Digitalt og autentisk

Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale

Kortversjon av prosjektrapport 01.03.2010

Riksarkivet 17.08.2010

Innhold

1. INNLEDNING

Dette er en forkortet versjon av rapporten fra prosjekt "Elmag-2" – Planlegging av *el*ektronisk *mag*asin for det digitalt skapte arkivmaterialet i Arkivverket. Rapporten ble framlagt 01.03.2010, og finnes i fullstendig form med 4 vedlegg[1] på Arkivverkets nettside[2]

Rapporten dannet basis for Riksarkivarens søknad om ABMU-midler til prosjektet "Arkivpakkestruktur for digitalskapt arkivmateriale", som ble startet i april 2010 i samarbeid med Oslo byarkiv, Bergen byarkiv, IKA Trøndelag og IKA Møre og Romsdal. Samarbeidsprosjektet med disse fire kommunale aktørene har nå fått navnet DIAS – Digital arkivpakkestruktur. Dette samarbeidsprosjektet har to hovedmål:

1) å definere en omforent struktur for arkivpakker som dekker behovene både på statlig og kommunal sektor
2) å utforme en kravspesifikasjon til et system som kan generere og vedlikeholde arkivpakker med en slik struktur.

DIAS-prosjektets vurdering av arkivpakkestruktur tar utgangpunkt i kap. 2 og 3.1 i denne rapporten. Disse delene tilsvarer kap. 2 og kap. 4.2 i den fullstendige Elmag2-rapporten fra 01.03.2010. Både den opprinnelige og den foreliggende forkortede versjonen av Elmag2-rapporten behandler imidlertid hele bredden av temaer med relevans for et digitalt depot:

- de fysiske lokalene, utrustningen og den øvrige materielle infrastrukturen for å lagre, administrere og vedlikeholde digitalt arkivmateriale,
- de digitale arkivobjektene som lagres og vedlikeholdes på installasjonene,
- rutineoppleggene for å bruke og vedlikeholde installasjonene, de administrative systemene og de lagrede objektene,
- prosessene knyttet til mottak og innlemmelse av digitalt arkivmateriale samt system- og rutineopplegget for å styre tilgang til lagret digitalt arkivmateriale

Denne forkortede versjonen av Arkivverkets Elmag2-rapport er tilpasset med sikte på DIAS-prosjektet, men bare ett av hovedkapitlene i den fullstendige Elmag2-rapporten er utelatt i kortversjonen. Det gjelder kap. 7 om oppfølging og videreføring av Elmag2.

[1] Vedleggene omfatter bl.a. to kravspesifikasjoner:

- *Kravspesifikasjon til Arkivverkets arbeid med digitalt skapt arkivmateriale* (vedlegg 1) med krav fra TRAC-standarden som vurderes som relevante for Arkivverkets depotfunksjoner, supplert med egen-definerte krav til håndteringen av autentisk arkivmateriale ved fremstilling og mottak av avleveringer.
- *Spesifikke krav til Arkivverkets forvaltning av digitalt skapt arkivmateriale i tilknytning til digitalt depot* (vedlegg 2), som er konsentrert om organisasjonsløsninger og rutiner med direkte relevans for Arkivverkets digitale depot. På disse punktene utdypes også spesifikasjonene fra vedlegg 1.

[2] http://www.arkivverket.no/arkivverket/Arkivbevaring/Elektronisk-arkivmateriale/Langtidslagring/Nytt-digitalt-depot

1.1 Sentrale begreper i rapporten

Autentisitet og integritet er begreper som ofte brukes litt om hverandre i arkivmiljøene. Med opprettholdt *integritet* menes at informasjon er bevart uendret. Integritet er imidlertid et intrikat begrep ved digital lagring. Årsaken er at digital informasjon kan være endret på fysisk nivå (teknisk representasjonsnivå) uten at det intellektuelle informasjonsinnholdet er endret. Dette gjør det nødvendig å skille mellom teknisk integritet og logisk *innholds-integritet.*

Autentisitet brukes i rapporten som betegnelse for at informasjonen og dens opphavs- og brukssammenheng er hva den utgir seg for å være. Autentisitetsbegrepet inkluderer integritet: For å være autentisk må informasjonen også være bevart uendret. Men dermed melder de problematiske aspektene ved integritetsbegrepet seg på ny, for tales det da om integritet i teknisk eller logisk forstand? Konklusjonen er at autentisitet er knyttet til et uendret *logisk* innformasjonsinnhold. Problemet er samtidig at opprettholdt integritet på fysisk nivå enkelt kan verifiseres – i motsetning til den form for integritet som begrenser seg til et uendret logisk innhold, jf. punkt 2.2.1 og 2.2.2, nedenfor.

Med *digitalt objekt* menes et objekt som omfatter en sekvens av bits med de tekniske metadata som gjør den til en meningsfull enhet, f.eks. en pdf-fil.

Med *arkivobjekt* menes noe mer spesifikt: et lagret objekt som lar seg identifisere som arkivmateriale, dvs. med den tilknyttede forståelsesinformasjon (logiske metadata) som gjør det kvalifisert til betegnelsen arkivmateriale. Rapporten opererer med arkivobjekter på ulike nivåer, som avleveringspakker, arkivpakker (bevaringspakker) og brukspakker etter OAIS-standardens modell, og som arkivdokumenter (”records”).

Med *transformert arkivobjekt* menes et arkivobjekt hvor egenskaper ved formen er blitt endret (konvertert) uten at dette berører selve informasjonsinnholdet og de metadata som knytter dette innholdet til en kontekst.

Migrering innebærer flytting av informasjon, men betegnelsen brukes ofte flertydig. Når det gjerne tales om ”migrasjonsstrategien” ved langtidsbevaring og ”migrert representasjon” i arkivpakker, inngår også en transformering (konvertering) av informasjon. I denne rapporten betyr migrering at informasjon blir flyttet (overkopiert) uten at det skjer en omformatering eller annen konvertering som endrer den fysiske representasjonen. Det dreier seg mao. om kloning av informasjon – til forskjell fra prosessen ved en transformering eller konvertering.

Med *sjekksum* (”hash-verdi”) menes et tall som beregnes på grunnlag av en bestemt sekvens av data (f.eks. en datafil) i henhold til en bestemt algoritme (anvisning). Summen brukes som grunnlag for fysisk integritetskontroll. Den som vil kontrollere at datafilen er uendret, beregner en ny sjekksum ved hjelp av den tilhørende algoritmen, og foretar en sammenligning med den opprinnelige. Dersom de to sjekksummene er like, er datafilens integritet bekreftet. Mye brukte algoritmer for dette formålet er MD5 og SHA.

2. KRAV TIL ARKIVVERKETS DIGITALE ARKIVDEPOT

2.1 Hovedutfordringer ved bevaring av digitalskapt arkivmateriale

Langtidsbevaring av digital dokumentasjon representerer store utfordringer, og medfører en rekke risikofaktorer. Flere ulike sikkerhetsbehov må ivaretas ved digital bevaring. Siden arbeidet med å motta og bevare digitale avleveringer startet i 1985, har Arkivverket særlig hatt fokus på tre slike behov: lagringssikkerhet, opprettholdt lesbarhet og opprettholdt forståelighet. I løpet av de seneste årene har imidlertid internasjonalt ledende arkivmiljøer også rettet en spesiell oppmerksomhet mot sikkerhetsaspektene knyttet til arkivmaterialets integritet og autentisitet.”Trustworthy repositories” er kommet til som et nytt begrep. Arkivdepotene må kunne dokumentere sin virksomhet etter kriterier som kvalifiserer til tillit, og deres arkivbestand må kunne bekreftes å være autentisk og pålitelig.

2.1.1 Lagringssikkerhet

Det tradisjonelle basiskravet ved digital lagring er sikkerhet for at informasjonen holdes digitalt intakt. Elektroniske lagringsmedier har kort levetid. Med regelmessige mellomrom – hvert 3.-10. år avhengig av typen medium – må all bevart informasjon overkopieres til nye, friske databærere for å sikre den digitale bestandigheten. Uten et slikt kontinuerlig vedlikehold vil informasjonen gå naturlig tapt etter relativt få år.

2.1.2 Opprettholdt lesbarhet

At informasjonen holdes digitalt intakt på lagringsmediet er likevel ikke nok til å sikre lesbarheten for ettertiden. Årsaken er de hyppige teknologiskiftene som kjennetegner IT-utviklingen, og som gjør at nye generasjoner av utstyr ikke kan tolke informasjon fra eldre. Teknologiendringene gjør det nødvendig å konvertere informasjon som skal bevares, til bærekraftige formater som er tolkbare for ulike generasjoner og typer av utstyr.

Mange har vansker med å forstå at digital informasjon ikke enkelt kan bevares i den form (og i de systemer) som den til daglig brukes. For arkivskapere vil det ofte også være en komplisert oppgave å foreta den transformering av informasjonen som kreves for å gjøre den langtidstilgjengelig, typisk ved avlevering til en depotinstitusjon. Depotinstitusjonen må for sin del være beredt til å foreta nye og samlede formatkonverteringer med visse mellomrom for at informasjonen skal overleve i en fortsatt lesbar form. Formatkontroll og periodisk formatkonvertering er nødvendige vedlikeholdsaktiviteter ved digital bevaring.

2.1.3 Opprettholdt autentisitet og forståelighet

I tillegg til å være teknisk tolkbar for leseutstyret må den digitale informasjonen være forståelig for ettertiden. For å være praktisk tilgjengelig må innholdet være tilknyttet forståelsesinformasjon, nærmere bestemt *tekniske metadata* for å fremstille informasjonen med korrekt struktur og oppsett, og *logiske metadata* for å knytte innholdet til sin opphavs- og brukssammenheng (kontekst). Kravene som tidligere er behandlet ovenfor, gjelder for all digital informasjon som skal bevares. Kravet om medfølgende kontekstopplysninger gjelder imidlertid spesifikt for arkivmateriale. Den bevarte informasjonen må bringe med seg *autentiserende* opplysninger om opphavs- og brukssammenhengen for å være forståelig *qua* arkivmateriale. Autentisitet er selve kardinalkravet til arkivmateriale. Arkiv-

dokumenter er unike produkter av handlinger og hendelser. Bare når de er tilknyttet opplysninger om sin opprinnelse og bruk "der og da", kan de ivareta sitt hovedformål: å dokumentere konkrete hendelser som vitnesbyrd og bevis.

Vi kan ikke bevare originale systemer, bare data (arkivinformasjon) fra dem. Data må hentes ut av registre og databaser ved å eksporteres. Men vitale elementer går tapt når vi slik begrenser oss til å ekstrahere informasjonen i seg selv. Det er for å kompensere for dette tapet at det også må produseres tekniske metadata som dokumenterer struktur og innholdselementer. Arkivinformasjon må bevares med tilknyttede tekniske metadata for at det skal være mulig å gjenskape den i et riktig og meningsfylt oppsett. Det vil da ikke være tale om å gjenskape et originalt system, men en opprinnelig *logisk* struktur. Uten slik teknisk dokumentasjon vil den bevarte arkivinformasjonen ha liten verdi.

2.1.4 Opprettholdt integritet

For å kunne tjene som dokumentasjon må bevart arkivinformasjon være pålitelig. Tillit til informasjonen og arkivdepotets håndtering av den er en helt avgjørende faktor. At digital informasjon enkelt kan kopieres og endres, medfører muligheter for manipulering og uautorisert endring av innhold i samtid og ettertid som ikke lett kan etterspores. Begrepet original blir problematisk i en digital verden hvor alt er kopier i en eller annen forstand. Autentisitet er likevel fortsatt det sentrale kriteriet. Et digitalt dokument kan være autentisk – være hva det utgir seg for – selv om det er tale om en kopi. Men det må da være bevart med opprettholdt integritet, dvs. med et uendret informasjonsinnhold. Integriteten må ivaretas uavbrutt gjennom alle faser av materialets livssyklus: gjennom den aktive arkivdannings-fasen, gjennom prosessen hvor informasjon selekteres og konverteres til en versjon for avlevering og gjennom de vedlikeholdsoperasjoner som senere utføres i et arkivdepot.

Integritetssikring er blant de mest krevende utfordringene ved langtidsbevaring av digitalt arkivmateriale, fremfor alt fordi det også trengs mekanismer som gjør det mulig å bekrefte at informasjonsinnhold er bevart uendret. Bak den dreide orienteringen mot temaet "trust" i de førende internasjonale arkivmiljøene ligger en økt erkjennelse av en viktig forskjell mellom papirbasert og digitalt skapt arkivmateriale. Papirarkiver kan bevares statiske, dvs. slik de ble skapt. Digitalt arkivmateriale er derimot i en prosess hvor det stadig, og på ulike måter blir transformert gjennom alle faser av sin livssyklus. Hvordan kan vi stole på at arkiv-materiale som gjentatte ganger er blitt omskapt, fortsatt er bevart autentisk? Og hvordan kan vi få dette verifisert? Mer konkret: Hvordan får vi bekreftet at de som har utført operasjoner på arkivmaterialet i faser med transformasjon, har gjort dette korrekt – og dessuten nøyd seg med med å foreta de endringer som hevdes å være utført?

2.1.5 Konfidensialitetssikring

Nyere arkivmateriale fra offentlig forvaltning kan være gradert eller taushetsbelagt, og krever tiltak for å hindre uautorisert tilgang og kopiering av data. Et digitalt depot må iverksette fysiske sikringstiltak og tilgangskontroll for å forvalte slikt materiale på en beskyttet måte i samsvar med fastsatte lovbestemmelser og eventuelle avtaler med private arkivskapere. Sensitive personopplysninger etter personopplysningsloven og taushets-belagt materiale etter forvaltningsloven og statistikkloven kan kreve beskyttelse i 100 år fra opphavsdato. Graderte opplysninger og andre typer materiale kan kreve kortere beskyttelsestid. Et digitalt depot må holde nitid oversikt over alt beskyttet materiale, og kunne identifisere hvilke bestemmelser som gjelder for hver enkelt forekomst. På grunnlag

av en slik statusoversikt må depotet også makte å frigi materiale for bruk på de tidspunkter som er fastsatt i de enkelte bestemmelser.

2.1.6 Langsiktige styringsforpliktelser

De elektroniske lagringsmediene representerer et ekstremt sårbart aspekt ved digital langtidsbevaring. Men rutinemessige overkopiering til nye, friske medier innebærer en kloning av informasjon, og resulterer ikke i en trinnvis svekkelse av den på linje med reprografi. Gitt at vi – og senere våre etterkommere – også utfører de øvrige vedlikeholds-oppgavene som er beskrevet ovenfor, så har digitalt lagret informasjon i seg selv en beskaffenhet som er egnet til å opprettholde tilgjengeligheten for ettertiden, også i et 1000-årsperspektiv. Dette krever langsiktige styringsforpliktelser og organisatorisk soliditet.

2.2 Spesielle utfordringer mht. å sikre integritet og autentisitet

Ved fremstillingen av arkivversjoner for avlevering foretas en selektering og omformatering av informasjon. Konvertering til standardiserte bevaringsformater kan også resultere i en forenkling og redusert funksjonalitet. Ny transformering av informasjonen kan senere være nødvendig som ledd i et arkivdepots vedlikehold av materialet. Det er derfor ingen enkel oppgave å bekrefte at digitalt arkivmateriale er bevart med opprettholdt integritet og autentisitet.

For digitalt arkivmateriale som skal tjene som vitnemål eller bevis, må påliteligheten være hevet over tvil. O. J. Simpson-saken i USA i 1995 brukes ofte som et eksempel på det motsatte. Retten underkjente Simpsons fingeravtrykk som bevis. Forsvareren klarte ikke å påvise at det var forfalsket. Det holdt imidlertid med å påvise at behandlingsrutinene kunne gitt politiet *mulighet* for manipulasjon. Normen blir altså at den som skaper og/eller forvalter informasjonen, selv må eliminere muligheter for tvil om dens autentisitet, eventuelt også om dens konfidensialitet. Det grunnleggende kravet til et arkivdepot blir å kunne bekrefte at materiale er bevart med et uendret informasjonsinnhold etter mottak. Dette gir ikke uten videre noen garanti for at innholdet er autentisk i seg selv, men i motsatt fall vil det være bevart som et autentisk falsum.

2.2.1 Praktiske forutsetninger for å utføre integritetskontroll

Utfordringene mht. å sikre integritet og autentisitet har sammenheng med at et lagret digitalt objekt eksisterer på flere ulike nivåer:

1) *som et fysisk objekt*, dvs. som tegn (representasjon av ”bits”) festet til et fysisk medium. IT-systemer forstår ikke bits uten fortolkende mellomledd
2) *som et logisk objekt*, dvs. som en enhet som kan gjenkjennes og prosesseres av programvare. Regler bestemmer oversettelsen mellom bits og logiske enheter (”formater”) som f.eks. tegn og tall
3) *som et konseptuelt (fattbart) objekt*, dvs. som en enhet som kan gjenkjennes og forstås av mennesker, f.eks. en bok, en kontrakt, et foto, et kart. Det konseptuelle dokumentet er det virkelige – for oss.

Integritetskontroll kan enkelt utføres på fysisk nivå. At strømmen av bits er uendret, kan bekreftes ved bruk av en sjekksum. Men digitale objekter konverteres jevnlig, og da endres bit-representasjonen. Det skjer selv om det konseptuelle innholdet er uendret. Endret bit-representasjon blir f.eks. konsekvensen når et Word 2000-dokument lagres uforandret i Word 2007-versjon. Langtidslagring av digital informasjon krever vedlikehold med

periodisk transformasjon av data. Vår strategi for bevaring – ofte kalt migrasjonsstrategien – bygger nettopp på forutsetningen om at informasjonen kan endre form uten (nødvendigvis) å tape sin innholdsintegritet.

Det er i det hele tatt umulig å langtidslagre et digitalt arkivdokument uten at noe element endres, konstaterer det internasjonale InterPARES-prosjektet. "There is no such thing as an uncorrupted record" (Luciana Duranti). Vi må nøye oss med å kreve at arkivdokumenter er bevart intakt og ukorrumpert *i alle vesentlige henseender*, nærmere bestemt:

- at det ikke har skjedd endringer som berører materialets identitet og innholdsintegritet,
- og at fravær av uakseptable modifikasjoner kan verifiseres.

Integritetskontroll ved langtidsbevaring må altså bygge på noe annet enn konstante bitstrømmer. Sjekksummer må utnyttes for alt de er verdt så langt det gjelder å bekrefte at informasjon som skal være *fysisk* uendret, faktisk også er det. Men det finnes i dag ingen tilsvarende enkel metode for å utføre integritetskontroll på konseptuelt nivå. At det ikke eksisterer mekanismer for integritetssikring av logisk informasjonsinnhold i seg selv, er en sterkt kompliserende faktor ved digital langtidsbevaring. For å kompensere for dette er det nødvendig å ty til arbeidskrevende rutiner. På konseptuelt nivå må egenskaper som kan bekrefte identitet og innholdsintegritet, defineres i metadata som er tilknyttet arkivdokumentene. Ved langtidsbevaring kreves derfor også egne vedlikeholdsaktiviteter for å dokumentere at og hvordan materialet har vært gjenstand for ubrutt integritetssikring.

2.2.2 Konsekvenser mht. metodegrunnlag og begrepsbruk

At informasjon på fysisk representasjonsnivå er bevart som en fullstendig bit-strøm, er en praktisk forutsetning for å bekrefte opprettholdt *integritet* ved bruk av en sjekksum. Arkivmiljøer internasjonalt, bl.a. det pågående arkivprosjektet InterPARES, lar denne praktiske forutsetningen være styrende for definisjonen av begrepet integritet. Etter denne definisjonen må informasjon ikke bare være uendret, men også være bevart fullstendig.

I denne rapporten inkluderer integritet – i likhet med InterPARES-definisjonen – et krav om fullstendighet når det brukes som et teknisk begrep. Men når det blir tale om logisk (konseptuell) *innholdsintegritet,* kan det stille seg annerledes. Betegnelsen innholdsintegritet brukes i rapporten om noe mer avgrenset: at det logiske informasjonsinnholdet (typisk tegn og tall) er bevart uendret. Logisk informasjonsinnhold kan være uendret selv om den fysiske representasjonen eller egenskaper ved den visuelle presentasjonen er blitt endret. Dette fremsto som uproblematisk så lenge det bare var tale om bevart tabellinformasjon fra registre og databaser. Men i dag kan audiovisuelle uttrykksmidler og andre egenskaper ved informasjonens form ha betydning for meningsinnholdet, herunder f.eks. farge og skrifttyper i ren tekstlig informasjon.

Ulike typer informasjon vil stille ulike krav til elementene som må være inkludert for at selve innholdet skal betraktes som uendret, men vurderingen vil være like avhengig av hva man i hvert tilfelle har som ambisjon å dokumentere med den bevarte informasjonen. Det kreves under enhver omstendighet at sammenhengen med en original er kjent. En slik sammenheng vil for eksempel gjerne være kjent når det gjelder still-bilder fra film og video. De utgir seg ikke for å være noen fullstendig original, men aksepteres for hva de er: svært ufullstendige, men likevel uforandrede fragmenter av originalen.

Autentisitet brukes gjerne i en folkelig og interdisiplinær betydning hvor det er mer eller mindre synonymt med ekthet. Men som arkivbegrep har autentisitet et spesifikt innhold. I tillegg til å være hva den utgir seg for, må arkivinformasjon være tilknyttet opplysninger om opphavs- og brukssammenhengen. InterPARES lar integritet inngå i autentisitets-begrepet[3]. Det gir god mening å kreve, som InterPARES, at informasjon må være bevart med opprettholdt integritet på fysisk nivå for å være autentisk i teknisk forstand. Men samles oppmerksomheten om logisk informasjonsinnhold, så disponerer dette for en større romslighet i begrepsbruken. For heller ikke et tradisjonelt papirdokument trenger å foreligge i en opprinnelig eller original form for å betraktes som autentisk – som en logisk enhet. Både avskrifter, karbonkopier uten brevhode, pergamentbrev med hull og mikro-filmutgaver kan bli vurdert som autentiske – under forutsetning av en kjent sammenheng med omgivelsene. Konklusjonen blir at et dokument *kan* være autentisk uten at dets fulle integritet er i behold. Men i desto større grad kreves informasjon om hvordan og i hvilken forstand dokumentet er autentisk.

Arkivinformasjon må kunne betraktes som autentisk selv om den er transformert, dvs. i tilfeller hvor egenskaper ved formen er blitt endret (konvertert) uten at dette berører selve informasjonsinnholdet og de metadata som knytter innholdet til en kontekst. En slik transformering hviler på forutsetningen om at autentisitet primært går på innhold, ikke på form og heller ikke på teknisk representasjon. Digitalt arkivmateriale må følgelig kunne konverteres og skifte form uten å tape sin autentisitet. Dette er et helt sentralt punkt i metodegrunnlaget for langtidsbevaring av arkivmateriale. Men den transformering av informasjon som fra tid til annen kreves ved digital langtidsbevaring, må stille desto strengere krav til dokumenterte prosesser, autentiserende metadata og medfølgende hjelpemidler for å bruke materialet.

I vår sammenheng er det primært tale om informasjon som er autentisk qua arkivmateriale. Den vil ikke nødvendigvis være autentisk i forhold til de handlinger og hendelser som den hevder å dokumentere. Det er i slike tilfeller at vi kan tale om et autentisk falsum.

2.3 Internasjonale standarder og "Best practices"

2.3.1 OAIS-standarden

I arkivmiljøer internasjonalt bygger så godt som all aktivitet for digital bevaring på prinsippene, terminologien og de funksjonelle beskrivelsene i *OAIS – Reference Model for an Open Archival Information System*[4]*,* som ble utformet av den amerikanske romfarts-organisasjonen CCSDS i 2002, og gjort til en internasjonal standard (ISO 14721) i 2003. Denne generelle standarden for arkivering definerer det konseptuelle rammeverket for et digitalt arkiv.

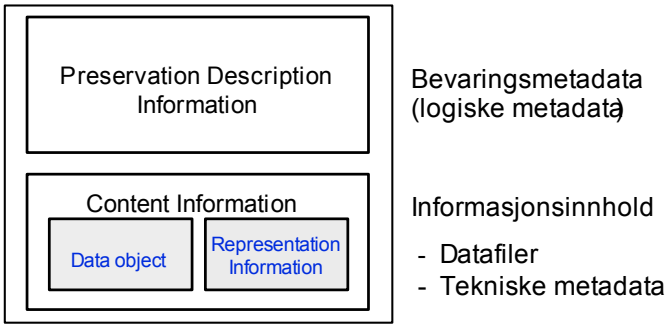
[3] Autentisitet kan innbefatte opprettholdt integritet, men ikke nødvendigvis omvendt. Informasjon med opprettholdt integritet vil ikke alltid være autentisk etter den arkivfaglige definisjonen av autentisitet, for et dokument kan være uendret uten å være tilknyttet (autentiserende) informasjon om en opphavs- og bruks-sammenheng. I praksis vil det si at dokumentet ikke kvalifiserer til betegnelsen arkivdokument.

[4] http://public.ccsds.org/publications/archive/650x0b1.pdf

OAIS er en modell for å innlemme, administrere og bruke bevart arkivmateriale i et depot. Den beskriver funksjoner, prosesser og informasjonsflyt i et digitalt depot med fokus på autentisitetssikring, og gir opplegg for vedlikehold innenfor rammen av et kontrollert miljø. OAIS-modellen konsentrerer seg om overordnede kategorier, og disse omfatter både de digitale bevaringsobjektenes konseptuelle og tekniske aspekter. Hvert bevaringsobjekt skal iht. OAIS lagres som en autonom og selvdokumenterende *arkivpakke,* permanent forbundet med alle tilhørende logiske og tekniske metadata. Slik skal informasjonen fortsatt kunne fremstilles som arkivmateriale, og slik skal den fortsatt være forståelig og autentisk som arkivmateriale.

Bevaringsobjektet i en OAIS-arkivpakke kan være et enkelt dokument eller et samlet datauttrekk fra en database. OAIS skiller mellom tre typer av arkivpakker: *Submission Information Package* (SIP) for bevaringsobjektet som mottas som aksesjon, *Archival Information Package* (AIP) for versjonen av det mottatte bevaringsobjektet som innlemmes tilrettelagt for bevaring i arkivdepotet, og *Dissemination Information Package* (DIP) for en AIP (eller flere) som gjøres tilgjengelig i bruksversjon. Når en ny AIP genereres av et arkivdepot, krever OAIS at den mottatte SIP-versjonen innlemmes i tillegg. For å muliggjøre en ettersporing av depotets operasjoner skal en opprinnelig SIP bevares uendret og integritetssikret – for alltid. Dette gjelder uavhengig av om den fortsatt er tolkbar.

Et bevaringsobjekt i en OAIS-arkivpakke har to grunnelementer: bevaringsmetadata *(Preservation Description Information)* og informasjonsinnhold *(Content Information)*. Informasjonsinnholdet har igjen to hovedelementer: (filer med) data *(Data Object)* og tekniske metadata *(Representation Information)*. Denne enkle OAIS-pakkemodellen bestående av bevaringsobjektets data, tekniske metadata for å fremstille dem og logiske (konseptuelle) metadata for å forstå dem, er illustrert i figuren til høyre.



OAIS-modellen har mangler, spesielt når det gjelder å håndtere arkivinformasjon fra Records Management-systemer. Men det er viktig å holde fast at OAIS ikke er en implementeringsmodell. Det er en referanse- og begrepsmodell med metadatakategorier, men uten konkrete forslag til metadata. Slike metadata og annen tilleggsfunksjonalitet foreslås imidlertid i en rekke oppfølgingsstandarder til OAIS. En av disse er TRAC-standarden, som følger opp anvisningene i OAIS om integritets- og autentisitetssikring. TRAC krever særlig omtale, og behandles i etterfølgende avsnitt. Andre viktige oppfølgingsstandarder til OAIS er:

- METS (Metadata Encoding & Transmission Standard)[5], som beskriver den indre strukturen i en arkivpakke og den "container" som omslutter pakken,
- XFDU[6] (ISO 13527: 2009), som spesifiserer en alternativ pakkestruktur, og

[5] http://www.loc.gov/standards/mets/

[6] http://public.ccsds.org/publications/archive/661x0b1.pdf

- PREMIS (Preservation Metadata: Implementation Strategies)[7], som definerer bevaringsmetadata for å støtte forståelighet, autentisitet og identitet.

2.3.2 TRAC-standarden

Digitale depoter verden over akkumulerer sterkt voksende informasjonsmengder, og må være gjenstand for krav som avspeiler deres forpliktelser. For å bli vurdert som pålitelige og tiltrodde, må slike depoter oppfylle definerte kriterier. Gjennom de siste årene er det gjennomført en rekke aktiviteter for å spesifisere slike kriterier, spesielt i USA og innenfor EU. Felles for aktivitetene er at de bygger på OAIS-standardens kategorier og funksjonsbeskrivelser. De sentrale utviklingsmiljøene har i samarbeid også formulert 10 hovedkriterier for et pålitelig digitalt depot: *Ten Core Principles of Trust Repository Design*[8].

Den amerikanske TRAC-rapporten fra 2007 – *Trustworthy Repository Audit and Certification - Criteria and Checklist*[9] – har vært gjenstand for stor oppmerksomhet, og brukes som referansedokument av de øvrige utviklingsmiljøene. Rapporten, som er utarbeidet av organisasjonen for forskningsbiblioteker og USAs riksarkiv i samarbeid, formulerer 90 kriterier som grunnlag for et sertifiseringsopplegg for digitale depoter. Utkast til en ISO-standard basert på TRAC ble i 2009 utarbeidet av romfartsorganisasjonen CCSDS, og lagt ut til høring[10].

For å oppnå en sertifisering etter kravene i TRAC må et digitalt depot være gjenstand for innsyn og evaluering. Det må selv aktivt kunne *dokumentere* og *demonstrere* sin evne til å oppfylle kravene, herunder krav som TRAC stiller til styringsforpliktelser og ansvarlighet, langsiktighet og organisatorisk levedyktighet, økonomi og finansiell bærekraft.

OAIS-standardens krav til integritets- og autentisitetssikring blir videreutviklet i TRAC. I sin administrasjon av digitale objekter må et depot kunne demonstrere at bevart informasjonsinnhold fortsatt samsvarer med opprinnelig mottatt innhold. Depotoperasjoner som resulterer i transformerte arkivpakker, må følgelig være ettersporbare. Opprinnelige arkivpakker må bevares, og det må finnes forbindelser mellom disse og senere transformerte versjoner.

For de fleste arkivdepoter – Riksarkivet inkludert – medfører TRAC en endret virkelighet med omsnudd bevisbyrde. Et godt renommé blir definitivt ikke nok for et depot, heller ikke et offentlig monopol. Depotet må selv eliminere grunnlaget for tvil eller spekulasjon om feil, uautoriserte endringer og andre uforsvarlige operasjoner som er *mulige* ved digital bevaring. Opprettholdt integritet må kunne bekreftes med verifiserende dokumentasjon. For å etterleve TRAC kreves et defensivt og forebyggende vedlikeholdsarbeid med kontinuerlig beskyttelse mot uautoriserte hendelser. Dette nødvendiggjør fulldokumenterte rutiner, loggføring av operasjoner på bevart materiale og sporing av endringer for å muliggjøre tilbakespoling til tidligere versjoner.

[7] http://www.loc.gov/standards/premis/

[8] http://content.yudu.com/Library/A10tra/PLATTERRepositoryPla/resources/8.htm.

[9] http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

[10] http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206520R1/Attachments/652x0r1.pdf

TRAC og et kommende sertifiseringsopplegg har som ambisjon å gi informasjonen i digitale depoter den samme umiddelbare pålitelighet som pengesedler fra minibanker. At arkivmateriale kan bekreftes å ha vært gjenstand for ubrutt integritetssikring fra og med mottak, er avgjørende for et arkivdepots pålitelighet og troverdighet. Men dette garanterer ikke at informasjonen er ekte og troverdig i seg selv. Fremstillingen av en arkivversjon for avlevering (SIP) er i denne forbindelse forbundet med særlig risiko, for i denne fasen er informasjonen både eksponert for feil og manipulasjon. At prosessen vanligvis også medfører en selektering og omformatering av informasjon, gjør det desto mer problematisk å validere avleveringspakkens informasjonsinnhold mot innholdet i det opprinnelige produksjonssystemet.

Integritetssikring i et produksjonssystem og under fremstillingen av en SIP er temaer som TRAC ikke behandler. TRAC (og OAIS-modellen) gir imidlertid rom for å dokumentere om og hvordan arkivmateriale er blitt integritets- og autentisitetssikret forut for en aksesjon.

2.4 Anbefalt rammeverk for digitalt depot

Prosjektet anbefaler følgende rammeverk for forvaltningen av arkivobjekter i Arkivverkets digitale depot:

1) OAIS brukes som modell for å innlemme, administrere og vedlikeholde arkivobjekter.
2) Integritetssikringen og dokumentasjonsrutinene i digitalt depot baseres på kravene i TRAC-standarden. Det settes som mål at Arkivverkets digitale depot skal være kvalifisert for en sertifisering etter kriteriene i TRAC.
3) TRAC suppleres med egendefinerte krav for å styrke autentisitets- og integritetssikringen av digitalt arkivmateriale ved fremstilling av avleveringer (SIP-pakker) og ved Arkivverkets mottak av det. Kravene spesifiseres i vedlegg 1-3 til rapporten.
4) Det legges til grunn at arkivpakker skal bruke METS som pakkeformat og PREMIS som standard for bevaringsmetadata.

2.5 Organisasjonsmessige forutsetninger

Kravene som stilles til et digitalt depot, har organisasjonsmessige aspekter som kan representere like store utfordringer for Arkivverket som de tekniske og utstyrsmessige. Forvaltningsoppgavene krever at det etableres roller for definerte funksjoner, at rollene fylles, og at rollene fungerer – enkeltvis og i samspill. Roller og rollenettverk må være definert innenfor 4 hovedområder:

1) informasjonssikkerhet (integritet og konfidensialitet)
2) systemdrift og teknisk vedlikehold
3) informasjonsforvaltning (mottak, testing og vedlikehold av arkivpakker)
4) tilrettelegging av materiale for bruk og brukere

2.5.1 Sikkerhetsorganisasjon

For at et sikkerhetsopplegg for digitalt depot skal fungere, må det finnes et apparat med stabstilknytning til virksomhetens ledelse for å følge opp tiltak. Verdiene som Arkivverket forvalter, og som tiltakene må sikre, er i all hovedsak immaterielle, og består av informasjon. De virksomhetskritiske truslene mot disse verdiene kan sammenfattes i følgende scenarier:

a) graderte opplysninger eller annen beskyttet informasjon blir kompromittert,
b) det lar seg ikke verifisere at informasjon er bevart med opprettholdt integritet,
c) bevart informasjon blir utilgjengelig og uleselig,
d) bevart informasjon går tapt eller ødelegges,
e) bevart informasjon blir manipulert, forfalsket eller endret på annen uautorisert måte.

Hendelser som faller under punkt a, vil representerer brudd på lovbestemmelser, men Arkivverket kan heller ikke tolerere de øvrige hendelsene ovenfor. De vil resultere i et tapt omdømme, og skade en sentral samfunnsfunksjon mer eller mindre uopprettelig.

Sikkerhetsorganisasjonen for digitalt depot må bygge på klargjorte ansvarsforhold og rapporteringsopplegg. For å håndtere sikkerhetstrusler må det være utarbeidet tre typer av dokumentasjon som samlet vil utgjøre internkontrollsystemet for digitalt depot:

1) *Styringsdokumentasjon* som definerer sikkerhetsmål og plikter som følger av bestemmelser i sikkerhetsloven, personopplysningsloven og andre relevante krav, og beskriver hvordan disse skal ivaretas gjennom intern organisering, ansvarsplassering og rolledefinisjoner
2) *Instrukser* som beskriver prosedyrer for gjennomføringen av sikkerhetstiltak
3) *Kontrolldokumentasjon* i form av rapporter, logger og sjekklister mv. som bekrefter at aktiviteter er utført iht. fastsatte instrukser og prosedyrer

2.5.2 Driftsorganisasjon

Et digitalt depot krever en profesjonell driftsorganisasjon. Kompetansen, arbeids-organisasjonen og bemanningen på driftssiden er avgjørende for mulighetene til å etablere Arkivverkets digitale depot. Etter prosjektets vurdering må IT-avdelingen ha minimum 3 medarbeidere med kompetanse og særskilt autorisasjon for drift av Arkivverkets lagrings-system. En av medarbeiderne må arbeide dedikert med systemet på heltid. De øvrige to må ha befatning med systemet jevnlig nok til å kunne tre inn i rollen som hovedansvarlig. Driftsorganisasjonen må ha definerte roller som:

1) systemansvarlig for installasjoner og driftsfunksjoner,
2) ansvarlig for sikkerhetskopiering (tape-roboter og off site-kopier),
3) operativt sikkerhetsansvarlig.

2.5.3 Organisasjon for informasjonsforvaltning

Informasjonsforvaltningen i tilknytning til digitalt depot omfatter all behandling av lagret materiale som arkivobjekter, dvs. prosessene ved mottak, testing, vedlikehold og tilgjengeliggjøring av arkivpakker. Informasjonsforvaltningen må ha definerte ansvarlige for følgende oppgaver:

1) sentral mottakskontroll,
2) testing av mottatt materiale,

3) generering og innlemmelse av arkivpakker i DSM, samt flytting av materiale mellom områder og mellom soner i lagringssystemet,
4) faglig vedlikehold av bevart arkivbestand.

Informasjonsforvaltningen må blant annet ivareta følgende funksjoner:

- verifisering og integritetssikring av arkivpakker (SIP) ved mottak,
- påføring av integritetssikrende sjekksummer ved generering av arkivpakker (AIP),
- kvalitetssikring av genererte arkivpakker (AIP) med vekt på komplett dokumentasjon (arkivbeskrivelse og andre bevaringsmetadata),
- etterkontroll og oppfølging av logger for utførte operasjoner i depot f.o.m. mottak av materiale,
- behandling av sensitive personopplysninger og gradert materiale,
- oppdatert dokumentasjon av rutineopplegg.

3. KONFIGURASJONSLØSNINGER

3.1 Organisering og integritetssikring av arkivpakker

Arkivobjektene i digitalt depot organiseres som selvberoende og selvforklarende arkivpakker (AIP-er) etter OAIS-modellen. TRAC-standarden og de spesifikke kravene i rapportens vedlegg 2 ligger til grunn for implementeringen. En fast struktur for arkivpakkene i digitalt depot er en sentral forutsetning for å etablere det kontrollerte miljø som kreves for å utføre vedlikehold og integritetssikring. En slik arkivpakkestruktur beskrives nedenfor, først som en logisk modell, og deretter som en implementeringsmodell. Den logiske modellen er til hjelp for forståelsen, men trenger tilpasninger for å være egnet for faktisk bruk. Implementeringsmodellen foretar slike tilpasninger.

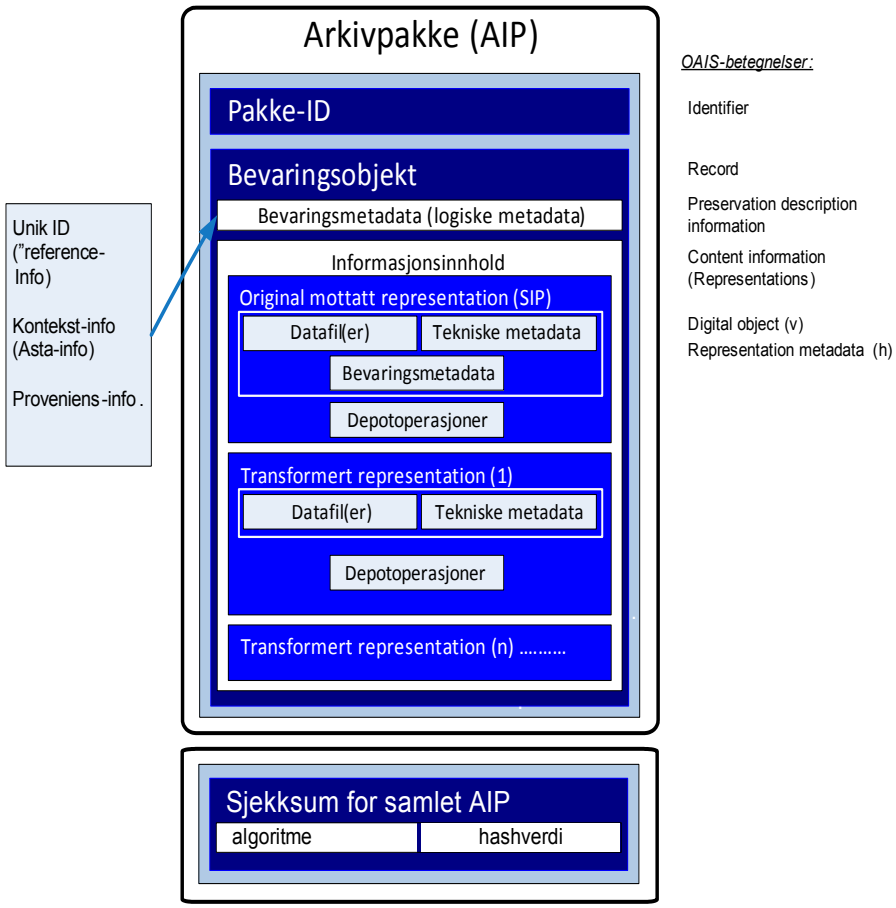
3.1.1 Logisk modell

Prosjektet foreslår å basere Arkivverkets arkivpakkestruktur på den logiske modellen som vises i figuren nedenfor. Modellen omfatter 3 seksjoner: 1) identifikator, 2) bevaringsobjektet med alle logiske og tekniske metadata og 3) samlet sjekksum for pakken

Bevaringsobjektet (2) har to hovedkomponenter: Logiske metadata (operasjons-, kontekst- og proveniensinformasjon) og informasjonsinnhold. Logiske metadata omfatter overordnede bevaringsmetadata som forutsettes å være rimelig stabile over tid. Men informasjonsinnholdet bestående av datafiler og tilhørende tekniske metadata må kunne bevares i flere versjoner. Operasjonene for å vedlikeholde materialet som er nødvendige i et arkivdepot, kan resultere i versjoner med et transformert innhold. Modellen må både gi rom for å bevare transformerte og opprinnelige versjoner av informasjonsinnholdet for å muliggjøre tilbakespoling og integritetskontroll etter utførte operasjoner. Det må dessuten finnes mekanismer for å verifisere at de ulike versjonenes informasjonsinnhold er bevart uendret. Mekanismene vil her være medfølgende sjekksummer fra arkivskaper ved avlevering og sjekksummer generert av Arkivverket ved mottak og senere operasjoner.

Det foreslås som fast rutine å bevare de to siste transformerte versjonene av informasjonsinnholdet i tillegg til den opprinnelige mottatte avleveringspakken (SIP).

Figur: Arkivpakke – logisk modell



Avleveringspakken (SIP) som danner grunnlaget for hver enkelt arkivpakke, må inkorporeres uendret i arkivpakken i den form den ble mottatt. En SIP skal ha tilknyttet en samlet sjekksum som er generert av arkivskaperen. Arkivverket skal under alle omstendigheter generere en egen sjekksum for hver SIP ved mottak. Alle filer som innlemmes av Arkivverket ved genereringen av arkivpsakken og ved senere vedlikeholdsoperasjoner, skal også være tilknyttet sjekksummer for å muliggjøre en verifisering av at informasjonen er bevart uendret på fysisk nivå (uendret bit-strom).

Hver arkivpakke skal dessuten ha en samlet sjekksum. Enhver endring av innhold i pakken – som når det foretas oppdateringer i Depotoperasjoner – vil kreve en regenerering av sjekksummen for den samlede pakken. Arkivpakkens samlede sjekksum vil derfor være et sentralt flagg mht. å spore endringer. Sjekksum for den samlede arkivpakken må lagres *utenfor* pakken – til forskjell fra de øvrige sjekksummene – fordi innlemmelsen av en samlet sjekksum i seg selv vil endre arkivpakkens bit-innhold.

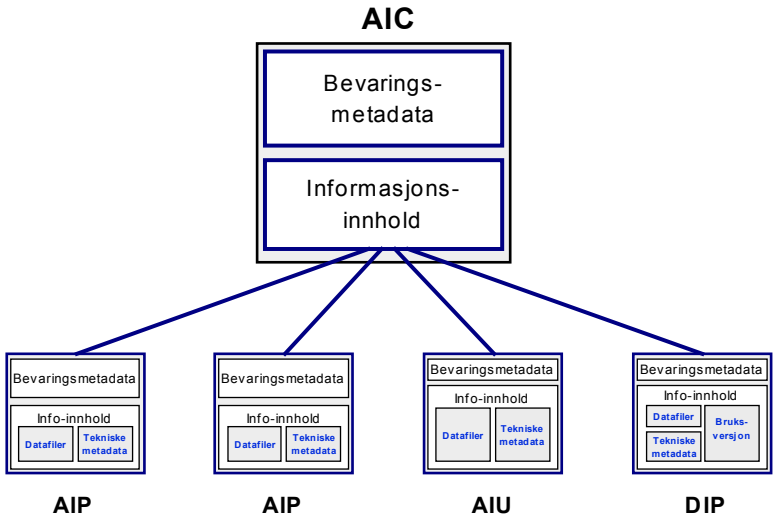
Kategorien "Depotoperasjoner" krever spesiell omtale. Den skal omfatte Arkivverkets testdata (herunder verifisering av medfølgende sjekksummer fra arkivskaperen), logger og dokumentasjon av operasjoner. Arkivverkets testing av en avlevering vil typisk resultere i endringer eller tilføyelser i en fil med tekniske metadata når denne medfølger fra arkivskaperen. Den nye metadatafilen skal da lagres i Depotoperasjoner – i tillegg til den opprinnelige, som fortsatt skal ligge (uendret) i SIP-delen. Dermed unngås en endring av bit-innholdet i SIP-delen, og vi slipper å generere en ny transformert versjon av hele

informasjonsinnholdet ved hver depotoperasjon. Det blir bare nødvendig å opprette en ny seksjon for en transformert representasjon i de tilfeller hvor også datafiler endres, dvs. ved en formatkonvertering av datafiler.

3.1.2 Implementeringsmodell

I den logiske arkivpakkestrukturen som beskrives foran, er OAIS-objektstrukturen realisert i en flerlags modell hvor AIP-er innbygges i AIP-er. Av praktiske grunner bør arkivpakke-strukturen likevel ikke implementeres fullt ut som en speiling av den logiske modellen. Den er monolittisk, og kan medføre svært store og lite håndterbare arkivpakker. Det er behov for en implementering hvor ulike versjoner av en arkivpakke på en løsere måte er forbundet med en overordnet. I stedet for én seksjonert pakke vil det altså være tale om en pyramide av enkeltpakker. Dette vil redusere prosesseringsproblemer som kan inntreffe ved generering av store arkivpakker, og forenkle kopiering til tape ved oppdatering.

OAIS gir også anvisninger om hvordan flere arkivpakker kan være tilknyttet en overordnet arkivpakkesamling – en AIC (Archival Information Collection) – i en nettverksstruktur. De underordnede enhetene kan være ordinære arkivpakker (AIP-er) eller grunnenheter med minimale egne bevaringsmetadata. De sistnevnte har betegnelsen Archival Information Unit (AIU) i OAIS. Tilknyttede AIP-er og AIU-er vil fremstå som innholdet (Content Information) i en AIC. En AIC har også egne bevaringsmetadata (PDI – Preservation Description Information) som beskriver hele samlingen av underliggende objekter. Oppbygningen av en AIC illustreres i figuren nedenfor.



Figur: Arkivsamlepakke (AIC)

Samlede sjekksummer for tilknyttede AIP-er og AIU-er kan lagres i AIC-en. Samlet sjekksum for AIC-en i seg selv, eventuelt for alle enheter som omfattes av en AIC, må også i dette tilfellet lagres et annet sted.

Den store fordelen med samlepakkemodellen er fleksibiliteten. Den gjør det også mulig å håndtere noen svakheter i OAIS og prosjektets logiske arkivpakkemodell. Mangler i OAIS åpenbarer seg når arkivversjoner fra Noark-systemer og annen "Records Management" (RM) innlemmes i arkivpakker. Fra RM-systemer er det helt essensielt å bevare autentiserende metadata, typisk journalinformasjon. I OAIS blir dette til Bevarings-metadata, men OAIS og prosjektets logiske modell blir for rigid når den forutsetter at

bevaringsmetadata i ethvert tilfelle kan ligge som fellesinformasjon på toppnivået i en flerlags arkivpakke. Fra RM-systemer krever hver transformert versjon *egne* bevarings-metadata. OAIS tar ikke høyde for at også Bevaringsmetadata – i det minste autentiserende metadata (som journalinformasjonen i Noark-systemer) – kan være avhengige av tilknyttede tekniske metadata for å fremstilles. Prosjektets implementeringsmodell – arkivsamlepakker basert på en AIC – gir den fleksibilitet som kreves for RM-systemer.

3.1.3 Bruk av implementeringsstandardene METS og PREMIS

METS (Metadata Encoding & Transmission Standard) er utviklet som et tillegg til OAIS. for å beskrive den indre strukturen i en arkivpakke-container og for å knytte sammen metadata og informasjonsinnhold. PREMIS (Preservation Metadata: Implementation Strategies) definerer bevaringsmetadata for å støtte autentisitet, identitet og forståelighet. Prosjektet foreslår å basere implementeringen av arkivpakkestrukturen på METS og PREMIS. Det svenske riksarkivet har tidligere valgt å ta i bruk disse standardene.

Det svenske riksarkivet vurderer nå å bruke arkivpakker med en ytre og indre METS-fil. En arkivpakke blir da bestående av to hoveddeler: (1) en *indre* METS-fil med tekniske metadata, bevaringsmetadata og katalogdata, og (2) en tar-fil med innholdsfiler og den medfølgende dokumentasjonen fra arkivskaperen. Sjekksummer for tar-delens innholds-filer lagres i METS-delen. En *ytre* METS-fil omslutter og beskriver de to hoveddelene. I den ytre METS-filen lagres også sjekksummer for den indre METS-delens metadata-filer. Sjekksummer for samtlige filer i en arkivpakke (unntatt den totale sjekksummen) kan dermed lagres som en del av pakken.

Det fremstår som aktuelt å adoptere elementer fra den svenske løsningsmodellen. Den går ut over kravene i OAIS, som strengt tatt bare omfatter integritetssikring av innholdsfiler, men den er i samsvar med kravene til en norsk løsning. Den norske løsningen, som må bygge på de forsterkede kravene i TRAC og de egendefinerte kravene i rapportens vedlegg 2, krever at også metadata integritetssikres med sjekksummer. Integritetssikrede metadata kreves definitivt for bevart arkivinformasjon fra Noark-systemer og andre RM-systemer.

3.1.4 Oppfølgingstiltak

Det er behov for å utvikle et program for generering av arkivpakkefiler i tilknytning til forvaltningssystemet for digitalt depot.

Det må vurderes om en AIP alltid skal være tilknyttet en overordnet AIC. Det er også behov for å vurdere hvordan sjekksumsikringen av AIC-er skal organiseres. Det må dessuten vurderes om samlet sjekksum for en AIC både skal genereres før og etter pakkingen til tar-format.

Bruk av METS og PREMIS krever en spesifisering av egne norske profiler (XML-skjemaer) for arkivpakker, som i Sverige. Det bør også være et siktemål å sende disse profilene til Library of Congress for registrering og publisering.

3.2 Lagringsløsning

Riksarkivet har hittil lagret digitalt skapt arkivmateriale på bortsettingsmedier. Magnetbånd (spolebånd) ble brukt fra starten i 1985. Siden 1993 har mediet vært optisk plate (CD-R). Bortsettingsmedier medfører tungvinte operasjoner både ved vedlikehold og bruk av materiale. Optiske medier har heller ikke den kapasitet som kreves for å håndtere overgangen til elektronisk arkiverte saksdokumenter i forvaltningen. Avleverings-bestemmelsene fra 2007 aksepterer disk som overføringsmedium. Etter en slik avlevering på disk kan et Noark-arkivuttrekk med elektroniske dokumenter måtte stykkes opp på flere hundre plater dersom Riksarkivet fortsatt bruker CD-R som bevaringsmedium. Situasjonen bedres ikke markert om mediet for langtidslagring oppgraderes til DVD eller Blu-ray.

Med hensyn til lagringskapasitet er optiske medier fullstendig distansert av de magnetiske mediene – disk og tape. En 2 TB disk lagrer like mye som 80 stk. ett-lags Blu-ray-plater. Langtidslagring på magnetiske medier er imidlertid forbundet med større risiko. Den "ultimate" risikofaktoren er elektromagnetiske pulser (EMP). Disse utløses av atom-bomber, men kan også skapes av dedikerte EMP-våpen uten å gjøre annen fysisk skade.

Magnetisk tape vil i dag typisk innebære bruk av tape-robot (tape-jukeboks). På markedet finnes tape-roboter med en kapasitet opp til 50 petabytes (PB), dvs. 50.000 TB. Med denne løsningen kan bruk av et bortsettingsmedium kombineres med flere av fordelene ved on-line lagring. Framhenting av informasjon og rutinemessig teknisk verifisering av det samlede datainnholdet kan utføres automatisert. Langtidslagring på tape vurderes fortsatt som sikrere og mindre sårbart enn lagring på disk. Kostnadene er også lavere. Det er beregnet at langtidslagring på disk koster 8 ganger mer pr. år. enn lagring på tape (i robot).

Fordelene ved langtidslagring på disk er mange. On-line tilgjengelighet effektiviserer vedlikeholdsaktivitetene ved langtidsbevaring. I tillegg åpnes helt nye muligheter for å tilrettelegge brukertjenester på bevart digitalt arkivmateriale. Man kan eventuelt nøye seg med å vektlegge sikkerheten som ligger i å bruke disklagring i tillegg til en annen teknologi, men disklagring er kostbart. For at dette alternativet skal være kostnadseffektivt og hensiktsmessig, må også de åpenbare praktiske fordelene ved on-line lagring utnyttes.

"State of the art" for lagring av store datavolumer er Storage Area Network-systemer (SAN). SAN-systemer kan også håndtere lagringsenheter på ulike teknologier, f.eks. disk, tape (roboter) og optiske medier i kombinasjon. Ett felles kommandosenter administrerer all lagret informasjon, og muliggjør en samlet overvåking av dataintegriteten på bit-nivå.

Prosjektet har ikke funnet grunn til å gå inn på en videre vurdering av hvilket lagrings-system som er det mest ideelle og økonomiske for digitalt skapt materiale. Riksarkivet har allerede anskaffet et SAN-system. Det kom på plass for å lagre master-filene til bestanden av digitalisert arkivmateriale (100 TB), som pr. i dag har et anslagsvis 200 ganger større volum enn det digitalt skapte materialet. Systemet, som er plassert i et sikret fjellmagasin, skal lagre 3 eksemplarer av hvert objekt på 2 ulike teknologier – ett eksemplar på disk og to på tape (tape-roboter). Når sentrale premisser for lagringen av det digitalt skapte materialet på denne måten er gitt, er prosjektets konklusjoner følgende:

- Løsningen med 3 kopier av hvert objekt på 2 ulike teknologier gir en tilfredsstillende lagringssikkerhet – som en rammeløsning.

- Lagringsteknologiene er i begge tilfeller magnetiske. Lagringssikkerheten styrkes imidlertid ved at magasinene er beliggende i fjell. Men som katastrofeberedskap må en komplett sikkerhetskopi også lagres eksternt.
- SAN-løsningen oppfyller krav til teknologiuavhengighet. Dataobjektene lagres teknologiuavhengig i den forstand at de senere kan flyttes til andre medier og teknologier.

3.3 Konfigurering av fysisk magasin

Riksarkivets nye fjellmagasin for sentral lagring av Arkivverkets digitalt skapte arkivbestand er fysisk sikret som et sperret område. Adgang må kreve sikkerhetsklarering, og være underlagt konfidensialitetskontroll. Lagringsløsningen med en versjon på disk og to kopier på tape må konfigureres med en inndeling i soner fordi ulike kategorier av materiale krever ulike forvaltningsregimer. Originalversjoner av avlevert og deponert arkivmateriale må lagres utilgjengelig for bruk i en lukket del (indre sone) av magasinet – Digitalt sikringsmagasin (DSM). Materiale fra DSM som tilrettelegges for Arkivverkets interne bruk, plasseres i en tilgjengelig ytre sone i magasinet. Utrustningen i den ytre sonen kan være installert i samme magasinrom som DSM, men direkte kommunikasjon mellom DSM og ytre sone må ikke være fysisk mulig.

Tilknytning til DSM skal skje via et dedikert nett. Medarbeidere med slik DSM-tilknytning skal ikke kunne kommunisere med annet utstyr. For disse må det etableres delte kontormiljøer for å sperre for kommunikasjon med ytre sone lokalt. Sentralmagasinets ytre sone skal imidlertid kunne aksesseres via Arkivverkets (åpne) lokale nettverk.

Også ytre sone krever områder med differensierte tilgangsrettigheter, brukerautorisasjon og konfidensialitetskontroll. Informasjonen må kunne inkludere materiale til bruk for publikums- og forskertjenester innenfor Arkivverkets publikumsarealer. Materialet i magasinets ytre sone kan imidlertid ikke være eksternt tilgjengelig. Digitalt arkivmateriale som skal gjøres tilgjengelig for eksterne brukere som en nettjeneste, må være plassert på egne servere utenfor digitalt depot – på linje med informasjonen på Digitalarkivet.

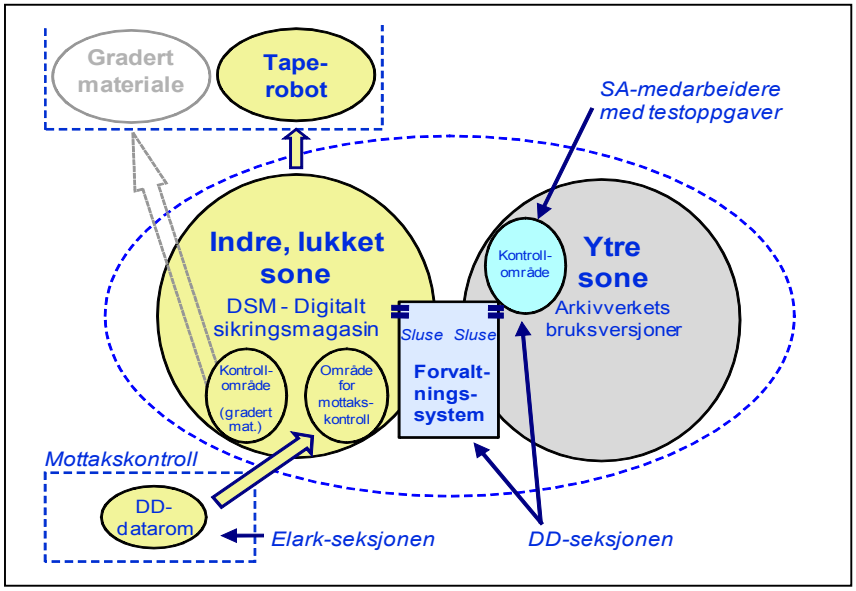
Gradert arkivmateriale må håndteres på dedikert og særskilt beskyttet utstyr i eget magasinrom – "innenfor" DSM, men uten kommunikasjon med DSM. Også i DSM er det imidlertid nødvendig med sikkerhetstiltak på nivå med dem som kreves for gradert informasjon. Det legges til grunn at lavgradert materiale også kan lagres i DSM. Det gjenstår å avklare med Nasjonal sikkerhetsmyndighet hvilket graderingsnivå DSM kan godkjennes for, men de nærmeste årenes deponeringer og avleveringer ventes uansett bare å innholde lavgradert digitalt arkivmateriale.

Konfigureringen av digitalt magasin illustreres i figuren nedenfor. Her vises også følgende andre områder i tilknytning til Arkivverkets digitale magasin:
- *DD-datarom* (eksisterende datarom for Seksjon for digitalt depot): Eget datarom i Riksarkivets administrasjonsbygning for sentral mottakskontroll av digitalt materiale. Sentrale mottakskontroll forutsettes utført av Elark-seksjonen i Riksarkivet.
- *Forvaltningssystem:* Eget system for å generere og vedlikeholde arkivpakker, for å hente arkivpakker inn/ut av DSM og for å flytte informasjon mellom DSM og ytre sone. Forvaltningssystemet og dets database ligger på et eget arbeidsområde i indre sone.

- *Kontrollområder:* Områder for lagring av materiale i fasen med testing etter utført mottakskontroll – inntil det kan innlemmes i DSM av forvaltningssystemet. Av hensyn til sikkerheten i DSM må et separat kontrollområde for testing som utføres av medarbeidere i statsarkivene (SA), ligge fysisk i ytre sone. Etter testing hentes materiale fra kontrollområdene til DSM av forvaltningssystemet.

Konfigurasjonen skal omfatte 2 stk. tape-roboter. Den ene plasseres i magasinet for gradert materiale. Den vil være forbundet med utstyret i sentralmagasinet, og skal ikke ha tilknytning til den dedikerte lagringsløsningen for gradert materiale i samme rom. Tape-robot nr. 2 plasseres i det sentrale magasinet. Denne er ikke inntegnet på figuren. Nåværende datarom for lagring av CD-er i Riksarkivets eldre fjellmagasin er heller ikke inntegnet. Dette magasinet forutsettes brukt til lagring av sikkerhetskopier på tape (tapeversjon nr. 3) inntil det foreligger en løsning for ekstern lagring av ("off site") sikkerhetskopier.

Figur: Digitalt magasin



Prikket linje markerer fysiske datarom. Indre og ytre sone er plassert i samme rom, men på fysisk skilte anlegg med tilknytning til to separate nettverk. Dedikert system for lagring av gradert materiale (planlagt) er plassert i eget rom som en indre del av indre sone. DD-rommet ligger ikke i fjellmagasinet, men er fysisk og logisk tilknyttet indre sone.

Mellom indre og ytre sone foreslås en sluse for å flytte og kopiere informasjon mellom de to sonene. Eksport til og fra DSM vil være en svært tungvint affære uten en slik løsning. Men sluseløsningen må fungere slik at den ikke punkterer sikkerheten i DSM. Den må kopiere informasjon mellom sonene på en strengt kontrollert måte, og hindre samtidige operasjoner i dem. Det foreslås en konfigurasjon hvor slusen styres fra det digitale depotets forvaltningssystem. Maskinen som kjører forvaltningssystemet utstyres med to nettverkskort med forbindelser til hver sin sone. Bare ett nettverkskort kan være aktivt om gangen.

Konfigurasjonen har også den fordel at den medfører atskilte operasjonsområder for teknisk drift og forvaltningen av informasjonsinnholdet (arkivobjektene) i DSM. Driftssiden (IT-avdelingen i Riksarkivet) forholder seg til SAN-administrasjonssystemet (ikke inntegnet på figuren). Seksjonen for digitalt depot (DD-seksjonen) forholder seg til forvaltningssystemet, og har inngang til DSM via forvaltningssystemets område.

3.3.1 Eget forvaltningssystem for digitalt depot

Det sentrale forvaltningssystemet i Arkivverkets digitale depot vil være en applikasjon med flere funksjoner. Det må utvikles spesielt for Arkivverket, eventuelt som en tilpasning av et eksisterende system. Systemet planlegges bl.a. å utføre følgende oppgaver:

- generere arkivpakker til DSM vha. eget pakkeprogram – herunder sjekksumgenerering
- oppdatere arkivpakker ved å hente dem ut av DSM og tilbakeføre dem til DSM med nygenererte sjekksummer etter oppdateringen
- generere informasjon til Asta-systemet og systemet for SAN-administrasjon og drift
- generere bruksversjoner av arkivmateriale i DSM for tilgjengeliggjøring i ytre sone
- gi oversikt over innhold og arkivpakkestrukturer i DSM
- hente data inn/ut av DSM og ytre sone, flytte data mellom DSM og ytre sone og styre sluseløsningen som skal hindre samtidige operasjoner i de to sonene.

3.4 Navngivingskonvensjon for logiske identifikatorer

Navngiving og adressering av arkivpakker i digitalt depot må være basert på bestandige identifikatorer. Det må genereres en unik ID for hver arkivpakke. Det foreslås å bruke en ID av samme type som for Arkivverkets digitaliserte materiale, f.eks. en streng på formatet ”aipååååmmddppnnn”, hvor ”aip” er et fast prefiks for AIP-er, ”ååååmmdd” pakkens produksjonsdato, ”pp” en medarbeiderkode (i de tilfeller ID tildeles manuelt) og ”nnn” et løpenummer innenfor datoen. Det anbefales å bruke ID-en som filnavn når arkivpakker ferdigstilles for lagring som tar-filer, f.eks. *”aip2011042400003.tar”*. På denne måten kan det automatisk etableres en URN (Uniform Resource Name) til pakken, f.eks. *URN:NBN:no-a1450-aip2011042400003.tar*. Denne kan igjen danne basis for en URL.

3.5 Samspillet mellom digitalt depot og Asta-systemet

Asta, Arkivverkets arkivinformasjonssystem, er primært et arkivbeskrivelsessystem og et verktøy for publikum. Systemet har ikke funksjonalitet for å administrere et digitalt depot av det kaliber som denne rapporten beskriver. Men på linje med annet arkivmateriale må arkivpakker i digitalt depot beskrives i Asta. Asta må både gi oversikt over ”originale” arkivpakker i indre sone (DSM) og bruksversjoner i den ytre. I det siste tilfellet vil Asta danne utgangspunktet for brukertjenester, og representere inngangen til arkivpakkene.

Arkivverkets Asta-system er plassert på en server som (logisk sett) tilhører den ytre sonen i digitalt depot. Astas beskrivelser av arkivpakkene må oppdateres i samspill med forvaltningssystemet i indre sone (DSM), som vil være den instans som genererer informasjon til Asta. Ved oppdatering av arkivbeskrivelser må det skje en synkronisering av informasjon mellom Asta, forvaltningssystemet, SAN-administrasjonssystemet og vedkommende arkivpakke i lagringssystemet. Forvaltningssystemet forutsettes da å ”melde” nygenererte og oppdaterte arkivbeskrivelser ved å legge dem ut som bestillinger til Asta. Oppdatering av overordnede opplysninger om arkiv og arkivskapere må imidlertid være styrt fra Asta-systemet. Disse må da meldes som bestillinger til forvaltningssystemet. Informasjonsutvekslingen mellom systemene må sluses mellom lagringssystemets skilte soner. Denne slusingen må styres av forvaltningssystemet i DSM.

Asta må referere til arkivpakkenes unike ID (URN), men opplysningene kan ellers begrenses til den overordnede arkivbeskrivelsen for arkivpakken som helhet. For bruksversjoner i ytre sone stiller det seg noe annerledes. Her vil det være behov for fyldigere dokumentasjon, spesielt om tilknyttede metadata. Registreringen av opplysninger i Asta bør kunne skje etter samme prinsipp som for materiale på CD-er i dag. Det vil si at systemet, delsystemet eller systemfunksjonen som har produsert arkivpakken, registreres som en *serie* under vedkommende arkivskapers felles *arkiv,* og at hver arkivpakke (AIP eller AIC) legges inn som en *underserie*. En bruksversjon av en arkivpakke (DIP) kan registreres i Asta som underserie til en underserie (her er det også mulig å bruke relasjonen "Versjon av" mellom arkivenheter i Asta). De to ulike typene av arkivpakker (AIP eller DIP) og deres soneplassering må uansett identifiseres i Asta.

4. PRODUKSJONSLINJER OG STYRINGSFUNKSJONER

4.1 Produksjonslinjer ved aksesjon og vedlikehold av arkivmateriale

Arkivverkets digitale depot krever fast organiserte produksjonslinjer i tilknytning til mottakskontroll, testing, innlemmelse og vedlikehold av arkivmateriale. I produksjonslinjene kreves definerte ansvarsområder og roller for å ivareta sentrale koordinerings-, tilsyns- og kvalitetssikringsfunksjoner.

4.1.1 Sentral mottakskontroll

Behovet for en ubrutt integritets- og konfidensialitetssikring gjør det nødvendig å etablere en sentral mottakskontroll i Riksarkivet. Dette gjelder også for materiale som skal testes av medarbeidere i statsarkivene. Mottaksinstansen skal kontrollere at en avleveringspakkes innhold og dokumentasjon er i samsvar med forutsetningene, verifisere medfølgende sjekksum for den samlede avleveringspakken og dokumentere behandlingsprosessen i Arkivverket. Umiddelbart ved mottak skal den dessuten generere en samlet sjekksum for hver avleveringspakke for å muliggjøre en senere verifisering av at informasjonsinnholdet er bevart uendret fra og med ankomst i Arkivverket.

4.1.2 Testing av mottatt materiale

Avleveringspakker som passerer den initielle kontrollen, hentes av forvaltningssystemet for DSM. En kopi plasseres på eget kontrollområde for testing. Systemet for prosesstyring må vise hvem som er ansvarlig for testingen, og vise testingens status.

Avleveringspakker skal testes og vurderes for godkjenning etter kriteriene i avleveringsbestemmelsene (§ 8-8). Dersom testingen resulterer i justerte eller oppdaterte tekniske metadata, skal denne dokumentasjonen lagres som tillegg til avleveringspakkens originale tekniske metadata. Dokumentasjon av testoperasjoner og -resultater integritetssikres med en samlet sjekksum av testeren. Testeren skal også utarbeide en oversikt over de filformater som forekommer i avleveringspakkens datainnhold.

Materiale som er godkjent etter testing, hentes tilbake fra kontrollområdet av forvaltningssystemet sammen med testdokumentasjonen. Koordinator skal også kvalitetssikre oppdateringen av arkivsystemet og prosesstyringssystemet.

4.1.3 Generering og vedlikehold av arkivpakker

Når en avleveringspakke (SIP) er ferdig testet og godkjent, skal forvaltningssystemet generere en arkivpakke (AIP/AIC). Kjernen i en ny arkivpakke vil være avleveringspakken supplert med dokumentasjon av depotoperasjonene ved mottakskontroll og testing. Genereringen av en arkivpakke vil bl.a. omfatte følgende delprosesser:

- tilrettelegging av bevaringsmetadata som skal inkluderes i pakken
- generering av en unik ID for pakken
- tilrettelegging av metadata om pakken for forvaltningssystemets database, for Asta og for SAN-administrasjonssystemet. (Metadata for Asta hentes fra to tilgjengelige kilder: arkivskapers dokumentasjon (i SIP) og prosesstyringssystemet for mottak og testing)
- beregning av alle gjenstående sjekksummer for ferdigstilte filer innenfor pakken
- ferdigstilling av pakken som en arkivfil (tar-fil)
- beregning av samlet sjekksum for pakken
- sammenknytning av pakken med andre pakker til en samlepakke (AIC)
- beregning av samlet sjekksum for samlepakken med tilhørende arkivpakker samt plassering av denne sjekksummen utenfor samlepakken

Alt vedlikehold og all oppdatering av bevarte arkivpakker styres fra forvaltningssystemet. Oppdatering skal skje ved at pakker hentes fra DSM til arbeidsområdet for forvaltningssystemet i den indre sonen. Oppdateringer må medføre generering av nye sjekksummer og synkronisering med Asta og SAN-administrasjonssystemet.

4.1.4 Prosesstrinn ved aksesjon

Tabellen nedenfor gir en forenklet oversikt over trinnene i behandlingsprosessen fra materiale mottas til genererte arkivpakker blir innlemmet i DSM.

	Oppgave/funksjon:	*Ansvarlig etter utført:*	*Lagrings-område:*
1	Materiale mottas som rekommandert post eller med bud, og registreres i arkivsystemet. (Ved elektronisk overføring vil Elark-seksjonen være mottaker og registreringsansvarlig).	Arkiv-tjenesten	(logisk) ytre sone
2	Arkivtjenesten bringer materialet til den sentrale mottaks-kontrollen.	Elark	DD-rom
3	Materialet kopieres til et eget (off-line) lagringsområde[11]. Omgående genereres deretter en samlet sjekksum for avleveringspakken, dvs. for mottatte datafiler og all tilhørende dokumentasjon. Operasjonene krever 2 autoriserte personer, og registreres i en egen logg. Person nr. 2 skal bekrefte den genererte sjekksummen og verifisere den.	Elark	DD-rom
4	Materialet settes i minimum 3 ukers karantene på eget (off-line) lagringsområde i påvente av virus-skanning.	Elark	DD-rom

[11] Dette vil ikke være nødvendig når materiale mottas på USB-disk. Sjekksumgenerering vil da kunne foretas på mottatt medium. Materialet kan også plasseres i karantene på mottatt medium., jf. punkt 4.

	Oppgave/funksjon:	*Ansvarlig etter utført:*	*Lagrings-område:*
5	Kvittering med opplysninger om behandlingen og antatt behandlingstid sendes arkivskaperen. Avleveringen registreres i prosesstyringssystem for mottak og testing (p.t. ArkiVente). Nødvendige opplysninger fremskaffes fra Asta.	Elark	DD-rom
6	Viruskontroll utføres etter karantenen.	Elark	DD-rom
7	Samlet sjekksum for avleveringspakken verifiseres og dokumenteres av 2 personer når den medfølger fra arkivskaperen. Verifisering av medfølgende sjekksummer *innenfor* pakken utføres som ledd i testingen, jf. pkt. 11-12.	Elark	DD-rom
8	Det foretas initiell kontroll av følgende: a) at avleveringspakken er tilfredsstillende autorisert av avgiveren, og b) at informasjonsinnholdet er korrekt og komplett iht. forutsetninger og avtaler.	Elark	DD-rom/ indre sone
9	Materialet kopieres til eget lagringsområde for mottakskontrollen i indre sone. Supplert med dokumentasjon fra mottakskontrollen legges materialet i "kø", klargjort for å hentes til testing av koordinator i DD.	DD	Indre sone/ m-området
10	Koordinator i DD henter materialet til forvaltningssystemets lagringsområde i indre sone. Ansvaret for arkivskaperkontakt og prosessdokumentasjon overtas av DD.	DD	Indre sone/ f-området
11	Materialet fordeles til testansvarlig i DD eller statsarkiv. I sistnevnte tilfelle overtar vedkommende statsarkiv også ansvaret for arkivskaperkontakt og dokumentasjon.	DD/ statsarkiv	Ytre sone/ kontroll-området
12	Gjennomført testing dokumenteres av testansvarlig, herunder evt. innhentede supplerende opplysninger. Tilleggsopplysninger og evt. justerte versjoner av tekniske metadata lagres som tillegg til det opprinnelige materialet, og sikres med sjekksummer.	DD/ statsarkiv	Ytre sone/ kontroll-området
13	Dersom avleveringen godkjennes etter testing, sendes godkjenningsbrev til arkivskaper. I motsatt fall begjæres nytt datauttrekk eller supplerende dokumentasjon. Interne registre oppdateres.	DD/ statsarkiv	Ytre sone/ kontroll-området
14	Det samlede materialet tilhørende en godkjent avlevering legges i "kø" på kontrollområdet, klargjort for innhenting av forvaltningssystemet. Ansvaret for oppdatering av interne registre følger materialet.	DD	Indre sone/ f-området
15	Materialet hentes til forvaltningssystemets arbeidsområde av koordinator i DD, og kvalitetssikres mht. dokumentasjonens kompletthet. Det skal bl.a. påses at det finnes opplysninger om filformater og tilgangsbestemmelser mm.	DD	Indre sone/ f-området
16	Det samlede materialet organiseres som en arkivpakke (AIP). Den genereres ved hjelp av forvaltningssystemet, og pakkes som en tar-fil. Samlet sjekksum for arkivpakken genereres 2 ganger: før pakkingen til tar-format – og etter.	DD	Indre sone/ f-området

	Oppgave/funksjon:	*Ansvarlig etter utført:*	*Lagrings-område:*
17	Arkivpakken tilknyttes en overordnet samlepakke (AIC). Dersom denne allerede eksisterer i DSM, hentes den ut, og oppdateres. Samlet sjekksum for tilknyttet AIP lagres i AIC. For ny eller oppdatert AIC genereres samlet sjekksum (2 versjoner) for lagring utenfor AIC.	DD	Indre sone/ f-området
18	Ved hjelp av forvaltningssystemet genereres beskrivelses-informasjon til Asta. Informasjon (til PDA-seksjonen) om aksesjon og tilvekst skal også genereres.	DD	Indre sone/ f-området
19	Ny arkivpakke (AIP) og ny eller oppdatert AIC innlemmes i DSM. Vha. forvaltningssystemet genereres informasjon til SAN-administrasjonssystemet i DSM, herunder samlet sjekksum i 2 versjoner for vedkommende AIC.	IT	Indre sone/ DSM

4.2 Konfidensialitetssikring og tilgangsstyring

Håndteringen av gradert materiale i DSM må følge bestemmelsene i sikkerhetsloven, men Arkivverkets egne sikkerhetsbehov gir grunn til å stille krav om at alt lagret materiale i DSM skal være beskyttet på nivå med gradert informasjon etter sikkerhetsloven. DSM må som helhet organiseres som et sperret område med den adgangskontroll som er nødvendig for å håndtere sikkerhetsgradert materiale.

Behandlingen av personopplysninger må oppfylle bestemmelsene i personopplysnings-loven (pol) og personopplysningsforskriften. Pol §§ 13 og 14 gir behandlingsansvarlige virksomheter pålegg om "planlagte og systematiske tiltak" for å sørge for "tilfredsstillende informasjonssikkerhet med hensyn til konfidensialitet, integritet og tilgjengelighet ved behandling av personopplysninger". Både informasjonssystemet og sikkerhetstiltakene skal være dokumentert. Dokumentasjonen av virksomhetens internkontroll skal dessuten være tilgjengelig for Datatilsynet og Personvernnemnda.

4.2.1 Håndtering av personopplysninger

Sensitive personopplysninger kan være gradert etter beskyttelsesinstruksen, men ikke alltid. Sensitive persondata som ikke er gradert, foreslås håndtert og lagret i digitalt depot på lik linje med materiale gradert FORTROLIG etter beskyttelsesinstruksen. Dette medfører ensartede behandlingsregler for sensitive personopplysninger og materiale gradert BEGRENSET, FORTROLIG og STRENGT FORTROLIG.

Materiale med sensitive personopplysninger foreslås klassifisert og merket på enhetsnivå på samme måte som gradert materiale. Det er behov for en årlig gjennomgang av dette materialet i DSM for å revurdere beskyttelsesbehovet for hver enkelt enhet.

Sensitive opplysninger etter personopplysningsloven er også taushetsbelagte etter forvaltningsloven. Men taushetsbelagte opplysninger etter forvaltningsloven kan omfatte annet enn sensitive personopplysninger. Prosjektet foreslår at alle typer taushetsbelagt materiale blir behandlet som sensitive, og håndtert på linje med materiale gradert BEGRENSET i digitalt depot.

4.3 Overvåking av installasjoner og prosesser

Det kreves definerte funksjoner for tilstands- og prosessovervåking og for rutinemessig logging av hendelser i tilknytning til:

1) *SAN-administrasjonssystemet,* som er den del av DSM som IT-avdelingen forholder seg til som driftsansvarlig
2) *Forvaltningssystemet,* som administreres av DD-seksjonen
3) *Mottakskontrollen,* som administreres av Elark-seksjonen
4) *Kontrollområder* for testing
5) *Ytre sone*, dvs. sonen for Arkivverkets brukskopier av arkivmateriale.

Disse funksjonene blir detaljert spesifisert i den fullstendige versjonen av rapporten.

5. TILGJENGELIGGJØRING AV ARKIVMATERIALE

Digitalt skapt arkivmateriale som skal være tilgjengelige for bruk i Arkivverket, må være plassert i det digitale magasinets ytre sone. Det vil da være tale om brukspakker (DIP) – dvs. bruksversjoner av arkivpakker i DSM – vanligvis i en tilrettelagt form. Eksempler på materiale som må finnes i versjoner klargjort for bruk i den ytre sonen, er uttrekk fra Noark-3-systemer som er relatert til avleverte sakarkiver i papirform til statsarkivene. Bruk av materialet i Arkivverket inkluderer publikums- og forskertjenester innenfor etatens publikumsarealer. Materiale som skal gjøres tilgjengelig for eksterne brukere som en nettjeneste, må imidlertid være plassert på egne servere utenfor digitalt depot.

Også den ytre sonen krever sikkerhetstiltak: områder med differensierte tilgangsrettigheter, brukerautorisasjon og konfidensialitetskontroll. Brukere forutsettes bare å ha lesetilgang, men som integritetssikring er dette ikke tilstrekkelig. Også bruksversjoner må integritetssikres med sjekksummer, og brukere bør ha mulighet for å verifisere sjekksummer.

5.1 Typer av bruksversjoner og brukertjenester

Det vil være behov for ulike typer av bruksversjoner. I mange sammenhenger vil en kopi av en arkivpakke fra DSM kunne tjene som brukspakke uten videre tilrettelegging. I andre tilfeller vil det være behov for å generere spesielt tilrettelagte brukspakker. Dessuten kan det være aktuelt å lage bruksversjoner som bygger på flere arkivpakker eller deler av pakker. En brukspakke må derfor alltid dokumentere sin sammenheng med objekter i en eller flere arkivpakker for at det skal være mulig å vurdere materialets autentisitet.

Det bør bli mulig for autoriserte brukere å fremhente bruksversjoner ved behov med utgangspunkt i Asta. Når Asta viser at det finnes en brukskopi i ytre sone, bør pakken kunne åpnes av en autorisert bruker i en dertil egnet brukertjeneste som er tilpasset for Noark-3, Noark-4, fagsystemuttrekk osv. På sikt bør brukere kunne sende en ”bestilling” til forvaltningssystemet dersom det ikke finnes en kopi av arkivpakken i den ytre sonen.