig helhetlig samfunnsdokumentasjon, har vårt prosjekt vært opptatt av å utrede hvordan vi med enkle metoder kan høste inn fra lokale organisasjoner.

Av praktiske årsaker ble styringsgruppa enige om at arkiver som avleveres gjennom prosjektet bevares hos Arbark. Ved prosjektets slutt ble disse arkivene overført til Digitalarkivet for langtidsbevaring (se kapittel 8).

I søknaden ble det presisert at prosjektet skulle jobbe med tilgjengeliggjøring av fil- og mappearkiver. Tilgjengeliggjøringssystemet Access to Memory er installert på Arbark, men systemet har ikke blitt brukt til tilgjengeliggjøring, av to grunner. For det første har det vært problemer med kommunikasjonen mellom Archivematica og Access to Memory, slik at det har vært vanskelig å få til overføringer mellom systemene. For det andre har både Arbark, MiA og NIA ventet på felles nasjonale løsninger på dette området. Alle tre institusjoner kommer til å bruke Digitalarkivet til tilgjengeliggjøring, da dette er en type fellesløsning vi er avhengige av for å nå ut til brukere. Visningsmodulen i Digitalarkivet var ikke klar til testing av fil-/mappearkiv innenfor tidsrammen av vårt prosjekt. Vi fikk derfor ikke jobbet med dette slik vi hadde ønsket. Men vi følger med på utviklingen av Digitalarkivet.

1.5. De politiske partiene

De politiske partiene ble valgt som *case* i dette prosjektet, da bevaring av arkivene deres er en sentral oppgave for alle tre institusjoner. De politiske partiene utgjør selve grunnlaget for det norske demokratiet, samtidig som de er utøvende aktører i den daglige demokratiske virksomheten. Både ut fra et perspektiv om en samlet samfunnsdokumentasjon og som en

[1] Arkivverket 2017a.
[2] Arkivverket 2019.

dokumentasjon på de demokratiske prosessene i Norge, har det avgjørende betydning at partienes funksjon som grasrotpolitiske verksteder og sentrale politikkutformende aktører blir dokumentert og sikret for ettertiden. Sett fra et grasrot- og demokratiperspektiv er partiaktiviteten på lokalt nivå særlig viktig å dokumentere. Gitt lokalpartienes karakter og virkeområde er det svært relevant å følge opp problemstillinger knyttet til deres arkivdanning med basis i lokalorganisasjonene selv.

En annen årsak til at nettopp partiorganisasjoner ble valgt, er at de har stor overføringsverdi. Siden de ligner på andre medlemsbaserte organisasjoner i privat sektor, kan både prosjektets deltakere og andre ha nytte av resultatene fra prosjektet.

Det er et klart behov for å styrke og samordne arbeidet med dokumentasjon av de politiske partienes virksomhet. I 2014 kartla Arbark arkivdanningen til sentralorganisasjonene i Arbeiderpartiet, Sosialistisk Venstreparti og Rødt. Her kom det fram at disse brukte fil-/mappearkiv til bevaring av dokumentasjon. Arbeiderpartiet hadde i tillegg papirarkiv. Verken Arbark, MiA eller NIA har hatt oversikt over arkivdanningen ute i lokallagene. Vår antakelse, basert på kontakt med noen partilag, har vært at mange er gått over til fullelektronisk arkivering i mappestrukturer. Det er da problematisk at vi ikke har hatt metoder for mottak og bevaring. Et hovedmål med prosjektet har vært å finne mottaksløsninger som gjør oss i stand til å bevare disse arkivene – både fordi det er vår oppgave som bevaringsinstitusjoner, men også av hensyn til målet om en samlet samfunnsdokumentasjon.

2.0. De politiske partienes digitale arkiver

For å få bred oversikt over arkivdanningen undersøkte prosjektdeltakerne ulike organisasjonsledd: Arbark har undersøkt Aps, SVs og Rødts sentrale partiorganisasjoner. MiA og NIA tok for seg lokale lag og partier på fylkesnivå i Akershus og Telemark. Arbeidsgruppa utarbeidet først en spørreundersøkelse. Undersøkelsen var bredspektret, siden den rettet seg mot en relativt variert brukergruppe. Vi brukte programmet LimeSurvey til å lage selve undersøkelsen.[3] Spørreundersøkelsen ble sendt ut til om lag 60 partikontakter via e-post. Deretter fulgte vi opp med intervjuer av arkivansvarlige.

2.1. Situasjonen i sentralorganisasjonene

2.1.1. Organisasjon, roller og arkivrutiner

De ansatte på partikontorene har egne funksjoner og arbeidsområder – daglig leder, politisk rådgiver, valgkampleder, organisasjonssekretær osv. De fleste ansatte har adgang til arkivene, men de følges mest opp av ansatte med særlig ansvar for arkiv. Selv om disse ikke nødvendigvis har stillingsbetegnelsen «arkivar», har de arkivkompetanse. Dette gjør at arkivsituasjonen i de sentrale partiorganisasjonene er relativt god. Arkivene følger gjerne en arkivnøkkel eller et system (typisk år og emne).

2.1.2. Arkivløsningene

Aps hovedarkiv er et papirarkiv, men de har også et fil-/mappearkiv på et fellesområde. De to arkivene speiler hverandre, men noe avvik er uunngåelig. I hovedsak er det snakk om et protokollarkiv. Årstallsmapper legges til fortløpende. Hovedklassene er 'Egen organisasjon', 'Andre organisasjoner', 'Den offentlige forvaltning' og 'Politisk virksomhet'. Problemet med at dokumenter spres ut over flere lagringsmedier og -lokasjoner er ikke i nevneverdig grad til stede hos Ap; arkivverdig materiale arkiveres i mappestrukturen. Utkast til vedtak og kopier arkiveres sjelden. Ap hadde på et tidspunkt et elektronisk sak-/arkivsystem. Erfaringene med

[3] For selve undersøkelsen, se «Vedlegg 01 – Spørreundersøkelse».

dette var dårlige, og de har ingen planer om å gå til innkjøp av lignende systemer. De har heller ingen planer om å bli fullelektroniske – papirarkivet er enn så lenge det «offisielle» arkivet.

SV er fullelektroniske. Papirarkivet er avlevert til Arbark og ordnet. Tidligere arkiverte de i en mappestruktur på et delt område, men nå bruker de SharePoint, som fungerer som saksbehandlings- og arkivløsning for de ansatte på partikontoret. Brukergrensesnittet minner om et elektronisk saksbehandlingssystem, men er mer dynamisk. Man kan for eksempel legge til årstallsmapper. Mappeinndelingen ligner på den i de andre partiene. I realiteten er arkivløsningen et mappearkiv. Spørsmålet er altså eksportmulighetene. Annet arbeid med SharePoint-plattformen har avklart at det finnes funksjonalitet for eksport av både dokumentbiblioteker og «kolonner» (metadata), men den innebygde funksjonaliteten har få utvalgsmuligheter.

Rødt har siden stiftelsen i 2007 hatt et fil-/mappearkiv på et delt område. De har ikke hatt papirarkiv. Mappeinndelingen er typisk for et politisk parti: Organisasjon, Administrasjon, Utvalg osv. Rødt mener selv at det mest interessante materialet i et historisk perspektiv ligger i «Organisasjon». Her har man for eksempel protokoller fra organene og beretninger fra lagene.

Litt forenklet kan man si at partiene har protokollarkiver. De vanligste filformatene er Word, PDF og Excel. Generelt er arkivsituasjonen i sentralorganisasjonene god. Arkivenes interne orden virker å være tilfredsstillende, sett fra et depot-ståsted. Det er lite personsensitive data i arkivene. Det kan finnes eksklusjonssaker, men disse omtales i nøytrale ordelag. Partiene har egne medlemsavdelinger som bruker eget system til inn- og utmelding. Dette vil neppe bli avlevert til Arbark. Personalarkivene er det ikke aktuelt å avlevere.

2.1.3. Kommunikasjon og sosiale medier

E-post blir i liten grad arkivert utenfor e-postsystemet. For SVs del finnes det en drag-and-drop-mulighet mellom e-postsystem og SharePoint, men den er lite brukt. Partiene mener selv at lite e-post er historisk interessant. Den mest interessante korrespondansen finnes mest sannsynlig i stortingsrepresentantenes domener. Når folk slutter, slettes e-postkontoen. Potensiell arkivverdig informasjon kan altså gå tapt.

Partikontorene sender jevnlig ut nyhetsbrev til lokallagsledere og/eller medlemmer, eksempelvis landsstyrevedtak eller annen viktig informasjon. Egne fagsystemer brukes til å sende ut brevene. Utsendingene arkiveres ikke på noe vis. Dette er arkivverdig materiale tilsvarende tradisjonelle rundskriv/nyhetsbrev, og vi bør finne måter for å høste dette. Arbark kan gå i dialog med partiene ved avlevering av annet materiale.

Mens nyhetsbrevene brukes til kommunikasjon innad i partiorganisasjonen, brukes sosiale medier aktivt i kommunikasjon utad. Sosiale medier bevares ikke på noen spesiell måte. Særlig informasjonen fra Facebook og Instagram må regnes som arkivverdig. Gjennom prosjektet har vi funnet en måte å høste dette materialet på (se kapittel 7). Slike avleveringer kan avtales med partiene i forbindelse med avlevering.

2.1.4. Audiovisuelt materiale

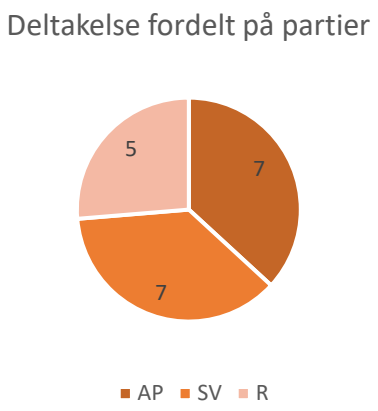
Partiene har egne grupper eller avdelinger som jobber med audiovisuelt materiale. Det er gjerne snakk om utadrettet virksomhet til bruk i valgkamp eller filming av landsmøter, altså materiale som bør bevares.[4] På grunn av dataomfang og plasshensyn lagres ikke dette i mappearkivene. Dette er kommunikasjonsavdelingenes domene, og arkivpersonale er lite involvert. Men sentralorganisasjonene er uansett interesserte i å avlevere slikt materiale til Arbark.

2.2. Oppsummering av funn på regionalt (fylkeskommunalt) og lokalt (kommunalt) nivå

I Akershus og Telemark var det tre partier fra regionalt nivå og 16 fra lokalt nivå som deltok i kartleggingen. 19 svarte på spørreskjemaet, mens 14 deltok på dybdeintervju.

Svarene fordelte seg på følgende partier:

[4] Se «Vedlegg 02 – Bevarings- og kassasjonsveiledning – sentrale partiorganisasjoner».



Funnene fra kartleggingen viste at det var store likheter mellom partienes digitalt baserte virksomhet og tilhørende elektroniske arkivdanning i Akershus og Telemark. Det er flere likheter mellom partiene på regionalt og lokalt nivå. Det er likevel slik at partiene på regionalt nivå i enkelte tilfeller har flere likhetstrekk med nasjonalt nivå. Dette kan skyldes at de jobber tettere opp mot nasjonalt nivå enn lokalt nivå.

2.2.1. Organisasjon, roller og arkivrutiner

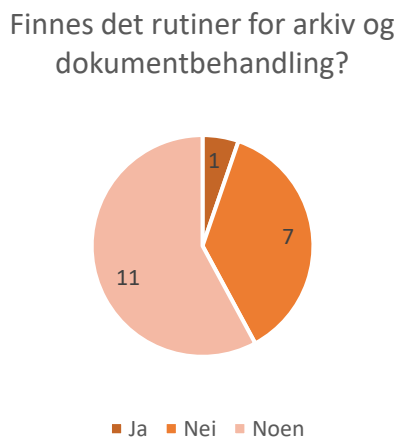
Standard oppbygning av et fylkes- eller lokallag er et valgt styre som leder lagets virksomhet. Styret består av leder, nestleder, sekretær, kasserer og varamedlemmer. Det varierer om partilagene har personer i styret til å fylle alle rollene. Politisk virksomhet på regionalt og lokalt nivå er i hovedsak frivillig arbeid. De tre fylkespartiene i denne undersøkelsen hadde derimot minst en fast ansatt. Dette skiller dem fra lokallagene.

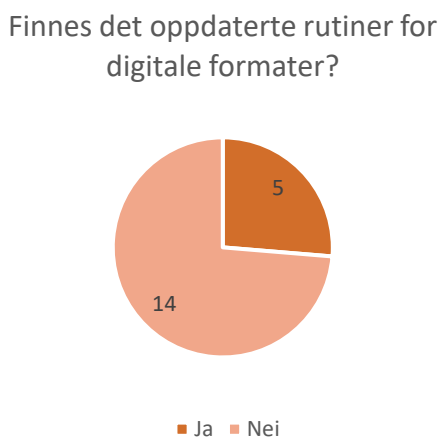
Kartleggingen viste at det er stor grad av frihet i hvordan dokumentbehandling håndteres i det enkelte lokallag. Det eneste unntaket er medlemsregistrering, der den nasjonale overbyggingen har betydning. Det skyldes krav til personvern (i personvernforordningen og personopplysningsloven), og det er tydelig at regionalt og lokalt nivå får opplæring i personvern. Det er også egne systemer for håndtering av medlemsregister.

Ansvar for dokumentforvaltningen er ikke alltid nedfelt i styrets arbeid. Hos noen av lagene er det ingen som har formelt ansvar for arkiv og dokumentasjon, mens hos andre er det sekretær eller leder. Noen partilag oppga at det var sekretær som har ansvar for innkallinger, protokoller og årsmeldinger. Andre oppga at det var leder som utarbeidet styrepapirer, men

sekretær hadde ansvar for protokoller og referater. Hos atter andre igjen varierte det hvem som var referent på møtene.

Det er med andre ord ikke entydige rutiner for dokumentbehandling. På lokalt nivå oppga flere å ha noen rutiner for arkiv og dokumentbehandling, men et fåtall har oppdaterte rutiner for digitale formater. Ingen av partiene, verken kommunale eller fylkeskommunale, har arkivplan. Det er i stor grad opp til det enkelte styret hvordan rutinene skal være. Det innebærer at nye medarbeidere lager sine egne systemer. Flere informanter oppga at det er vanskelig å finne tilbake til dokumentene fra tidligere perioder. I tillegg oppga flere partier at de fortsatt hadde papirarkiv. Digital utvikling, tilfeldigheter og at nye personer kommer til oppgis som årsaker til at arkiveringssystemet har endret seg over tid.





Partilagene svarer veldig ulikt på om de har oversikt over og vet hvor mange arkivverdige dokumenter som ikke blir bevart. Det fremkommer av kartleggingen at det er stor usikkerhet blant lagene om hva som faktisk er arkivverdig. Det var flere partilag som ikke hadde rutiner for å sikre at det ikke arkiveres kopier. Når det gjelder sletting av dokumenter fra arkivet oppga lagene enten at de ikke slettet noe selv, eller at de ikke vet fordi dokumentene er lagret på private PC-er og sletting blir deretter. Lagene har i liten grad retningslinjer eller kunnskap om lover og forskrifter om hva som skal arkiveres. Mange er derimot bevisst på personvernproblematikk og opphavsrett.

Kartleggingen viste at i den digitale arkivdanningen flyter dokumenter på en annen måte enn i den analoge verden. Styreroller blir også påvirket: Tidligere var arkivering i hovedsak en sekretærjobb, mens det nå ofte ser ut til å være leders ansvar. Begrunnelsen som oppgis er at leder allerede sitter på all dokumentasjon. I all hovedsak viste intervjuene at man i

arkivsammenheng hadde en relativt grei forståelse før digitaliseringen kom. Nå som stadig flere er mer eller mindre digitale i arbeidsprosessen, er ikke alle like oppdaterte på *hva* som skal bevares og ikke, *hvordan* materialet skal bevares, eller *hvorfor*. Men dette gjelder ikke for alle. Noen partilag har et veldig bevisst forhold til dokumentasjon, elektroniske arkiver og til skylagring. Ett partilag har til og med en dedikert IT-ansvarlig. Flertallet opplever derimot at disse tingene er vanskelige, spesielt det å ta vare på e-post og sosiale medier.

2.2.2. Arkivløsningene

Kartleggingen viste at det i stor grad er personavhengig hvordan arkivdanningen foregår i det enkelte lag. Mange av partilagene sorterte mappene etter årstall, saker og møtetyper (årsmøter, medlemsmøter, styremøter osv.). Det ble periodevist ryddet i strukturen ved at nye mapper ble lagt til etter behov. Enkelte av partilagene vedlikeholdt og ryddet oftere i arkivet. De fleste mente at strukturen fungerte hensiktsmessig. Flere påpekte at dette var fordi de ikke hadde store mengder dokumenter.

Fire av de 19 spurte partilagene oppga i spørreskjemaet at de bruke elektronisk saksbehandlingssystem. Her har nok noen misforstått hva et elektronisk saksbehandlingssystem er, ettersom ingen av lagene faktisk har et system av typen Acos Websak. Misforståelsen er litt interessant og tilsier at depotarkivarer bør være tydelige i kommunikasjonen med arkivskapere. Videre svarte alle lagene nei på spørsmål om de integrerte dokumentasjon fra mappestrukturen til andre systemer. Lagene produserte også dokumenter som ikke ble lagret i mappestrukturen. E-post, innhold fra sosiale medier og diverse som ulike medlemmer har liggende på private PC-er er eksempler på dette.

Filformatene partilagene benytter er hovedsakelig Word, PDF, Excel og JPEG. Det ble også nevnt filformater som PNG, CR2, MXF, MP4, AI og AVI. Verktøyene som brukes er i hovedsak Microsoft Office- og Google-produkter, slik som Microsoft 365, OneDrive og Google Drive. SharePoint og Dropbox brukes også til lagring. Et lag nevnte at MittAp ble benyttet for å lagre protokoller og referater.

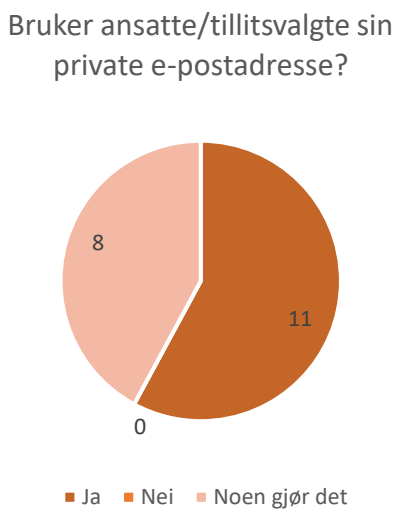
Til medlemsregister bruker lagene HyperSys og MittAp, men dette er systemer administrert av partiet sentralt. Noen partilag bruker også disse systemene for å kommunisere med medlemmer.

I dybdeintervjuene oppga flere av partilagene at de viktigste dokumentene lagres i skytjenester, men at disse dokumentene også kunne ligge på den enkeltes PC. Under *Arbeidsmetoder og verktøy* svarte nesten alle lagene at de benytter seg av Microsoft- eller Google-baserte skyløsninger. Ett partilag brukte programvaren Jottacloud, en norsk skylagringstjeneste. I varierende grad ble filene lagret på filservere eller i skylagring, men i de fleste tilfeller lagret informantene dokumentasjonen på deres respektive datamaskiner.

2.2.3. Kommunikasjon (e-post) og sosiale medier

Kun én informant mente at organisasjonen hadde et bevisst forhold til hva som bevares av e-post. Enkelte partilag oppga at viktige e-poster og vedlegg ble lagret i mappestrukturen, men stort sett arkiverer ikke lagene e-post.

En stor utfordring er at flere bruker privat e-postadresse i partipolitisk arbeid. Det gjelder selv om partilaget har tilgang til en egen partiadresse. Vanlige e-postløsninger er Outlook, Gmail, Hotmail og Online.



Av andre kommunikasjonsløsninger ble det nevnt bruk av Skype eller andre videokonferanseløsninger til intern elektronisk kommunikasjon og bruk av Facebook Messenger eller andre direktemeldingsløsninger.

Kartleggingen viste at sosiale medier ble brukt aktivt av alle lag, både på regionalt og lokalt nivå. Sosiale medier er særlig brukt i forbindelse med arrangementer, valgkamp og stands. Partiene benyttet sosiale medier for å få ut informasjon til medlemmer og sympatisører, de deler leserinnlegg og saker som er relevant for det politiske arbeidet. De brukte også sosiale medier i valgkamp for å fremme politiske meninger og fortelle om politiske saker i lokalmiljøet.

Alle partilagene brukte Facebook. Facebook ble brukt til å skrive om arrangementer og gi informasjon om møter; de deler nyhetsartikler, kronikker, leserinnlegg, videoene de har laget og innhold fra partiet sentralt; de deler saker de fremmer; de kommenterer saker og informerer om saker de får gjennomslag for. Noen av lagene oppga at de har en offentlig Facebook-profil som de bruker til å spre informasjon og en intern profil for medlemmer hvor det legges ut invitasjoner til møter. Imidlertid er det slik at mange benytter privat Facebook-profil til å fronte saker som egentlig gjelder partiet. En del partilag oppga at de hadde en hovedansvarlig/redaktør knyttet til sosiale medier.

Seks av partilagene svarte at de brukte Instagram og ett lag oppga Twitter. Instagram ble brukt for å vise bilder fra arrangementer og saker som pågår, presentere kandidater og i forbindelse med kampanjer og valgkamp. En informant oppga at de bruker Instagram Stories.

Hvor ofte partilagene la ut innhold på sosiale medier varierte. Inntrykket var at de fleste brukte sosiale medier aktivt og at de publiserte flere ganger i uken. Innhold fra sosiale medier arkiveres stort sett ikke i fil-/mappearkivet, med unntak av ett partilag.

2.2.4. Audiovisuelt materiale

Partilagene produserer audiovisuelt materiale. Lagene bruker fotografi til å dokumentere møter, aksjoner, arrangementer og valgkampaktivitet. I tillegg har de portrettfoto av kandidater og illustrasjonsfoto til nettartikler, partiprogram og brosjyremateriale. Det varierte hvem av medlemmene i laget som tar bildene. Flere brukte bilder fra sentralorganisasjonens bildebase. De færreste oppga at de brukte bilder hentet fra internett.

Flere partilag oppga at de produserte filmmateriale. Det kunne være i forbindelse med valgkamp, rekruttering, filmer av åpne medlemsmøter, snutter med politisk innhold, appeller til bruk i sosiale medier, eller filmer som omhandler politiske gjennomslag eller merkedager. Det varierer hvem av medlemmene som lager filmene. Youtube ble brukt for å legge ut politiske filmer og digitale arrangementer (for eksempel 17. mai-feiring i koronaperioden). Når det gjaldt filmproduksjon viste det seg at partimedlemmene var veldig bevisst opphavsrett.

Partilagene oppga at de sjelden tok i bruk lyd i seg selv. Enkelte fortalte at de jobbet med eller ønsket å starte podkast.

Noe foto og film ble lagret i mappearkiv, men som oftest ble materialet liggende hos den som hadde produsert det. For øvrig blir materiale liggende der det blir publisert, for eksempel Facebook. Enkelte lag oppfordret medlemmer til å sende inn bilder som lagres på private enheter.

2.2.5. Fokus på sikkerhet

Kartleggingen viste at partiene på regionalt og lokalt nivå er opptatt av sikkerhet knyttet til medlemslister og personopplysninger. Informantene var også opptatt av passordbeskyttelse i forbindelse med skylagring. Samtidig har ikke styrene nødvendigvis tydelige rutiner for overlevering av passord ved utskiftninger i styret. Ettersom mange bruker privat PC, er

datasikkerheten nødvendigvis opp til den enkelte. Kartleggingen identifiserte heller ikke rutiner for sikkerhetskopiering. Det betyr at det var stor risiko for tap av dokumenter ved for eksempel PC-krasj, noe flere hadde opplevd. Flere partilag oppga at de fortsatt tok papirutskrifter av de viktigste protokollene, og lagret disse. Det tyder på at partiene ikke føler seg helt trygge på at teknologiske løsninger vil bevare partiets virksomhet over tid.

2.2.6. Avlevering av arkiver

Prosjektet søkte om støtte over tre år. I siste søknad presiserte vi at vi i siste prosjektperiode skulle jobbe med reelle arkivavleveringer. Det innebar å motta digitalt skapte arkiver fra partilag og teste ut metodene vi hadde utarbeidet de to første årene. Kartleggingen tidlig i prosjektet hadde vist at mange partilag var interessert i å avlevere arkiver under forutsetning av at det var sikkert. Det kom frem at informantene var usikre på arkivdepotinstitusjonenes kompetanse innenfor teknologi og personvern (klausuleringsrutiner). Det skulle vise seg å bli mer komplisert å få avlevert arkiv enn forventet.

Det ble gjort omfattende innsats fra MiA og NIA på lokalt nivå mot de politiske partiene. Til tross for at flere av de aktuelle partiene tidligere hadde levert sine papirarkiver til Arbark, var det vanskelig å få dem til å avlevere digitalt skapt arkiv. Prosjektdeltakere var også med på styremøter for å informere om prosjektet og betydningen av avlevering. Oppsummert ble det avlevert fire lokale arkiver. Ingen arkiver kom inn fra regionalt nivå.

2.2.7. Oppsummering regionalt og lokalt nivå

Kartleggingen viste at det er lavt fokus på arkiv på lokalt nivå, mens det er noe høyere på regionalt nivå. At politisk aktivitet lokalt er basert på frivillighet, påvirker resultatet. Ansatte, eller i hvert fall ressurser til å jobbe med arkiv, er viktig for å sikre bevaring av arkivene. Kartleggingen tyder på stor sannsynlighet for tap av digitalt skapte dokumenter som både kan ha informasjons- og juridisk verdi for fremtiden. Det er opplagt behov for arkivarer og depotinstitusjoner som er synlige i samfunnet, og som kan håndtere digitalt skapt privatarkiv. Dette innebærer kompetanse, teknologiske og økonomiske ressurser for å sikre arkivene, og ikke minst mulighet til å drive god veiledning. For øvrig vises det til pågående

forskningsartikkel (som også var et mål i prosjektet) som vil gå nærmere inn på funnene i kartleggingen.

2.3. Bevaring og kassasjon

På møtene ble det klart at det var stor usikkerhet i organisasjonene når det gjelder hva som skal bevares og ikke. Dette gjelder naturlig nok særlig i mindre organisasjoner, men også til en viss grad i de store.

Arbeidsgruppa utarbeidet derfor veiledninger for bevaring og kassasjon; én veiledning for sentralorganisasjonene og én for lokal- og fylkeslag. Veiledningene ble laget med utgangspunkt i 1) tidligere mottatt arkivmateriale fra partiorganisasjoner, 2) en funksjonsanalyse av politiske partier, 3) det svenske Tjänstemäns & akademikers arkivs «Struktur för dokumenthandteringsplan och gallringsanvisningar 2.0».[5] Veiledningene kan forhåpentligvis være til hjelp for arkivskaperne, men først og fremst vil de bli brukt i depot i forbindelse med ordning.

For selve veiledningene, se vedlegg til denne rapporten.

[5] https://tam-arkiv.se/tam-6-struktur-for-dokumenthanteringsplan-och-gallringsanvisningar-2-0/ (link hentet 7. oktober 2022)

3.0. Avleveringsmetoder og periodisering

I kartleggingsfasen var prosjektet i dialog med arkivskaperne angående avlevering av digitalt skapt arkivmateriale. Vi ønsket å benytte muligheten prosjektet ga til å gjennomføre avlevering/mottak, for å kunne teste og vurdere metodene vi kom fram til. Noen arkivskapere var interessert i å foreta avlevering. Det var fil-/mappearkivene som var aktuelle å levere, ettersom det er her det mest arkivverdige materialet befinner seg (protokoller med saksdokumenter osv.). I det følgende beskriver vi hvordan slikt materiale kan mottas, samt mulige løsninger med tanke på periodisering.

3.1. Avlevering på portabel disk

Det er vanlig at arkivskapere avleverer portable disker med arkivinformasjon til arkivinstitusjon. Minnepenn eller harddisk er de mest utbredte lagringsmediene. Dette er kanskje den enkleste avleveringsmetoden sett fra depotarkivarens ståsted: Arkivinformasjonen overføres bare fra minnepenn/harddisk til mottaksområde, der den håndteres videre.

3.2. Avlevering fra skytjeneste

Mange arkivskapere har arkivene sine i forskjellige skyløsninger – Google Drive, OneDrive, Dropbox osv. Disse har innebygde eksportfunksjoner, slik at filer og mapper kan hentes ut. Inne i skyløsningene er materialet typisk organisert på år.

Noen arkivskapere som leverte til prosjektet sendte helt enkelt en e-post-link med tilgang til skyløsningen. Dermed kunne vi fra egen PC huke av mappene inne i løsningen og eksportere dem. Eksporten kommer som en zip-fil.

Avlevering fra skyløsning ble også gjennomført ved at medlemmer i arbeidsgruppa dro ut til arkivskaper: Arkivskaper logget seg på skyløsningen, valgte i dialog med oss mappene som skulle leveres, og foretok nedlasting. Zip-fil ble overført til en portabel disk vi hadde med. Arkivinformasjonen lå i viruskarantene på denne disken i tre uker, før den ble lagt inn på arbeidsområde på PC, hvor vi kunne jobbe videre med sikring av dataene.

3.3. Utkopiering hos arkivskaper

I noen tilfeller vil det være aktuelt å dra ut til arkivskaper og selv foreta utkopiering av fil-/mappearkiv (altså fra Windows-mapper, ikke skyløsning). Arkivar gjør da en utvelgelse av filer og mapper, og kopierer over til medbragt disk eller PC.

Her oppstår imidlertid spørsmål knyttet til periodeskille i mappearkivet.

3.4. Periodisering

Periodisering er sjelden et problem ved eksport fra skyløsninger, da mappene her oftest er organisert på år (ikke etter emne). For eksempel: 2013/Styrende organer/Styremøter/Protokoll 23.06.2013.doc. Windows-baserte mappearkiver er derimot ofte ordnet etter et system hvor årstallsmappene først dukker opp et stykke inn i mappestrukturen. Når arkivskaper skal avlevere igjen i framtiden risikerer vi å motta de samme filene en gang til. Etter arbeidsgruppas mening er det nødvendig med en form for periodisering, nettopp fordi dette vil gjøre det enklere ved framtidige avleveringer.

Arbeidsgruppa har kommet fram til tre måter å foreta periodisering på:

Alternativ 1: Hele arkivet avleveres på minnepenn eller lignende. I depot brukes et kopiprogram (for eksempel Beyond Compare[6]) til å hente ut filene skapt mellom eksempelvis 2005-2015. Vi får da et såkalt mykt periodeskille (2016-2022), slik at arkivskaper ikke foretar endringer i det historiske arkivet hos seg (arkivskaper gjør sjelden endringer i dokumenter som er flere år gamle). I slike tilfeller må det komme klart fram i en eventuell avleveringsavtale at arkivperioden er 2005-2015, og at materiale mellom 2015 og 2022 er å anse som ikke avlevert (blir slettet i depot). Så kan arkiv om noen år igjen avleveres på samme måte – arkivperioden som avleveres da vil være 2015 til for eksempel 2025.

Alternativ 2: Arkivskaper oppretter historisk arkivdel før avlevering; en kopi av overordnet mappestruktur, en *blueprint* av grunnstrukturen i arkivet, uten årstallsmapper og dokumenter. Historisk arkiv legges i en mappe som eksempelvis får navnet «Historisk 2005-2022». Denne

[6] https://www.scootersoftware.com/

avleveres til depot, mens en kopi av blueprinten blir det nye aktive arkivet hos arkivskaper. Blueprinten kan på et senere tidspunkt avleveres på samme måte. På denne måten unngår vi nye avleveringer med filer vi allerede har mottatt. Arkivskaper må ha en versjon av det avleverte arkivet hos seg, slik at de har nødvendige dokumenter tilgjengelig (i hvert fall i noen år). Vi anser dette som en bedre løsning enn alternativ 1.

Alternativ 3: Arkivinstitusjon gjør selv alternativ to. Noen arkivskapere foretrekker at arkivinstitusjonen gjør periodiseringen for dem. Det vil innebære at programmet Beyond Compare brukes til å lage periodeskillet: Vi kopierer arkivets mappestruktur (unntatt filene) og oppretter en ny arkivdel/*blueprint* med identisk mappestruktur («Aktivt arkiv 2022 - »). I tilfeller hvor arkivskaper kun har Windows-mappearkiv (og altså ikke leverer fra skyløsning) vil periodisering hos arkivskaper skje samtidig som arkivet avleveres/hentes: Beyond Compare tas med på en portabel disk; derfra åpner vi programmet, foretar kopi av mappestruktur og plasserer blueprint i mappen «Aktivt arkiv 2022 -».

I alle alternativene nevnt ovenfor kan det skje en aktiv utvelgelse i periodiseringsprosessen: Alle mapper og filer trenger ikke nødvendigvis bli med til depot; dersom det er tid kan arkivaren, gjerne sammen med arkivskaper, gjøre en vurdering av hva som ikke bør med til depot.

Med alternativ 2 og 3 ville det være logisk å skrivebeskytte historisk arkivdel, slik at arkivskaper ikke foretar endringer i dokumentasjon som egentlig er avlevert til depot. Så kunne arkivskaper heller kopiere filer fra historisk arkiv over til aktivt arkiv ved behov, eventuelt lage snarveier. Problemet er at noen arkivskapere avleverer for å kunne frigjøre lagringsplass, akkurat som med papirarkivene. En annen årsak til sletting er bedre oversikt (mindre informasjon, mindre å lete i). Uansett årsak vil arkivskaperne altså slette gamle filer de mener de ikke lenger har bruk for, og dersom mappen er skrivebeskyttet går åpenbart ikke dette. For noen vil dermed skrivebeskyttelse være uaktuelt. Over tid er konsekvensen av dette at eldre dokumenter vil forsvinne hos arkivskaper, etter hvert som behovet blir borte. Arbeidsgruppa anser ikke dette som et stort problem, siden arkivet jo vil være avlevert og sikret i depot. Og at eldre arkiv etter hvert ikke blir tilgjengelig hos arkivskaper (enten det er på PCer eller fysiske arkivrom), er selvsagt ikke noe nytt. Når partier får behov for gammel dokumentasjon kan de kontakte arkivinstitusjon, slik som før. (De vil også etter hvert kunne benytte Digitalarkivet.) Men det er likevel et poeng at partiene sitter på dokumentasjonen en stund, så det ikke blir mange henvendelser å håndtere for arkivinstitusjonen.

For detaljert beskrivelse av hvordan programmet Beyond Compare kan brukes til å periodisere (alternativ 3), se «Vedlegg 05 – Bruk av Beyond Compare til periodisering».

4.0. Mottak av digitale arkiver

Arkivinstitusjoner må gjøre en del operasjoner ved mottak for å sikre at arkivinformasjonen kan bevares i et langtidsperspektiv. I dette kapitlet går vi nærmere inn på disse operasjonene, bakgrunnen for dem og hvorfor de er viktige. Vi ser også på sentrale faglige standarder.

4.1. Sentrale utfordringer knyttet til langtidsbevaring

Fra 2010 til 2012 gjennomførte Arkivverket sammen med andre aktører prosjektet «DIAS – Digital arkivpakkestruktur». Målet var å etablere et felles rammeverk for digitale arkivdepoter. I prosjektrapporten pekes det på fire hovedutfordringer ved digital langtidsbevaring: lagringssikkerhet, opprettholdt lesbarhet, opprettholdt autentisitet og opprettholdt integritet.[7]

Lagringssikkerhet handler både om hvem som skal ha tilgang til informasjonen, samt at informasjonen må holdes digitalt intakt. Fysiske lagringsmedier har begrenset levetid. Det oppstår slitasje på både mekaniske komponenter og lagringssubstrat under bruk, og ved frakoblet oppbevaring vil de være utsatt for flere typer påvirkning, som oksidasjon, stråling og magnetisk «smitte»/«bitråte». Samlet sett gjør dette at garantert levetid for slike medier er begrenset. Under optimale lagringsforhold kan de lagrede dataene overleve noen år, men under suboptimale forhold kan det skje datatap bare etter måneder eller uker. Derfor må arkivdokumenter overføres fra sårbare medier, slik som minnepenner, til en sikrere instans. For en arkivinstitusjon vil det si et digitalt magasin med begrenset tilgang.

Opprettholdt lesbarhet handler om at vi må ha oppdatert maskin- og programvare som kan framstille dataene. Det skjer hele tiden en utvikling av applikasjonene som åpner filer. Det kommer for eksempel jevnlig nye versjoner av Microsoft Word, programvaren som tolker og åpner Word-filer. Over tid kan dette føre til at nyere versjoner av programvare ikke klarer å tolke og åpne eldre filer. For at filinnhold skal overleve teknologiskifter må filene migreres til nye formater, og dataene må flyttes over til ferske lagringsmedier og plattformer.

Opprettholdt autentisitet og *opprettholdt integritet* må ses i sammenheng. Ettertiden må gjøres i stand til å forstå arkivmaterialet, og man skal kunne stole på at informasjonen er ekte.

[7] Riksarkivaren 2012: 12.

Det betyr at det må tilføres opplysninger om arkivets innhold og dets opphav. Samtidig skal det ikke skje uautoriserte endringer i filene; informasjonsinnholdet må være bevart uendret.

Det er disse fire utfordringene vi i all hovedsak jobber ut fra i arbeid med langtidsbevaring.

Den beste metoden for integritetssikring (at informasjonsinnholdet i filene blir bevart uendret) er verifisering ved hjelp av sjekksummer. En sjekksum er en unik rekke med bokstaver og tall som beregnes på grunnlag av det samlede bit-innholdet i ei fil. Ved arkivmottak kan depottjenesten ved hjelp av programvare beregne sjekksummer for filene, og dersom man på et senere tidspunkt vil kontrollere at en fil er uendret, kan man beregne ny sjekksum og sammenligne denne med den opprinnelige. Dersom de to sjekksummene er identiske, har filen ikke blitt endret.

Merk at det er informasjonsinnholdet – teksten – i filen som bevares uendret: grunnet teknologisk utvikling kan ikke ei fil bevares uendret over tid, filer må viderekonverteres til det til enhver tid mest aktuelle langtidslagringsformatet. Når et dokument konverteres fra et format til et annet, endres bit-strømmen, selv om det tekstlige innholdet i dokumentet er det samme. For å kunne påvise at informasjonsinnholdet er uendret, må man derfor lage nye sjekksummer for konverterte filer. Begge sjekksumrapportene bevares sammen med arkivdokumentene.

Merk også at verifisering med sjekksummer har noen åpenbare begrensninger: de forteller ikke *hva* som er endret, *hvem* som har gjort endringen og *hvorfor*, eller *når* endringen har skjedd. Sjekksummene sier kun *at* noe har endret seg.

4.2. OAIS – Open Archival Information System[8]

Et viktig hjelpemiddel i digitalt depot er standarder og bestepraksis. Den såkalte OAIS-modellen – *Open Archival Information System* – har lenge vært en grunnsten i det digitale arkivfaget. Både internasjonalt og i Norge har OAIS vært toneangivende når det gjelder bevaring av digitale objekter og tilknyttede prosesser. Den er først og fremst et rammeverk og sier lite konkret om hvordan et arkivdepot kan gjøre ting i praksis. Men det meste av standarder og konkrete metoder er basert på prinsippene i OAIS, for eksempel mottaks- og bevaringssystemer som Archivematica og RODA, samt arkivstandarder som Noark 5.

[8] Delkapitlet er basert på Lavoie 2014.

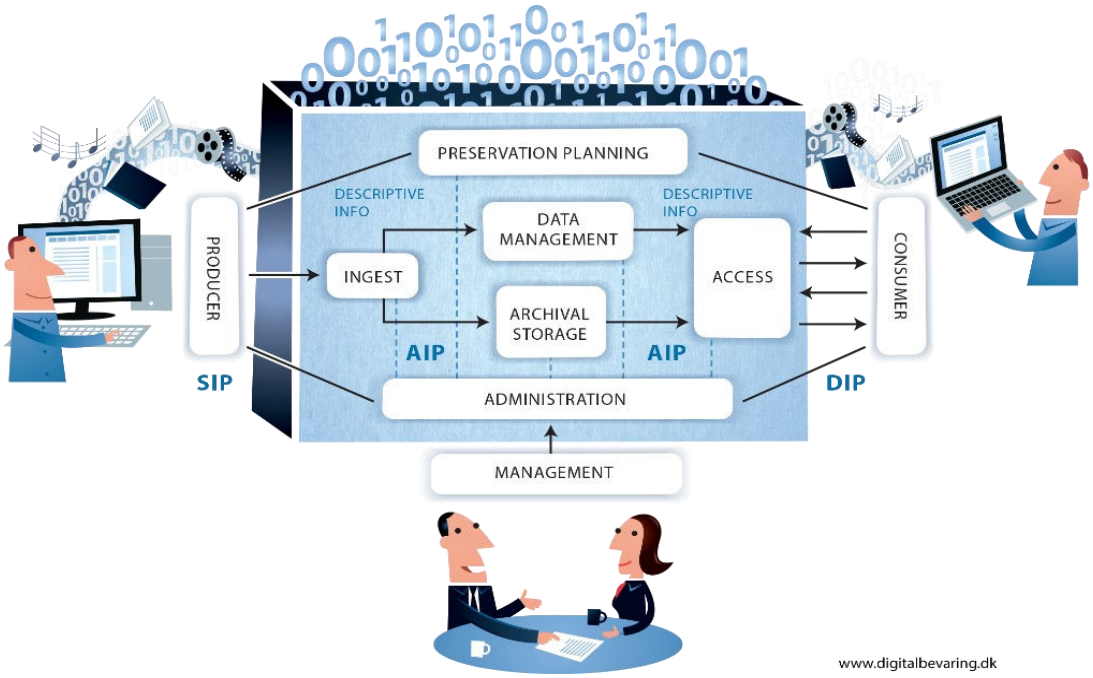
Grunnleggende i OAIS er prosessmodellen. Arkivskaper (*producer* i OAIS-termer) avleverer til OAIS-arkivet (arkivinstitusjon), hvor det er flere prosesser:

- Mottak (*ingest*)
- Lagring (*archival storage*)
- Dataforvaltning (*data management*), altså å holde oversikt over arkivmaterialet gjennom metadata med mer
- Bevaringsplanlegging (*preservation planning*), dvs. funksjonen for bevaringsstrategi og oppdateringer av den
- Tilgang (*access*), altså levering av materiale til brukere (*consumers*)
- Administrasjon (*administration*)

Hensikten med en slik prosessmodell er å beskrive alle elementene som inngår i et fullstendig depot, alle prosessene som omhandler håndtering og bevaring av digitalt materiale, og ivaretakelse av autentisitet og integritet i disse prosessene. Modellen er splittet opp i ulike funksjoner. I Norge har DIAS-standarden vært den praktiske implementeringen av OAIS.[9]

Pakkemodellen er også sentral. En avlevering eller et datasett kan være i forskjellige pakker i forskjellige faser av behandlingen: SIP (Submission Information Package), AIP (Archival Information Package) og DIP (Dissemination Information Package). SIP er avleveringspakken som blir levert til depotinstitusjon, AIP er den fullstendige bevaringspakken som langtidsbevares, og DIP er en tilgjengeliggjøringspakke som blir tilgjengelig for brukere. En mottatt og godkjent SIP transformeres til en AIP. Forskjellen ligger egentlig i funksjonen pakken tjener, der en AIP typisk vil ha mer metadata (for eksempel arkivbeskrivelse) enn SIP. Pakkene består gjerne også av forskjellige filformater; for eksempel kan en SIP inneholde Word-filer, som i transformasjonsprosessen til AIP konverteres og lagres som PDF/A. En DIP kan ha et utvalg informasjon fra en (eller flere) AIPer, og ha andre formater igjen, som er mer egnet for fremvisning.

[9] Arkivverket 2017b.

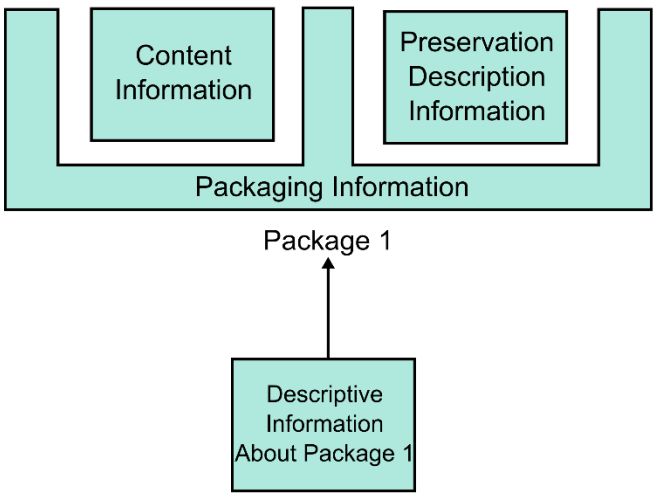


Jørgen Stamp, Creative Commons 2.5

En pakke (SIP, AIP eller DIP) består av to hoveddeler: selve innholdet (*content information*) og bevaringsbeskrivende informasjon (*preservation description information, PDI*). *Content information* består av selve dataobjektet, men også av representasjonsinformasjon – informasjon nødvendig for å gjengi/vise og forstå innholdet. Formatbeskrivelse er et eksempel på representasjonsinformasjon. I *preservation description information* inngår *fixity* (sjekksum og digitale signaturer med mer), tilgangsrettigheter, referanseinformasjon (unik ID), proveniensinformasjon (historikken til innholdet, herunder bevaringshistorikk) og kontekstinformasjon (innholdets relasjoner til annet innhold).

Det finnes også overordnet pakkeinformasjon, som skal binde sammen pakkens ulike deler til en logisk pakke (AIP). Beskrivende informasjon støtter gjenfinning av pakker.

Selve modellen (forenklet) ser slik ut:



Hentet fra Mess, Creative Commons Share Alike 4.0.

4.3. Andre standarder

Andre sentrale standarder ved digital bevaring er *Metadata Encoding and Transmission Standard* (METS), *Preservation Metadata: Implementation Strategies* (PREMIS), *Encoded Archival Description* (EAD) og *Encoded Archival Context – Corporate bodies, Persons and Families* (EAC-CPF). Standardene vedlikeholdes av Library of Congress, The Society of American Archivists og Statsbiblioteket i Berlin.

METS er en standard for beskrivende, administrativ og strukturell metadata vedrørende digitale objekter.[10] Selve METS-dokumentet er en spesifikasjon (i form av et XML-skjema) som inneholder sju seksjoner: METS header, Descriptive metadata, Administrative metadata, File section, Structural map, Structural links og Behavioral section. METS er en åpen standard som er inkorporert i en rekke digitale bevaringssystemer, deriblant Archivematica og RODA.

PREMIS er den internasjonale metadatastandarden for langtidsbevaring av digitale objekter.[11] Den definerer et sett av metadata som er nødvendige for å kunne bevare digital informasjon over tid: Object, Event, Agent og Rights. Disse elementene angir informasjon om selve filene

[10] https://www.loc.gov/standards/mets/
[11] https://www.loc.gov/standards/premis/

(teknisk og opphavsmessig), hva som har skjedd med dem, hvem som har gjort handlingene (institusjon/arkivar) og eventuelle (juridiske) rettigheter knyttet til materialet.

I Archivematica er PREMIS-standarden implementert i filen METS.xml.[12] I Archivematica-behandlede arkivpakker finner man altså METS- og PREMIS-dataene i én og samme fil:


```
1  <?xml version='1.0' encoding='UTF-8'?>
2  <mets:mets xmlns:mets="http://www.loc.gov/METS/" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xs
3    <mets:metsHdr CREATEDATE="2021-04-15T12:54:24"/>
4    <mets:dmdSec ID="dmdSec_1">
5      <mets:mdWrap MDTYPE="PREMIS:OBJECT">
6        <mets:xmlData>
7          <premis:object xmlns:premis="http://www.loc.gov/premis/v3" xsi:type="premis:intellectualEntity" xsi:schemaLocation="http://www.loc.gov/premis/v3 http
8            <premis:objectIdentifier>
9              <premis:objectIdentifierType>UUID</premis:objectIdentifierType>
10             <premis:objectIdentifierValue>d6451ccc-fb79-47d2-be65-e977e1c1f59d</premis:objectIdentifierValue>
11           </premis:objectIdentifier>
12           <premis:originalName>LO_i_Oslo-d6451ccc-fb79-47d2-be65-e977e1c1f59d</premis:originalName>
13         </premis:object>
14       </mets:xmlData>
15     </mets:mdWrap>
16   </mets:dmdSec>
17   <mets:amdSec ID="amdSec_1">
18     <mets:techMD ID="techMD_1">
19       <mets:mdWrap MDTYPE="PREMIS:OBJECT">
20         <mets:xmlData>
21           <premis:object xmlns:premis="http://www.loc.gov/premis/v3" xsi:type="premis:file" xsi:schemaLocation="http://www.loc.gov/premis/v3 http://www.loc.g
22             <premis:objectIdentifier>
23               <premis:objectIdentifierType>UUID</premis:objectIdentifierType>
24               <premis:objectIdentifierValue>1a1df844-9444-48cb-919b-4bf20409796d</premis:objectIdentifierValue>
25             </premis:objectIdentifier>
26             <premis:objectCharacteristics>
27               <premis:compositionLevel>0</premis:compositionLevel>
28               <premis:fixity>
29                 <premis:messageDigestAlgorithm>sha256</premis:messageDigestAlgorithm>
30                 <premis:messageDigest>204eaacc9ee04590d24470b37f509d09431d40d300d38a31402e35e888dce0a3</premis:messageDigest>
31               </premis:fixity>
32               <premis:size>52078</premis:size>
33               <premis:format>
34                 <premis:formatDesignation>
35                   <premis:formatName>Acrobat PDF/A - Portable Document Format</premis:formatName>
```


EAD – *Encoded Archival Description* – er en faglig standard for logisk arkivbeskrivelse i OAIS-baserte arkivpakker. Den er XML-varianten av den generelle arkivbeskrivelsesstandarden *General International Standard Archival Description* (ISAD(G)) og inneholder samme felter som denne.

EAC-CPF – *Encoded Archival Context – Corporate bodies, Persons and Families* – er standarden for beskrivelse av arkivskapere og andre aktører i OAIS-baserte arkivpakker. Den er XML-varianten av aktørbeskrivelsesstandarden *International Standard Archival Authority Record for Corporate bodies, Persons and Families* (ISAAR(CPF)).

4.4. Filformater og strategier

Filformater og konvertering er som vi har sett sentralt i digitalt depot (for mer om filer, se «Vedlegg 06 – Om filer og filtyper»). I motsetning til papirmateriale, som er svært bestandig,

[12] XML-formatet, *Extensible Markup Language*, er et åpent tekstformat som organiserer data i en hierarkisk struktur der merker gir informasjon om innhold. Formatet er også godt egnet til bevaring av metadata, og brukes derfor ofte til dette i digitale depoter.

har digitale data kort levetid. Langtidsbevaring krever derfor aktivt vedlikehold, ved at informasjon migreres fra gamle formater og plattformer til nye. Dette er en løpende oppgave, som vil måtte gjentas med ujevne intervaller i hele arkivets levetid.

For å minimere denne jobben er det lurt å velge bevaring i åpne filformater, altså formater der definisjonen er publisert i *public domain* slik at formatet ikke er patentbeskyttet og/eller rettighetsbelagt, men publisert under en lisens som garanterer at formatet og alle avarter av det alltid skal være fritt og gratis tilgjengelig. Eksempler på åpne formater er PNG, PDF og HTML. Samtidig bør man velge formater som har bred støtte i arkivverdenen og lang forventet levetid, slik som PDF/A. PDF/A regnes som mer bestandig enn andre formater nettopp fordi det er åpent og plattformuavhengig. PDF/A er da også godkjent arkivformat i Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver, også kjent som Riksarkivarens forskrift.[13] Selv om denne forskriften først og fremst er rettet mot offentlige arkiver, er den et nyttig verktøy for oss i privat sektor.

For dokumenter som inneholder tekst angir forskriften PDF/A og TIFF som godkjente arkivformater.[14] Arkivene som omfattes av dette prosjektet inneholder i stor grad Microsoft Word-, PDF- og Excel-filer. Microsofts .doc og .docx-formater er proprietære, altså produsenteide formater som er bundet opp til en leverandør/et selskap. Disse bør man konvertere vekk ifra. I PDF- og Word-filer samsvarer formatteringen inne i dokumentene godt med PDF/A, ettersom PDF/A er utviklet spesielt for arkivering av slike dokumenter. PDF/A-standarden skiller seg fra vanlig PDF ved å forby egenskaper som vil gjøre filen uegnet for langtidsbevaring, for eksempel kryptering og «font linking» (dvs. å referere til eksterne skrifttyper i motsetning til skrifttyper som er inkludert i selve dokumentet). Det er altså gode grunner for, og naturlig, å konvertere Word og PDF til PDF/A. Det er mer komplisert med konvertering av Excel-regneark til PDF/A. Et regneark vil kunne ha dimensjoner eller innhold som ikke støttes av PDF/A-formatet, og konvertering vil dermed kunne være mangelfull eller feile helt.[15] Likevel er PDF/A etter Riksarkivarens forskrift gjeldende arkivformat for regneark.[16]

---

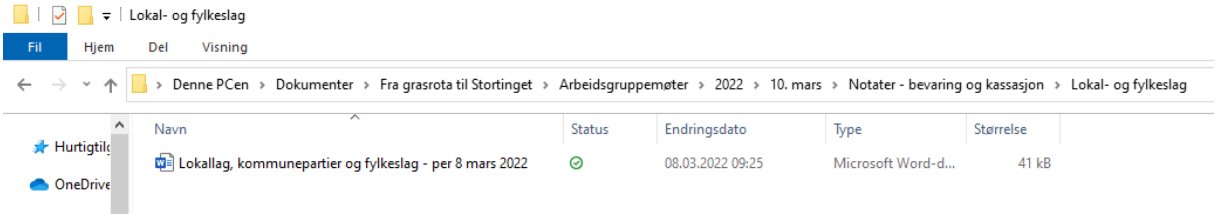
[13] Riksarkivarens forskrift 2017.
[14] Riksarkivarens forskrift § 5-17.
[15] Arbark har hittil konvertert Excel til PDF/A, men originalfilen bevares også. Om dette er den beste strategien med hensyn til Excel er fortsatt uvisst. En bedre løsning ville sannsynligvis vært å overvåke filene og få beskjed når et format er i ferd med å skiftes ut, slik at vi kunne migrert videre. Men et slikt system er ikke på plass. Arbark fortsetter inntil videre praksisen med å konvertere til PDF/A og samtidig bevare original Excel-fil.
[16] Riksarkivarens forskrift § 5-17. XML aksepteres også.

Selv om det ligger bilde-filer (JPEG, PNG) i partienes mappearkiver, inneholder arkivene prosjektet har mottatt i liten grad slike filer. Arbeidsgruppa har derfor fokusert lite på bildeformater.[17]

4.5. Særlige utfordringer med fil-/mappearkiver

4.5.1. Lange filstier

Windows har visse begrensninger på antall tegn i filstien til et dokument.



Microsofts filsystem NTFS[18] tillater maksimalt 260 tegn i hele filstien, fra og med første tegn i stien (f.eks. C:\) til og med filnavn. Dersom den fullstendige stien overskrider 260 tegn, får man problemer ved åpning og kopiering av filen.

Dette skyldes den såkalte MAX_PATH-variabelen i NTFS-filsystemet, som setter en restriksjon på 260 tegn.[19] Alle nyere Windows-versjoner har denne grensen.[20] Dette kan få konsekvenser for oss i depot. Problemet oppstår i arkiver hvor man har lange dokumenttitler, hvor det i tillegg er mange mapper nedover i mappehierarkiet.

I digitalt depot er man avhengig av programvare til å løse forskjellige oppgaver. Aktuell programvare som kjører i Windows vil også kunne rammes av max antall tegn-problemet. Det samme gjelder ved kopiering fra en disk til en annen.

Men i de aller fleste mappearkiver vil ikke dette være noe problem. Det skal litt til å komme opp i 260 tegn. Arkivskapere har stort sett relativt fungerende systemer med korte, konsise titler, og et ikke altfor komplekst mappehierarki. Men prosjektdeltakerne opplever at noen

[17] Men Archivematica (se kapittel 5) kan brukes til å håndtere og konvertere bildefiler i mappearkiver.
[18] NTFS – New Technology File System.
[19] https://www.howtogeek.com/179173/ask-htg-why-is-windows-reporting-this-folder-is-too-long-to-copy/ (link hentet 7. oktober 2022.)
[20] Det nåværende Long Filenames-systemet med 255 tegn ble introdusert i Windows 95. Før Windows 95 var det en grense på åtte tegn per filnavn.

arkivskapere bruker omstendelige titler for å beskrive dokumenter presist. Problemet blir ytterligere aktuelt hvis man i depot legger til ekstra mapper på overordnede nivåer for å standardisere pakkestrukturen. Derfor trengs løsninger. Slik arbeidsgruppa ser det, er tre løsninger aktuelle:

1. Forkorte mappe- og filnavn i filstien. Dette er svært tvilsomt rent arkivfaglig, særlig filnavn. Dersom man endrer filnavn kan også metadataene for dokumentet endres (hvis man flytter fila mellom disker), og det er overhodet ikke ønskelig. Vi regner det likevel som akseptabelt å forkorte mappenavn, så lenge endringer dokumenteres i en depotlogg.
2. Oppløse mappehierarkiet og putte alle dokumentene i en «objects»-mappe. Arkivfaglig er det ikke ideelt, men hvis man tar vare på det opprinnelige mappehierarkiet (for eksempel i en CSV-fil) går det an.
3. I Windows 10 og nyere versjoner er det muligheter for å instruere Windows til å godta stilengder opp til litt over 32 000 tegn.[21] Dette fungerer dog ikke automatisk i alle applikasjoner, det bare åpner for at alle applikasjonene kan ha muligheten til å gjøre det, dersom de bruker de «riktige» Windows-funksjonene til å utføre fil- og mappeoperasjoner. For å fjerne begrensningen:
   1) Åpne Group Policy Editor (skriv inn gpedit.msc i søkefeltet)
   2) Naviger til: *Local Computer Policy > Computer Configuration > Administrative Templates > System > Filesystem > NTFS*
   3) Velg *Enable NTFS long paths*

Arbeidsgruppa regner det første alternativet som best. Alternativ 2 – oppløsning av mappehierarkiet – er litt voldsomt, da det kanskje bare er noen få dokumenter som skaper problemer. Alternativ 3 er ikke gunstig, siden slike endringer i Windows kan få utilsiktede konsekvenser.

Ved arkivmottak kan programmet TreeSize[22] brukes til søk etter lange filstier. For beskrivelse av bruk av TreeSize, se «Vedlegg 07 – Rutinebeskrivelse for laging av arkivpakker med Archivematica».

[21] https://mspoweruser.com/ntfs-260-character-windows-10/
[22] https://www.jam-software.com/treesize_free?ca=1 Arbark bruker betalingsløsningen.

4.5.2. Bevaring av opprinnelige metadata

Alle filer bærer med seg essensiell metadata, slik som filstørrelse, format og tidspunkt for når filen er opprettet og endret.

I arkivsammenheng er bevaring av disse metadataene utfordrende, både i forbindelse med avlevering og konvertering:

- Avlevering: Ved kopiering fra en disk til en annen kommer opprinnelig metadata med, så avlevering på minnepenn går fint. Men ved overføring via e-post eller nedlasting fra skyløsning blir opprettelsesdato den datoen filene lastes ned til lokal disk. Dersom dokumentdato (dato for når dokumentet er ferdigstilt) ikke er skrevet inn i selve dokumentet mangler dermed helt essensiell beskrivelsesinformasjon, noe som åpenbart er problematisk rent arkivfaglig. I partienes arkiver er kanskje ikke dette et stort problem, siden disse i hovedsak er protokollarkiver og dato oftest inngår i filnavnet. Fra løsninger som SharePoint finnes det også muligheter for bevaring av opprinnelig metadata ved nedlasting.[23]
- Konvertering: Ved konvertering fra et filformat til et annet opprettes det en ny fil, og teknisk metadata for originalfil blir ikke nødvendigvis med over til konvertert fil. Igjen, dette er kanskje ikke et stort problem i rene protokollarkiver. Men andre typer dokumenter – arbeidsnotater, evalueringer, forslag osv. – har ikke alltid påført dato. Faktisk er datering, enten inne i dokumentet eller som del av filnavn, heller unntaket enn regelen. Da vil opphavsinformasjon mangle. Det er noe av grunnen til at arkivdepot må bevare både originalfiler og konverterte filer. Samtidig vil en hensiktsmessig mappestruktur være til hjelp: mappestrukturen er i seg selv en arkivbeskrivelse og vil som regel inneholde årstallsangivelse for mappene. Prosjektets arbeidsgruppe har også utarbeidet veiledninger hvor vi anbefaler at arkivskaper alltid skriver forfatter og dokumentdato inne i dokumentet.

[23] Se for eksempel: https://andrewwarland.wordpress.com/2021/07/02/metadata-persistence-when-records-are-moved-in-sharepoint/; https://techcommunity.microsoft.com/t5/sharepoint/preserve-created-date-when-downloading-files-from-sharepoint/m-p/3050586

4.6. Beskrivelse av «manuell» mottaksrutine hos Arbark før prosjektstart

Utfordringene beskrevet under punkt 4.1 blir nødvendigvis bestemmende for hvilke mottaksoperasjoner som må gjennomføres ved sikring og behandling av arkivdataene.

Etter arbeidsgruppas mening bør følgende handlinger foretas ved mottak:

- Integritetssikring/generering av sjekksummer
- Viruskontroll
- Identifisering av filformater
- Beskrivelse av arkivinformasjon og kontekst
- Ordning og eventuell sletting av filer
- Filkonvertering
- Overføring til sikkert lagringsområde

Arbark benyttet en stund en manuell mottaksrutine der disse bevaringsoperasjonene ble gjort en etter en, med forskjellig programvare. Selv om rutinen ikke lenger følges, var den utgangspunktet for metoden som arbeidsgruppa har utarbeidet gjennom prosjektet (se neste kapittel). Den er derfor beskrevet nedenfor. Da rutinen ble utviklet ble forskjellige verktøy prøvd ut, både lisensiert og gratis programvare. Alle verktøyene kjørte i Windows, bortsett fra Access to Memory, som er installert på Linux/Ubuntu-plattform.

Prosedyren var slik at mottatte arkiver ble lagt inn på et dedikert mottaksområde på lokal PC, hvor de ble håndtert.

Ved arkivmottak ble det opprettet en depotlogg; en kort beskrivelse av alle jobbene som ble utført fra mottak til og med overføring til digitalt magasin. Både programvare som ble brukt til de forskjellige mottaksoperasjonene, og datering av disse, ble dokumentert. Depotloggen ble bevart sammen med arkivet, i en overordnet mappe: «repository_operations».

Ved mottak laget arkivaren også et notat i Notepad, «Notat om avlevering». Her inngikk Asta-referanser (aksesjonsnr., aktørnr., arkivnr.), dato for avlevering, antall filer, dataomfang, fagansvarlig og en kort beskrivelse av hvordan arkivet hadde blitt avlevert. Dette notatet ble bevart sammen med arkivet, i mappen «descriptive_metadata». Årsaken til at slike metadatafiler skal bevares sammen med arkivinformasjonen, er OAIS' krav om selvbeskrivende arkivpakker: man skal ikke trenge å gå noe annet sted for å forstå materialet.

Videre innebar rutinen bruk av programmet Fixity AVP[24] til å lage sjekksummer. Fixity produserte en rapport med oversikt over alle filene i arkivet og tilhørende sjekksum. Rapporten ble plassert i «repository_operations»-mappen, sammen med depotloggen. Denne integritetssikringen skal gjøres så tidlig som mulig, ettersom det er viktig å kunne påvise at det ikke har skjedd endringer i dokumentene.

Etter integritetssikring lå arkivmaterialet i viruskarantene på mottaksområdet i tre uker, før arkivaren kjørte virussjekk. (At karantenetiden er akkurat tre uker er litt tilfeldig, men det har å gjøre med oppdatering av antivirusprogrammet: Antivirusprodusentene trenger noe tid til å oppdatere programvaren sin for å oppdage og håndtere nye virus. Vi anser tre uker som tilstrekkelig for slik oppdatering.) Dersom det blir funnet virus- eller malware-forekomster i det mottatte materialet vil det aktuelle antivirus- eller antimalware-programmet håndtere infeksjonen i henhold til eget regelsett. Dersom de aktuelle infiserte filene ikke lar seg rense, vil de flyttes til et eget karanteneområde, og må håndteres manuelt i ettertid. Alle deteksjoner logges.

Arkivinstitusjoner bruker gjerne programmer som DROID[25] til å identifisere filformater. Arbark har i liten grad brukt slike programmer, da filformatene som kommer inn stort sett er Word, PDF, Excel, Power Point, PNG og JPEG. Arbarks arkivarer har derfor nøyd seg med en visuell sjekk, og skrevet filformater inn i arkivbeskrivelsen.

Arbark bruker Asta til arkivbeskrivelse, men i mottaksrutinen inngikk også bruk av Access to Memory, som kan brukes til å lage beskrivelser som følger standardene ISAD(G) og ISAAR. Disse beskrivelsene kan eksporteres slik at de blir del av arkivpakken, som EAD.xml og EAC_CPF.xml. Dette for å gjøre pakken selvbeskrivende.

Ordning av mappestruktur og sletting av filkopier (eller annet som ikke hører hjemme i arkivet) har i varierende grad blitt gjort ved Arbark, avhengig av hvor godt ordnet arkivet har vært i utgangspunktet. Programmet TreeSize har blitt brukt til arkivbegrensing og til søk etter lange filstier (se «Vedlegg 07 – Rutinebeskrivelse for laging av arkivpakker med Archivematica»).

[24] https://www.weareavp.com/fixity-pro-release-2020/
[25] *Digital Record Object Identification*, verktøy utviklet av det britiske nasjonalarkivet for identifisering av filformater: https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/

Neste steg i rutinen var konvertering til arkivformat. Programmet Neevia Document Converter Pro[26] ble brukt til å konvertere produksjonsformater til PDF/A. I konverteringsprosessen kopierer programmet original mappestruktur, og konverterte filer plasseres i riktige mapper.[27] Både originalfiler og konverterte filer ble bevart.

Ved konvertering endres bit-strømmen i filene. Etter konvertering ble derfor nye sjekksummer produsert, slik at man fikk sjekksum-rapporter både for originaler og konverterte filer.

Ferdigbehandlede arkiver som er overført til digitalt magasin er organisert i en pakkestruktur som ligner på Arkivverkets DIAS-standard[28]: Selve mappearkivet ligger i mappen «content», mens arkivbeskrivelse og «Notat om avlevering» ligger i «descriptive_metadata». Depotlogg, sjekksum-rapport og eventuell annen relevant informasjon fra mottaksarbeidet ligger i «administrative_metadata/repository_operations». Originalfilene og medfølgende sjekksummer ligger i en egen mappe (samme mappenivå som ferdigbehandlet arkiv). Rotmappe har angivelsen arkivskapernavn + ytterår.

På tross av mye arbeid med å utarbeide en hensiktsmessig mottakspraksis ved Arbark, har det vært mangler ved rutinen som her er beskrevet. Retningslinjene har ikke alltid blitt fulgt like nøye, dermed har det blitt variasjoner i arkivpakkene. Dette skyldtes delvis at det kom inn arkiver mens rutinen ble utarbeidet. I tillegg kunne arkivarene ta i bruk nye verktøy, uten at disse nødvendigvis umiddelbart kom inn i rutinebeskrivelsen. Det skal og sies at prosessen er noe tungvint, i og med alle de forskjellige programmene som har blitt brukt. I begynnelsen av dette prosjektet testet arbeidsgruppa ut Arbarks mottakspraksis, og konkluderte med at den var for tungvint. Vi begynte derfor å se etter mer helhetlige systemer, basert på standardene beskrevet i dette kapitlet, som kunne gjøre jobben mer automatisert.

[26] https://neevia.com/products/dcpro/
[27] For detaljer om konvertering med Neevia, se vedlegg 2 til Arbarks prosjektrapport fra 2019, «Farvel DBS. Utredning av metoder for datahøsting fra private databasesystemer».
[28] Se Riksarkivaren 2012: 22.

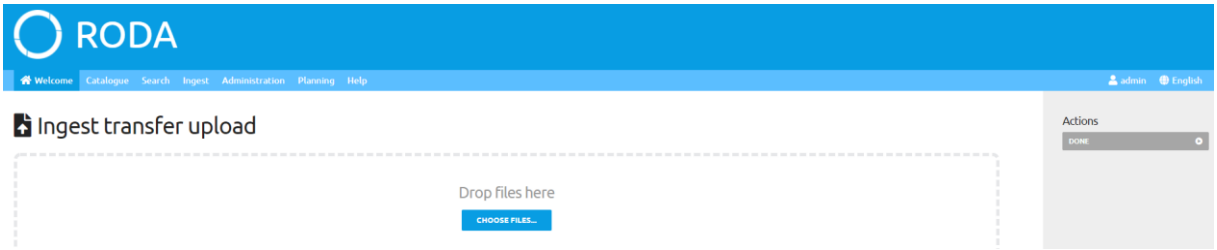
5.0. Archivematica og RODA

To systemer som forenkler mottaksjobben er Archivematica og Repository of Authentic Digital Records (RODA). I dette kapitlet tar vi nærmere for oss prosjektets erfaringer med de to systemene.

5.1. RODA

RODA er en nettleser-basert mottaks- og bevaringsløsning som samsvarer med OAIS.[29] RODA støtter også andre sentrale standarder som METS, PREMIS og EAD. Systemet har funksjonalitet til å automatisere bevaringshandlinger, slik som virussjekk, filidentifisering, sjekksum-generering, innpakking (i BagIt[30]- eller E-ARK[31]-formatet) m.m. Brukergruppa internasjonalt er voksende. Systemet har åpen kildekode, som innebærer at det kan lastes ned og installeres gratis. Siden RODA kjører på Linux-plattform trengs Linux-kompetanse for installasjon.

Arbark installerte RODA og testet versjon 3.3.1. RODA krever innlogging via nettleser. De sentrale funksjonene for å lage arkivpakker er: *Catalogue*, *Transfer* og *Process*. *Catalogue* er oversikten over arkivdataene i depot; det er her arkivpakkene havner når de er ferdige. *Transfer* er det midlertidige lagringsområdet for arkivpakkene mens de produseres. *Process* er jobbene som kjøres for å lage arkivpakker.



EAD-standarden er implementert i RODA, så i *Catalogue*-fanen kan man beskrive arkivmaterialet på en standardisert måte, ved å fylle ut de feltene en ønsker (EAD er som nevnt basert på ISAD(G)).

[29] https://www.roda-community.org/#welcome
[30] https://en.wikipedia.org/wiki/BagIt
[31] https://www.eark-project.com/

Videre kan filene lastes opp i *Ingest* > *Transfer*. Eventuelt kan arkivaren laste ned programtillegget RODA-In på lokal PC og bruke det til å lage SIP-pakker. RODA-In kan lastes ned fra GitHub.[32] Det finnes gode visuelle brukerveiledninger for RODA-In inne i programmet.

Når pakken er lastet opp til *Ingest* > *Transfer* huker man av hvilke bevaringshandlinger som skal gjennomføres. Når jobben er utført ligger AIP i *Catalogue*. Ettersom RODA forholder seg til METS og PREMIS, blir bevaringsbeskrivende og teknisk metadata bevart som del av arkivpakken. Altså får man XML-baserte loggfiler med oversikt over hvilke operasjoner RODA har utført.

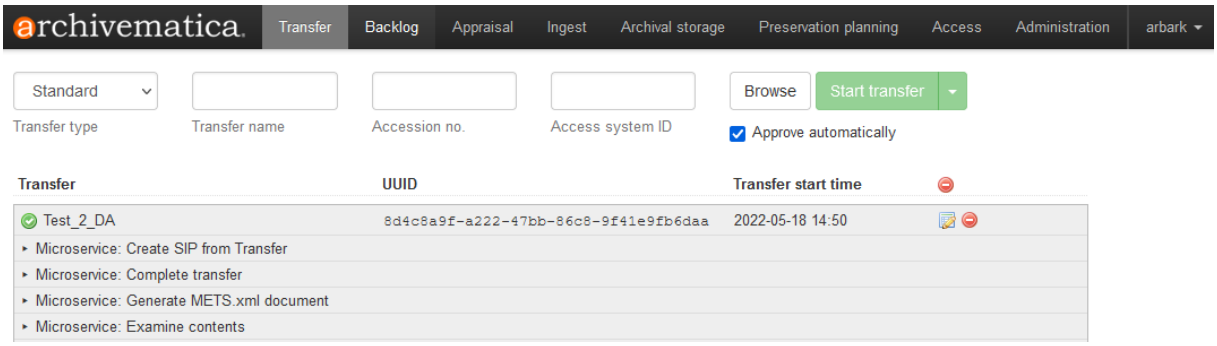
5.2. Archivematica

Archivematica har mange likhetstrekk med RODA: det er nettleser-basert, har åpen kildekode, samsvarer med OAIS og automatiserer bevaringshandlingene (i Archivematica kalles disse for mikrotjenester).[33] Disse bevaringshandlingene utføres av underliggende programmer som er integrert med Archivematica, og som i likhet med resten av systemet har åpen kildekode.

En testversjon av Archivematica (versjon 1.11) ble installert på Arbark i mai 2020. MiA og NIA har hatt fjerntilgang til denne installasjonen. På installasjonen ble det satt opp et eget «transferområde»; en egen mappe i filsystemet på Archivematica-serveren som depottjeneste har tilgang til gjennom pålogging. Herfra lastes filarkiv opp til Archivematica. Dette for å ha bedre kontroll med opplastingen og gjøre prosessen mer strømlinjeformet.

Brukergrensesnittet i Archivematica består av forskjellige faner. Disse utgjør en «pipeline» som arkivmaterialet sendes gjennom. Når materialet er ferdigbehandlet havner arkivpakken i lagringsløsningen *Archival storage*.

[32] https://github.com/keeps/roda-in/releases
[33] https://www.archivematica.org/en/



I fanen *Transfer* laster man opp mappe fra transferområdet. Archivematica kjører så forskjellige automatiske mikrotjenester, deriblant virussjekk og sjekksum-generering, før pakken sendes videre til *Ingest*. Her fortsetter mikrotjenestene, blant annet konvertering fra produksjonsformater (Microsoft Word, Excel, Power Point) til arkivformat (PDF/A).

Standardversjonene av Archivematica og RODA konverterer ikke til PDF/A, men det skal være mulig å få programmene til å gjøre denne jobben. Mens vi aldri klarte å få RODA til å konvertere til arkivformat, er dette en relativt enkel jobb i Archivematica: Vi installerte først konverteringsverktøyet Unoconv på Archivematica-serveren. Inne i fanen *Preservation planning*, under *Format policy registry – Tools* la vi til Unoconv. I samme fane, under *Normalization – Commands* la vi deretter inn egendefinerte kommandoer for konverteringen, basert på dokumentasjonen til Archivematica og dokumentasjonen til Unoconv, samt ved å sammenligne med andre forhåndsdefinerte kommandoer. Under *Normalization – Rules* satte vi deretter regler for hva som skulle skje med forskjellige filformater, for eksempel at 'Microsoft Word Document 97-2003' skulle konverteres til 'PDF/A-2b'.[34]

Slik kan vi administrere og definere regler for konvertering inne i mottakssystemet.

Etter at mikrotjenestene er fullført i *Ingest* havner arkivet i *Archival storage*. Som med RODA inneholder ferdig prosesserte Archivematica-pakker oversikt over de tekniske operasjonene systemet har utført. I Archivematica bevares denne informasjonen i filen METS.xml. Denne inneholder også teknisk metadata for både originalfiler og konverterte filer. Om opprettelsesdato for originalfil kommer med avhenger av avleveringsmetode; hvis filene er lastet ned fra skyløsning blir opprettelsesdato lik dato for nedlasting (se punkt 4.5.2).

[34] PDF/A-2b - versjon av PDF/A som er godkjent arkivformat etter Riksarkivarens forskrift § 5-18(2).

5.3. Sammenligning av systemene

For bedre å kunne vurdere hvor gode systemene er, foretok arbeidsgruppa en sammenligning av RODA og Archivematica: Det ble utarbeidet en kravspesifikasjon for mottakssystem, med *må*-, *bør*- og *kan*-krav (se vedlegg 8 og 9). Her kom Archivematica best ut. Vurderingen er at RODA er mindre brukervennlig, særlig når det gjelder opplasting av pakker og en noe forvirrende katalog.[35]

Visuelt er grensesnittet til RODA fint, men programmet er ikke nødvendigvis veldig intuitivt. For eksempel: Når man skal laste opp en pakke i *Transfer*, er det ikke en knapp som heter «Upload». Man må trykke og lete seg fram til opplastingsknappen (og andre funksjoner), noe som svekker brukervennligheten.

RODA har likevel noen fordeler i forhold til Archivematica, for eksempel muligheten til arkivbeskrivelse i *Catalogue*-fanen. Denne funksjon finnes ikke inne i Archivematica, som baserer seg på beskrivelse med Access to Memory.

Installasjon av begge systemene har utfordringer: I Archivematicas tilfelle har noen versjoner hatt feil i installasjonsprosedyren, slik at vi har måttet kompensere ved å enten rokere på rekkefølgen komponentene installeres i, eller gjøre manuelle konfigurasjonsendringer etter installasjon. For RODAs vedkommende tilbys bare installasjon i Docker, og dette blir noe begrenset i forhold til hvordan vi ønsker å konfigurere og bruke systemet. Det største problemet var likevel at vi ikke klarte å få RODA til å konvertere til arkivformat.

5.4. Archivematica som mottakssystem på Arbark

Konsekvensen av denne sammenligningen ble at Arbark har bestemt seg for å ta i bruk Archivematica i mottaksarbeidet framover. Archivematica vil foreløpig kun bli benyttet ved mottak av fil- og mappearkiver. Siden første installasjon har systemet blitt oppgradert til nyere versjoner. Installasjon kan gjøres på flere måter. Arbark har installert manuelt.

Archival storage, Archivematicas lagringsområde, fungerer bare som et lager og er på ingen måte tilfredsstillende som digitalt magasin. Arbark vil derfor kun bruke Archivematica til å

[35] Særlig sistnevnte skyldes muligens at vi ikke rakk å bli vant til brukergrensesnittet, men det i seg selv forteller like fullt noe om brukervennlighet.

lage arkivpakkene: Etter at Archivematica har produsert pakkene vil disse bli eksportert fra *Storage* og overført til Arbarks digitale magasin. Pakkene eksporteres som 7zip-filer og lagres i BagIt-pakkeformatet. Framover vil Arbark også bevare arkivpakkene i Digitalarkivets bevaringsløsning, se kapittel 8.

5.5. Tilleggsoperasjoner

Archivematica har sine begrensninger. Selv om Arbark kun vil bruke det som mottakssystem, gjør ikke systemet alt. Prosjektets arbeidsgruppe har derfor vurdert en del tilleggsoperasjoner som må gjøres manuelt før materialet lastes opp til Archivematica. Dette er for en stor del prosedyrer som er hentet fra Arbarks tidligere mottaksrutine (se punkt 4.6).

Det var en del arbeid med å få på plass operasjonene før opplasting til Archivematica, da systemet på visse områder er ganske rigid. For eksempel må sjekksum-rapporten[36] som lages før opplasting være lagret i et bestemt format, og formatteringen inne i denne filen må følge Archivematicas spesifikasjoner. Dette gjorde at vi ikke kunne bruke Fixity AVP, verktøyet Arbark tidligere har brukt til å generere sjekksummer. Arbeidsgruppa fant heller ikke et nytt program som samsvarte med Archivematicas krav. Vi så oss derfor nødt til å lage et eget verktøy (Checksummer). Arbeidsgruppa måtte også lage et eget mappeoppsett for opplasting til Archivematica.

Før prosessering i Archivematica utføres følgende oppgaver:

- Umiddelbart etter mottak lages det førstegangs sjekksummer for originalfilene.
- Arkivar oppretter dokumentasjonsfiler som lagres sammen med arkivdataene: «Depotlogg» og «Notat om avlevering». I depotloggen dokumenterer arkivaren hva som gjøres med arkivet underveis i mottaksprosessen – ordning, sletting av filer, angivelse av hvilken programvare som brukes til de forskjellige jobbene, datering av operasjonene osv. «Notat om avlevering» skal inneholde den samme informasjonen som aksesjonsføres i Asta, og referanser dit.

[36] Se «Trinn 7: Lage nye sjekksummer» i vedlegg 7, «Rutinebeskrivelse for laging av arkivpakker med Archivematica».

- Arkivbeskrivelse ved hjelp av Access to Memory. Arkiv- og aktørbeskrivelse fylles ut inne i systemet, før de eksporteres som EAD.xml og EAC_CPF.xml, og lagres sammen med arkivdataene (i mappen «descriptive_metadata»).
- Dersom filer har blitt slettet, eller det har blitt foretatt ordning, må det lages nye sjekksummer.
- Deretter overføres arkivpakken til Archivematicas transferområde, for opplasting og prosessering i Archivematica.

For detaljert beskrivelse av disse operasjonene, se «Vedlegg 07 – Rutinebeskrivelse for laging av arkivpakker med Archivematica».

6.0. E-post

6.1. E-post – the good and the bad

E-post er i dag en av mange kommunikasjonskanaler. Det er både fordeler og ulemper med e-post. Noen e-postsystemer tillater brukere å redigere lagrede e-poster, som allerede er sendt eller mottatt.[37] Altså kan man gå inn og gjøre endringer i ettertid. Integriteten, et sentralt arkivhensyn, faller da i stor grad vekk. Håndtering av e-post er ofte noe den enkelte bruker står for selv, som kan føre til at e-poster slettes og bevares basert på den enkelte brukers vaner. IT-ansatte og/eller leverandører kan sette regler for hvor lenge e-poster bevares (basert på alder, akkumulering eller maks størrelse på kontoen), uten å ta hensyn til mer arkivorienterte oppbevaringsperioder. Hva som skjer med e-postene til ansatte som slutter, er ofte opp til IT-ansatte. De kan slettes helt eller bevares, men de bevares ikke nødvendigvis lenge. Epostkontoorganiseringen i seg selv kan være problematisk, da den kan føre til at et saksforhold er spredt rundt omkring i ulike e-postkasser uten at det er noen åpenbar måte å sammenkoble e-postene på.[38]

Men e-post har også fordeler. Man kan snakke om «ubevisst arkivdanning»: ansatte gjør 'bare' jobben sin, der det å sende og motta e-post (informasjon/dokumentasjon) er en del av jobben. Samtidig bevarer e-postsystemet sendte og mottatte e-poster automatisk. Systemet bevarer også bevis på samarbeid: Transaksjonene som utgjør samarbeidet blir bevart i e-postsystemet og blir dermed bevis på dette samarbeidet. Man får innblikk i arbeidsprosesser og dokumentasjonen som utveksles.[39] *Public Record Office of Victoria* (Australia) dro følgende konklusjon etter samtaler med etterforskningsbyråer av ulike slag: «Disse organisasjonene påpekte at selv om dokumentene i saksbehandlingssystemene var viktige, fantes de nødvendige bevisene i e-postsystemet.»[40] Sett slik kan e-post definitivt ha en bevaringsverdi.

[37] Waugh 2014.
[38] Waugh 2014.
[39] Waugh 2014.
[40] Findlay 2021.

6.2. E-post i partiene - situasjonsbeskrivelse

De sentrale partiorganisasjonene har egne e-postdomener og offisielle e-postadresser for sine ansatte og tillitsvalgte. Partiene har også en generell e-postadresse. Arkivinstitusjoner har altså mange e-postkontoer å forholde seg til. E-poster arkiveres i liten grad i organisasjonenes fil-/mappearkiv. De lagres på hver enkelt e-postkonto samt på den generelle postadressa. I tillegg bruker stortingspolitikere sine egne kontoer som de får tildelt på Stortinget. Siden e-postkontoer ofte slettes når man slutter i jobben/som tillitsvalgt, går eventuell arkivinformasjon tapt.

I fylkes-/regionpartiene bruker ansatte partiets domene. Det varierer om styremedlemmer og medlemmer på fylkestinget bruker partidomene, privat e-postadresse eller adresse utdelt fra fylket. Med noen unntak blir ikke e-post bevart utenfor e-postsystemet (noen overfører e-post til mappearkivet).

Lokallagene har i svært liten grad egne ansatte, og følgelig har man heller ikke eget domene. De tillitsvalgte bruker private e-postadresser. I noen lag blir vedlegg som blir vurdert som interessante lagret i mappestrukturen.

6.3. Bevaring av e-post – det pragmatiske valget

Bevaring av e-post har lenge vært et problem. Ofte har praksis vært å legge e-post over i et dedikert arkivsystem, et såkalt *separate record repository*,[41] der bruker selv legger til noe metadata, mens andre metadata genereres automatisk. Bruker har selv trolig liten interesse av å legge til metadata og arkivere dem i arkivsystemet, hvis e-postene allerede ligger i e-postsystemet klare for gjenfinning. Dette på tross av at gjenfinning i e-postsystem kan ta tid og slik sett koster ressurser.

Det er rimelig å anta at mye korrespondanse ikke brukes etter en kort saksbehandlingsfase. Her kan man dels snakke om «kast-over-gjerdet»-mentalitet.[42] Man er ferdig, og med noen få unntak trenger man ikke lenger å forholde seg til e-postdokumentasjonen.

[41] Lappin et. al 2021: 252-253.
[42] Riis 2021.

I de politiske partiene vil det heller ikke være snakk om Noark-basert bevaring (i særlig stor grad). Innkjøp av Noark-godkjente systemer vil trolig koste en god del mer enn partiene synes det er verdt. Gitt at den økonomiske støtten svinger noe med valgresultater, kan kostnader til et sak-/arkivsystem falle litt bak på prioriteringslisten.

Der partiene *har* funnet tid og ressurser til et dedikert system, er innen medlemsforvaltning. Årsaken er at partiene er avhengige av å ha kontroll over medlemssituasjonen og kontingentinnbetalinger. Effektive systemer for inn- og utmelding og fakturering er helt nødvendig, fordi oppgavene er så ensartede og volumet så stort. Fagsystem som tilfredsstiller arkivkrav vil da være en såkalt *intervention in business application* (fagsystem med arkiv), et konsept den amerikanske arkivviteren David Bearman stod bak.[43]

Det er noe annet med e-post. Her er det tydelig at partiene ikke har funnet det ressursmessig effektivt å ha dedikerte fagsystemer eller eget sak-/arkivsystem for e-postkorrespondanse. I den grad det er gjort en bedømming her (bevisst eller ubevisst) er konklusjonen at e-post (og andre meldingstjenester) kan håndteres i sitt opprinnelige system. Eventuelt blir en liten andel, hovedsakelig vedlegg, overført til fil-/mappearkivet, som jo i og for seg kan ligne på en sak-/arkivmodell.

En slik ordning går under fellesbetegnelsen *in place recordkeeping*. Man håndterer korrespondansen i sin opprinnelige applikasjon. Tanken er at det ikke finnes kostnadseffektive metoder for rutinemessig overføring til andre strukturer.[44] De to andre modellene ser ikke ut til å passe like bra.

Kunstig intelligens kan i nær framtid gjøre at dette bildet endrer seg, det vil si at man automatisk tar beskjeder med mer fra sine opprinnelige systemer og lagrer dem i en arkivstruktur eller -system tilpasset de hensyn både dokumentasjonsforvaltningen og eventuelt arkivdepotet måtte ha. Bøygen for virksomheter blir da kanskje oppstartskostnaden; er det verdt å implementere kunstig intelligens på for eksempel e-poster? Det er et spørsmål virksomhetene må ta stilling til. Moderne e-postløsninger har allerede noen funksjoner for å understøtte arkivering.[45]

---

[43] Lappin et. al 2021: 252-253 & 263-264.
[44] Lappin et. al 2021: 263.
[45] Se for eksempel: https://www.nucleustechnologies.com/blog/compliance-in-office-365/ og https://learn.microsoft.com/nb-no/microsoft-365/compliance/mip-easy-trials?view=o365-worldwide

6.4. Bevaring og e-post

In place recordkeeping kan sies å være utbredt i privat sektor. Hvor bevisst implementeringen av denne modellen er hos arkivskaperne (her partier), kan diskuteres. Det er ikke usannsynlig at dette bare skjedde i overgangen fra analogt til digitalt arkiv. Fil-/mappearkiv har nok i større grad blitt sett på som «ordentlig arkiv», siden tekstdokumenter var kjent fra papirverdenen. Når det gjelder e-postbevaring må arkivarer stille spørsmålet: Hvilke e-poster skal bevares i en arkivinstitusjon? I en partiorganisasjon er det som andre steder trolig en miks av private og virksomhetsrelaterte e-poster i enhver innboks/utboks. Bevaringsinstitusjonen kan gå for alle e-poster i hele organisasjonen, men her åpner det seg muligheter for å ta vekk kopier (interne e-poster vil være i minst to e-postkasser, i tillegg kan vedlegg også være i fil- og mappearkivet).

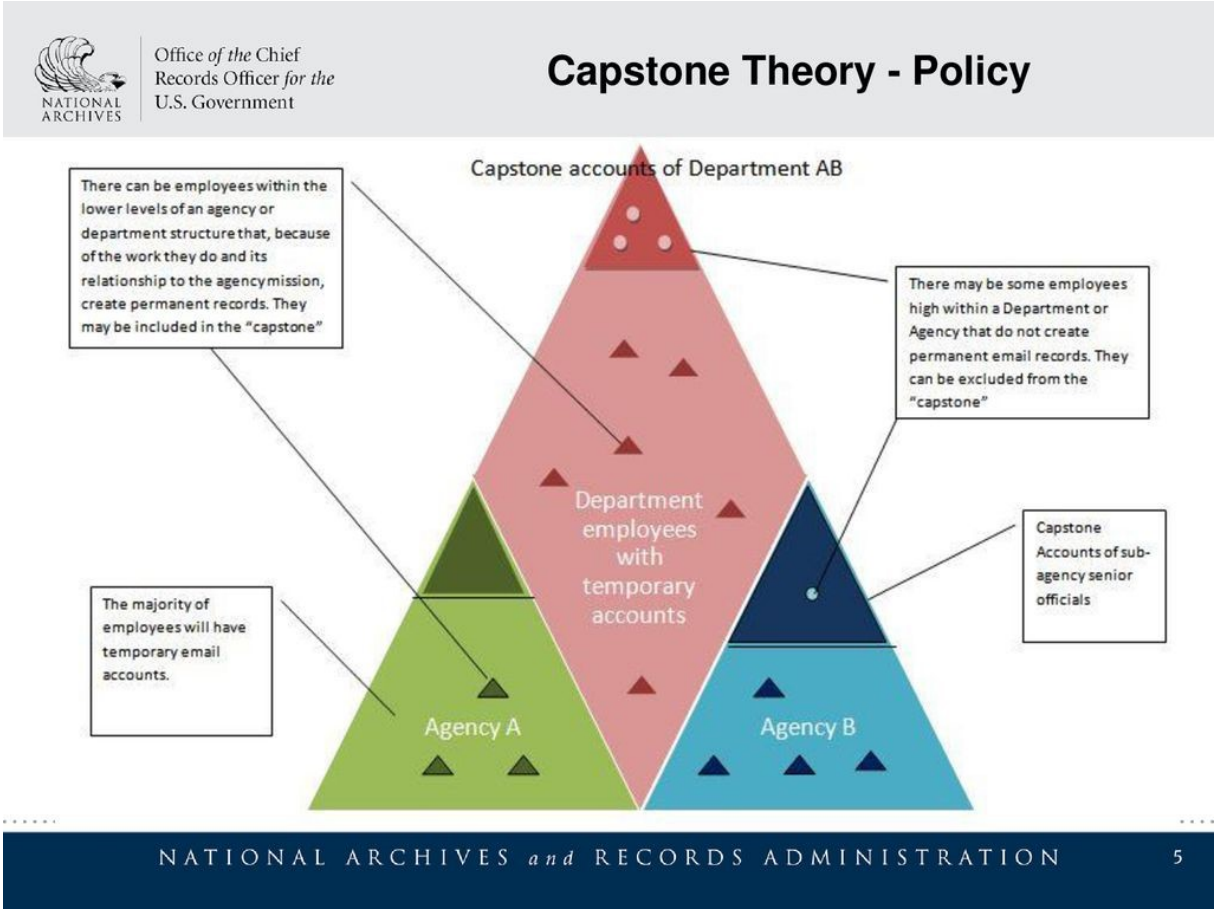
Ved å bevare alle e-poster (eventuelt med arkivbegrensning av private e-poster og midlertidige beskjeder) kan man få en stor mengde som gir et bredt innsyn. Størrelsen i seg selv er ikke et *lagrings*problem. Derimot kan mengden gjøre at den krever mer ressurser, ikke minst med tanke på hardware og kraft, jf. e-avfall og bærekraft, for å forvalte frem til avlevering, vedlikehold i depot og til å søke. Spesielt førstnevnte hensyn er viktig, siden private arkivskapere kan gjøre nærmest som de vil med eget materiale. En ekstra kostnad er ikke ønskelig.

Fra et depotperspektiv er det en investering å opprettholde tilgang og meningsbærende innhold i lang tid. Men det er ikke nødvendig å bevare alt. Nye e-poster vil ikke nødvendigvis gi ny innsikt, men kan repetere informasjon som allerede er kjent. For et arkivdepot er det viktig å ha en relasjon til denne grensen, selv om den nok er «flyttet» med tanke på stordataprosjekter, så depotet kan søke ut ny dokumentasjon fra områder som samlingen ikke dekker så godt.

6.4.1. Capstone

I det foregående har vi mellom linjene argumentert for en mer konsentrert og spisset innsamling av e-post. En konkretisering av en slik strategi er den såkalte Capstone-

metodikken.[46] *Capstone* ble laget av *National Archives and Records Administration* (NARA) i USA, spesifikt Jason Baron, ansvarlig direktør for rettsprosesser/rettstvister (*litigation*).[47] Kjernen i metodikken er å finne nøkkelpersoner («Capstone officials») i en gitt organisasjon som mottar og sender viktig e-post. Disse e-postkontoene blir så valgt ut for langtidsbevaring i depot. Hver utvalgte e-postkonto blir en egen aggregering, en serie om man vil.



Bilde fra NARA som illustrerer *Capstone*.

Arbeidsgruppa vil understreke at en viss deduplisering går an hvis man ser på epostkontoene som noe samlet. Arkivdepotet trenger for eksempel kun den ene e-posten fra en Capstone-konto til en annen, ikke begge to (kopi) som opprinnelig lagres i hver sin konto (ut- og innboks). Det er snakk om virksomhetens e-post, ikke personarkiv.

Basert på en slik metodikk mener arbeidsgruppen å ha identifisert tre typiske kontoer som kan egne seg for bevaring:

1) Partileder

[46] National Archives and Records Administration 2015.
[47] Lappin et. al 2021: 261-263.

2) Nestleder
3) Partisekretær

Til sammen utgjør dette partiledelsen; funksjonene med mest formell makt. Fordi deres maktutøvelse internt i partiet og storsamfunnet ellers har størst konsekvenser i hele organisasjonen, gir det mening å bevare disse kontoene. I tillegg er ledelsen involvert i mange forskjellige saker, man kan si at de ofte opererer på tvers av funksjoner. De deltar i utvalg, jobber med politikkutvikling, organisasjonsutbygging, kommunikasjon med mer. Gitt at en ledergruppe «hopper litt frem og tilbake» med tanke på hva de gjør, gir materialet et godt tverrsnitt av partiorganisasjonens ulike saker og arbeidsprosesser. Da i tillegg til et mer «vanlig» eller «offisielt» fil-/mappearkiv.

Leder og nestleder vil ha det øverste ansvaret for virksomheten. Deres posisjon på toppen er verdt å bevare. De er også ansiktet til partiet utad. Deres tankegang og prioriteringer har mye å si. Partisekretærer har en mer tradisjonell administrativ (leder)posisjon internt, men kan også ha ansvar for politikkutvikling. En slik rolle tilsier da mer innblikk i det indre livet i en organisasjon enn de to andre, som jo må være mer utadrettede, ikke minst mot pressen og opinionen generelt. Her kan man snakke om arkivteoretikeren T. R. Schellenbergs skille mellom *dokumentasjonsverdi*, der man først og fremst er opptatt av fakta om virksomheten og arbeidsprosesser (internt fokus), mens *informasjonsverdi* går mer på fenomener utenfor organisasjoners fire vegger som arkivskaper kommer i kontakt med.[48] Til sammen tilsier dette en bred dekning av ulike aspekter rundt et politisk parti. Dog, skillet her er ikke nødvendigvis skarpt.

Analysen her taler for bevaring av disse tre kontoene (eventuelt fire hvis partiet har to nestledere). De bør ses i sammenheng med tanke på sletting av kopier (mer arkivbegrensning enn ren kassasjon). Her bør også fil-/mappearkivet tas i betraktning. Et typisk eksempel er at utsendelse av dokumenter går til ledelsen via e-post, mens de også finnes i fil-/mappearkivet.

Det bør også være mulighet for fjerning av mer «midlertidig» eller «triviell» informasjon, men her må det utvikles noe mer presise kriterier i samarbeid med arkivskaper. Det kan også være aktuelt med andre kassasjonskriterier hvis arkivskaper og arkivinstitusjon kommer til enighet om slike. Grensen mellom privatperson og politiker kan også være aktuell å se på med tanke på bevaring eller kassasjon. Klausuler er også viktige, fordi de kan gi private arkivskapere ekstra trygghet for at e-poster ikke blir for «åpne» for allmennheten (i alle fall

[48] Valderhaug 2011: 77.

ikke for tidlig). Her kan arkivinstitusjon og arkivskaper sette opp rimelige kriterier, blant annet med utgangspunkt i International Council on Archives yrkesetiske retningslinjer.[49]

6.4.2. Programmer

For å få til et regime for inntak av e-post, har vi testet en del programmer. ePADD har flere sterke sider. Programmet har åpen kildekode og er basert på fire moduler: *appraisal, processing, discovery* og *delivery*. De to første har med *ingest*/mottak å gjøre, de to siste *access*/tilgjengeliggjøring.

ePADD bruker *natural language processing* (NLP) for å lage indekser som brukes i de to siste modulene. Forutsatt at språket er US engelsk plasserer ePADD ting i kategorier. For eksempel blir Yale plassert i universitetskategorien, slik at 'universiteter' blir en kategori eller fasett man kan bruke i søk hvis Yale (og andre universiteter) forekommer som ord i e-postarkivet. NLP blir her da en teknologi som muliggjør uthenting av data fra e-poster for å konstruere automatisk genererte indekser.

Et interessant moment med *discovery*-modulen er at man søker i sladdet materiale. Dette gjør at man kan tilgjengeliggjøre et e-postarkiv på nett, men bare noen elementer eller enkeltord er søkbare, resten er sladdet. Et annet moment mer på inntakssiden er merking av e-poster for klausulering eller kassasjon.

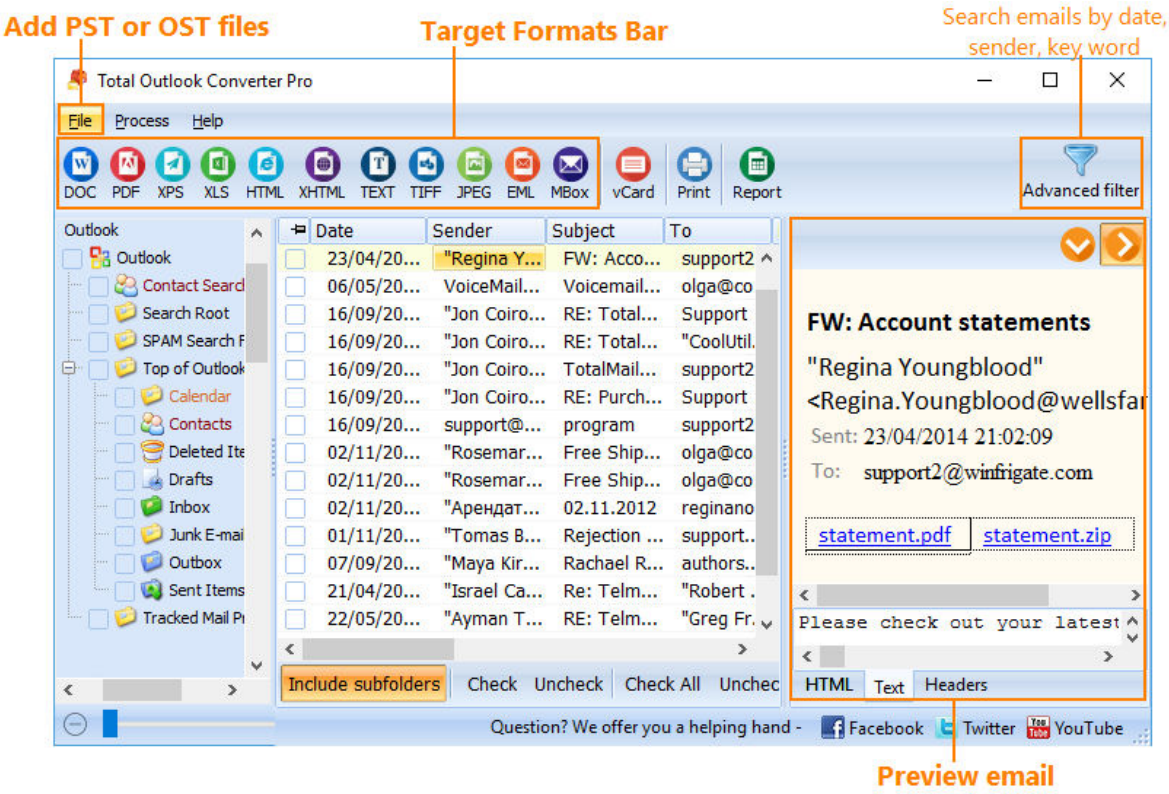
Arbeidsgruppa var også interessert i om *appraisal*-modulen kunne brukes til å utføre mottaks- og ordningsarbeid. Men ved installasjon var det store problemer med stabilitet. I begge periodene vi jobbet med programmet fikk vi feilmeldinger. Etter noen forsøk på å komme til bunns i feilene, ble det tatt en avgjørelse om at arbeidsgruppa ikke kunne bruke flere ressurser på programmet. Selv om vi synes ePADD er interessant, hjelper ikke det når programmet ikke har den nødvendige driftsstabiliteten.

Likevel kan NLP være interessant for arkivsektoren. En fordel med kategorier/fasetter er at brukeren ikke må finne søkeord selv. Visualisering er et annet interessant moment: ePADD kunne vise grafer over korrespondenter over tid, så man fikk et inntrykk av hvem mottakeren kommuniserte mest og minst med.

[49] International Council on Archives 1996.

Et annet program vi har hatt betydelig mer suksess med, er Total Outlook Converter Professional (TOCP) fra CoolUtils. Programmet har et mindre virkefelt enn ePADD. Der ePADD er et helthetlig system for e-postbevaring og tilgjengeliggjøring, dreier TOCP seg kun om eksport og konvertering fra Outlook-kontoer. Outlook-datafiler (pst-filer[50]) konverteres til PDF/A. Dette inkluderer vedlegg og lenker (siden den lenker til blir bevart, men siden kan forandres eller tas vekk mellom tidspunktet e-posten blir sendt/mottatt og konvertert). TOCP kan konvertere vedlegg separat eller konvertere slik at vedlegg og e-post blir én datafil. Selv om TOCPs funksjoner er nedskalert i forhold til ePADD, er programmet lettere å bruke. Det kan også ekstrahere ut såkalt intern e-postmetadata (les: er i filen) som forsvinner ved konvertering til PDF/A. Det er absolutt en styrke i den grad man vil gjøre noe med slik metadata. I tillegg er programmet relativt rimelig med tanke på lisenspris.[51]

At TOCP er proprietær programvare og kun håndterer Microsofts format for e-post (Outlook), er et minus. Samtidig er det i stor grad Microsofts programvare som brukes i Norge, noe som gjør denne begrensningen litt enklere å akseptere.



Bilde fra *CoolUtils*. *Targets formats* sier hva man kan konvertere til.

[50] Personal Storage Table-filer lagrer personlig informasjon fra e-postsystem, slik som meldinger og kontakter.
[51] 160 US dollar per lisens, andre lisensvalg kan gjøres.

6.4.3. Bevaringsverdig e-post i mindre partiorganisasjoner – en anbefaling

De fleste arkivskaperne som omfattes av dette prosjektet bevarer per i dag ikke e-post på annen måte enn i selve epostsystemet. Årsaken er sannsynligvis at det er søkbart så lenge det ligger i systemet, og dermed mest lettvint. Samtidig kan det være krevende å gjøre en bevaringsvurdering, både rent arkivfaglig, men også med tanke på tidsbruk.

Den enkleste måten å bevare e-post på, er overføring til mappestruktur. Det er i hvert fall arbeidsgruppas vurdering. En del gjør da også dette i dag. For mindre partilag anbefaler vi derfor at e-post som er en viktig del av en styresak kopieres inn i et saksdokument. Dokumentet kan for eksempel være i Word-format. Både teksten i e-posten og metadata (avsender/mottaker, tittel, tidspunkt osv.) må med. Dokumentet kan ligge som vedlegg til protokollen i mappearkivet. Eventuelt kan deler av e-posten limes rett inn i protokollen.

For større organisasjoner vil Capstone-metodikken være mer hensiktsmessig.

7.0. Sosiale medier

De politiske partiene er utadvendte organisasjoner som bruker sosiale medier aktivt. Det innebærer at det ligger arkivverdig informasjon både på Facebook, Twitter og Instagram. Arkivverket har jobbet en stund med datahøsting fra sosiale medier. Arkivverket samarbeider med medieovervåkningstjenesten Retriever om å dokumentere de politiske partienes aktivitet på Facebook, Twitter og Instagram.[52] I tillegg til partiorganisasjonene høster de fra de personlige (offentlige) profilene til partilederne. Etter hvert har man også begynt å samle inn fra stortingsrepresentantene.[53] Informasjonen høstes regelmessig.

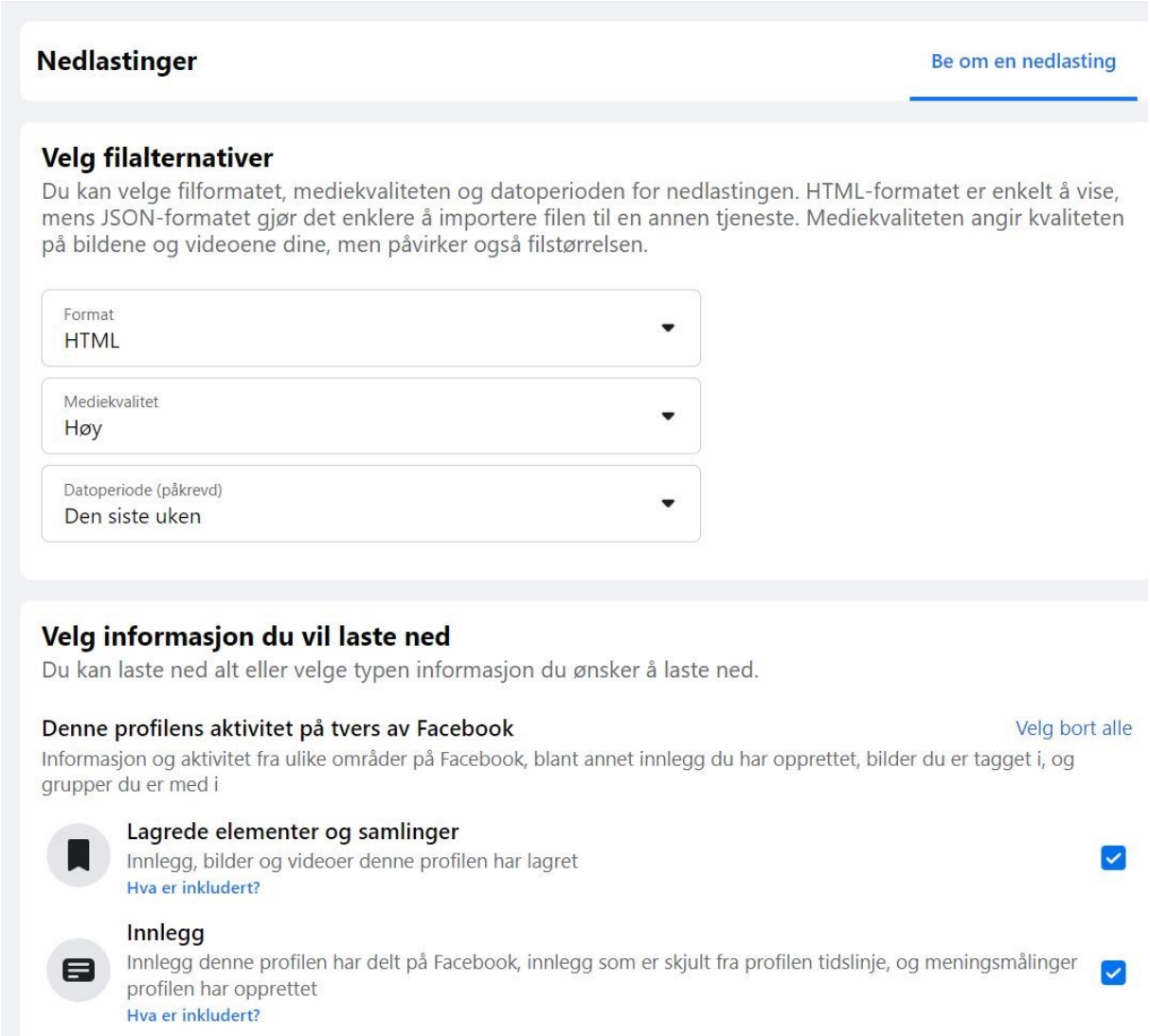
Det er svært positivt at Arkivverket gjør denne jobben. MiA, NIA og Arbark har ikke ressurser til å drive løpende høsting. Men siden Arkivverket kun høster fra sentrale partiorganisasjoner, har det vært et mål i vårt prosjekt å finne metoder for høsting og sikring av lokale organisasjoners SoMe-kontoer. Slik kan vi gi et bidra til en samlet samfunnsdokumentasjon. Utover prosjektet kan slike metoder selvsagt også anvendes på andre arkivskapere vi har ansvar for.

7.1. Nedlasting fra nettleser

Data fra SoMe-kontoer kan enkelt eksporteres direkte fra nettleser. For prosjektdeltakerne er slike eksporter den mest realistiske måten å få tak i data på. Nedlasting er passordbeskyttet, slik at man må være logget på som administrator. Både Facebook, Instagram og Twitter har slike nedlastingsmuligheter, og både nedlastingsprosess og resultat er svært lik. Facebook-data er likevel mer komplekse, så arbeidsgruppa har derfor fokusert mest på dette.

Ved nedlasting fra Facebook kan man velge format: Enten JSON, som er godt egnet for langtidslagring, eller visningsformatet HTML. Ved nedlasting velger man også hva som skal med i eksporten (innlegg, videoer, kommentarer osv.). Det er mulig å sette datointervall.

[52] Arkivverket 2017a.
[53] Arkivverket 2019.



Eksport fra Facebook kommer som en zip-fil. Ved utpakking foreligger en mappestruktur med dataene fra profilen:

Navn	Endringsdato
activity_messages	05.10.2022 12:01
bonuses	05.10.2022 12:01
collaborations	05.10.2022 12:01
comments_and_reactions	05.10.2022 12:01
community_profiles	05.10.2022 12:01
events	05.10.2022 12:01
facebook_gaming	05.10.2022 12:01
files	05.10.2022 12:01
friends_and_followers	05.10.2022 12:01
fundraisers	05.10.2022 12:01
groups	05.10.2022 12:01
live_audio_rooms	05.10.2022 12:01
messages	05.10.2022 12:01
other_activity	05.10.2022 12:01
pages_and_profiles	05.10.2022 12:01
places	05.10.2022 12:01
polls	05.10.2022 12:01
posts	05.10.2022 12:01
profile_information	05.10.2022 12:01
saved_items_and_collections	05.10.2022 12:01
short_videos	05.10.2022 12:01
stories	05.10.2022 12:01
index.html	05.10.2022 12:01

Nedlasting via nettleser er altså enkel måte å få tak i dataene på. Men metoden har sine svakheter. Man må ha en egen organisasjonsprofil. Dersom man laster ned fra privat profil kommer det med persondata som sannsynligvis ikke skal bevares i arkivdepot. Samtidig er det en viss fare for at en del mindre organisasjoner ikke har egne organisasjonsprofiler, men at sideadministratorer logger på sin vanlige private profil og opererer derfra. I slike tilfeller får man ved eksportering en sammenblanding av privat info og organisasjonens data.

Gjennom prosjektperioden gjorde arbeidsgruppa flere tester med nedlasting fra egne private Facebook-profiler, samt institusjonsprofiler. Våren 2020 kom ikke kommunikasjon med eksterne med, kun egne (private/institusjonens) kommentarer. Men i nye tester høsten 2022 ble kommentarer fra eksterne med. Slike endringer skjer etter Facebooks eget forgodtbefinnende. Ved nedlasting må derfor innholdet inspiseres. Når det på et tidspunkt blir aktuelt å tilgjengeliggjøre Facebook-data, må institusjonene vurdere hva som faktisk kan gjøres tilgjengelig, og hvilke tilgangsbegrensninger som eventuelt skal gjelde.

Det er uvisst om lukkede grupper (altså grupper med egne administratorer) kommer med, ettersom vi ikke har testet på profiler med slike. Når det gjelder innholdet i nedlastingene er arbeidsgruppas vurdering at det uansett er innleggene («posts») som er mest verdifulle. Disse er en viktig del av organisasjonens virksomhet utad, og denne funksjonen må bevares for ettertiden.

Siden eksporten kommer som en mappestruktur, går det selvsagt an å slette mapper som ikke trenger å bevares *etter nedlasting* (foreta arkivbegrensing). Det vil sannsynligvis uansett være lurt, da flere mapper er tomme (kun inneholder en ren tekstfil, «no-data.txt»).

For Arbark, NIA og MiA er disse eksportene – både fra Facebook, Twitter og Instagram – den enkleste måten å høste på. Avtale med arkivskaper kan inngås i forbindelse med at annet materiale avleveres. Arkivskaper kan enten gjøre nedlastingen selv, eller med bistand fra oss i depot.

For beskrivelse av nedlasting fra Facebook, se «Vedlegg 10 – Fremgangsmåte ved Facebook-nedlasting». Merk at Facebooks visuelle grensesnitt til stadighet endres, slik at slike veiledninger raskt blir utdaterte.

7.2. Automatisk netthøsting – *Webrecorder*

Selv om nedlasting via nettleser er prosjektets foretrukne løsning, har arbeidsgruppa også sett litt på netthøsting. Vi har prøvd ut flere av verktøyene fra Webrecorder-prosjektet.[54] Webrecorder er en rekke open source-programmer som kan brukes til å høste nettsider og gjenskape dem. Alle er tilgjengelige på plattformen GitHub. Medlemmer i arbeidsgruppa har prøvd ut ArchiveWeb.page-programmet, ReplayWeb.page og Browsertrix Crawler.[55]

ArchiveWeb.page er en pakkeløsning som muliggjør høsting direkte gjennom nettleser. Dataene som høstes lagres i WARC- eller WACZ-formatet lokalt på maskinen.[56] Etter nedlasting kan sidene gjenskapes ved hjelp av ReplayWeb.page. Browsertrix Crawler kan på sin side brukes til automatisk høsting.

---

[54] https://webrecorder.net/
[55] https://webrecorder.net/tools
[56] For mer om disse formatene, se linker: WARC (Web ARChive): https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml#:~:text=The%20WARC%20format%20is%20a,from%20the%20World%20Wide%20Web; WACZ (Web Archive Collection Zipped): https://github.com/webrecorder/wacz-format

Webrecorder er interessant som netthøstingsverktøy, men er sannsynligvis ikke ideelt til høsting av sosiale medier, blant annet fordi det innebærer en del manuelt arbeid. Browsertrix Crawler-verktøyet kan høste automatisk, men det krever kjennskap til kommandolinje og Docker-teknologi,[57] noe som er problematisk. Vi har foreløpig heller ikke erfaring med lagringsformatet WARC.

7.3. DigiFacket-prosjektet

Arbarks svenske søsterinstitusjon, Arbetarrörelsens arkiv och bibliotek,[58] har deltatt i et prosjekt hvor man har samlet inn nettsider og sosiale medier. Medlemmer i Grasrota-prosjektets arbeidsgruppe dro i mai 2022 til Sverige for å høre om prosjektet, og om dette var arbeid vi kunne dra nytte av.

I prosjektet, DigiFacket, ble det utviklet software som automatisk ved gitte intervaller henter og lagrer informasjon fra fagbevegelsens nettsider og sosiale medier. Deretter blir materialet indeksert slik at det blir søkbart og mulig å bruke i forskningssammenheng. Denne softwaren har nå institusjonene som deltok i prosjektet, deriblant svenske Arbark, tilgjengelig. Men i dag er det i realiteten kun nettsider som høstes, ikke SoMe-informasjon. Det viste seg nemlig at crawlerne (programmene som foretar høstingen) stoppes av Facebook, som krever passord ved datainnsamling.

På samme studietur besøkte vi også Tjänstemän & akademikers arkiv (TAM Arkiv), en annen arkivinstitusjon som bevarer arkiv fra svensk fagbevegelse.[59] De deltok også i DigiFacket-prosjektet, men har senere utviklet en egen løsning for automatisert høsting. Også deres løsning høster kun nettsider, av samme grunn som svenske Arbark. Når det gjelder SoMe har TAM-arkivet basert seg på samme strategi som oss i Grasrota-prosjektet: Arkivskaper laster selv ned SoMe-data via nettleser. TAM-arkivet har laget en skriftlig veiledning for nedlasting, som er publisert i det svenske tidsskriftet *Arkiv* (Arkiv nr. 2/2022).

[57] https://en.wikipedia.org/wiki/Docker_(software)
[58] https://www.arbark.se/sv/
[59] http://www.tam-arkiv.se Mens svenske Arbark bevarer arkiv etter organisasjoner innenfor det svenske LO-systemet, har TAM Arkiv bevaringsansvar for fagorganisasjoner utenfor LO.

Dette gjorde oss enda sikrere på at vi har valgt riktig strategi. Følgelig er konklusjonen at arkivskaper med administrator-rettigheter selv foretar nedlasting fra sosiale medier (eventuelt med bistand fra oss).[60]

[60] Framover kan det for Arbarks del likevel bli aktuelt å se mer på netthøsting. Det er nemlig visse typer nett-materiale Arbark med fordel kunne høstet på egen hånd, selv om Nasjonalbiblioteket som kjent gjør dette. Særlig interessant er materiale man tradisjonelt har karakterisert som 'trykksaker', altså dokumenter som arkivskapere legger ut på nettsidene sine.

8.0. Digitalarkivet

I søknaden skrev vi at prosjektet skulle jobbe med tilgjengeliggjøring av digitale arkiver. Planen var å teste ut løsninger for tilgjengeliggjøring mot slutten prosjektet, etter at mottaksløsninger var utprøvd. Her var selvsagt Digitalarkivets modul interessant, da både Arbark, NIA og MiA er avhengige av fellesløsninger som Digitalarkivet for å nå ut til arkivbrukere. Det viste seg imidlertid at vi var for tidlig ute; visningsmodulen var ikke klar til testing med mappestrukturer innenfor vår prosjektperiode. Dermed får tilgjengeliggjøring i Digitalarkivet heller bli utforsket i et senere prosjekt.

Bevaringsløsningen til Digitalarkivet derimot ble utviklet raskere enn forutsett. Siden Digitalarkivet/Arkivverket var representert i referansegruppa, begynte vi å diskutere muligheten for å prøve ut bevaringsmodulen innenfor prosjektperioden.

8.1. Langtidsbevaring – Archivematicas begrensninger

Archivematica kan ikke alene brukes som bevaringsløsning. Selv om prosjektdeltakerne har fått erfaring med Archivematica som mottakssystem, løser ikke det alle utfordringer knyttet til langtidsbevaring. Som vi har beskrevet i denne rapporten egner Archivematica seg til å lage bevaringspakker, men systemet har ingen tilfredsstillende løsning for bevaring, for eksempel når det gjelder overvåking av filformater. *Archival storage* i Archivematica er egentlig bare et lager der pakkene havner, og på ingen måte godt nok som fullverdig digitalt magasin. Archivematica er heller ikke ment å fungere som magasinløsning; det er derimot meningen at systemet integreres med et bevaringssystem, for eksempel ArchivesSpace, jfr. Archivematicas brukermanual.

I prosjektet har vi i stedet eksportert arkivpakkene fra *Archival storage* og lagret dem som zip-filer i Arbarks digitale magasin. (Prosjektets styringsgruppe ble tidlig enige om at NIA og MiA kunne bevare arkiver mottatt i prosjektperioden i Arbarks digitale infrastruktur.) Men Arbarks magasin er langt fra ferdigutviklet. Samtidig er dette kun ment å fungere på mellomlang sikt, altså i en del år, inntil en mer permanent fellesløsning kommer på plass. Digitalarkivet er en slik permanent fellesløsning.

8.2. Testing av arkivopplasting til Digitalarkivet

Siden alle tre institusjoner kommer til å benytte seg av Digitalarkivet, ønsket arbeidsgruppa å teste opplasting, nå som løsningen faktisk var på plass. Vi ble enige med Digitalarkivet/Arkivverket om et pilotprosjekt, der arbeidsgruppa skulle få anledning til å gjøre seg kjent med opplastingsløsningen, og deretter laste opp reelle arkivdata.

På møter med Digitalarkivet ble det enighet om følgende: Arbeidsgruppa skulle få tilgang til et eget testområde i Digitalarkivets produksjonsløsning, der vi kunne foreta testoverføringer over en periode. I denne perioden skulle det kun lastes opp konstruerte testdata, ikke reelle arkiver. Konkret innebar dette workshoper i arbeidsgruppa, hvor vi lastet opp så mange arkivpakker vi ønsket. Dermed fikk vi viktig erfaring med overføringsprosesser og nødvendige verktøy.

I forberedende møter med Digitalarkivet ble trinnene i overføringen gjennomgått. Arbeidsgruppa fikk også tilsendt et dokument med en detaljert beskrivelse av prosessen. Overføring til Digitalarkivet forutsetter nedlasting og bruk av verktøyene Uploader og Arkade. Uploader er programmet som brukes til selve opplastingen, samme program som brukes for opplasting av digitaliserte dokumenter. Arkade er Arkivverkets egetutviklede verktøy for testing av arkivuttrekk. I denne sammenheng er det viktige at programmet kan opprette arkivpakker i Digitalarkivets påkrevde pakkeformat. Før opprettelse av pakke legger arkivaren inn data om pakken, slik som navn på arkivskaper, tidsperiode for arkivet, ansvarlig arkivinstitusjon, med mer. Arkade er gratis og kan lastes ned fra Arkivverkets nettsider.

Selve overføringen er veldig enkel og intuitiv: Først brukes Arkade til å pakke arkivene i riktig format, DIAS-formatet. Arkade genererer i tillegg til selve arkivpakken en «overføringsfil» (dvs. en XML-beskrivelse av pakken), som sendes som e-post-vedlegg til en kontaktperson i Arkivverket. Etter at denne har kontrollert informasjonen, mottar man en invitasjonslenke. Arkivet kan så overføres med Uploader. Etter fullført opplasting sender Digitalarkivet en bekreftelse via e-post om at overføringspakke er lastet opp. Dokumentet vi fikk av Digitalarkivet/Arkivverket gir en god beskrivelse av disse stegene.

8.3. Opplasting av reelle data til bevaring i Digitalarkivet

Gjennom prosjektet avleverte flere partilag sine digitale arkiver. Ved mottak ble disse behandlet med Archivematica. Avtalen med Digitalarkivet var at disse skulle overføres til Digitalarkivet før prosjektets avslutning. Før overføring lagde styringsgruppa en avtale med Digitalarkivet/Arkivverket om bevaring av disse arkivene.

Som nevnt er selve opplastingsjobben veldig enkel. Det problematiske var pakkeformatet: DIAS-formatet er en forutsetning for opplasting til Digitalarkivet. Archivematica-pakkene er på sin side basert på BagIt-formatet. Siden formatene ikke er kompatible kunne vi ikke overføre Archivematica-behandlede pakker direkte til Digitalarkivet. Én mulighet hadde vært å laste opp «rådataene» (SIP), dvs. de originale arkivpakkene slik de ble avlevert av lokallagene. Men dette anså arbeidsgruppa som uaktuelt, da vi ikke ville få med metadata, arkivbeskrivelser og Asta-referanser fra mottaksprosessen. Konvertering til arkivformat inngikk heller ikke i avtalen med Digitalarkivet, så arkivdokumentene ville blitt langtidsbevart i produksjonsformater, og det ville vært uheldig rent arkivfaglig. Vi ønsket med andre ord å laste opp AIPene.

For å få det til ville vi forsøke å lage en automatisert ompakkingsrutine av BagIt/Archivematica-pakker til DIAS. Det innebar å utarbeide et skript for ompakking der denne jobben kunne skje automatisk, ved at metadata-filer og mapper fra BagIt ble plassert i en DIAS-struktur. Før skriptet ble laget gjorde vi en sammenligning av pakkeformatene, og vurderte hvor data fra BagIt passet inn i DIAS. Dette var ikke en stor jobb: Ettersom begge er digitale pakkeformater, har BagIt og DIAS i utgangspunktet en del likheter. I tillegg ligger prosjektets versjon av BagIt nært opptil DIAS, siden arbeidsgruppa har sett til Arkivverkets DIAS-rapport og løsningene derfra.

Da ompakkingsrutinen var utarbeidet ble den prøvd ut i Digitalarkivets testområde. I testingen gikk alt etter planen – DIAS-pakkene fikk overført nødvendige data fra BagIt, disse ble plassert riktig, samtidig som overflødige filer ble fjernet. Men ved endelig opplasting gikk det ikke helt som vi ønsket: Etter ompakking med Arkade ble en tar-fil plassert inne i content-mappen. Denne tar-filen inneholder Archivematicas DIAS-lignende mappestruktur (altså den skript-genererte filen). Resultatet ble altså litt mer komplekst enn ønskelig. Men alle dataene er med, så pakkene kan uansett være gode utgangspunkt for tilgjengeliggjøring (når det eventuelt blir aktuelt). Etter prosjektet vil Arbark fortsette å se mer på denne ompakkingen.

For teknisk informasjon om ompakking, se vedlegg 12 og 13.

9.0. Avslutning

9.1. Vilje til avlevering

Gjennom prosjektperioden var det vanskelig å få arkivskaperne med på avlevering. Under kartleggingen var flere positive, men etter hvert så de ut til å miste interessen. Det var kun de mest interesserte som var villige til å levere digitalt skapt materiale. Slik sett kan arkivbevaring bli veldig personavhengig. Det kan være flere grunner til at arkivskapere nøler med avlevering. Vi går ikke inn på dette i detalj her, men det er såpass interessant at noen mulige årsaker bør nevnes.

I mindre organisasjoner er ikke nødvendigvis interessen for arkiv veldig stor. Folk som har verv på frivillig basis har også begrenset med tid til arkivspørsmål. Kompetansen varierer, slik at mange nok er usikre på avleveringsprosessen og om arkivet er «i god nok stand» til å leveres til arkivinstitusjon. Manglende tiltro til at eget arkiv har interesse for ettertiden, kan også være en faktor.

Uansett kan vi konstatere at det var vanskeligere enn antatt å få tak i arkivene. Sannsynligvis vil dette fortsette å være en utfordring for institusjoner som våre. Per februar 2023 jobber medlemmer i arbeidsgruppa og styringsgruppa med en artikkel som tar for seg problematikken. Artikkelen skal etter planen publiseres i Norsk museumstidsskrift.

9.2. Metoder for mottak

Det sentrale i prosjektet har vært kompetansebygging. Deltakerne skulle gjøres i stand til å begynne med mottak av digitale arkiver. Metodene måtte være forankret i arkivfaglige prinsipper og basere seg på anerkjente standarder. Som konkret løsning var Archivematica tidlig en åpenbar kandidat. Arbeidsgruppa jobbet derfor mye med å bygge opp rutiner rundt dette systemet. Sent i prosjektet ble det klart at Digitalarkivet også var klar til å brukes til mottak og bevaring. En klar forskjell på metodene er at med Archivematica gjør depotarkivar mye av mottaksarbeidet selv, slik at man har full kontroll med prosessen. Mens med Digitalarkivet overlates tekniske operasjoner til Arkivverket.

Konklusjonen for Arbark sin del er å fortsette med Archivematica som mottakssystem også etter prosjektperioden. Digitalarkivet kan likevel brukes som digitalt magasin for arkivpakkene som lages med Archivematica, selv om det må jobbes mer med konvertering fra BagIt til DIAS. For MiA og NIA sin del er det derimot aktuelt å bruke Digitalarkivet både som mottaksløsning og magasin. Erfaringene fra prosjektet er at begge disse løsningene fungerer godt.

9.3. Bevaring og kassasjon

I kartleggingsdelingen ble det tydelig at arkivskaperne ønsket veiledning angående bevaring og kassasjon. De to b/k-veiledningene som ble utarbeidet, er konkrete verktøy som vil bli brukt av institusjonene i forbindelse med ordning. Kanskje vil partilagene selv benytte dem når de skal avlevere til Arbark, MiA eller NIA. Forhåpentligvis kan veiledninger motvirke arkivskapernes vegring mot å levere fra seg arkiv.

9.4. E-post

Å finne løsninger for e-postbevaring har som ventet vært utfordrende. Særlig det at mange bruker privat e-post er problematisk. Det vil ikke bli aktuelt å høste fra private kontoer. Det enkleste vil være å komme med tydelige anbefalinger om å bevare arkivverdig e-post i mappearkivene. I hvilken grad dette vil etterfølges, vil nok variere. Mye e-post i mindre organisasjoner vil nok aldri komme til arkivdepot.

For større organisasjoner er Capstone-metodikken nyttig. Det er ikke vanskelig å finne nøkkelroller i politiske organisasjoner. Utover prosjektet kan metoden også brukes på andre typer organisasjoner.

Det er betryggende at det finnes programvare som kan brukes til å håndtere e-post. Total Outlook Converter er brukervennlig og konverterer e-post og vedlegg til arkivformat. Selv om programmet begrenser seg til Outlook, finnes det tilsvarende programmer for andre e-postsystemer.

9.5. Sosiale medier

Det er ikke aktuelt for våre institusjoner å høste sosiale medier løpende. I hvert fall ikke foreløpig. Den beste metoden her er at administrator logger seg på profilen og laster ned dataene manuelt. Slik har vi altså en løsning for høsting og bevaring. Tilgjengeliggjøring av SoMe-data er fortsatt uavklart.

9.6. Samarbeid og framdrift

Prosjektet har vært et samarbeid mellom tre institusjoner; én arkivinstitusjon og to museer. Som en middels stor arkivinstitusjon har Arbark kapasitet til å jobbe med digitale arkiver. Før prosjektet hadde Arbark over flere år jobbet med dette, som konsekvens av at arkivskaperne har blitt fullelektroniske. MiA og NIA hadde før prosjektet få ressurser til å jobbe med disse spørsmålene. Noe av gevinsten med et tett samarbeid har vært at Arbark har kunnet dele sin kompetanse med andre institusjoner.

At samarbeidspartnerne har kunnet jobbe dedikert med el-arkiv i tre år, har ført til ytterligere kompetanseheving: Vi har fått større forståelse for de IT-tekniske og de grunnleggende arkivfaglige utfordringene vi står overfor. I tillegg har selvsagt utprøving av forskjellige metoder og systemer vært viktig. Det har vært naturlig å installere «tyngre» systemer hos Arbark, slik som ePADD, Archivematica og RODA. MiA og NIA har fått fjerntilgang til flere av disse. Testing har vært nyttig, selv om programmene ikke nødvendigvis har blitt del av institusjonenes verktøykasse.

Koronapandemien fikk en viss innvirkning på prosjektets framdrift. Det var først og fremst kartleggingen som ble rammet. Å få til møter med partiene var en periode vanskelig. Men prosjektet var såpass bredt anlagt at arbeidsgruppa kunne gjøre andre ting imens. Det ble for eksempel lagt innsats i metodeutvikling og utprøving av programvare. Slik sett var det hele tiden progresjon, på tross av koronakrisen. Men gjennom hele prosjektet la pandemien begrensninger på reisevirksomheten. Det ville nok blitt flere ekskursjoner i normale tider.

Oppsummert har samarbeidet mellom prosjektdeltakerne fungert svært bra. Å diskutere og sparre med andre som holder på med det samme har vært verdifullt. Det kan senere bli aktuelt å sette i gang nye samarbeidsprosjekter.

9.7. Formidling

Det har vært en prioritet å formidle resultater og funn fra prosjektet i relevante fora. I prosjektperioden har vi holdt følgende presentasjoner og innlegg:

- Webinar i regi av Arkivverket («Bevaring av digitalt skapte privatarkiv – praksis og erfaringer»), 23. mars 2021.
- Årsmøte i Arbeiderbevegelsens arkiv i Telemark, 18. september 2021.
- Nettverksmøte for Finnmarksarkivene, 27. januar 2022.
- Styremøte i Bærum arbeiderparti, 7. mars 2022.
- Erfaringsutveksling med Arbetarrörelsens arkiv och bibliotek (12. mai 2022) og TAM Arkiv (13. mai 2022), Stockholm.
- Privatarkivkonferansen i Bergen, 15. november 2022.
- Nettverksmøte for Finnmarksarkivene, Tromsø, 26. januar 2023.

Litteratur

Arkivverket (2017a, 24. august). *Vi høster politiske partiers aktivitet på sosiale medier* (pressemelding). www.arkivverket.no. Lenke hentet 7. oktober 2022: https://www.arkivverket.no/nyheter/vi-hoster-politiske-partiers-aktivitet-pa-sosiale-medier

Arkivverket (2017b, 5. juni). *DIAS – Digital arkivpakkestruktur*. www.arkivverket.no. Lenke hentet 16. januar 2023: https://www.arkivverket.no/forvaltning-og-utvikling/regelverk-og-standarder/dias-prosjektet-digital-arkivpakkestruktur

Arkivverket (2019, 4. april). *Høsting av sosiale medier (pilot SoMe)*. www.arkivverket.no. Lenke hentet 7. oktober 2022: https://www.arkivverket.no/arkivutvikling/utviklingsprosjekter/avsluttede-prosjekter/hosting-av-sosiale-medier

Findlay, C. (2021). Recordkeeping theory, models & strategies and today's workplace. *Recordkeeping Roundtable*. Lenke hentet 7. oktober 2022: https://rkroundtable.org/2021/04/11/recordkeeping-theory-models-strategies-and-todays-workplace/#more-87762

International Council on Archives (1996). *Yrkesetiske retningsliner for arkivarar*. Lenke hentet 7. oktober 2022: https://www.ica.org/sites/default/files/ICA_1996-09-06_code%20of%20ethics_NO.pdf

Lappin, J., Jackson, T., Matthews, G. & Ravenwood, C. (2021). Records management models in an era of partial automation. *Archival Science 21*, 243-266. https://doi.org/10.1007/s10502-020-09354-9

Lavoie, B. (2014). *The Open Archival Information System (OAIS) Reference Model: Introductory Guide (2nd Edition)*. Lenke hentet 7. oktober 2022: https://www.dpconline.org/docs/technology-watch-reports/1359-dpctw14-02/file

National Archives and Records Administration (2015). *White Paper on The Capstone Approach and Capstone GRS*. www.archives.gov. Lenke hentet 7. oktober 2022: https://www.archives.gov/files/records-mgmt/email-management/final-capstone-white-paper.pdf

Riis, L. (2021, 10. oktober). *Hva kreves av framtidens arkivarer?* www.documaster.com. Lenke hentet 7. oktober 2022: https://www.documaster.com/no/blogg/fremtidens-arkivar

Riksarkivaren (2012). *Vel bevart! Rapport fra samarbeidsprosjektet DIAS – Digital arkivpakkestruktur*. Lenke hentet 7. oktober 2022: https://www.arkivverket.no/forvaltning-og-utvikling/regelverk-og-standarder/dias-prosjektet-digital-arkivpakkestruktur

Valderhaug, G. (2011). *Fotnote eller tekst? Arkiv og arkivarar i det 21. hundreåret*. ABM-Media AS.

Waugh, A. (2014). Email – a bellwether records system. *Recordkeeping roundtable*. Lenke hentet 7. oktober 2022: https://rkroundtable.org/2014/06/30/email-a-bellwether-records-system/

Lover og forskrifter

Riksarkivarens forskrift (2017). *Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver* (FOR-2022-01-27-132). Lovdata. https://lovdata.no/dokument/SF/forskrift/2017-12-19-2286?q=riksarkivarens%20forskrift