Digitalt og autentisk

Planlegging av ny depotløsning for Arkivverkets digitalt skapte arkivmateriale

Prosjektrapport

Riksarkivet 01.03.2010

Innhold

1. INNLEDNING

Rapporten er utarbeidet av prosjektet "Elmag-2" – Planlegging av *el*ektronisk *mag*asin for det digitalt skapte arkivmaterialet i Arkivverket. Prosjektarbeidet startet i oktober 2008, og ble avsluttet i februar 2010.

Prosjektrapporten har to deler. Vedlegg med skjematiserte kravspesifikasjoner og en sikkerhets- og sårbarhetsanalyse (vedlegg 1-4) er samlet i rapportens del 2.

1.1 Prosjektmål

Rammen for prosjektet ble definert av Riksarkivaren 07.10.2008 (sak 08/19731). Følgende mål ble definert 22.01.2009 i den justerte versjonen av prosjektplanen (sak 08/21078):

> *Prosjektet skal legge fram forslag til en konkret plan for å etablere et elektronisk magasin for Arkivverkets digitalt skapte arkivmateriale som kan gi samlet oversikt og kontroll over dette materialet, og muliggjøre et samlet vedlikehold. Lagringssystemet og de prosedyrer og rutineopplegg som etableres i tilknytning til det, skal sikre det digitale materialets opprettholdte lesbarhet og autentisitet, beskytte det mot alle former for uautorisert endring og hindre uautorisert tilgang.*

I beskrivelsen av prosjektets bakgrunn og formål fremhever prosjektplanen 3 momenter:

- Nært forstående deponeringer/ avleveringer av elektroniske dokumentarkiver krever et magasinsystem med en helt annen dimensjonering enn Arkivverkets nåværende CD-baserte lagring.
- Det er behov for å utnytte muligheter som har åpnet seg for å behandle, vedlikeholde og overvåke Arkivverkets elektroniske arkivbestand på en sikrere og mer effektiv måte enn tidligere.
- Det er behov for å implementere kriteriene for "trusted repositories" fra nyere internasjonalt standardiseringsarbeid.

1.2 Gjennomføringen av prosjektet

Prosjektarbeidet er gjennomført med følgende organisering:

Prosjektleder:	Trond Sirevåg
Prosjektgruppe:	Hans Fredrik Berg, Terje Pettersen-Dahl, Anthony Lærdahl og Stian Skindlo
Ressursgruppe:	Tor Breivik, Lars Nygaard og Petter Svendsen
Styringsgruppe:	Ivar Fonnes, Anne Mette Dørum, Hans-Herman Fischer og Ole Gausdal

Ressursgruppens oppgave har vært å vurdere forslag fra prosjektgruppen i en videre sammenheng og med særlig vekt på sikkerhet og internkontroll. De to gruppene har hatt felles møter i sluttfasen av prosjektet.

Prosjektet bygger på resultatene fra Elmag-prosjektet (senere kalt Elmag-1) som ble avsluttet med en foreløpig rapport 25.01.2008: "Elektronisk magasin - Versjon 1" (sak 07/18114). Dette prosjektets hovedkonklusjon var at det må bygges opp en sikkerhetsorganisasjon i Riksarkivet før det kan etableres et elektronisk magasin. Riksarkivaren oppnevnte 13.11.2009 en sikkerhetsorganisasjon for Riksarkivbygningen.

Etter den opprinnelige planen av 28.10.2008 var rammen for prosjektet å levere forslag om etablering av et elektronisk magasin. I januar 2009 ble rammen utvidet til også å omfatte den konkrete planleggingen. Sluttføringen av prosjektet, som opprinnelig var satt til høsten 2009, ble i august 2009 forskjøvet til 31.12.2009. Samtidig ble det besluttet å inkludere bruksversjoner av bevart arkivmateriale og en beskrivelse av organisasjons- og arbeidsrutiner i prosjektplanen. Disse senere justeringene i planen ble foretatt da det ble avklart at en magasinløsning etter prosjektets modell ikke ville kunne realiseres i 2010.

I prosjektet har strukturen i arkivpakker basert på OAIS-modellen vært et viktig tema fordi denne strukturen også vil bestemme utformingen av magasinløsningens sentrale forvaltningssystem. Strukturen i arkivpakker var emne for et eget prosjektseminar i Riksarkivet i juni 2009 med deltakelse av ekspertise fra det svenske riksarkivet. Temaet ble videre diskutert på Nordisk arkivakademi i Boden 10.-11. november 2009.

Prosjektarbeidet medførte en erkjennelse av at problematikken omkring arkivpakkestruktur og systemforvaltning også nødvendiggjør et samarbeid med sentrale aktører på kommunal sektor. I oktober 2009 gjorde Riksarkivaren avtale med Oslo byarkiv, Bergen byarkiv, IKA Trøndelag og IKA Møre og Romsdal om å søke om midler fra ABM-utvikling til å gjennomføre et samarbeidsprosjekt i 2010. Søknad om støtte til dette prosjektet – "Arkivpakkestruktur for digitalskapt arkivmateriale" – ble sendt 15.10.2009. Søknaden ble meldt innvilget av ABM-utvikling 22.02.2010.

Samarbeidsprosjektet om arkivpakkestruktur skal etter planen omfatte to hovedaktiviteter i 2010: en omforent struktur som dekker behovene både på statlig og kommunal sektor, og en kravspesifikasjon til et system som kan generere og vedlikeholde arkivpakker med en slik struktur. Uviklingen av programvare på grunnlag av denne kravspesifikasjon – det vil for Arkivverkets vedkommende si basiselementet i magasinløsningens forvaltningssystem – ligger utenfor rammen for samarbeidsprosjektet, og antas først å kunne skje i 2011. Arkivverkets forvaltningssystem vil også kreve viktige funksjoner i tillegg til programvaren som utvikles i kjølvannet av samarbeidsprosjektet med de kommunale aktørene.

Arkivverkets realisering av en lagringsløsning etter prosjektets modell vil derfor måtte skje trinnvis. Da blir det desto viktigere at det foreligger en oversikt over komponentene som må være på plass for å implementere modellen. På møte i styringsgruppen 02.12.2009 ble prosjektet bedt om å utarbeide en oversikt over nødvendige oppfølgingstiltak og konkrete forslag til nye prosjektløp etter sluttføringen av Elmag-2 31.12.2009. Slike tiltak for oppfølging og videreføring behandles i rapportens kapittel 7.

1.3 Sentrale begreper i rapporten

Begrepet *digitalt depot* brukes gjennomgående i rapporten. Det har et videre begrepsinnhold enn digitalt magasin, og samler i seg alle aspektene som rapporten behandler:

- de fysiske lokalene, utrustningen og den øvrige materielle infrastrukturen for å lagre, administrere og vedlikeholde digitalt arkivmateriale,

- de digitale objektene som lagres og vedlikeholdes på installasjonene,
- systemene for å administrere de fysiske installasjonene og tilhørende infrastruktur,
- systemene for å administrere den lagrede informasjonen som digitale objekter og som arkivobjekter,
- rutineoppleggene for å bruke og vedlikeholde installasjonene, de administrative systemene og de lagrede objektene,
- prosessene knyttet til mottak, testing og innlemmelse av nytt arkivmateriale,
- system- og rutineopplegget for å gi tilgang til lagret digitalt arkivmateriale.

Autentisitet og integritet er begreper som ofte brukes litt om hverandre. Med opprettholdt *integritet* menes at informasjon er bevart uendret. Flere definisjoner, bl.a. den som brukes av det pågående internasjonale arkivprosjektet InterPARES[1], presiserer at informasjonen også må være bevart fullstendig. I denne rapporten inkluderer integritet et krav om fullstendighet når det brukes som et teknisk begrep. At informasjon på fysisk nivå er bevart som en fullstendig bit-strøm, er en praktisk forutsetning for å bekrefte opprettholdt integritet ved bruk av en sjekksum (jf. nedenfor).

I rapporten brukes også betegnelsen *innholdsintegritet*, og da om noe mer avgrenset: At det logiske informasjonsinnholdet (typisk tegn og tall) er bevart uendret og fullstendig. Innholdsintegritet er et logisk begrep. Logisk informasjonsinnhold[2] kan være uendret selv om den fysiske representasjonen eller egenskaper ved den visuelle presentasjonen er blitt endret. Dette fremsto som uproblematisk så lenge det 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.

Autentisitet brukes i rapporten som betegnelse for at informasjonen og dens opphavs-sammenheng[3] er hva den utgir seg for å være. InterPARES inkluderer integritet i autentisitetsbegrepet. 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 mikrofilmutgaver kan bli vurdert som

[1] http://www.interpares.org/ip2/ip2_terminology_db.cfm

[2] Her er det viktig å presisere at OAIS-standarden bruker informasjonsinnhold ("Content Information") i en spesialisert betydning hvor tekniske aspekter er inkluderet. Informasjonsinnhold i OAIS består av to komponenter: Data (informasjonen i seg selv) og representasjonen for å fremstille dem, jf. punkt 2.2.1 og 4.2.1,ff.

[3] InterPARES tilstreber interdisiplinære definisjoner, og definerer autentisitet allment og vidt som "a record that is what it purports to be, and that is free from tampering or corruption". Begrepet nærmer seg dermed den vanlige folkelige definisjonen, hvor det brukes mer eller mindre synonymt med ekthet. Med dette bidrar InterPARES tilsynelatende til å utvanne autentisitet som spesifikt arkivbegrep. For å knytte en "record" til en opphavssammenheng anvender imidlertid InterPARES et tilleggsbegrep: Identity.

autentiske – under forutsetning av en kjent sammenheng med omgivelsene[4]. 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.

Informasjon med opprettholdt integritet vil heller 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 brukssammenheng. I praksis vil det si at dokumentet ikke kvalifiserer til betegnelsen arkivdokument.

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 dokumenteres. I så fall er arkivdokumentet bevart som et autentisk falsum.

Med *arkivobjekt* menes et lagret objekt som lar seg identifisere som arkivmateriale, dvs. med den tilknyttede forståelsesinformasjon (logiske metadata) som gjør det kvalifisert til betegnelsen arkivmateriale. Dette i motsetning til et *digitalt objekt*, som omfatter en sekvens av bits med de tekniske metadata som gjør den til en meningsfull enhet, f.eks. en pdf-fil. 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. En slik transformering hviler på forutsetningen om at autentisitet primært går på innhold, ikke på form og heller ikke på teknisk representasjon. Digitalt arkivmateriale kan følgelig konverteres og skifte form uten å tape sin autentisitet. Men en slik transformering stiller strenge krav til dokumenterte prosesser, autentiserende metadata og medfølgende hjelpemidler for å bruke materialet.

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 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 (Message-Digest algorithm 5) og SHA (Secure Hash Algorithm). Ulike bitstrømmer *kan* gi samme sjekksum, men sannsynligheten avtar med økende sjekksumlengde.

[4] Anders Bo Nilsen: *Elektroniske arkivaliers autenticitet. Form eller innhold – i teori og praksis*. ARKIV nr. 6 – 2001.

2. KRAV TIL ARKIVVERKETS DIGITALE ARKIVDEPOT

Kapitlet behandler sikkerhetsaspektene og de sentrale teknologiske og metodologiske forutsetningene for å etablere et digitalt arkivdepot. Det gjennomgår nyere internasjonale standarder for arkivbevaring med særlig vekt på løsninger for å sikre materialets integritet og autentisitet. Det spesifiserer på denne bakgrunn krav til Arkivverkets håndtering av digitalt skapt arkivmateriale, og avsluttes med å sette opp krav til den overordnede konfigureringen av Arkivverkets digitale depot.

2.1 Utfordringer og krav ved bevaring av digitalskapt arkivmateriale

Langtidsbevaring av digital dokumentasjon representerer store utfordringer, og medfører en rekke risikofaktorer. En oversikt over de mange og ulike sikkerhetsbehovene som må ivaretas ved digital bevaring, følger nedenfor.

a) Lagringssikkerhet
Basiskravet ved digital lagring er sikkerhet for at informasjonen holdes digitalt intakt. Elektroniske lagringsmedier har kort levetid. Med regelmessige mellomrom må all bevart informasjon på slike medier overkopieres til nye databærere for å sikre den digitale bestandigheten. Overkopiering må skje hvert 3. til 5. år ved lagring på disk, og hvert 5. til 10. år ved lagring på tape eller optiske plater. Uten et slikt vedlikehold vil informasjonen gå naturlig tapt etter relativt få år. Elektroniske medier medfører imidlertid også risiko for tekniske lagringsfeil. All informasjon må derfor sikkerhetskopieres. Lagringen må dessuten være gjenstand for rutinemessig kontroll. For å øke sikkerheten benyttes fortrinnsvis ulike lagringsteknologier for originalmateriale og kopiversjoner. Langtidsbevaring av informasjon på elektroniske medier er følgelig ressurskrevende. Oppgaven innbefatter aktive vedlikeholdsrutiner som ikke tåler kontinuitetsbrudd.

b) 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. På elementært nivå dreier lesbarhet seg om maskin- og programvarens evne til å tolke digitalt kodede data (en bit-strøm) som meningsfylt informasjon. Teknologiendringene gjør det nødvendig å konvertere informasjon som skal bevares, til bærekraftige formater som er tolkbare for ulike generasjoner og typer av utstyr. Senere må man også være beredt til å foreta nye og samlede formatkonverteringer med visse mellomrom for at informasjonen skal overleve i en fortsatt lesbar form. Rutinemessig formatkontroll og periodisk formatkonvertering er en nødvendig del av vedlikeholdsaktivitetene ved digital langtidsbevaring.

c) Opprettholdt 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 ha 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 kontekst-opplysninger 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 arkiv-materiale. Arkivdokumenter 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.

d) Opprettholdt integritet
For å tjene som dokumentasjon må den bevarte arkivinformasjonen være pålitelig. Tillit til informasjonen og arkivdepotets håndtering av den er en helt avgjørende faktor. At digital informasjon enkelt kan kopieres og endres, medfører også 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 også være bevart med opprettholdt integritet, dvs. med et uendret innhold[5]. Integriteten må ivaretas uavbrutt gjennom de migreringer, konverteringer og øvrige vedlikeholdsoperasjoner som utføres i et arkivdepot, og det bør finnes mekanismer for å bekrefte dette. Integritetssikring er blant de mest krevende utfordringene ved langtids-bevaring av digitalt arkivmateriale.

e) Sikret konfidensialitet
Nyere arkivmateriale fra offentlig forvaltning kan være gradert eller taushetsbelagt, og krever tiltak for konfidensialitetssikring, dvs. 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 taushetsbelagt 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.

Når det spesielt gjelder kravene om lesbarhet og forståelighet, er det behov for å tilføye at opprettholdt tilgjengelighet til arkivinformasjon nå også er blitt et lovbestemt rettighets-krav. Tilgjengelighet til personopplysninger kom til som en helt ny bestemmelse i personopplysningsloven § 13, jf. også § 2-12 i tilhørende forskrift. Kravet vil også ha gyldighet for avleverte arkivversjoner av registermateriale med personopplysninger.

Som det fremgår av oversikten ovenfor, medfører langtidsbevaring av digital informasjon mange former for risiko. Informasjonen må rutinemessig kopieres til nye "friske" medier. Men digital kopiering innebærer en kloning av informasjon, og resulterer ikke i en trinnvis

[5] Uendret innhold må i dette tilfellet forstås logisk, og ikke som en uendret fysisk bit-strøm. Konverteringer vil kunne endre bit-strømmen.

svekkelse av den på linje med reprografi. Gitt at bevaring skjer på en standardisert og riktig måte, og gitt at vi – og senere våre etterkommere – utfører de vedlikeholdsoppgaver som kreves, 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.

2.1.1 Lagringsløsning - Alternative teknologier

Sikkerhetsaspektene er fundamentale, men langtidslagring på elektroniske medier har også andre viktige aspekter som økonomi, tilrettelegging for effektivt vedlikehold og hensiktsmessighet med tanke på bruk og brukertjenester.

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). Sikkerhetskopi har i begge tilfeller blitt bevart på samme medium. Bortsettingsmedier medfører tungvinte operasjoner både ved vedlikehold og bruk av materiale. Mediene må i hvert tilfelle fremhentes og monteres. Optiske medier har heller ikke den kapasitet som kreves for å håndtere overgangen til elektronisk arkiverte saksdokumenter i forvaltningen. De nye avleveringsbestemmelsene fra 2007 aksepterer disk som overføringsmedium. Etter en slik disk-avlevering 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 – og nyere minneenheter uten bevegelige deler, som alle gjennomgår en kontinuerlig revolusjon med doblet kapasitet annethvert år. Disker har i dag en kapasitet på inntil 2 TB (terabytes), og tape (LTO-5) kan gi plass for 1,6 TB. Til sammenligning lagrer Blu-ray 25 GB (gigabytes), eventuelt 50 GB ved to-lags lagring. Det vil si at en 2 TB disk tilsvarer kapasiteten på 80 stk. ett-lags Blu-ray-plater[6].

Langtidslagring på magnetiske medier er imidlertid forbundet med større risiko. Den "ultimate" risikofaktoren er elektromagnetiske pulser (EMP). Disse utløses av atombomber, men kan også skapes av dedikerte EMP-våpen uten å gjøre annen fysisk skade. Statens arkiver i Danmark har vurdert det som for risikofylt å bruke utelukkende magnetisk lagring, og valgt optiske plater i jukeboks (DVD, og på sikt Blu-ray) som hovedteknologi for langtidsbevaring. Det svenske Riksarkivet baserer seg på den annen side helt ut på magnetisk tape, og da ut fra en samlet vurdering av lagringssikkerhet og økonomi.

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 online lagring. Både 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).

[6] Kapasiteten på Blu-ray tilsvarer 35 CD-R-plater a 700 MB (megabytes). Én enkelt 2 TB disk har mao. samme lagringskapasitet som 2800 CD-R-plater.

Fordelene ved langtidslagring på disk er mange. On-line tilgjengelighet effektiviserer all administrasjon, ikke minst de mange vedlikeholdsaktivitetene ved langtidsbevaring. I tillegg åpnes helt nye muligheter for å tilrettelegge brukertjenester og faktisk bruk av 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 større bestander av digital informasjon er Storage Area Network-systemer (SAN). Denne typen lagringsnett kan håndtere lagringsenheter på ulike teknologier, herunder roboter for tape og optiske medier. Ett felles kommandosenter administrerer all lagret informasjon, og muliggjør en samlet overvåking av dataintegriteten på bit-nivå. Høyhastighets dataoverføring brukes mellom lagringsenhetene. Data kan sikres ved å dupliseres på flere forskjellige teknologier[7]. Driftssikkerheten ivaretas typisk ved redundant kjøling, redundant strømforsyning og nødaggregat for strøm. SAN-systemer har mekanismer for sjekksumkontroll som gjør det mulig å detektere endringer i datainnhold, og gir dermed grunnlag for teknisk integritetssikring. Systemene har også funksjoner for automatisert verifisering av data etter overkopiering (migrering) mellom medier. Riksarkivet anskaffet i 2009 deler av et SAN-system for å lagre master-filene til bestanden av digitalisert arkivmateriale.

2.1.2 Spesielle utfordringer mht. å sikre integritet og autentisitet

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

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.

[7] I et SAN-system vil typer av lagret materiale også kunne ligge utelukkende på tape (tape-roboter). Når både original- og kopiversjoner ligger på tape, er det ekstra viktig å unngå bruk av tape-kassetter fra identiske produksjonsserier.

Utfordringene mht. å sikre integritet og autentisitet har sammenheng med at et lagret digitalt objekt eksisterer på flere ulike nivåer[8]:

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[9] – 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 bit-strømmers konstans. 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. 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 uavbrutt integritetssikring.

---

[8] InterPARES-prosjektet: *The State of Digital Preservation: An International Perspective. Documentation Abstracts* (2002). http://www.interpares.org/ip1/ip1_dissemination.cfm?proj=ip1&cat=pu-cp-r
Jf. her spesielt Kenneth Thibodeau: *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*.

[9] Betegnelsen er ikke helt heldig i de sammenhenger hvor det foretas en konvertering av informasjon i tillegg til migreringen, jf. behandlingen av begreper i underkapittel 1.3.

2.2 Internasjonale standarder og "Best practices"

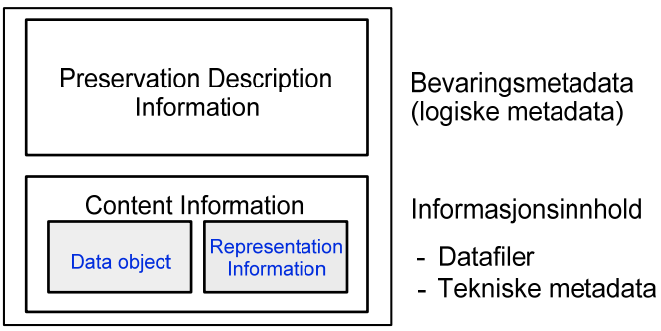
2.2.1 OAIS-standarden

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*[10] – som ble utformet av den amerikanske romfarts-organisasjonen CCSDS (Consultative Commitee for Space Data Systems) 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.

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 en opprinnelig SIP 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.



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

OAIS-modellen har mangler, spesielt når det gjelder å håndtere arkivinformasjon fra Records Management-systemer[11]. Men det er viktig å holde fast at OAIS ikke er en implementeringsmodell. Det er en referanse- og begrepsmodell med metadata-kategorier, 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[12], som beskriver den indre strukturen i en arkivpakke og den "container" som omslutter pakken,
- XFDU[13] (ISO 13527: 2009), som spesifiserer en alternativ pakkestruktur, og
- PREMIS[14], som definerer bevaringsmetadata for å støtte forståelighet, autentisitet og identitet.

2.2.2 TRAC-standarden

"Trusted digital repository" er kommet til som et nytt internasjonalt begrep. Digitale depoter verden over akkumulerer sterkt voksende informasjonsmengder, og må være gjenstand for krav som avspeiler deres forpliktelser. For at slike depoter skal bli vurdert som pålitelige og tiltrodde, må de oppfylle definerte kriterier. Gjennom de siste årene er det gjennomført en rekke aktiviteter for å spesifisere slike kriterier, spesielt i USA og innenfor EU[15]. 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*[16].

Den amerikanske TRAC-rapporten fra 2007 – *Trustworthy Repository Audit and Certification - Criteria and Checklist*[17] – har vært gjenstand for stor oppmerksomhet, og

[11] Slike mangler i OAIS omtales i punkt 4.2.2.1, nedenfor.

[12] METS (*Metadata Encoding & Transmission Standard) er spesifisert av Digital Library Federation, og vedlikeholdes av Library of Congress)*, http://www.loc.gov/standards/mets/

[13] XFDU (XML Formatted Data Unit Structure and Construction Rules) er spesifisert av romfarts-organisasjonen CCSDS, http://public.ccsds.org/publications/archive/661x0b1.pdf

[14] PREMIS (*Preservation Metadata: Implementation Strategies*) er utviklet av Online Computer Library Center (OCLC) og Research Libraries Group (RLG), og vedlikeholdes av Library of Congress, http://www.loc.gov/standards/premis/

[15] Først ute var de amerikanske bibliotekorganisasjonene Research Libraries Group (RLG) og Online Computer Library Centre (OCLC) med rapporten *Trusted Digital Repositories: Attributes and Responsibilities* (2002). I Europa gjennomføres programmer i England (DCC – the UK Digital Curation Centre), i Tyskland (nestor-prosjektet) og i Nederland (Koninklijke Bibliotheek). Et eget EU-prosjekt – DRAMBORA (*Digital Repository Audit Method Based on Risk Assessment*) – har utviklet en metodikk for selvevaluering i store og små digitale depoter med vekt på å gjøre risikofaktorene ved arkivering forståelige og håndterbare, jf. http://www.repositoryaudit.eu/

[16] *Ten Core Principles*: http://content.yudu.com/Library/A10tra/PLATTERRepositoryPla/resources/8.htm.

[17] *Trustworthy Repository Audit and Certification – Criteria and Checklist* (2007), utarbeidet av RLG – Research Libraries Group og NARA – National Archives and Records Administration, http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

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[18].

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.

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.

[18] CCSDS: *Audit and Certification of Trustworthy Digital Repositories* (review stage for ISO standard), http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206520R1/Attachments/652x0r1.pdf

Integritetssikring i et produksjonssystem og under fremstillingen av en SIP er temaer som TRAC ikke behandler[19]. TRAC (og OAIS-modellen) gir imidlertid rom for å dokumentere om og hvordan arkivmateriale er blitt integritets- og autentisitetssikret forut for en aksesjon.

2.3 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 integritets-sikringen av digitalt arkivmateriale ved fremstilling av avleveringer (SIP-pakker) og ved Arkivverkets mottak av det. Kravene behandles i punkt 2.3.1, nedenfor, og spesifiseres i sin helhet i vedlegg 1-3.
4) Det legges til grunn at arkivpakker skal bruke METS som pakkeformat og PREMIS som standard for bevaringsmetadata, men det anbefales å avvente et endelig valg til de praktiske implikasjonene er tilfredsstillende klarlagt, jf. punkt 4.2.3.

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 en lagringsløsning – et SAN-system – for Arkivverkets digitaliserte materiale, 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, nærmere bestemt på disk og tape (LTO). Dupliserte disker sikrer mot utstyrsfeil, men dersom det blir feil i data på den ene disken, vil feilen også bli duplisert til den andre. Versjonene på tape (tape-robot) er sikkerhetskopier, hvorfra data kan overføres til disk ved feil på diskene. 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 vil gi 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.

[19] To OAIS-relaterte standarder for fremstilling av en SIP er spesifisert av CCSDS:
- ISO 20652-2004: *Producer – Archive Interface Methodology Abstract Standard* – "PAIMAS", http://public.ccsds.org/publications/arc hive/651x0b1.pdf
- *Producer-Archive Interface Specification* – "PAIS", (review stage for ISO standard) http://mailman.ccsds.org/pipermail/moims-dai/attachments/20060602/8ee057c5/paimas_implementationWB_04-0001.obj

Disse standardene synes å basere seg på at informasjonsinnholdet i en SIP er helt identisk med produksjons-systemets, og kan valideres mot dette.

- SAN-løsningen oppfyller krav til teknologiuavhengighet. Dataobjektene lagres teknologiuavhengig i den forstand at de senere kan flyttes til andre medier og teknologier. Det antas at SAN-systemets registre og indekser også vil kunne eksporteres med opprettholdte referanser til de fysiske og logiske objektene som er lagret.

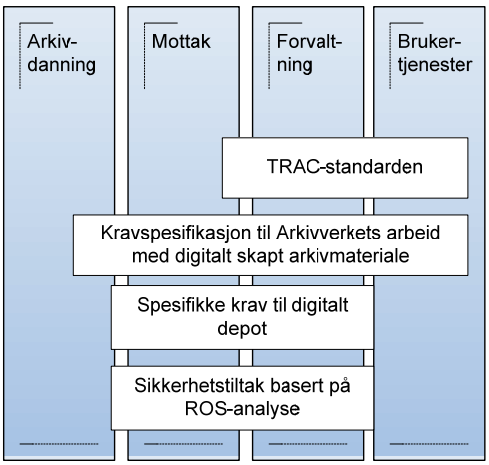
2.3.1 Spesifiserte krav til forvaltningen av arkivobjekter

Prosjektet har utarbeidet to detaljerte kravspesifikasjoner som følger som vedlegg til rapporten. Den første – *Kravspesifikasjon til Arkivverkets arbeid med digitalt skapt arkivmateriale* (vedlegg 1) – inneholder krav fra TRAC-standarden som vurderes som relevante og gyldige for Arkivverkets depotfunksjoner. I tillegg supplerer den TRAC med krav til håndteringen av autentisk arkivmateriale ved fremstilling og mottak av avleveringer. Denne kravspesifikasjonen er generell, og omfatter de fleste aspektene ved Arkivverkets arbeid med digitalt skapt arkivmateriale.

Det andre kravsettet – *Spesifikke krav til Arkivverkets forvaltning av digitalt skapt arkivmateriale i tilknytning til digitalt depot* (vedlegg 2) – er konsentrert om organisasjonsløsninger og rutiner med direkte relevans for Arkivverkets digitale depot. På disse punktene utdypes også den første spesifikasjonen. Tilrettelegging for avlevering i arkivdanningsfasen og brukertjenester på avlevert materiale er temaer som ikke behandles i "Spesifikke krav" – til forskjell fra den første spesifikasjonen.

Prosjektet har også utarbeidet en ROS-analyse: *Sikkerhetstiltak basert på en risiko- og sårbarhetsanalyse av system-, drifts- og rutineopplegg i tilknytning til Arkivverkets lagringssystem for digitalt skapt arkivmateriale* (vedlegg 3). Tiltakene som foreslås som ledd i denne oversikten, representerer en spissing av "Spesifikke krav".

Figur: Kravspesifikasjoner



Figuren viser dekningsområdene for de ulike settene med krav, og hvordan de bygger på hverandre – fra TRAC og nedover.

(vedlegg 1)

(vedlegg 2)

(vedlegg 3)

Kravspesifikasjon til Arkivverkets arbeid med digitalt skapt arkivmateriale (vedlegg 1) forutsetter og legger opp til en revisjon av gjeldende avleveringsbestemmelser for elektronisk arkivmateriale fra 2007. Den antas også å være aktuell som utgangspunkt for en egen forskrift med krav til depoter som forvalter digitalskapt arkivmateriale fra offentlige organer.

2.3.2 Lagringskonfigurasjon

SAN-løsningen i Riksarkivet er ikke ferdig utbygd. Det gjenstår bl.a. å anskaffe et SAN-administrasjonssystem. Den endelige konfigureringen av lagringsløsningen er heller ikke helt avklart. Det er bare anskaffet én tape-robot. Objekter lagres nå på to teknologier, men foreløpig bare i to eksemplarer: ett på disk (duplisert) og ett på tape. Det er behov for å holde fast ved det opprinnelige målet om to kopier på tape i tillegg til versjonen på disk. Med dagens konfigurasjon vil imidlertid dette måtte bety to kopier produsert av det samme robot-systemet. Prinsipalt bør lagringssystemet være konfigurert med 2 uavhengige tape-roboter, som ved Nasjonalbiblioteket i Rana.

Et tilfredsstillende utbygd system for tape-lagring er viktig. Slik det fremstår, er det on-line lagring på disk som har initialisert læresetningen om lagring på minimum to teknologier. Hvor tape brukes som eneste medium, som i det svenske riksarkivet, lagres også alle kopier på tape. Med et videre utbygd lagringssystem bestående av flere tape-roboter bør lagring utelukkende på tape i visse sammenhenger også være et alternativ for Arkivverket. Det finnes typer av materiale hvor dette kan være formålstjenlig, eventuelt etter en innledende periode med disklagring. Eksempler er master-filer til digitalisert materiale og voluminøse digitalskapte arkiver som bevares deponert (jf. forestående deponering av pantebokmateriale fra Statens kartverk). Det er ikke noe praktisk behov for å holde denne typen materiale kontinuerlig svivende på disk. Tape-lagring er betydelig rimeligere, sikrere, og – for visse typer materiale – praktisk og hensiktsmessig nok.

2.3.3 Ekstern sikkerhetskopi

En komplett sikkerhetskopi på tape må oppbevares ”off site” på trygg geografisk avstand fra Riksarkivet. Prinsipalt vil prosjektet foreslå at det produseres et 3. sett med tape-kopier for ekstern oppbevaring, slik at Arkivverkets fortsatt har én kopiversjon for hver av de to tape-robotene. Spørsmålet må imidlertid vurderes videre med vekt på de praktiske implikasjonene ved å produsere et 3. sett med tape-kopier.

I møte med Nasjonalbiblioteket er det tidligere gjort intensjonsavtale om utveksling av sikkerhetskopier mellom Riksarkivet og NB Rana. Rom 2E i den eldre delen av Riksarkivets fjellmagasin (i dag magasin for bestanden av CD-er) forutsettes å kunne brukes til lagring av utvekslede kopier. 2E må ominnredes for dette formålet, og rommet må tilknyttes fysisk adgangskontroll.

Inntil en ordning med ekstern oppbevaring kommer i stand, foreslås rom 2E også brukt til lagring av Arkivverkets egne off site-kopier.

2.3.4 Lagringsbehovet for digitalt skapt arkivmateriale

De drøyt 700 avleveringene og deponeringene med digitalskapt arkivmateriale som Arkivverket har mottatt siden 1985, beregnes bare å utgjøre ca. 0,4 TB (400 GB) totalt. Materialet omfatter nesten utelukkende tabellinformasjon, og tar derfor liten plass etter dagens mål. Omleggingen til elektronisk arkivering i forvaltningen vil etter hvert gjøre avleveringene betydelig mer voluminøse. Prosjektet har tidligere estimert lagringsbehovet for digitalt skapt arkivmateriale for de kommende 5 år til totalt 4 TB. Anslaget baseres på 40 Noark-4-uttrekk med elektroniske dokumenter á 20 GB pr. år. Det digitalt skapte

materialet vil dermed fortsatt utgjøre et beskjedent volum sammenholdt med Arkivverkets digitaliserte materiale.

Utviklingen kan imidlertid medføre overraskelser når det gjelder lagrede volumer – og har i dette tilfellet også gjort det. Det er nylig annonsert at Statens kartverk i 2011 skal deponere elektronisk pantebok med opplysninger opp til 2010 og med etterskannede dokumenter tilbake til 1988. Volumet av deponeringen vil være 45 TB, dvs. anslagsvis det dobbelte av de lagrede master-filene til Arkivverkets digitaliserte kirkebøker.

Lagring (deponering) av digitalt arkivmateriale som en vertstjeneste for eksterne instanser representerer en annen usikkerhetsfaktor når det gjelder kapasitetsbehov. Det annonseres i Kulturdepartementets digitaliseringsmelding (2009) at Nasjonalbiblioteket og Arkivverket skal tilby slike sentrale vertstjenester for kultursektoren. Det vil her ventelig både være tale om ulike typer digitalisert materiale og arkivmateriale som er digitalt skapt.

2.3.5 Konfigurering og sikring av fysisk magasin

Magasinløsningen må konfigureres med en inndeling i soner fordi ulike kategorier av materiale krever ulike forvaltningsregimer. Det er nødvendig å skille mellom digitalisert og digitalt skapt materiale, og mellom ulike kategorier av digitalt skapt materiale. I det siste tilfellet er det behov for å holde følgende kategorier atskilt:

a) gradert materiale som er avlevert eller deponert (”permanent lager”),
b) ugradert arkivmateriale som er avlevert eller deponert (”permanent lager”),
c) mottatt materiale i fasen med godkjenningsvurdering (brukertildelt sone for testing),
d) kopiversjoner av avlevert materiale som legges ut til Arkivverkets interne bruk (herunder for statsarkivene).

Originalversjoner av ugradert arkivmateriale (kategoriene b-c) må lagres i en lukket del av magasinet – Digitalt sikringsmagasin (DSM) – mens bruksversjoner (d) kan ligge i en mer tilgjengelig ytre del. Gradert arkivmateriale (a) må i utgangspunktet håndteres på dedikert utstyr i eget magasinrom. Men også i DSM er det nødvendig med sikkerhetstiltak på nivå med dem som kreves for gradert informasjon. Det må kunne legges til grunn at lavgradert materiale også kan lagres i DSM.

Følgende sikringstiltak forutsettes for magasinlokalene som inngår i digitalt depot (jf. også illustrasjon med figur under punkt 4.1):

1) *Datarom (”SAN-rom”) for DSM og ”ytre sone” i nytt fjellmagasin*
   Rommet – og dermed både informasjonen i DSM og ”ytre sone” – sikres med fysisk adgangskontroll. Prinsipalt skal tilgang kreve to personers nøkkelkort. Dette kan være problematisk i lys av ressurssituasjonen på personalsiden, og spørsmålet må derfor vurderes av Riksarkivets sikkerhetsorganisasjon. Direkte kommunikasjon mellom DSM og ytre sone skal ikke være fysisk mulig. Kontorarbeidsplasser med nettilknytning til DSM skal heller ikke kunne kommunisere med annet utstyr.

2) *Kombinert datarom for tape-robot og eget utstyr for lagring av gradert informasjon*
   Rommet, som ligger vegg i vegg med SAN-datarommet i nytt fjellmagasin, sikres med egen adgangskontroll. Det skal kreves 2 personer for å oppnå tilgang. Server

med eget disksystem for lagring av gradert informasjon skal ha eget administrasjonssystem og egen tape-backup, og ikke ha kommunikasjonslinjer til annet utstyr, heller ikke til tape-roboten i samme rom. Tape-roboten skal være forbundet med lagringssystemet (DSM) i SAN-rommet.

3) *Rom 2E i eldre fjellmagasin for bevaring av "off site"-kopier på tape*
   Dette rommet sikres med fysisk adgangskontroll (mangler i dag). Dersom tapebestanden innbefatter gradert informasjon eller sikkerhetskopier fra eksterne organer, skal det kreves 2 personer for å oppnå tilgang.

4) *DD-seksjonens datarom (testrom med safe) i administrasjonsbygningen (underetg.)*
   Rommet, som skal brukes til mottakskontroll av avleveringspakker før de innlemmes i DSM, tilknyttes det lukkede nettverket for digitalt sikringsmagasin (DSM) og sikres med fysisk adgangskontroll (mangler i dag). Det er behov for å avklare om sikringstiltakene også gir rom for mottakskontroll og test av gradert materiale, eventuelt bare materiale med laveste gradering. Inntil dette må slik kontroll utføres i kombinert magasin for tape-robot og gradert materiale.

Prosjektet legger til grunn at DSM også skal kunne lagre lavgradert materiale, – med den konsekvens at et dedikert magasin bare trengs for materiale med høyere gradering. I dagens digitale arkivbestand forekommer bare beskjedne innslag av materiale gradert fortrolig etter beskyttelsesinstruksen. Avleveringer med høyere gradert materiale er heller ikke ventet de nærmeste årene. Ut fra disse forutsetningene vil prosjektet foreslå at man inntil videre venter med å ta i bruk det dedikerte magasinet for gradert arkivmateriale, jf. punkt 4.1, nedenfor. Magasinet må imidlertid holdes klargjort for installasjon av nødvendig utstyr.

2.3.6 Forholdet mellom lagringsløsninger for digitalskapt og digitalisert materiale

Originalversjonene (master-filene) av Arkivverkets digitaliserte arkivmateriale ligger allerede lagret i DSM. Dette materialets volum (p.t. ca. 100 TB) gjør at det digitalt skapte materialet bare vil utgjøre en marginal del av DSM, i det minste innledningsvis.

Prosjektet legger til grunn at de to typene av materiale skal lagres helt atskilt på egne partisjoner i DSM. For det digitaliserte materialet benyttes også et eget forvaltningssystem og et separat, lukket IP-nett mot DSM. Fysisk vil det ikke være mulig å aksessere det digitalt skapte materialet via inngangen for det digitaliserte materialet – og omvendt. De to typene av materiale må imidlertid være underlagt samme type sikkerhetstiltak i DSM. All lagret informasjon i DSM må i praksis behandles på nivå gradert.

Den digitaliserte arkivbestanden omfatter også materiale som er transkribert. Muligheten må holdes åpen for at det kan være behov for å forvalte transkribert materiale sammen med det digitalt skapte.

3. ORGANISASJONSMESSIGE FORUTSETNINGER

Kapitlet gjennomgår de overordnede organisatoriske tiltak som kreves ved etableringen av et digitalt depot. Krav formuleres til en samlet organisasjon for sikkerhetsarbeidet, til en organisasjon for systemdrift og til en organisasjon for å forvalte den lagrede bestanden av digitale objekter som arkivmateriale.

Kravene til et digitalt depot som stilles i kapittel 2, 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

Nedenfor behandles funksjoner og roller som kreves under de tre første av disse områdene. Det fjerde området – bruk og brukertjenester – behandles i kapittel 6.

3.1 Krav til sikkerhet og sikkerhetsorganisasjon

Arbeidsprosessene i tilknytning til digitalt depot krever organiserte sikkerhetstiltak. For at et sikkerhetsopplegg for digitalt depot skal fungere, må det finnes et apparat med stabs-tilknytning til virksomhetens ledelse for å følge opp tiltak. Riksarkivaren oppnevnte 13.11.2009 den formaliserte sikkerhetsorganisasjon for Riksarkivbygningen som må være på plass når et digitalt depot tas i bruk.

Sikkerhetstiltakene for digitalt depot utgjør likevel bare en del av Arkivverkets opplegg for IT-sikkerhet, og de representerer et av flere områder som omfattes av Riksarkivbygningens sikkerhetsorganisasjon. Typer av tiltak, som å ivareta personvern i tilknytning til det bevarte digitale arkivmaterialet, må dessuten håndteres i en videre sammenheng.

3.1.1 Trusselbildet

Digitalt depot omfatter utstyr, men verdiene som forvaltes 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.

Det brede trusselbildet i forbindelse med et digitalt depot behandles i ROS-analysen som følger som vedlegg 3 til rapporten.

3.1.2 Dokumentert internkontroll

For å definere og håndtere sikkerhetstrusler må det være utarbeidet tre typer av dokumentasjon som skal ligge til grunn for informasjonsbehandlingen i 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.

Samlet vil denne dokumentasjonen utgjøre internkontrollsystemet for digitalt depot. Rapportens vedlegg 1-3 er bidrag til dokumentasjonen under punkt 1 og 2 ovenfor.

Det er behov for å utarbeide en egen sikkerhetshåndbok for digitalt depot som samler i seg all nødvendig styringsdokumentasjon, instrukser med prosedyrebeskrivelser og fastsatt rammeverk for produksjon av kontrolldokumentasjon. Den foreliggende dokumentasjonen bør innarbeides i en slik sikkerhetshåndbok, og suppleres med tiltak som foreløpig er mangelfullt dekket, nærmere bestemt:

- organiseringen av sikkerhetstiltakene,
- ivaretakelsen av spesifikke plikter ved informasjonsbehandlingen som følger av lov- og forskriftsbestemmelser, jf. også delkapittel 5.2,
- spesifiserte instrukser og prosedyrebeskrivelser
- definerte ansvarsområder og roller i tilknytning til digitalt depot.

En *beredskaps- og katastrofeplan* må inngå i samlingen av instrukser og prosedyre-beskrivelser. Så langt tiltak i beredskaps- og katastrofeplanen lar seg teste ut i praksis, er det også nødvendig at dette blir gjort. Gjenoppretting av informasjon på grunnlag av tape-kopier er blant tiltakene som må testes ut i praksis.

Det forutsettes en årlig gjennomgang og revisjon av opplegget for internkontroll.

3.1.3 Ansvarsområder for sikkerhet

Sikkerhetsorganisasjonen for digitalt depot må bygge på klargjorte ansvarsforhold:

1) Arkivverkets ledelse må ha det overordnede ansvaret for utformingen av sikkerhets-tiltak og kontrollen med at virksomheten utføres i tråd med tiltakene. Kanaler og rutiner for rapportering til Arkivverkets ledelse må være definert.

2) Det operative ansvaret for at sikkerhetsarbeidet utføres i tråd med fastsatte retningslinjer, instrukser og prosedyrer, må være definert. Dette operative ansvaret foreslås ivaretatt slik:
   a) Fysisk sikring av datarom delegeres til lederen for Administrasjonsavdelingen.
   b) IT-sikkerhet i digitalt depots installasjoner og i driften av dem delegeres til lederen for IT-avdelingen med systemansvarlig for digitalt depot som utførende.
   c) Sikkerheten for arkivforvaltningen av informasjon i og i tilknytning til digitalt depot delegeres til lederen for Depotavdelingen med lederen for DD-seksjonen som utførende.

   Roller og ansvarsområder knyttet til systemdrift og informasjonsforvaltning blir videre detaljert i henholdsvis punkt 3.2.2 og punkt 3.3.2.

3) Det må være definert egne roller med ansvar for hvert av følgende funksjonsområder:
   a) Adgangskontroll, herunder autorisasjon av brukere og oppfølging av operasjonslogger for faktisk tilgang.
   b) Internkontroll spesifikt knyttet til personopplysninger, jf. nærmere definerte krav i punkt 5.2.3.
   c) Internkontroll spesifikt knyttet til gradert informasjon, jf. nærmere definerte krav i punkt 5.2.2.
   d) Oppfølgingen av fastsatte kontrollrutiner og avviksrapportering i samsvar med utarbeidet kontrolldokumentasjon.

4) Det må finnes et rammeverk med regler for rapportering fra linjeorganisasjonen til Riksarkivbygningens sikkerhetsorganisasjon. Regler for avviksrapportering fra fastsatte sikkerhetskrav må være særskilt definert. Medarbeidere fra statsarkivene og Elark-seksjonen i Riksarkivet som utfører sikkerhetsrelaterte oppgaver i digitalt depot, må være underlagt sikkerhetsregimene som IT-avdelingen og Depotavdelingen er operativt ansvarlige for, og i slik sammenheng rapportere til disse enhetene.

3.2 Krav til driftsorganisasjon

Et digitalt depot krever en profesjonell driftsorganisasjon. Kompetansen, arbeidsorganisasjonen og bemanningen på driftssiden er avgjørende for mulighetene til å etablere Arkivverkets digitale depot.

3.2.1 Bemanningsbehov

Etter prosjektets vurdering må IT-avdelingen ha minimum 3 medarbeidere med kompetanse og særskilt autorisasjon for drift av Arkivverkets lagringssystem. 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.

3.2.2 Roller og funksjonsområder

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.

Driftsorganisasjonen må blant annet ivareta følgende funksjoner:

- vedlikehold og utskiftning av hardware-komponenter iht. utviklingsprogram,
- migrering av data til nye disker iht. fastsatt utskiftingsplan (inkludert verifisering),
- synkronisering og verifisering av kopiversjoner på tape,
- monitorering/overvåking av logger for genererte og verifiserte sjekksummer og logger for endrede opplysninger i DSM,
- rapportering til Riksarkivbygningens sikkerhetsorganisasjon på grunnlag av monitorerings- og loggingsfunksjoner,
- dokumentasjon av systemopplegg og driftsrutiner.

3.3 Krav til 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.

3.3.1 Bemanningsbehov

For IT-avdelingen dreier det seg om oppgaver som kommer i tillegg til dagens. For DD-seksjonen og Elark-seksjonen dreier det seg om endrede oppgaver, og prosjektet har ikke funnet grunn til å vurdere bemanningsbehovet for disse to enhetene. For DD-seksjonen innebærer imidlertid omleggingen at hele den nåværende arkivbestanden på CD-er må konverteres og innlemmes i digitalt depot. En full oppgradering av bestanden til ny struktur og ny dokumentasjonsstandard etter mønster av det nylig avsluttede konverteringsprosjektet ved Statens arkiver i Danmark vil etter DD-seksjonens egne beregninger samlet kreve 20 årsverk.

3.3.2 Roller og funksjonsområder

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.

Disse ansvarsområdene blir videre detaljert under punkt 5.1.6. Inntil den eldre digitale arkivbestanden på CD-er innlemmet i DSM, kreves også en ansvarlig for gjennomføringen og kvalitetssikringen av programmet for konvertering.

Informasjonsforvaltningen må blant annet ivareta følgende funksjoner:

- verifisering og "frysing" av arkivpakker (SIP) ved mottak,
- koordinering av testing som utføres av medarbeidere i statsarkivene,
- 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,
- flytting av materiale mellom soner i lagringssystemet,
- behandling av sensitive personopplysninger og gradert materiale,
- oppdatert dokumentasjon av rutineopplegg.

4. KONFIGURASJONSLØSNINGER

Tre hovedaspekter ved organiseringen av lagringsløsningene i digitalt depot beskrives mer detaljert i dette kapitlet: den fysiske og logiske magasinkonfigurasjonen, strukturen i det digitale sikringsmagasinets arkivpakker som skal bevares som integritetssikrede enheter med alle tilhørende metadata, og utvekslingen av arkivbeskrivelser mellom lagrings-systemet og arkivinformasjonssystemet Asta.

4.1 Konfigurering av magasinløsning

Prosjektet baserer seg på følgende konfigurasjon med 3 magasinrom i Riksarkivets fjellanlegg og et eget testrom i administrasjonsbygningen:

- et soneinndelt sentralmagasin i Riksarkivets nye fjellmagasin
- et tilstøtende og særskilt beskyttet magasinrom i nytt fjellmagasin for dedikert lagring av gradert arkivmateriale
- magasin (2E) i eldre fjellanlegg for "off site" lagring av sikkerhetskopier
- et eget, sikret datarom i administrasjonbygningens underetasje for mottakskontroll av arkivmateriale.

Det legges til grunn at konfigurasjonen skal omfatte 2 stk. tape-roboter, og at den ene av disse (som allerede er anskaffet) skal plasseres i magasinet for gradert materiale. Denne roboten vil være forbundet med utstyret i sentralmagasinet, og skal ikke ha tilknytning til den dedikerte lagringsløsningen for gradert materiale i samme rom[20]. Tape-robot nr. 2 plasseres i det sentrale magasinet.

Den overordnede inndelingen av digitalt depot i en indre og en ytre sone er beskrevet under punkt 2.3.5. Denne delingen for å skille mellom originalversjoner av arkivmateriale i den indre sonens digitale sikringsmagasin (DSM) og bruksversjoner i ytre sone, er både fysisk og logisk. Den fysiske utrustningen for de to lagringsløsningene forutsettes plassert i samme rom – sentralmagasinet – men direkte kommunikasjon mellom sonene skal likevel ikke være mulig.

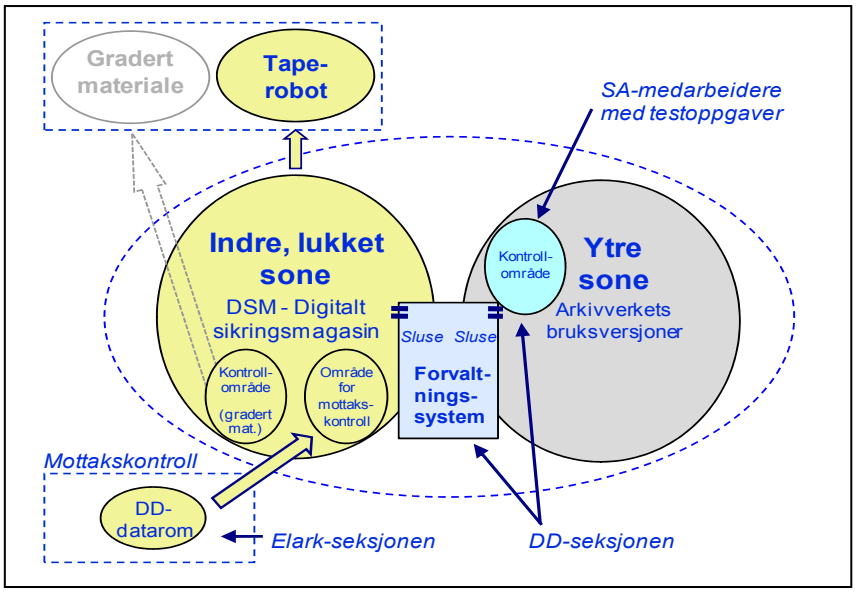
Prosjektet definerer DSM – Digitalt sikringsmagasin – som et område i indre sone som utelukkende lagrer arkivpakker. Områder for dedikerte tilretteleggings- og vedlikeholds-oppgaver som også er plassert i indre sone, jf. punkt 4.1.1 – 4.1.3, ligger etter en streng definisjon utenfor DSM. Det skal finnes "sporveksler" som styrer og detekterer trafikk mellom de ulike områdene i indre sone.

[20] På sikt antas serveren for det graderte materialet å kunne tilknyttes tape-roboten. Dette krever at roboten settes opp med logiske tape-biblioteker som er uleselige for andre IP-adresser enn de spesifiserte. IP-adresser kan dessuten sperres slik at det kreves en egen fysisk kobling for å nå det logiske biblioteket. Denne modellen kan eventuelt også gjøre det mulig å benytte tape-roboten til kopiering av informasjon fra installasjonene i det tilstøtende datarommets ytre sone.

Tilknytning til indre sone skal skje via et dedikert nett. Foruten datarommet (testrommet) i administrasjonsbygningens underetasje skal bare autoriserte medarbeidere i IT-avdelingen og Seksjon for digitalt depot (DD-seksjonen) i Riksarkivet ha slik netttilknytning til den indre sonen. For disse medarbeiderne må det også etableres delte kontormiljøer for å sperre for ekstern kommunikasjon med indre sone lokalt. Informasjonen i sentralmagasinets ytre sone skal imidlertid kunne aksesseres via Arkivverkets (åpne) lokale nettverk.

En magasinkonfigurasjon etter planløsningen ovenfor står fast, men både av ressursmessige og praktiske grunner foreslår prosjektet at man inntil videre venter med å ta det dedikerte magasinet for gradert materiale i bruk, jf. punkt 2.3.5, ovenfor. Også DSM er sikret for å muliggjøre lagring av gradert informasjon. 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. Materialet antas også å ha et beskjedent omfang.

Figur: Digitalt magasin



Prikket blå 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.

Om kontrollområder: se punkt 4.1.3.

4.1.1 Sluse mellom indre og ytre sone

Det er behov for en sluse mellom indre og ytre sone for å flytte og kopiere informasjon mellom de to sonene. Uten en slik løsning vil eksport til og fra DSM være en svært tungvint affære. Men sluseløsningen må da 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.

Prosjektet vil foreslå en konfigurering hvor sluseløsningen styres fra det digitale depotets forvaltningssystem, jf. figur ovenfor. Slusingen er basert på at maskinen som kjører forvaltningssystemet, har to nettverkskort med forbindelser til hver sin sone. Bare ett nettverkskort kan være aktivt om gangen.

Sluseløsningen gjør at det nok vil knytte seg en viss usikkerhet til Nasjonal sikkerhetsmyndighets (NSM) vurdering av sikkerheten for gradert informasjon. Etter prosjektets vurdering reduseres ikke sikkerheten i indre sone. Denne sonen kan fortsatt defineres som

et sperret område med den adgangskontroll som er nødvendig for å håndtere sikkerhetsgradert materiale. Dette gjelder under forutsetning av:

- at personale med adgang til indre sone er autorisert for informasjonen i området, og
- at informasjonen er separert fra åpne nett og tilsvarende beskyttet i kontormiljøene med tilknytning til indre sone.

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

4.1.2 Eget forvaltningssystem for digitalt depot

Det sentrale forvaltningssystemet i Arkivverkets digitale depot vil være en applikasjon med flere funksjoner. Systemet må utvikles spesielt for Arkivverket, eventuelt som en tilpasning av et eksisterende system. Det svenske riksarkivet har tatt i bruk et slikt system for administrasjon av digitale arkivobjekter. Det er basert på åpen kildekode, og representerer dermed et alternativ som utgangspunkt for en tilpasning.

Forvaltningssystemet og dets database forutsettes plassert i indre sone. Det vil ha et eget arbeidsområde. Flytting av materiale inn og ut av dette området må kunne styres og detekteres. Forvaltningssystemet planlegges 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 digitalt depots driftssystem (systemet for SAN-administrasjon)
- generere informasjon til Asta-systemet
- generere bruksversjoner av arkivmateriale i DSM for tilgjengeliggjøring i ytre sone
- gi oversikt over innhold og arkivpakkestrukturer i DSM (ved hjelp av Asta og forvaltningssystemets database)
- vise statusinformasjon (f.eks. formatoversikter), utvalgte logger (f.eks. for tilgang) og rapporter fra DSM
- hente data inn/ut av DSM og ytre sone samt flytte data mellom DSM og ytre sone
- styre sluseløsningen som skal hindre samtidige operasjoner i de to sonene.

Forvaltningssystemet ligger fysisk i indre sone. Asta-systemet er plassert i Arkivverkets åpne nettverk, og tilhører dermed logisk den ytre sonen. Dynamikken ved sluseløsningen kommer til uttrykk ved at det bare er forvaltningssystemet som kan skape og administrere trafikk av Asta-informasjon mellom sonene.

4.1.3 Egne områder for mottakskontroll og testing

Prosjektet foreslår en sentral mottakskontroll for digitale avleveringer og deponeringer, jf. punkt 5.1.2. Initiell kontroll og verifisering ved mottak planlegges utført i DD-seksjonens

datarom (testrom). Dette datarommet skal være tilknyttet det digitale magasinets indre sone via samme lukkede nett som kontormiljøene i DD-seksjonen. Rommet skal være fysisk sikret som et sperret område, og adgang skal kreve sikkerhetsklarering.

Til bruk for mottakskontrollen skal det også finnes et dedikert lagringsområde i indre sone. Materiale skal overføres hit, klargjort for testing, når det har passert den initielle kontrollen ved mottak. Til dette området skal det også kunne defineres dedikerte tilgangsrettigheter. Personer som er autorisert for oppgaver i tilknytning til mottakskontrollen, skal ordinært ha tilgang til lagringsområdet for mottak i indre sone uten å ha tilgang til andre områder og annen informasjon i den indre sonen.

I lagringssystemet må det dessuten finnes dedikerte områder for å gjennomføre testing av avleveringspakker. For ugradert materiale foreslås et felles kontrollområde for DD-seksjonen og utpekte medarbeidere i statsarkivene som skal utføre testing. Dette kontrollområdet må ligge i ytre sone. Statsarkiv-medarbeiderne vil ikke ha tilgang til den indre sonen og DSM. Dette er styrende for plasseringen i ytre sone.

Kontrollområdet for ugradert arkivmateriale i ytre sone skal være beskyttet på linje med informasjonen i DSM, men med den forskjell at det kan aksesseres fra de aktuelle statsarkivene over kryptert linje. Testing skal være basert på fjernaksess. Nedlasting av informasjon til lokal PC skal ikke være mulig. Dette innebærer at all nødvendig test-programvare må finnes på kontrollområdet. Løsningen hindrer at Arkivverkets datanettverk blir blokkert ved nedlasting av store avleveringer. Med et sentralt testmiljø unngås dessuten spesielt strenge krav til kontormiljøet hos den enkelte statsarkiv-medarbeider – med tilhørende lokale tiltak og prosedyrer for å følge opp lovbestemmelser om betryggende oppbevaring av taushetsbelagt og fortrolig informasjon.

All testing av gradert arkivmateriale skal skje i Riksarkivet. I dette tilfellet brukes et dedikert kontrollområde innenfor indre sone. Materiale med en høyere gradering vil – når den tid kommer – måtte testes på utstyret for gradert materiale i dedikert magasin hvor det også skal lagres. Dette gjelder innenfor rammene av den konfigurasjon som her er beskrevet, men det kan tenkes smidigere løsninger når behovet etter hvert melder seg.

Bare forvaltningssystemet skal kunne plassere arkivobjekter på kontrollområdene for testing. Eksport eller nedlasting fra kontrollområdene skal bare kunne skje ved at informasjonen hentes av forvaltningssystemet. Autoriserte brukere skal kunne slette informasjon på områdene. Sletting av informasjon på kontrollområdene forutsettes logget, i likhet med all import og eksport i regi av forvaltningssystemet.

4.1.4 Tilgangsgrupper

Med utgangspunkt i de overordnede konfigurasjonsløsningene som beskrives foran, foreslås følgende generelle tilgangsgrupper i Arkivverkets digitale depot:

	Tilgang til område:	*Tilgang for rolle:*
1.	DSM og utrustningen i ytre sone	System- og driftsansvarlig (IT)
2.	Utstyr i dedikert magasin for gradert materiale	System- og driftsansvarlig (IT)
3.	DSM	Vedlikeholdsansvarlig for arkivmateriale
4.	Dedikert magasin for gradert materiale	Test- og vedlikeholdsansvarlig for gradert arkivmateriale
5.	DD-datarom og eget kontrollområde i indre sone	Ansvarlig for mottakskontroll
6.	Kontrollområde i ytre sone	Ansvarlig for testing av ugradert arkivmateriale
7.	Kontrollområde i indre sone	Ansvarlig for testing av gradert arkivmateriale
8.	Ytre sone (åpen del)	Medarbeidere med rett til å aksessere bruksversjoner av digitalt arkivmateriale
9.	Dedikerte brukerområder i ytre sone	Medarbeidere autorisert for aksess til særskilt beskyttede bruksversjoner

Tilgangsrettigheter skal kunne kombineres i profiler. For DD-seksjonen vil en typisk kombinasjon være c + f. Foruten tilgangsgruppene ovenfor kreves dedikerte tilgangsrettigheter for roller som koordinerings- og tilsynsansvarlige, jf. punkt 3.2.2 og 5.1.6.

4.2 Organisering og integritetssikring av bevart arkivmateriale

Dette delkapitlet følger opp anbefalingene i kapittel 2 om å organisere arkivobjektene i digitalt depot som selvberoende og selvforklarende arkivpakker 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 fast arkivpakkestruktur muliggjør også samspill med et sentralt overvåkings- og administrasjonssystem.

Nedenfor beskrives en slik arkivpakkestruktur, først som en logisk modell, og deretter som en implementeringsmodell. Den logiske modellen er til hjelp for forståelsen. Den trenger imidlertid tilpasninger for å være egnet for faktisk bruk. Implementeringsmodellen foretar slike tilpasninger, men er ellers å betrakte som et supplement til den logiske modellen.

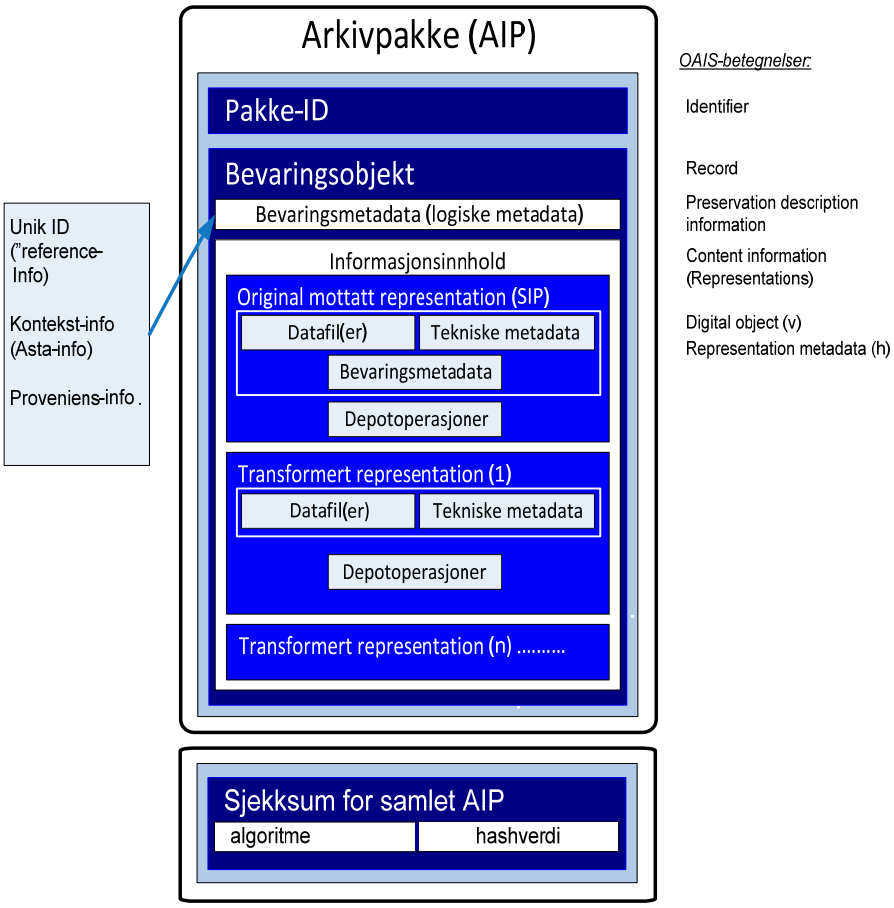
4.2.1 Organisering av arkivpakker: logisk modell

Prosjektet foreslår å basere Arkivverkets arkivpakkestruktur på den logiske modellen som vises i figuren nedenfor. En slik struktur innbefatter metadata-kategoriene i OAIS-modellen, og har 3 hovedkomponenter:

1) identifikator
2) bevaringsobjektet med alle logiske og tekniske metadata
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. Modellen må gi rom for transformerte versjoner av informasjons-innholdet i tillegg til de opprinnelig mottatte versjonene. Det siste sikrer at opprinnelig mottatt informasjonsinnhold blir bevart med intakte mekanismer for integritetssikring, og gir dermed tilbakespolings- og kontrollmuligheter. Mekanismene kan være medfølgende sjekksummer fra arkivskaper ved avleveringen og sjekksummer generert av Arkivverket ved mottak, eventuelt bare de sistnevnte, jf. punkt 5.1.2.

Figur: Arkivpakke – logisk modell [21]



[21] Figuren er inspirert av Filip Boudrez: *Digital containers for shipment into the future* (2005), som beskriver implementeringen av arkivpakker i Antwerpen byarkivs digitale depot som ledd i EU-prosjektet eDavid, jf. http://www.expertisecentrumdavid.be/docs/digital_containers.pdf.

Som en fast rutine i Arkivverkets digitale depot foreslås det å bevare de to siste transformerte versjonene av informasjonsinnholdet i tillegg til den opprinnelige mottatte avleveringspakken (SIP).

Sjekksummen for den samlede arkivpakken må lagres *utenfor* arkivpakken – til forskjell fra pakkens øvrige sjekksummer. Dette er nødvendig fordi innlemmelsen av en samlet sjekksum i seg selv vil endre arkivpakkens bit-innhold. Samlet sjekksum for de enkelte arkivpakker forutsettes lagret i SAN-administrasjonssystemet.

Kategorien "Depotoperasjoner" krever spesiell omtale. Den skal omfatte Arkivverkets testdata, logger og tilleggsopplysninger. En avleveringspakke – en SIP (Submission Information Package) – skal alltid bevares eksakt slik den ble mottatt. Integritetssikringen av opprinnelig mottatt informasjonsinnhold gjør det nødvendig å holde Arkivverkets operasjoner atskilt fra dette. Det er i denne sammenheng grunn til å poengtere at Arkivverkets testing av en avlevering typisk vil resultere i endringer eller tilføyelser i en ADDML-fil med tekniske metadata når denne medfølger fra arkivskaperen. Den nye ADDML-filen skal da alltid 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.

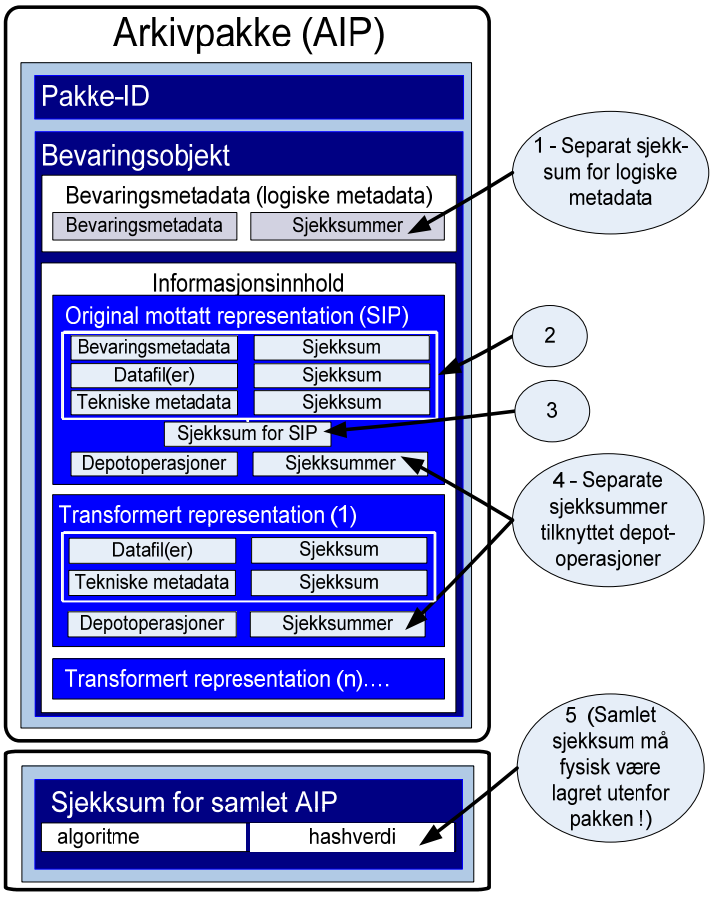
At endrede tekniske metadata kan lagres i Depotoperasjoner, innebærer at det bare blir nødvendig å opprette en ny seksjon for en transformert representasjon i de tilfeller hvor også datafiler endres. Det vil i praksis si at fullstendige transformerte representasjoner bare kreves ved en formatkonvertering av datafiler.

Ved mottak skal Arkivverket selv generere en samlet sjekksum for avleveringspakken, og verifisere eventuelle medfølgende sjekksummer fra arkivskaperen, jf. kravene i vedlegg 2, punkt 2. Dokumentasjon av Arkivverkets verifisering og sjekksummer som Arkivverket selv genererer ved mottak, skal lagres i Depotoperasjoner, mens arkivskaperens sjekk-summer skal bero i SIP-delen. SIP-delen med original mottatt representasjon skal omfatte den komplette informasjonen som ble mottatt fra arkivskaperen, mao. også arkivskaperens logiske metadata og annen medfølgende dokumentasjon ved avleveringen. "Bevarings-metadata" – også med nødvendige Asta-opplysninger – kan dermed ajourholdes og kompletteres uavhengig av den opprinnelige informasjonen fra arkivskaperen.

4.2.1.1 Nærmere om bruk av sjekksummer

Sjekksummer brukes som mekanisme for å bekrefte at informasjonen har vært gjenstand for uavbrutt integritetssikring gjennom alle stadier av bevaring. En sjekksum kan imidlertid bare bekrefte opprettholdt integritet i de tilfeller informasjonen er bevart uendret på fysisk nivå (uendret bit-strøm). Figuren nedenfor gir en mer detaljert fremstilling av prosjektets forslag mht. bruken av sjekksummer i Arkivverkets arkivpakker, fortsatt med utgangspunkt i den *logiske* arkivpakkemodellen.

Figur: Arkivpakke med sjekksummer



Kommentarer til figuren:

1) Bevaringsmetadata *kan* ha egen sjekksum. Men fordi arkivskaperens opprinnelige logiske metadata blir integritetssikret sammen med den opprinnelige representasjonen i arkivpakkens SIP-del, er dette strengt tatt ikke nødvendig. Dette gjør Bevaringsmetadata til en seksjon for komplettering og ajourhold av opplysninger. Hovedkravet blir da logging av endringer, og det bør vurderes videre om det er mer hensiktsmessig å bruke andre metoder enn sjekksummer.

2) En SIP *kan*[22] inneholde separate sjekksummer for bevaringsdata, datafiler og tekniske metadata. I Noark-5-avleveringer vil også hver enkelt medfølgende dokumentfil ha egen sjekksum. Bare arkivskaperens originale sjekksummer skal

[22] Formuleringen må nødvendigvis bli løs på dette punktet inntil det blir fastsatt bestemmelser om hvordan avleveringer/ deponeringer skal være organisert som arkivpakker og integritetssikret med sjekksum(mer) av arkivskaperen.

inngå i en bevart SIP. I en transformert representasjon vil nye sjekksummer være generert av Arkivverket, men originale sjekksummer, f.eks. i dokumenter fra Noark-5-systemer, kan fortsatt være bevart intakt i tillegg.

3) En SIP kan også ha tilknyttet en samlet sjekksum som er generert av arkivskaperen. Den skal da omfatte all inkludert informasjon i avleveringen: Datafiler, tekniske metadata, bevaringsinformasjon og annen dokumentasjon. Når Arkivverkets mottakskontroll genererer en slik samlet SIP-sjekksum, jf. punkt 5.1.2 om sentral mottakskontroll, skal den lagres i "Depotoperasjoner", jf. nedenfor. Senere transformerte representasjoner trenger ikke en tilsvarende samlet sjekksum på seksjonsnivå.

4) Depotoperasjoner skal inneholde sjekksummer for ulike typer av opplysninger:
   - for Arkivverkets integritetssikring av en SIP ved mottak,
   - for Arkivverkets testdata og annen utfyllende tekniske dokumentasjon i tilknytning til en SIP,
   - for reviderte/kompletterte tekniske metadata som er generert av Arkivverket i tilknytning til en SIP,
   - for senere operasjoner foretatt i depot

   Det vil altså kunne variere om en representasjons tekniske metadata med tilhørende sjekksummer er plassert innenfor en SIP-seksjon eller i Depotopplysninger. I Depotopplysninger vil disse opplysningene dessuten kunne finnes i flere versjoner. Arkivpakken må derfor ha et flagg som lokaliserer og identifiserer hva som er *gyldige* tekniske metadata med tilhørende sjekksummer.

5) Hver arkivpakke skal ha tilknyttet 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 detektere endringer. Den samlede sjekksummen kan genereres på ordinær måte, eller alternativt, som en sum av sjekksummene som finnes i vedkommende arkivpakke. I det siste tilfellet vil det være tale om en enklere form for kontrollsum[23]. Det må utprøves i praksis hvilket av alternativene som er det mest hensiktsmessige.

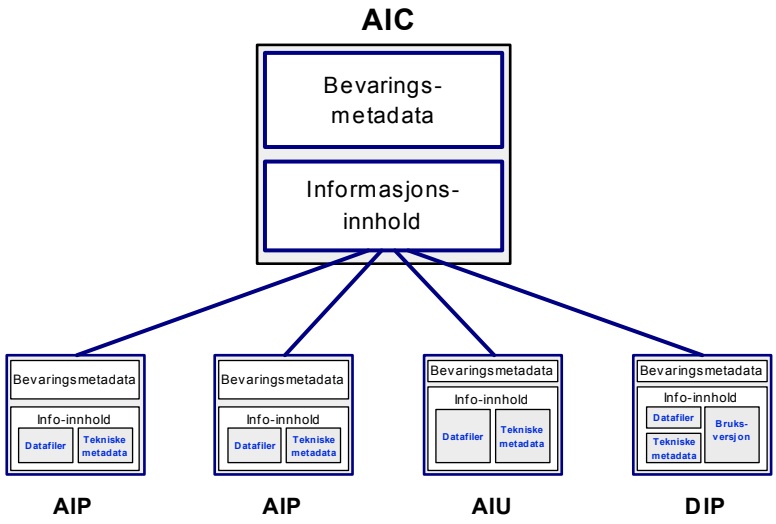
Det digitale sikringsmagasinet (DSM) må være sikret mot at en sjekksum erstattes med en ny, eventuelt mot at dette skjer i kombinasjon med en uautorisert endring av informasjon. All endring av informasjon i DSM forutsettes logget. Generering av sjekksummer logges særskilt, likeledes all verifisering av sjekksummer. Dette må også gjelde sjekksummene som genereres og verifiseres i eget mottaksrom og innenfor de dedikerte kontrollområdene for mottak og testing av avleveringer.

[23] I denne sammenheng bør en modell med hierarkisk genererte sjekksummer vurderes. En slik modell planlegges for Noark-5-systemer. Sjekksum for hvert enkelt (saks)dokument lagres her i den overordnede filen arkivstruktur.xml (som inneholder metadata i form av journalinformasjon). Sjekksum for hver overordnet fil lagres ett nivå opp i filen avlevering.xml, og sjekksummen for denne lagres ytterligere ett nivå opp, i filen info.xml. Fordelen med denne hierarkiske modellen er det ikke blir nødvendig å generere eller verifisere en samlet pakkesjekksum hver gang. Ulempen er at en samlet pakkesjekksum ikke kan genereres eller verifiseres i en enkelt operasjon.

4.2.2 Organisering av arkivpakker: 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 og andre grunner som vil fremgå av etterfølgende punkt 4.2.2.1, bør arkivpakkestrukturen likevel ikke implementeres fullt ut som en speiling av den logiske modellen. Denne monolittiske modellen 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. Dette vil redusere prosesseringsproblemer som kan inntreffe ved generering av store arkivpakker. Et annet viktig praktisk moment er at kopieringen til tape ved oppdatering av pakker vil forenkles.

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. OAIS-betegnelsen for arkivpakker av varianten grunnenhet er Archival Information Unit (AIU). 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)

På linje med prosjektets logiske arkivpakkemodell kan det beregnes en samlet sjekksum for alle enheter som omfattes av en AIC. Sjekksummen må også i dette tilfellet lagres et annet sted, dvs. i SAN-administrasjonssystemet eller forvaltningssystemet for digitalt depot. Samlede sjekksummer for de enkelte tilknyttede AIP-er og AIU-er vil derimot kunne lagres i AIC-en. Bruk av AIC-er vil dermed medføre de samme fordeler som prosjektets logiske arkivpakkemodell. Gjennom organiseringen av arkivobjektene unngås unødig repetisjon av informasjon[24]. Kompleksiteten reduseres og oversiktligheten i digitalt depot bedres ved at systemadministrasjonen bare trenger å "se" overordnede enheter.

[24] På dette punktet kunne det stilles spørsmål ved den logiske modellens legalisme i forhold til OAIS. Iht. OAIS punkt 4.1.1.5 skal det genereres en ny AIP dersom bevaringsmetadata eller informasjonsinnhold oppdateres. Ved en streng fortolkning med større vekt på det bokstavelige enn det funksjonelle blir en flernivå arkivpakkestruktur i strid med kravet i OAIS – om det skulle bety så mye. Implementeringsmetoden som bygger på bruk av AIC-er, er derimot helt ut legalistisk i forhold til OAIS.

Den store fordelen med samlepakkemodellen er fleksibiliteten. Et eksempel på dette er den enkle muligheten for også å knytte en klargjort bruksversjon av en arkivpakke – en Dissemination Information Package (DIP) – til en overordnet AIC[25].

4.2.2.1 Håndtering av arkivobjekter fra Noark-systemer

Samlepakkeløsningens fleksibilitet gjør det også mulig å håndtere noen svakheter og mangler ved OAIS-modellen. Mangler ved OAIS åpenbarer seg når arkivversjoner fra Noark-systemer og annen "Records Management" (RM) skal innlemmes i arkivpakker:

- Kategorien Bevaringsmetadata (PDI – Preservation Description Information) er unyansert. Hovedfokuset er på bevaringshistorikken. Men fra RM-systemer er det helt essensielt å bevare autentiserende metadata, typisk journalinformasjon. Her har OAIS ingen annen kategori å tilby enn "sekken" Bevaringsmetadata.
- OAIS tar ikke høyde for at Bevaringsmetadata – i det minste autentiserende metadata – kan være avhengige av tilknyttede tekniske metadata for å fremstilles. Tekniske metadata (Representation Information) i OAIS er utelukkende knyttet til selve datainnholdet (Content Information). Men OAIS setter heller ingen sperre mot en alternativ kobling.

Kategoriene i OAIS får som resultat at journalinformasjon fra Noark-4 og Noark-5 går til Bevaringsmetadata. Dokumentene som journalinformasjonen er metadata til, går til Datainnhold (Content Information), og skjemaene til alle medfølgende xml-filer går til Tekniske metadata (Representation Information). Journalinformasjon fra Noark-3 vil på tilsvarende måte gå til Bevaringsmetadata, mens den tilknyttede strukturbeskrivelsen (Noark-3-standardens) går til Tekniske metadata. For Noark-3-systemer vil datainnholdet (Content Information) være på papir, og dermed ligge utenfor arkivpakken.

I Norge er en journal et formelt dokument som per definisjon har et informasjonsinnhold. Det kan derfor fremstå som underlig at journalinformasjon ikke lar seg kategorisere som innhold i arkivpakken. Dette skyldes at OAIS-modellen ikke gir rom for å håndtere journalopplysninger *både* som metadata til dokumenter og som informasjon i seg selv. Journalrapportene (inkludert Offentlig journal f.o.m. Noark-5) og de øvrige rapportene som skal avleveres fra Noark-systemer, vil imidlertid ha sin naturlige plassering i kategorien Datainnhold (Content Information). At en journalrapport kan betraktes som den formelle journalen, bidrar til å balansere bildet.

PREMIS utbygger og differensierer bevaringsmetadata-kategorien i OAIS. PREMIS bør nyttes til å etablere en fastere konvensjon for håndteringen av autentiserende metadata generelt og Noark-informasjon spesielt i arkivpakker.

Også den logiske modellen under punkt 4.2.1 vil i praksis være for trang for RM-systemer fordi hver transformert representasjon i dette tilfellet trenger egne bevaringsmetadata. Den logiske modellen blir for rigid når den forutsetter at bevaringsmetadata i ethvert tilfelle kan

[25] Strengt tatt gir ikke OAIS rom for å bevare en DIP. Iht. OAIS skal en slik bruksversjon av en AIP fremstilles og gjøres tilgjengelige "on the fly" – for deretter å makuleres etter bruk. Kompetente arkivmiljøer overser OAIS på dette punktet, og undrer seg over CCSDS-aktørenes uvitenhet om hvor arbeidskrevende det kan være å fremstille en bruksversjon.

ligge som fellesinformasjon på toppnivået i en flerlags arkivpakke. Men prosjektets implementeringsmodell – arkivsamlepakker basert på en AIC – gir den fleksibilitet som er nødvendig for RM-systemer. Det blir mulig å veksle mellom å knytte ”økonomiske” AIU-er og "fullverdige" AIP-er til en klase av arkivobjekter.

4.2.3 Bruk av implementeringsstandardene METS og PREMIS

METS (Metadata Encoding & Transmission Standard) er utviklet som et tillegg til OAIS. Den er et implementeringsalternativ for å beskrive den indre strukturen i en arkivpakke-”container” og for å knytte sammen metadata og informasjonsinnhold. PREMIS (Preservation Metadata: Implementation Strategies) er et implementeringsalternativ som definerer bevaringsmetadata, herunder metadata for å støtte autentisitet, identitet og forståelighet. PREMIS har egen ”container”, men kan også innordnes i METS. Begge standardene er i utgangspunktet beregnet for bibliotekverdenen, og vedlikeholdes av Library of Congress i USA.

Det svenske riksarkivet har valgt å ta i bruk både METS og PREMIS. Prosjektet arrangerte i juni 2009 et seminar med Karin Bredenberg fra det svenske riksarkivet for å avklare veivalg og implikasjoner. Aktuelle problemstillinger ble senere diskutert på Nordisk arkivakademi i Boden i november 2009. Prosjektets foreløpige konklusjoner etter denne kontaktvirksomheten er følgende:

1) Det medfører flere fordeler å basere implementeringen av OAIS på METS (som container) og PREMIS. Videre utviklingsarbeid og senere vedlikehold kan da også trekke på andre miljøer som bruker disse standardene.

2) For å gjøre implementeringen enklest mulig bør følgende løsninger velges, som i Sverige:
   a) PREMIS knyttes samlet til METS (innbygges i digiprovMD).
   b) Alternativet med å konvertere alle binærfiler (ikke-XML-filer) til Base64 velges ikke.
   c) Som container-filformat velges tar-format[26].
   d) Kategorien dmdSec i METS skal ikke tas i bruk (for Descriptive information).

Det svenske riksarkivet vurderer nå å bruke arkivpakker med en ytre og indre METS-fil. Løsningen innebærer at en arkivpakke blir bestående av to hoveddeler: (1) en *indre* METS-fil med tekniske metadata, bevaringsmetadata (i PREMIS) og katalogdata (i EAD/EAC[27]), og (2) en tar-fil med innholdsfiler og den medfølgende dokumentasjonen fra arkivskaperen. Sjekksummer for tar-delens innholdsfiler 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.

[26] tar – *tape archive format*. Utbredt eldre format (oppr. Unix) som brukes til å pakke inn filer i en samlefil (tar-fil) for distribusjon eller arkivering.

[27] EAD – *Encoded Archival Description*: Standard for utveksling av arkivinformasjon som vedlikeholdes av Library of Congress i samarbeid med The Society of American Archivists, http://www.loc.gov/ead/
EAC – *Encoded Archival Context*: Standard for informasjon om arkivskapere (”authority encoding”) og arkivers brukskontekst. EAC kan brukes som en utvidelse av EAD, jf.: http://eac.staatsbibliothek-berlin.de/

Den svenske implementeringen av METS 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.

Det fremstår også som aktuelt å adoptere et annet element i den svenske løsningsmodellen. Det gjelder praksisen med å kopiere den indre METS-filen – PREMIS-, EAD-, EAC- og (eventuelt) ADDML-dokumentasjonen inkludert – inn i arkivinformasjonssystemet Arkis (tilsvarende norske Asta). Alle metadata om arkivpakkene blir dermed tilgjengelige for brukere.

4.2.4 Oppfølgingstiltak

Det er behov for å utvikle et program for generering av arkivpakkefiler i tilknytning til forvaltningssystemet for digitalt depot, jf. punkt 4.1.2. Programmet bør også håndtere andre oppgaver som skal ivaretas av forvaltningssystemet

Ved utviklingen av forvaltningssystemet bør det også treffes valg mellom ulike alternativer som vil foreligge ved generering og lagring av samlede sjekksummer for arkivpakker. Valgene, som særlig må vurderes i lys av løsningenes praktiske aspekter, er følgende:

1) Det må avklares om en AIP alltid skal være tilknyttet en overordnet AIC. I motsatt fall vil regelen måtte bli at samlet sjekksum for en AIP skal lagres i overordnet AIC når en AIC finnes, og i SAN-administrasjonssystemet (og eventuelt i forvaltningssystemet, jf. nedenfor) når den mangler.

2) Lagringen av samlede sjekksummer i systemet for SAN-administrasjon er nødvendig for å kunne verifisere arkivpakkers opprettholdte integritet i DSM, men det gjenstår å vurdere om, og i hvor stor utstrekning det er behov for å lagre sjekksummer i forvaltningssystemets database i tillegg.

3) SAN-administrasjonssystemet vil bare identifisere tar-filer, og ikke kunne ”se” objektene innenfor hver arkivpakke. Samlet sjekksum for en AIC eller AIP må følgelig være generert etter tar-pakkingen for at SAN-administrasjonssystemet skal kunne verifisere at den er bevart uendret i DSM. Men en slik verifisering etter utpakking av tar-filen krever en samlet sjekksum som er generert *før* pakkingen til tar-format. Om det siste strengt tatt er nødvendig, krever en videre vurdering.

4) Det er behov for å vurdere hvordan sjekksumsikringen av AIC-er skal organiseres. Alternativene vil her være å la sjekksummen omfatte en AIC som isolert enhet eller en AIC med alle tilknyttede AIP-er, eventuelt begge deler. En AIC-sjekksum som omfatter alle tilknyttede enheter, kan videre beregnes på alternative måter: som en generert sjekksum av de tilknyttede enhetenes sjekksummer eller som en ren sammenstilling (sekvensiell sammenkobling) av disse.

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.

4.3 Navngivingskonvensjon for logiske identifikatorer

Navngivingen og adresseringen av arkivpakker i digitalt depot må være basert på identifikatorer som både er effektive og bestandige. Det må genereres en unik ID for hver arkivpakke. Prosjektet går inn for 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 (5 siffer "nnnnn" dersom medarbeiderkode er unødvendig). Det er ikke noe mål i seg selv at en ID skal ha et logisk innhold. Det viktige er at den er unik.

Det anbefales å bruke den unike ID-en som filnavn når arkivpakker ferdigstilles for lagring som tar-filer, eksempelvis *"aip2011042400003.tar"*. På denne måten kan det automatisk etableres en URN[28] til pakken, f.eks. *URN:NBN:no-a1450-aip2011042400003.tar*. Denne kan igjen brukes til å bygge opp en URL. I de tilfeller adressen til en URL er tilgjengelig, kan en arkivpakke i digitalt depots ytre sone hentes frem ved hjelp av webteknologi. En slik URL kan være: *http://urn1/URN:NBN:no-a1450-aip2011042400003.tar,* der "urn1" er et lokalt, virtuelt domenenavn.

Tilsvarende navn foreslås for sjekksummer som blir beregnet for arkivpakker som helhet, og plassert i xml-filer, eksempelvis: *"aip2011042400003.xml"*.

Koblingen mellom de logiske URN-adressene og deres fysiske diskplassering lagres i en enkel databasetabell med to kolonner, en for den unike delen av URN-en, og en for den fysiske lagringen (f.eks. /*data3/urnobjects/ kb/Roll-NOR1-0365/kb20050811020353.jpg*). Både praktisk og sikkerhetsmessig er det – som i eksemplet – svært hensiktsmessig å bruke den unike URN-delen som objektets filnavn. Hvis objekter flyttes fysisk i lagringssystemet, eller hvis nye objekter lagres, vil URN-tabellen automatisk bli oppdatert påfølgende natt. Endrede og nye partisjoner i lagringssystemet gjennomløpes, og koblingene i tabellen ajourføres. Bruk av de unike URN-delene som fil- eller katalognavn gjør altså at koblingen mellom objektenes logiske og fysiske adresse alltid kan gjenopprettes etter feil eller ulykker.

SAN-administrasjonssystemet håndterer den fysiske lagringen av arkivpakker ved å plassere tar-filer på disk og tape. Tabellen som kobler arkivpakkenes ID (URN) med den fysiske plasseringen, forutsettes å være lokalisert i SAN-administrasjonssystemet. Men koblingstabellen må også kunne hentes frem av forvaltningssystemet ved å spørre fra dette systemet mot SAN-administrasjonssystemet. En alternativ løsning vil eventuelt være å plassere koblingstabellen i forvaltningssystemets database slik at arkivpakker kan

[28] En URN (Uniform Resource Name) er en *permanent, globalt unik, logisk* adresse til et digitalt objekt: en enkeltfil, en katalog eller annen samlefil (f.eks. en tar-fil), eventuelt også en informasjonspakke (AIP eller DIP) i tråd med OAIS-modellen. Elementene i en URN presenteres som en samlet streng, f.eks.: `URN:NBN:no-a1450-aip2011042400003.tar`. Her er de første 11 tegnene faste for en norsk URN, de neste 6 er faste og unike for Arkivverket, mens de etterfølgende 16 (inntil suffikset tar) er unike for hvert digitalt objekt i Arkivverket. En URN brukes på Internett ved å "innbakes" som tillegg i en URL (www). Det kreves noe egen programvare for å fremhente og vise en URN innbakt i en URL, men den er enkel å utvikle. URN kan også brukes web-uavhengig vha. spesialutviklet programvare

fremhentes uavhengig av administrasjonssystemet. Prosjektet betrakter SAN-plasseringen som den mest naturlige.

4.4 Samspillet mellom digitalt depot og Asta-systemet

På linje med annet arkivmateriale må alle arkivpakker i digitalt depot beskrives i Asta, Arkivverkets arkivinformasjonssystem. Asta må både gi oversikt over "originale" arkivpakker i indre sone (DSM) og bruksversjoner i den ytre. I det siste tilfellet vil Asta-systemet danne utgangspunktet for brukertjenester, og representere inngangen til de aktuelle arkivpakkene.

Asta 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. Det mangler også funksjoner for å styre prosessen ved mottakskontroll og testing av arkivpakker. Til denne oppgaven brukes i dag et eget system av eldre dato. Oppgaven kan komme til å bli videreført i Asta eller et tilknyttet dedikert system, men dette er foreløpig uavklart. Asta bør imidlertid enkelt kunne suppleres med funksjoner for å styre statusskiftet fra deponering til avlevering. For digitalt arkivmateriale som er deponert, er dette en nødvendig administrativ funksjon.

4.4.1 Problemet med skilte soner

Arkivverkets Asta-system er installert på en server som (logisk sett) vil tilhøre den ytre sonen i digitalt depot. Astas arkivbeskrivelser av arkivpakkene i digitalt depot må genereres og oppdateres i samspill med enheter i indre sone (DSM), først og fremst med forvaltningssystemet, 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. En slik informasjonsutveksling på tvers av skilte soner er svært problematisk. Ett og samme Asta-system vil ikke kunne være tilgjengelig i begge soner. Sluseløsningen som er beskrevet under punkt 4.1.1, åpner imidlertid for nødvendig styrt transport av informasjon mellom sonene.

Slusingen av informasjon mellom Asta og DSM vil måtte skje satsvis. Dette setter i seg selv grenser for omfanget av informasjonsutvekslingen og for kompleksiteten i synkroniseringsfunksjonene mellom de to sonene. Det er en av flere grunner til at Asta ikke bør beskrive innholdet i arkivpakker detaljert, men begrense seg til det overordnede nivået (AIC-pakkenivået).

Utvekslingen av beskrivelsesinformasjon mellom soner vil også hemmes av at Astas database ikke lar seg aksessere av et eksternt system. Interaktiv registrering er den eneste eksisterende metoden for innlegging av data i Asta. I løpet av 2010 vil imidlertid systemet få en egen importmodul. Asta vil dermed kunne oppdateres med katalogdata som er generert ved hjelp av andre verktøy.

4.4.2 Beskrivelsen av arkivpakker i Asta

Asta må referere til arkivpakkenes unike ID (URN). Men for øvrig legges det til grunn at opplysningene i Asta skal begrenses til den overordnede arkivbeskrivelsen for arkivpakken

som helhet. Det skal ikke kopieres eller spesifiseres opplysninger fra forvaltningssystemet og SAN-administrasjonssystemet om tilhørende metadatafiler, testdokumentasjon, kopier på tape, transformerte versjoner og andre depotoperasjoner. Men opplysninger om fremstilte bruksversjoner av arkivpakker (DIP) må registreres i Asta – når de finnes i DSM.

For bruksversjoner i ytre sone stiller det seg noe annerledes. Her vil det være behov for fyldigere dokumentasjon, spesielt om tilknyttede metadata. På dette punktet anbefales en løsning etter mønster av den svenske, hvor hele METS-filen kopieres inn i arkiv-informasjonssystemet, jf. punkt 4.2.3, siste avsnitt.

Registreringen av opplysninger i Asta forutsettes å skje etter de samme prinsipper som for materiale på CD-er i dag. Det vil si at datasystemet, 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*. Mapper innenfor underserien – henholdsvis for Metadata, Dokumentasjon, Data, Dokumenter og Rapporter – forutsettes i liten grad brukt for arkivpakker beroende i DSM, men det må gis opplysninger om eventuelle feil, mangler ved dokumentasjonen eller andre spesielle forhold. For digitalt skapt materiale er det ellers særlig viktig at arkivbeskrivelsen inneholder opplysninger om hvilke funksjoner og typer av informasjon som materialet omfatter. Systemnavn kan sjelden gi holdepunkter om dette.

En bruksversjon av en arkivpakke (DIP) kan registreres i Asta som underserie til en underserie. Men det kan også være et alternativ å bruke den definerte relasjonen ”Versjon av” mellom arkivenheter i Asta. De to ulike typene av arkivpakker (AIP eller DIP) og deres soneplassering må uansett identifiseres i Asta. Muligheten for at en DIP kan bygge på en kombinasjon av flere AIP-er, gjør dette desto mer nødvendig. Valget mellom de alternativene måtene å registrere bruksversjoner av arkivpakker på i Asta – eller like gjerne: samstemmingen av disse alternativene til fast konvensjon for registrering – krever en videre vurdering basert på mulighetene som faktisk vil foreligge i Asta.

4.4.3 Referanse til fysisk plassering i lagringssystemet

Arkivpakker som er plassert i indre sone (DSM), skal ikke kunne aksesseres via Asta. Asta trenger derfor ikke å referere til lagringsadresser (tar-filer) i DSM, men Asta må referere til den enkelte URN for at forvaltningssystemet skal kunne identifisere objektet i DSM. Brukspakker i ytre sone skal imidlertid kunne aksesseres via Asta, og for slike bruksversjoner må Asta også referere til lagringsadressen.

CD-ene som i dag brukes som medium for langtidslagring, registreres i Asta som *lagrings-enheter*. Men prosjektet ser ingen grunn til å operere med lagringsenheter i Asta etter at det digitale magasinet er tatt i bruk. Lagringsenhetene er *fysiske enheter*. I likhet med tradisjonelle arkivesker kan CD-er ofte ha et N:M-forhold til arkivenhetene, og en av deres viktigste egenskaper er angivelsen av hylleplassering. DSM vil i praksis utgjøre én felles lagringsenhet for hele det digitalt skapte arkivmaterialet. SAN-administrasjonssystemet vil automatisk holde rede på den fysiske plasseringen av arkivpakker på disk og tape. Det vil være særdeles upraktisk å oppdatere Asta med slik informasjon, og Asta vil heller ikke kunne nyttiggjøre seg den.

Asta er nå i stedet i ferd med å få mekanismer (databasetabeller med tilhørende program-funksjoner) for å knytte ett eller flere digitale objekter (filer o.a.) direkte til en arkivenhet ved hjelp av en URN-basert URL, altså i et 1:1- eller 1:M-forhold. URN-en er en permanent og globalt unik identifikator for filen. URL-en gir mulighet til å hente frem filen fra datalageret ved hjelp av web-teknologi. Det er et naturlig 1:1-forhold mellom en arkivpakke (AIP) som en logisk arkivenhet og den tilhørende tar-filen.

Prosjektet legger til grunn at en arkivenhet i Asta skal ha en permanent logisk referanse til den tilknyttede arkivpakken i lagringssystemet i form av en URN. Til bruksversjoner i ytre sone forutsettes Asta å ha en referanse til tar-filen, f.eks. i form av en URL.

4.4.4 Synkronisering av beskrivelsesinformasjon mellom enheter

Synkroniseringen med informasjon i forvaltningssystemet, SAN-administrasjonssystemet og arkivpakkene i lagringssystemet ville vært svært problematisk dersom beskrivelses-informasjonen i Asta skulle speile de interne elementene i arkivpakker og depot-operasjonene som tidvis utføres i dem. Men også når arkivbeskrivelsen i Asta er mindre detaljert, vil en koordinert oppdatering av de ulike enhetene med beskrivelsesinformasjon være en krevende oppgave.

Genereringen av beskrivelsesinformasjon forutsettes å skje i forvaltningssystemet (innenfor DSM) som ledd i genereringen av arkivpakker. All senere oppdatering av beskrivelsesinformasjon i SAN-administrasjonssystemet og DSMs arkivpakker vil også styres fra forvaltningssystemet. Både nygenererte og oppdaterte arkivbeskrivelser må "meldes" til Asta ved å legges ut som bestillinger til dette systemet. Dette inkluderer statusopplysninger som skal "flagges" i Asta. Oppdatering av overordnede opplysninger om arkiv og arkivskapere må imidlertid være styrt fra Asta-systemet. Disse må da meldes som bestillinger til forvaltningssystemet.

På sikt kan det bli åpnet for at forvaltningssystemet får direkte lese- og skrivetilgang til databasen i Arkivverkets Asta-system, men dette vil kreve en betydelig programutvikling og ansvarsavklaringer, og løsningen er neppe realistisk med det første. I mellomtiden må utvekslingen av arkivpakke-metadata mellom forvaltningssystemet og Asta skje ved interaktiv bruk av de to systemenes brukergrensesnitt, eller – fortrinnsvis – ved hjelp av de nye import- og eksportmodulene som planlegges implementert i Asta i løpet av 2010. Format for import og eksport av slike katalogdata kan med fordel være EAD og EAC – de standardiserte utvekslingsformatene for henholdsvis ISAD(G) og ISAAR (CPF)[29] – men det kan også være SQL (INSERT og UPDATE).

Informasjonsutvekslingen mellom systemene må under enhver omstendighet sluses mellom lagringssystemets skilte soner. Slusingen må styres av forvaltningssystemet i DSM. Det betyr at opplysninger som er lagt "i bestilling" mellom de aktuelle systemene, både må hentes inn til DSM og flyttes ut til ytre sone av forvaltningssystemet.

[29] ISAD(G) – *General International Standard Archival Description*. Standard for arkivbeskrivelse utgitt av The International Council of Archives (ICA/CIA), jf.: http://www.ica.org/sites/default/files/isad_g_2e.pdf
ISAAR(CPF) – *International Standard Archival Authority Record for Corporate Bodies, Persons, and Families*. Standard for beskrivelse av arkivskapere utgitt av ICA, jf.: http://www.ica.org/en/node/30230

5. PRODUKSJONSLINJER OG STYRINGSFUNKSJONER

Dette kapitlet behandler produksjonslinjer i tilknytning til mottakskontroll, testing og innlemmelse av arkivmateriale i digitalt depot. Behovet for konfidensialitetssikring av gradert materiale og sensitive personopplysninger vurderes, og håndteringsregler for slikt materiale foreslås. Kapitlet behandler dessuten funksjoner for overvåking av installasjoner og prosesser som må være på plass for å ivareta informasjonssikkerheten i digitalt depot.

5.1 Produksjonslinjer ved aksesjon og vedlikehold av arkivmateriale

Rammeverket for rutineopplegget i digitalt depot defineres i *vedlegg 2*.

5.1.1 Prosesstrinn ved aksesjon

Tabellen nedenfor gir en forenklet oversikt over trinnene i behandlingsprosessen fra materiale mottas til genererte arkivpakker blir innlemmet i DSM. Oversikten refererer til figuren under punkt 4.1, foran. Prosessbeskrivelsen er forenklet ved å utelate eventualiteter som at materialet kan kreve oppfølging i forhold til arkivskaperen under prosessen, eller bli nektet godkjenning. Flere slike "hva/hvis"-situasjoner gjennom prosessen gjør det også nødvendig å beskrive journalføringen og andre interne registreringsfunksjoner mer nyansert enn i oversikten, som i hovedsak nøyer seg med å definere ansvaret for register-føringen og dokumentasjonsoppgavene til enhver tid.

	Oppgave/funksjon:	*Ansvarlig etter utført:*	*Lagrings-område:*
1	Materiale mottas som rekommandert post eller med bud, og registreres i ePhorte. (Skjer overføringen elektronisk, 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[30]. Omgående genereres deretter en samlet sjekksum for avleveringspakken, dvs. for mottatte datafiler og all tilhør-ende 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

[30] Dette vil ikke være nødvendig når materiale mottas på USB-disk. Sjekksumgenerering vil da kunne foretas på mottatt medium. I dette tilfellet vil det også være naturlig å plassere materialet 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. punkt 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. Mottatte datafiler, arkivskapers dokumentasjon og dokumentasjon fra mottakskontrollen legges i "kø", klargjort for å hentes til testing av koordinator i DD. Testansvarlig kan evt. være oppgitt av Elark.	DD	Indre sone/ m-området
10	Koordinator i DD henter materialet til forvaltningssystemets lagringsområde i indre sone. Ansvaret for arkivskaperkontakt og ePhorte/ArkiVente-oppdatering 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 ePhorte/ArkiVente-oppdatering.	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, tilgangsbestemmelser og evt. deponeringstid.	DD	Indre sone/ f-området

	Oppgave/funksjon:	*Ansvarlig etter utført:*	*Lagrings-område:*
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
17	Arkivpakken tilknyttes en overordnet samlepakke (AIC). Dersom denne allerede eksisterer i DSM, hentes den ut, og oppdateres. I motsatt fall genereres en ny AIC. 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, jf. punkt 4.4.4, foran. Ved denne operasjonen skal det også genereres nødvendig informasjon til PDA-seksjonen om aksesjon og tilvekst.	DD	Indre sone/ f-området
19	Ny arkivpakke (AIP) og ny eller oppdatert AIC innlemmes i DSM. Ved hjelp av forvaltningssystemet genereres informasjon til SAN-administrasjonssystemet i DSM, herunder samlet sjekksum i 2 versjoner for vedkommende AIC.	IT	Indre sone/ DSM

Ved genereringen av sjekksummer for samlede pakker under punktene 16-19 vil det foreligge alternative rutineopplegg som forventes avklart ved utviklingen av forvaltnings-systemet for digitalt depot, jf. punkt 4.2.4. Dersom det skulle bli åpnet for at en AIP også kan innlemmes i DSM uten å være forbundet med en overordnet AIC, må samlet sjekksum for en slik frittstående AIP lagres i SAN-administrasjonssystemet. Utviklingen av forvaltningssystemet forutsettes også å avklare flere andre valg mellom rutinealternativer:

- om samlede sjekksummer skal lagres i forvaltningssystemets database i tillegg til SAN-administrasjonssystemets,
- om samlede sjekksummer skal genereres i 2 versjoner: før og etter pakking til tar-format,
- om en AIC-sjekksum skal omfatte AIC-en som isolert enhet, eller inkludere alle til-knyttede AIP-er. Når flere AIP-er inkluderes, foreligger også alternative genererings-metoder, jf. punkt 4.2.4.

5.1.2 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. I tillegg til å kontrollere at en avleveringspakkes innhold og dokumentasjon er i samsvar med forutsetningene, skal denne instansen verifisere med-følgende sjekksum for den samlede avleveringspakken og dokumentere behandlings-prosessen.

Umiddelbart ved mottak skal det dessuten genereres en samlet sjekksum for hver avleveringspakke for å muliggjøre en senere verifisering av at informasjonsinnholdet er bevart uendret fra og med ankomst i Arkivverket. Denne operasjonen skal følge faste prosedyrer, og foretas uavhengig av om materialet er integritetssikret med sjekksummer av

arkivskaperen. Alt mottatt materiale ved avleveringen skal omfattes av sjekksummen. Sjekksum-genereringen, verifiseringen av den og overføringen av det mottatte materialet til et lagringsområde skal attesteres av 2 autoriserte personer, og registreres i en egen logg.

Det er behov for et kompetent, profesjonelt apparat for å håndtere verifisering og generering av sjekksummer, og for å dokumentere operasjonene ved mottak. Apparatet krever flere autoriserte personer. Det er også et viktig moment at mottakskontroll og testing ikke utføres av de samme personene. Prosjektet legger derfor til grunn at den sentrale mottakskontrollen skal utføres av autorisert personale fra Elark-seksjonen i Riksarkivet.

Avleveringspakker må skannes for virus. Det er nødvendig å plassere alt mottatt materiale i karantene – minimum i 3 uker – før viruskontrollen gjennomføres. Dette sikrer at antivirus-programmet er oppdatert i forhold til ny, ondartet programvare.

Mottakskontrollens generering av en samlet sjekksum kan ikke vente til virus-skanningen er gjennomført etter karantenetiden. En slik forsinkelse vil senere kunne skape tvil (eller gi påskudd for tvil) om prosessen har gitt åpning for uautoriserte operasjoner på materialet[31]. Rutineopplegget må derfor baseres på en løsning hvor sjekksum-generering foretas umiddelbart ved mottak på frittstående utstyr uten risiko for å infisere annen utrustning og andre lagringsområder.

Etter integritetssikring og viruskontroll foretas en innledende kontroll av avleverings-pakkens innhold:

a) Det kontrolleres visuelt at innholdet er komplett og overensstemmende med bestemmelser og avtaler.
b) Det foretas en verifisering av avleveringspakkens samlede sjekksum når denne følger med fra arkivskaperen. En ny sjekksum skal da genereres og sammenholdes med arkiv-skaperens. Også denne operasjonen skal attesteres av 2 autoriserte personer og loggføres.

Det skal særskilt dokumenteres hvordan arkivskaperen har integritets- og autentisitetssikret materialet. Dette vil kunne ha stor betydning for ettertidens vurdering av materialets pålitelighet i seg selv.

Materiale som passerer den initielle kontrollen, legges med tilhørende dokumentasjon og operasjonslogger fra kontrollen i kø for å hentes av forvaltningssystemet.

Mottakskontrollen skal sende kvittering for det mottatte materialet til arkivskaperen eller avhenderen. Det forutsettes at Elark-seksjonen (v/mottakskontrollen) er Arkivverkets sentrale kontaktpunkt mot arkivskapere, men DD-seksjonen eller eventuelt vedkommende statsarkiv må være ansvarlig for kontakten gjennom prosessen med testing og godkjenning. Desto viktigere er det at kvitteringen for mottak informerer om ansvarsplasseringen ved Arkivverkets videre behandling av materialet.

Mottakskontrollen forutsettes også å ha det sentrale ansvaret for prosesstyringssystemet for mottak og testing. Det må fortsatt finnes et system tilsvarende dagens ArkiVente for dette

[31] Dersom en avleveringspakke også er integritetssikret med sjekksummer av arkivskaperen – og disse fortsatt er verifiserbare – vil det uansett ikke være grunnlag for tvil.

formålet, jf. det pågående arbeidet med å utforme en kravspesifikasjon i Betty-prosjektet. Ved mottak av en avleveringspakke må opplysninger om arkivskaperen, arkivet og vedkommende arkivserie legges inn dette systemet. Asta vil i mange tilfeller allerede ha denne informasjonen. I motsatt fall må mottakskontrollen sørge for at opplysningene blir lagt inn interaktivt i Asta, og at de deretter blir registrert i prosesstyringssystemet.

I fasen med testing og godkjenning må oppdateringen av prosesstyringssystemet være DD-seksjonens og eventuelt vedkommende statsarkivs ansvar. Prosesstyringssystemet må være plassert i ytre sone av digitalt depot, men kunne aksesseres av mottakskontrollen i DD-datarommet (via separat nettilknytning til ytre sone). Smidigere former for informasjons-utveksling med Asta enn i dagens ArkiVente og mulighetene for at et nytt system kan ha tilknytning direkte til Asta, er spørsmål som ventes avklart i kjølvannet av Betty-prosjektet.

5.1.3 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. Det må også vise testingens status.

Avleveringspakker skal testes og vurderes for godkjenning etter kriteriene i avleverings-bestemmelsene § 8-8. Dersom testingen resulterer i justerte eller oppdaterte tekniske metadata, f.eks. en supplert ADDML-beskrivelse eller en ADDML-beskrivelse som er konvertert fra format 7.3 til 8.2, skal denne dokumentasjonen lagres som tillegg til avleveringspakkens originale tekniske metadata. Testeren skal også utarbeide en oversikt over de filformater som forekommer i avleveringspakkens datainnhold. Dokumentasjon av testoperasjoner og testresultater skal lagres sammen med eventuelle justerte tekniske metadata og supplerende informasjon som måtte være innhentet fra arkivskaperen, og integritetssikres med en samlet sjekksum av testeren.

Testeren skal utarbeide brevutkast om godkjenning eller avvisning av avleveringspakker etter testingen. Beslutninger om godkjenning eller avvisning skal kvalitetssikres ved behandling i Depotavdelingens linjeorganisasjon. Dette gjelder også for testing som utføres av statsarkiv-medarbeidere, med mindre beslutningsmyndigheten er blitt formelt delegert til vedkommende statsarkiv av Riksarkivaren.

Materiale som er godkjent etter testing, hentes tilbake fra kontrollområdet av forvaltnings-systemet sammen med testdokumentasjonen og annen tilleggsdokumentasjon. I denne forbindelse skal også oppdateringen av ePhorte og prosesstyringssystemet kvalitetssikres av DD-seksjonens koordinator.

5.1.4 Generering av arkivpakker (AIP) og samlepakker (AIC)

Når en avleveringspakke (SIP) er ferdig testet og godkjent, skal forvaltningssystemet gene-rere en arkivpakke (AIP) på grunnlag av denne. Kjernen i en ny arkivpakke vil være den originale avleveringspakken supplert med depotoperasjoner (jf. logisk modell, punkt 4.2.1) i form av dokumentasjon fra Arkivverkets mottakskontroll og testing.

Generering av arkivpakker forutsetter at det er utviklet en egen applikasjon for dette formålet, jf. punkt 4.1.2. Genereringen vil naturlig være en interaktiv prosess hvor noen metadata hentes maskinelt, mens andre registreres manuelt på grunnlag av tilgjengelige kilder. Prosessen ved genereringen av en arkivpakke vil ha 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
- tilrettelegging av metadata om pakken for Asta.
- tilrettelegging av metadata om pakken for SAN-administrasjonssystemet.
- beregning av alle gjenstående sjekksummer for ferdigstilte filer innenfor pakken og plassering av sjekksummene i andre filer enn dem de skal integritetssikre
- ferdigstilling av pakken som en arkivfil (tar-fil)
- beregning av samlet sjekksum for pakken og plassering av denne sjekksummen utenfor pakken, dvs. i en samlepakke (AIC), jf. nedenfor
- sammenknytning av pakken med andre pakker til en samlepakke (AIC)
- beregning av samlet sjekksum for samlepakken med tilhørende arkivpakker
- plassering av denne sjekksummen utenfor samlepakken

Operasjonene logges i en egen fil eller i forvaltningssystemets database. Samlede sjekksummer kan enten lagres i SAN-adminstrasjonssystemet eller i forvaltningssystemet. Valget mellom disse to alternativene bør bygge på et fyldigere bilde av de praktiske aspektene.

Ved generering av en ny arkivpakke forutsettes metadata for Asta å kunne hentes fra to tilgjengelige kilder: fra arkivskapers dokumentasjon i avleveringspakken (SIP), og fra prosesstyringssystemet for mottak og testing.

5.1.5 Vedlikehold av bevarte arkivpakker

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

Henting av arkivpakker til forvaltningssystemets arbeidsområde og lagringen (tilbake-plasseringen) av pakker i DSM logges.

5.1.6 Definerte roller i produksjonslinjene

Produksjonslinjene ovenfor krever definerte ansvarsområder og roller for å ivareta viktige koordinerings-, tilsyns- og kvalitetssikringsfunksjoner. Disse foreslås definert slik:

	Rolle/ansvarsområde	*Plassering*
1	Avtaler om avlevering/deponering. Kontakt med arkivskapere om prosedyrer ved overføring	Elark
2	Mottak av avleveringer og initiell mottakskontroll – (1) Rolle som operasjons- og dokumentasjonsansvarlig og (2) rolle for å attestere operasjoner	Elark
3	Håndtering av sikkerhetsgradert arkivmateriale – Rolle som ansvarlig i forhold til system for internkontroll	DD
4	Håndtering av annet beskyttet materiale (sensitive personopplysninger, taushetsbelagt materiale og gradert materiale etter beskyttelses-instruksen) – Rolle som ansvarlig i forhold til system for internkontroll	DD
5	Testing, godkjenningsvurdering og dokumentasjon av avleveringer – Rolle som ansvarlig for kvalitetssikring og dokumentasjon	DD
6	Administrasjon/koordinering av statsarkivmedarbeideres testing	DD
7	Kopiering mellom ytre sone/kontrollområdet og DSM – Rolle som operasjons- og tilsynsansvarlig	DD
8	Vedlikehold av arkivpakker i DSM – Rolle som forvaltningsansvarlig for bevart arkivbestand	DD
9	Etterkontroll av logger og rapporter knyttet til generering/verifisering av sjekksummer og utførte endringer av opplysninger i DSM – Rolle som tilsynsansvarlig	DD

Rollene under punkt 5-7 og 9 er tidligere i dette delkapitlet beskrevet som funksjoner som ivaretas av DD-seksjonens koordinator.

5.2 Konfidensialitetssikring og tilgangsstyring

5.2.1 Overordnede krav og behov

For deler av det digitalt skapte arkivmaterialet må Arkivverkets digitale depot ivareta kravene til konfidensialitet. Materiale i original versjon bevares innenfor et lukket område – Digitalt sikringsmagasin (DSM) – og vil bare være tilgjengelig for dedikerte medarbeidere i DD-seksjonen og IT-avdelingen. Men det er viktig å understreke at DSM også vil omfatte materiale som er deponert. Dette materialet skal definitorisk være utilgjengelig for bruk. At materiale bevares med status som deponert, innebærer mao. at kravet om konfidensialitet gjelder uavhengig av innholdet.

Digitalt skapt materiale som gjøres tilgjengelig for Arkivverkets egen bruk, skal lagres i det digitale depotets ytre sone som kopier av originaler i DSM eller som tilrettelagte bruks-versjoner av dem. Men også den ytre brukersonen må være et beskyttet område med tilfredsstillende krav til konfidensialitetssikring.

Gradert materiale etter beskyttelsesinstruksen finnes i et svært begrenset omfang i Arkiv-verkets nåværende digitalskapte arkivbestand. Sikkerhetsgradert materiale forekommer

foreløpig ikke, men Arkivverket må være beredt til å motta slikt materiale – både materiale som er sikkerhetsgradert i sin helhet, og elektroniske sakarkiver med forekomster av sikkerhetsgraderte dokumenter. Nasjonal sikkerhetsmyndighet (NSM) kan imidlertid komme til å konkludere at dagens samlede ugraderte arkivbestand må beskyttes på nivå Begrenset iht. sikkerhetsloven når den lar seg aksessere samlet i digitalt depot. Muligheten bør være nevnt fordi sikkerhetsmyndigheten (den gang FO/Sikkerhetsstaben) vurderte departementenes samlede ugraderte informasjon slik da regjeringskvartalets stamnett ble etablert tidlig på 1990-tallet.

Arkivverkets egne sikkerhetsbehov gir grunn til å stille krav om at alt lagret arkivmateriale i DSM skal være beskyttet på nivå med gradert informasjon etter sikkerhetsloven. Praktiske behov kommer i tillegg, og disse veier ekstra tungt. For vi unngår med dette å ta graderte dokumenter ut av sin opprinnelige sammenheng når de forekommer sporadisk i sakarkiver. Men det gjenstår å avklare hvilket graderingsnivå DSM kan godkjennes for. Muligheten for å etablere et eget, dedikert magasin for materiale med høyere gradering enn Begrenset må derfor holdes åpen, men spørsmålet krever ingen snarlig avklaring. Avleveringer med høyere gradert informasjon antas ikke å være aktuelt de nærmeste årene.

Behovet for å bevare materiale i sin opprinnelige sammenheng gjelder også sakarkiver som kopieres til Arkivverkets egen bruk i det digitale magasinets ytre sone. Det er grunn til å poengtere at også den ytre sonen *kan* være tilfredsstillende beskyttet for lagring av materiale med lavere gradering. En sentral forutsetning vil da være tilgang basert på brukerautorisasjon og adgangskontroll. Særskilte beskyttelses- og tilgangsregler må dessuten knyttes til det graderte materialet. Informasjon med laveste gradering krever primært beskyttelse mot uvedkommende (konfidensialitetssikring). Tilgang på dette nivået er ikke betinget av sikkerhetsklarering, men må bygge på et tjenstlig behov for den aktuelle informasjonen. Sikkerhetskravene i en ytre sone som inneholder lavgradert informasjon, er likevel blant spørsmålene som må tas opp med NSM.

5.2.2 Håndtering av gradert materiale

I Arkivverkets digitale magasin forutsettes materiale gradert FORTROLIG og STRENGT FORTROLIG iht. beskyttelsesinstruksen å bli håndtert på samme måte som materiale gradert BEGRENSET iht. sikkerhetsloven. Dette vil være i samsvar med beskyttelses-instruksen § 12. Ettersom det ikke er etablert en egen godkjennings- og tilsynsordning for håndteringen av gradert materiale etter beskyttelsesinstruksen, kan slik informasjon dermed behandles etter det opplegg som er fastsatt for materiale med laveste gradering etter sikkerhetsloven.

Håndteringen av gradert materiale i DSM må følge bestemmelsene i sikkerhetsloven med forskrifter. Etter forskrift om informasjonssikkerhet § 5-15 skal et informasjonssystem være sikkerhetsgodkjent før det kan benyttes til håndtering av sikkerhetsgradert informasjon. Sikkerhetsgodkjenningen skal gjennomføres på grunnlag av et sikkerhets-konsept, definert operasjonsmåte, sikkerhetsdokumentasjonen og verifikasjon av sikkerhetstiltak. Godkjenning skal etter forskriftens § 5-16 foretas av NSM eller eventuelt av virksomhetens egen leder i tilfeller hvor det er tale om enkle informasjonssystemer.

Arkivverkets digitale depot kan nok betegnes som et enkelt informasjonssystem når det graderte materialets omfang og sikkerhetsnivå legges til grunn. Dette vil i alle fall gjelde i en innledende fase. Muligheten foreligger dermed for å ta lagringsløsningen i bruk uten

forhåndsgodkjenning fra NSM. Prosjektet vil likevel foreslå at det søkes om godkjenning fra NSM, ikke minst på grunn av behovet for en langsiktig planlegging av magasinløsningen. Det forelås å ta kontakt med NSM allerede i den videre planleggingsfasen – slik NSM selv anbefaler.

Følgende dokumentasjon må foreligge som grunnlag for NSMs vurdering:

- et overordnet sikkerhetskonsept
- et systemteknisk konsept
- en godkjenningsstrategi

For selve prosessen med godkjenning må det deretter utarbeides sikkerhetsdokumentasjon for digitalt magasin som spesifisert i forskrift om informasjonssikkerhet §§ 5-22 til 5-26. Denne dokumentasjonen skal omfatte kravspesifikasjon for sikkerhet (KSS), tempest-risikovurdering og implementeringsbeskrivelser i form av en detaljert driftsinstruks og en brukerinstruks.

5.2.2.1 Forslag til håndteringsregler

Prosjektet foreslår følgende håndteringsregler for gradert materiale i digitalt depot:

1) DSM som helhet organiseres som et sperret område med den adgangskontroll som er nødvendig for å håndtere sikkerhetsgradert materiale, jf. bestemmelsene i forskrift om informasjonssikkerhet §§ 6-7 til 6-12.
2) Materiale gradert FORTROLIG og STRENGT FORTROLIG iht. Beskyttelsesinstruksen håndteres som materiale gradert BEGRENSET iht. Sikkerhetsloven.
3) Gradert materiale i DSM skal være særskilt merket på enhetsnivå. Det skal foretas en årlig gjennomgang av materialet for å revurdere beskyttelsesbehovet for hver enkelt enhet.
4) Bare personale i DD-seksjonen og IT-avdelingen med definerte oppdaterings- og vedlikeholdsoppgaver skal være tilknyttet DSM. Disse medarbeiderne må være sikkerhetsklarert og autorisert for det sikkerhetsnivå som det graderte materialet i DSM tilsier.
5) DSM skal kunne aksesseres fra de aktuelle kontormiljøene i DD-seksjonen og IT-avdelingen, men informasjonen skal være fysisk separert fra åpne nett også på kontornivå. Lokal nedlasting eller kopiering av graderte opplysninger fra DSM skal ikke være mulig.
6) For å hindre uvedkommendes tilgang til opplysninger i DSM skal de aktuelle kontormiljøene i DD-seksjonen og IT-avdelingen beskyttes med særlige sikringstiltak.
7) Innlemmelse og oppdatering av informasjon i DSM skal bare kunne foretas fra det sentrale styrings- og forvaltningssystemet for DSM.
8) Mottakskontroll av alle avleveringer til Arkivverket skal foretas i DD-seksjonens datarom i underetasjen. Det forutsettes at rommet utstyres med adgangskontroll. Testmaskinen i datarommet må være satt opp slik at alle områder med skrivetilgang blir slettet når maskinen slås av. Alternativt kan det benyttes RAM-disk som bare lagrer i minnet. Overføringsmedier med mottatte avleveringer som inneholder gradert materiale skal mellomlagres i datarommets safe inntil sletting eller destruksjon kan foretas på forskriftsmessig måte.

9) Testing av gradert materiale skal bare utføres av spesielt autoriserte medarbeidere i DD-seksjonen, og skje på et dedikert kontrollområde innenfor DSM.

Prosjektet har vurdert håndteringsløsninger for høyere gradert materiale som eventuelt vil kreve et dedikert og særskilt sikret datarom i Riksarkivets fjellmagasin, men foreløpig ikke gått videre med dette.

5.2.3 Håndtering av personopplysninger

Behandlingen av personopplysninger må oppfylle bestemmelsene i personopplysningsloven (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.

Forskriftens kapittel 2 spesifiserer regler og plikter knyttet til sikkerhetsledelse, sikkerhetsstrategi, risikovurdering, sikkerhetsrevisjon, avvikshåndtering, organisering, personell, taushetsplikt, fysisk sikring og sikring av konfidensialitet, tilgjengelighet og integritet mm. På de fleste av punktene kreves dokumenterte opplegg. Følgende overordende krav stilles:

- det må formuleres sikkerhetsmål og en sikkerhetsstrategi som viser formålet med å behandle personopplysninger og tiltak som iverksettes for å oppnå sikkerhetsmålene,
- det skal jevnlig undersøkes om sikkerhetsstrategien og de organisasjonsrutiner og arbeidsinstrukser som bygger på den, gir god nok informasjonssikkerhet som resultat.

Datatilsynet skal når som helst kunne be om å få fremlagt dokumentasjon om opplegget for internkontroll. For Arkivverket vil det representere en omfattende oppgave å få på plass et fulldokumentert kontrollopplegg. Internkontrollen kan heller ikke begrenses til avlevert digitalskapt arkivmateriale. Etter bestemmelsene må den omfatte behandlingen av elektroniske personopplysninger i etatens totale virksomhet.

5.2.3.1 Arkivverkets identifikasjons- og informasjonsplikt

Når problemstillingene i forhold til personopplysningsloven avgrenses til avleverte elektronisk arkiver, er det først grunn til å fastslå at det meste av dette materialet omfatter personopplysninger. Det er derfor også behov for å klargjøre om personopplysningslovens § 18 – Rett til innsyn – har gyldighet for det avleverte registermaterialet som Arkivverket forvalter. Iht. pol § 18 har enhver rett til å få vite hva slags behandling av personopplysninger en behandlingsansvarlig foretar, og iht. pol § 20 skal den behandlingsansvarlige av eget tiltak varsle den registrerte når personopplysninger samles inn fra andre enn den registrerte selv.

Bestemmelsene i pol § 20 gjør det nødvendig å avklare spørsmålet om Arkivverkets plikt til å varsle personer når opplysninger om dem hentes ut til brukertjenester. Prosjektets konklusjon er at en slik opplysningsplikt ikke hviler på Arkivverket, men på den som samler inn informasjonen, jf. pol §§ 20 og 21.

Et annet spørsmål som aktualiseres, er Arkivverkets informasjonsplikt som følge av bestemmelsen i pol § 18, 1. ledd om retten til innsyn. Prosjektets konklusjon på dette

punktet er at Arkivverket vil måtte etterleve lovens innsynskrav – med de begrensninger som følger av pol § 23. En slik plikt vil ha viktige praktiske implikasjoner, og innebærer at Arkivverket i det minste må kunne *identifisere* avlevert digitalt registermateriale som inneholder personopplysninger. Materiale med sensitive personopplysninger (helse, religion, seksuell tilhørighet, bøter og kriminelt rulleblad mv.) må kunne identifiseres særskilt. I praksis krever dette et system med klassifisering av alt avlevert digitalt register-materiale mht. personopplysninger.

5.2.3.2 Forslag til håndteringsregler

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.

5.3 Overvåking av installasjoner og prosesser

Oversikten nedenfor omfatter særlig loggingsfunksjoner, men den har også funksjoner for varsling. I begge tilfeller er det behov for å definere rapporter som skal produseres – rutinemessig og i tilknytning til bestemte hendelser – til bruk for Riksarkivbygningens sikkerhetsorganisasjon.

5.3.1 SAN-administrasjonssystemet

SAN-administrasjonssystemet er den del av DSM som IT-avdelingen vil forholde seg til som driftsansvarlig. En tentativ oversikt over SAN-administrasjonssystemets informasjonselementer og egenskaper følger som vedlegg 4.

I systemet for SAN-administrasjon kreves følgende funksjoner for tilstandsovervåking og sporing av prosesser:

a) kontroll av at alle arkivpakker som skal eksistere i systemet – og utelukkende disse – faktisk finnes i systemet
b) kontroll av filsystemets konsistens
c) logging av alt som skjer på filsystemet
d) verifisering av utført migrering (disk-overkopiering)
e) verifisering av kopiering til tape-versjoner

f) logging av brukeres tilgang til digitale objekter
g) logging av kopiering som foretas av digitale objekter (av forvaltningssystemet)
h) logging av utførte operasjoner på digitale objekter (endringer, tilføyelser)
i) logging ved innlemmelse/import av nye objekter (arkivpakker) i systemet
j) logging av objekter/arkivpakker som eksporteres til forvaltningssystemet
k) logging av sjekksummer som genereres på alle nivåer i systemet
l) logging av sjekksummer som verifiseres på alle nivåer i systemet
m) rutinemessig (periodisk) verifisering av arkivpakkenes samlede sjekksummer – med tilknyttet varslingsfunksjon for uoverensstemmelser.

Overvåkingsfunksjoner som omfatter status for arkivobjekter i DSM, vil være basert på eksporterte opplysninger fra det digitale magasinets forvaltningssystem ved generering og oppdatering av arkivpakker som lagres i DSM, jf. punkt a og k. Funksjonene under punkt k kan alternativt være lokalisert i forvaltningssystemet.

5.3.2 Forvaltningssystemet

Forvaltningssystemet administreres av DD-seksjonen. Her kreves:

a) endringslogging i forvaltningssystemets database
b) overvåking av databasens integritet
c) overvåking av arkivobjekter i sonene, herunder område for mottakskontroll og kontrollområder til bruk ved test av avleveringer
d) logging av eksport til, og import fra kontrollområdet i ytre sone
e) logging ved generering av arkivpakker: ved pakking til tar-filer og ved generering av pakkesjekksummer (jf. punkt 5.3.1, k)
f) logging av import/eksport mellom indre sone og ytre sones brukerområder
g) logging av import/eksport mellom områder innenfor indre sone
h) logging av all informasjon som kopieres (av autoriserte brukere) fra forvaltningssystemets arbeidsområde
i) separat logging av skiftet sonetilknytning (bruk av sone-svitsj)
n) varsling om formater i DSMs arkivpakker som er foreldet og ute av bruk - eller trues av foreldelse (basert på oversikt over samtlige forekommende formater).

Det vil dessuten være naturlig å importere prosessinformasjon fra mottakskontrollen og testområdene, jf. nedenfor, for å lagre opplysningene i forvaltningssystemets database.

5.3.3 Mottakskontroll

Mottakskontrollen administreres av Elark-seksjonen. Her kreves:

a) overvåking av objekter som kontrolleres (identifikasjon og liggetid mm.), herunder objekter som innledningsvis plasseres i karantene
b) logg for verifisering av medfølgende sjekksummer
c) logg for generering av sjekksummer i tilknytning til mottak (attestert logg)
d) beslutnings- og ferdigstillelseslogg

e) logg for eksport fra DD-datarom til eget mottaksområde i DSM

5.3.4 Kontrollområder for testing

Kontrollområdet til bruk ved testing av gradert materiale er lokalisert innenfor DSM, og aksesseres fra forvaltningssystemet. Kontrollområdet til bruk ved testing av ugradert materiale er lokalisert i ytre sone. I begge områder kreves

a) overvåking av objekter under testing (identifikasjon og liggetid mm.)
b) logg for generering av nye sjekksummer
c) beslutnings- og ferdigstillelseslogg
d) logg for informasjon som er slettet

5.3.5 Ytre sone (sone for bruksversjoner)

Magasinets ytre sone skal omfatte brukskopier av arkivpakker til Arkivverkets interne bruk, eventuelt også til bruk for forskere og andre med autorisasjon for å aksessere definerte objekter fra Arkivverkets publikumsarealer. Områder med egne tilgangs-bestemmelser og egen brukerautorisasjon skal kunne defineres i ytre sone.

Innhold i arkivpakker i ytre sone skal ikke kunne endres. Kontrollfunksjonene kan dermed begrenses til følgende

a) kontroll av at alle arkivpakker som skal eksistere i denne del av systemet – og bare disse – faktisk finnes der
b) kontroll av filsystemets konsistens
c) logging av brukeres tilgang til arkivobjekter i ytre sone
d) logging av kopiering som foretas av digitale objekter (nedlasting til lokal PC) – i den grad slik kopiering tillates

Ytre sone vil ha et eget administrasjonssystem. IT-avdelingen vil være ansvarlig for driften av systemet. DD-seksjonen vil i utgangspunktet ha ansvaret for tilpasningen og over-føringen av bruksversjoner til systemet. DD-seksjonen fremstår imidlertid ikke som en naturlig ansvarlig instans for bruk og brukerautorisasjon.

6. TILGJENGELIGGJØRING AV ARKIVMATERIALE

Dette kapitlet nøyer seg med å skissere en ramme for bruk og brukertjenester på avlevert, digitalt skapt arkivmateriale som i originalversjon beror i Digitalt sikringsmagasin (DSM).

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. Typiske 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. Den krever områder med differensierte tilgangsrettigheter, brukerautorisasjon og konfidensialitetskontroll. Tilgang til materiale må logges. Brukere forutsettes bare å ha lesetilgang til materiale, men som integritets-sikring er dette ikke tilstrekkelig. Også bruksversjoner må integritetssikres med sjekksummer, og brukere bør ha mulighet for å verifisere sjekksummer.

Det forutsettes et sikkerhetsnivå som kan gjøre det forsvarlig å kopiere taushetsbelagt materiale til den ytre sonen. Det gjelder også når formålet med kopieringen er å klargjøre materialet for videre bruk[32]. Det gjenstår å avklare med NSM i hvor stor grad den ytre sonen kan omfatte materiale som krever særskilt beskyttelse. Men Arkivverkets ambisjon bør være at materiale opp til Begrenset iht. sikkerhetsloven kan ligge i denne sonen, gitt at det blir definert klare brukerroller og etablert tilfredsstillende systemer for bruker-autorisasjon og tilgangskontroll.

6.1 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 anvendes som en brukspakke uten noen videre tilrettelegging. I andre tilfeller vil det være behov for å generere spesielt tilrettelagte brukspakker. Det kan dessuten være aktuelt å tilrettelegge bruksversjoner som bygger på flere arkivpakker eller deler av arkivpakker. En brukspakke må derfor alltid dokumentere sin sammenheng med objekter i en eller flere arkivpakker for at det skal være mulig for brukere å vurdere materialets autentisitet, jf. underkapittel 1.3 om sentrale begreper.

[32] Dette alternativet kan imidlertid kreve at materialet må hentes tilbake til forvaltningssystemet for å ferdigstilles med genererte sjekksummer og URL-adresse.

Det bør bli mulig for autoriserte brukere å fremhente bruksversjoner ved behov med utgangspunkt i Asta. Når det fremgår av Asta 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 det også bli mulig for brukere å sende en "bestilling" til forvaltningssystemet dersom det ikke finnes en tilgjengelig kopi av arkivpakken i den ytre sonen. Dette vil likevel ikke kunne medføre en automatisk kopiering av pakken til den ytre sonen. Slusingen ut av DSM må styres manuelt av en autorisert medarbeider, relevante sjekk-summer må verifiseres i den ytre sonen etter utkopieringen og Asta må oppdateres med en URL til plasseringen i ytre sone.

I tilfeller hvor arkivpakker i den ytre sonen blir slettet fordi de er ute av bruk, må dette logges og Asta-systemet oppdateres.

6.2 Apparat for brukertjenester

Det kreves et apparat i krysningspunktet mellom DD-seksjonen og Seksjon for bruker-tjenester for å tilrettelegge bruksversjoner av arkivpakker.

For å betjene brukere kreves et eget apparat. I Riksarkivet vil dette naturlig være lokalisert i Seksjon for brukertjenester. Denne seksjonen fremstår også som den naturlige instans for brukerautorisasjon.

En prioriteringsinstans for brukertjenester på digitalskapt materiale – både i indre sone av digitalt depot og på nett – hører naturlig hjemme på etatsnivå i Arkivverket.

7. OPPFØLGING OG VIDEREFØRING AV PROSJEKTET

Her beskrives oppfølgingstiltak som er nødvendige for å implementere Arkivverkets digitale depot med utgangspunkt i prosjektets konklusjoner og forslag. Kapitlet gir en oversikt over gjenstående oppgaver, og inneholder forslag om spesifiserte prosjektløp i 2010.

7.1 Hovedutfordringer ved implementeringen av digitalt depot

Implementeringen av digitalt depot vil representere store utfordringer for Arkivverket både ressursmessig og organisatorisk. De organisatoriske utfordringene består i å få opp en profesjonell driftsorganisasjon, en egnet forvaltningsorganisasjon og en aktiv sikkerhets-organisasjon.

En sentral forutsetning for at implementeringen skal lykkes, er at prosjektløpene som omfatter den konkrete realiseringen av Arkivverkets digitale depot, blir skikkelig forankret i enhetene som skal ha ansvaret for drift og forvaltning, nærmere bestemt i IT-avdelingen og DD-seksjonen, og på et mer avgrenset område i Elark-seksjonen. Det er avgjørende for realiseringen at det bygges opp en profesjonell driftsorganisasjon. Her kreves ikke bare en tilfredsstillende dimensjonert enhet i det øyeblikk Arkivverkets digitale depot er klart til bruk, men en trinnvis opprustning basert på at IT-avdelingen har en medarbeider som kan arbeide dedikert med implementeringsprosjektene og utformingen av detaljerte drifts-rutiner.

I 2010 vil det være behov for å behandle uavklarte spørsmål og gjenstående planleggings-oppgaver, jf. punkt 7.2 og 7.3. Konkrete implementeringsprosjekter, jf. punkt 7.4, må gjennomføres parallelt med disse aktivitetene.

7.2 Spørsmål som krever videre avklaring

Det er nødvendig å avklare med Nasjonal sikkerhetsmyndighet i hvor stor utstrekning og under hvilke forutsetninger gradert informasjon tillates lagret i DSM.

Flere spørsmål i forbindelse med konfigureringen av et digitalt depot krever en avklaring av praktiske konsekvenser og implikasjoner før det kan trekkes konklusjoner eller treffes valg mellom alternativer. En slik videre vurdering kreves på 4 sentrale punkter:

1) valg av overordnet struktur og pakkingsformat (container-format) for arkivpakker, jf. punkt 4.2.3 og 4.2.4, foran
2) valg av SAN-administrasjonssystem
3) utviklings- og tilpasningsmulighetene i forhold til Asta-systemet
4) valg av modell for utveksling av tape-kopier med Nasjonalbiblioteket.

Vurderingen og anskaffelsen av et SAN-administrasjonssystem vil hvile på IT-avdelingen, men den bør skje i nær kontakt med prosjektaktiviteter som spesifiserer det digitale depotets forvaltningssystem, jf. nedenfor.

7.3 Planleggings- og tilretteleggingsoppgaver

Det er behov for oppfølgingstiltak på en rekke områder:

1) Spesifisering av en nasjonal standard for arkivpakkestruktur, bruk av sjekksummer og funksjonalitet for generering og vedlikehold av arkivpakker, jf. punkt 1.2
2) Spesifisering og utvikling av et forvaltingssystem for digitalt depot, jf. punkt 4.1.2 og 5.3.2
3) Sikkerhetsdokumentasjon til Nasjonal sikkerhetsmyndighet (NSM), jf. punkt 5.2.2,
   a) som innledende vurderingsgrunnlag for NSM,
   b) som grunnlag for en godkjenningsprosess
4) Dokumentasjon for internkontroll iht. kravene i personopplysningsloven m/forskrift, jf. punkt 5.2.3
5) Implementering av sikkerhetsopplegg for digitalt depot, herunder:
   a) merking av gradert/beskyttet informasjon i digitalt magasin
   b) sikring av kontormiljøer
   c) autorisasjon av personer for roller
6) Samlet dokumentasjon av system-, drifts- og rutineopplegg for digitalt depot
7) Revisjon av statlige avleveringsbestemmelser for elektronisk arkivmateriale bl.a. med krav til avleveringspakker (SIP)
8) Ny forskrift med krav til digitale depoter som oppbevarer offentlig arkivinformasjon
9) Samarbeid med Nasjonalbiblioteket for å utrede samarbeid om lagring som vertstjeneste, jf. Kulturdepartementets digitaliseringsmelding (2009).

Mellom flere av disse tiltakene vil det være en innbyrdes avhengighet. Den endelige konfigureringen av digitalt depot, og dermed også implementeringen av sikkerhet (5) og dokumentasjonen av system-, drifts- og rutineopplegg (6), vil måtte avvente utfallet av dialogen med Nasjonal sikkerhetsmyndighet (3). Spesifiseringen av forvaltningssystemet for digitalt depot (2) må på sentrale punkter avvente resultatene fra prosjektet for å utforme nasjonale standarder (1). Også arbeidet med forskriftene (7 og 8) vil ventelig måtte ta utgangspunkt i dette prosjektet.

7.4 Forslag til prosjektløp for 2010

7.4.1 Prosjekt om spesifisering og generering av arkivpakker

ABM-utvikling har nylig vedtatt å gi økonomisk støtte til det planlagte samarbeids-prosjektet mellom Riksarkivaren og sentrale kommunale aktører i 2010. Prosjektet skal spesifisere en generell norsk arkivpakkestruktur, og bruke denne som basis for å utarbeide en kravspesifikasjon til programvare for generering og oppdatering av arkivpakker mm. Prosjektplanen forutsetter at Arkivverket skal bidra med 2 årsverk i 2010. Elmag-prosjektet vil foreslå 4 prosjektdeltakere fra Arkivverket i dette samarbeidsprosjektet:

prosjektlederen fra Elmag-2, to fra Riksarkivet (DD- og Elark-seksjonen) og en fra statsarkivene.

7.4.2 Prosjekt for implementering av Arkivverkets digitale depot

Det foreslås et hovedprosjekt med 3 delprosjekter for å implementere Arkivverkets digitale depot på grunnlag av anbefalingene fra Elmag-2 og kommende forslag fra samarbeids-prosjektet med de kommunale aktørene:

1) Et delprosjekt med forankring i IT-avdelingen og med følgende mål:
   a) etablering av en driftsorganisasjon for digitalt depot med detaljerte rutine-beskrivelser og instrukser
   b) planlegging og gjennomføring av gjenstående utstyrsanskaffelser i digitalt depot, bl.a. SAN-administrasjonssystem og gjenstående tape-robot
   c) utprøving av tape-roboter med vekt på verifiseringsfunksjonene og synkroniseringen av lagrede objekter mot disksystemet

   Dette delprosjektet vil være avhengig av en betydelig konsulentbistand

2) Et delprosjekt med forankring i DD-seksjonen og med følgende mål:
   a) utarbeide detaljerte rutinebeskrivelser og instrukser for forvaltningen av digitalt arkivmateriale i digitalt depot
   b) ha ansvaret for utviklingen av forvaltningssystemet for digitalt depot, dvs. for programutvikling basert på kravspesifikasjon fra samarbeidsprosjektet med de kommunale aktørene supplert med funksjonskrav som er spesifikke for Arkivverket

   I tillegg til DD-seksjonen forutsettes prosjektet å ha deltakere fra Elark-seksjonen og minimum ett statsarkiv, spesielt i forbindelse med oppgave a. Oppgave b vil kreve betydelig konsulentbistand.

3) Et delprosjekt for implementering av sikkerhetsopplegg i digitalt depot
   a) for å fremstille nødvendig dokumentasjon for NSM
   b) for å kommunisere med NSM om sikkerhetsopplegget i digitalt depot, og få NSMs vurdering av Riksarkivets eksisterende SAN i en ny depotløsning
   c) evt. for å samarbeide med NSM ved en godkjenningsvurdering av sikkerhetsopplegget i digitalt depot

   Dette delprosjektet foreslås forankret i Riksarkivbygningens sikkerhetsorganisasjon.

Det foreslås et hovedprosjekt for å koordinere og følge opp aktivitetene i delprosjektene. Hovedprosjektet forutsettes også å koordinere aktiviteter mellom Arkivverkets implementeringsprosjekter og eksterne prosjekter. I tillegg til prosjektsamarbeidet om arkivpakkestruktur med de kommunale aktørene vil videre kontakt og samarbeid om løsninger med det svenske riksarkivet være en viktig oppgave. Denne kontaktvirksomheten forventes også å være til støtte for samarbeidsprosjektet med de kommunale aktørene.

7.4.3 Forskriftsprosjekter

Det forslås egne prosjekter for å utarbeide utkast til reviderte avleveringsbestemmelser for elektronisk arkivmateriale og ny forskrift med krav til digitale depoter som oppbevarer offentlig arkivinformasjon. En revisjon av avleveringsbestemmelsene er blant annet nødvendig for å få inn nye krav til organiseringen og genereringen av avleveringspakker. Revisjonen av avleveringsbestemmelsene foreslås forankret i DD-seksjonen. Forankringen av de to forskriftsprosjektene vurderes videre.

7.4.4 Samarbeidsprosjekter med Nasjonalbiblioteket

Et samarbeidsprosjekt med Nasjonalbiblioteket om lagring som vertstjeneste er allerede under planlegging. Det er også nødvendig å vurdere hvilke interne organisasjonstiltak en slik tjeneste vil kreve i Arkivverket.

I tillegg forutsettes et prosjekt for å planlegge utveksling av tape-kopier mellom Nasjonalbiblioteket og Arkivverket. Dette arbeidet forelås forankret i IT-avdelingen.