Oppdatert 26.08.2011

Kravspesifikasjon til forvaltningssystem for DIAS-arkivpakker

Dette dokumentet har to deler: en del med generelle krav til et forvaltningssystem (del 1) og en systematisk del (del 2), hvor krav er knyttet direkte til hver enkelt av aktivitetsfasene ved håndteringen av DIAS-arkivpakker i et digitalt depot.

Del 1 representerer DIAS-prosjektets generelle kravspesifikasjon til et forvaltningssystem for arkivpakker i et digitalt depot. Det er da tale om et depot som bygger på OAIS-modellen. Formålet med *del 2* er å konkretisere og anskueliggjøre kravene. Av denne grunn skilles det i del 2 også mellom hva som skal skje i hver av aktivitetsfasene mht. funksjoner og rutiner, og hva som kreves av forvaltningssystemet i vedkommende fase.

Også de generelle kravene i del 1 har to deler. Den primære og innledende delen omfatter krav til et system begrenset til håndteringen av *arkivobjekter*. Administrasjon av lagring – av medier og av "bits" og filer som *digitale objekter* – kreves i tillegg[1]. Et separat lagringssystem kan forutsettes å bli brukt for dette formålet, men ikke nødvendigvis. Et forvaltningssystem for arkivobjekter kan også være kombinert med grunnfunksjoner for lagringsadministrasjon. Med tanke på en slik mulighet er del 1 supplert med basale tilleggskrav til lagrings-administrasjon (punkt 18 - 26) – som en opsjon.

Del 1: Generelle krav til et forvaltningssystem

Forvaltningssystemet skal kunne utføre følgende funksjoner:

1. Kontrollere definerte arbeids- og lagringsområder utenfor forvaltningssystemets egen database
   - Autorisere brukertilgang til disse områdene
   - Logge opprettelse, oppdatering og sletting av filer på områdene
   - Kontrollere (sammenligne) filer fra arbeidsområder med originalversjoner på et dedikert "kontrollområde" som utelukkende systemet har tilgang til (forvaltningssystemets eget arbeidsområde).
2. Lagre nøkkelinformasjon om arkivpakker og om tilknyttede utførte operasjoner på arbeidsområder og systemets kontrollområde i en tilhørende database.
3. Hente en avleveringspakke – en DIAS-strukturert SIP eller en annen vilkårlig sammen-setning av filer/objekter – inn til et definert arbeidsområde, foreta utpakking av filer og lagre filer på arbeidsområdet.
4. Generere sjekksummer for filer/objekter og for samlede pakker.
5. Verifisere sjekksummer for filer/objekter og for samlede pakker.

[1] DIAS-modellen skal også kunne brukes ved off-line lagring på CD/DVD, og i dette tilfellet vil det bare kreves et enkelt system for lagringsadministrasjon.

6. Generere tar-filer for vilkårlige grupperinger av objekter.
7. Generere en arkivpakke basert på basert på DIAS-METS, DIAS-PREMIS, EAD og EAC-CPF – alternativt som en bevaringspakke (AIP), tillegg til bevaringspakke (AIU), brukspakke (DIP) og samlepakke (AIC). Prinsipielt skal det også kunne genereres en avleveringspakke (SIP).
8. Eksportere informasjon fra og om en arkivpakke fra systemets sentrale kontrollområde til:
   - et eksternt system (typisk: et sentralt arkivinformasjonssystem)
   - et eventuelt eget lagringssystem for digitalt depot.
9. Importere nøkkelinformasjon om en arkivpakke (herunder pakkens samlede sjekksum) til forvaltningssystemets egen database.
10. Lagre/plassere arkivpakke (AIP, AIU, AIC, DIP) i depot ved å skrive til vilkårlig valgt lagringsmedium– også i flere eksemplarer på ulike lagringsteknologier.
11. Hente ut (kopiere) arkivpakke fra depot til systemets sentrale kontrollområde for oppdatering.
12. Logge innlegging av arkivpakker i depot, uthenting (kopiering) av pakker til sentralt kontrollområde, oppdatering av pakker og kopiering av objekter ut av sentralt kontroll-område (ved bruk av terminalaksess).
13. Gi samlet oversikt over lagrede arkivpakker i depot – organiseringen av pakker i AIC-er, den enkelte pakkes ID (URN) og type, opprettelsestidspunkt, "flagg" som viser om pakken er en aktiv (gjeldende) versjon, tilgangsbestemmelser, klausul og eventuell sikkerhetsmerking.
14. Gi mulighet for fremfinning av arkivpakker i depot.
15. Gi mulighet for å styre tilgang til lagrede arkivpakker i depot (adgang til å kopiere til arbeidsområdet). Tilgang skal kunne styres individuelt for de enkelte arkivpakker på grunnlag av definert tilgangskategori.
16. Gi mulighet for å sperre for uthenting/kopiering av definerte arkivpakker fra depot på grunnlag av definert tilgangskategori.
17. Gi mulighet for å produsere rapporter, eksempelvis: oversikt over lagrede pakker i depot, pakker med sensitive personopplysninger, utført verifisering av sjekksummer og utførte operasjoner ved generering og oppdatering av pakker.

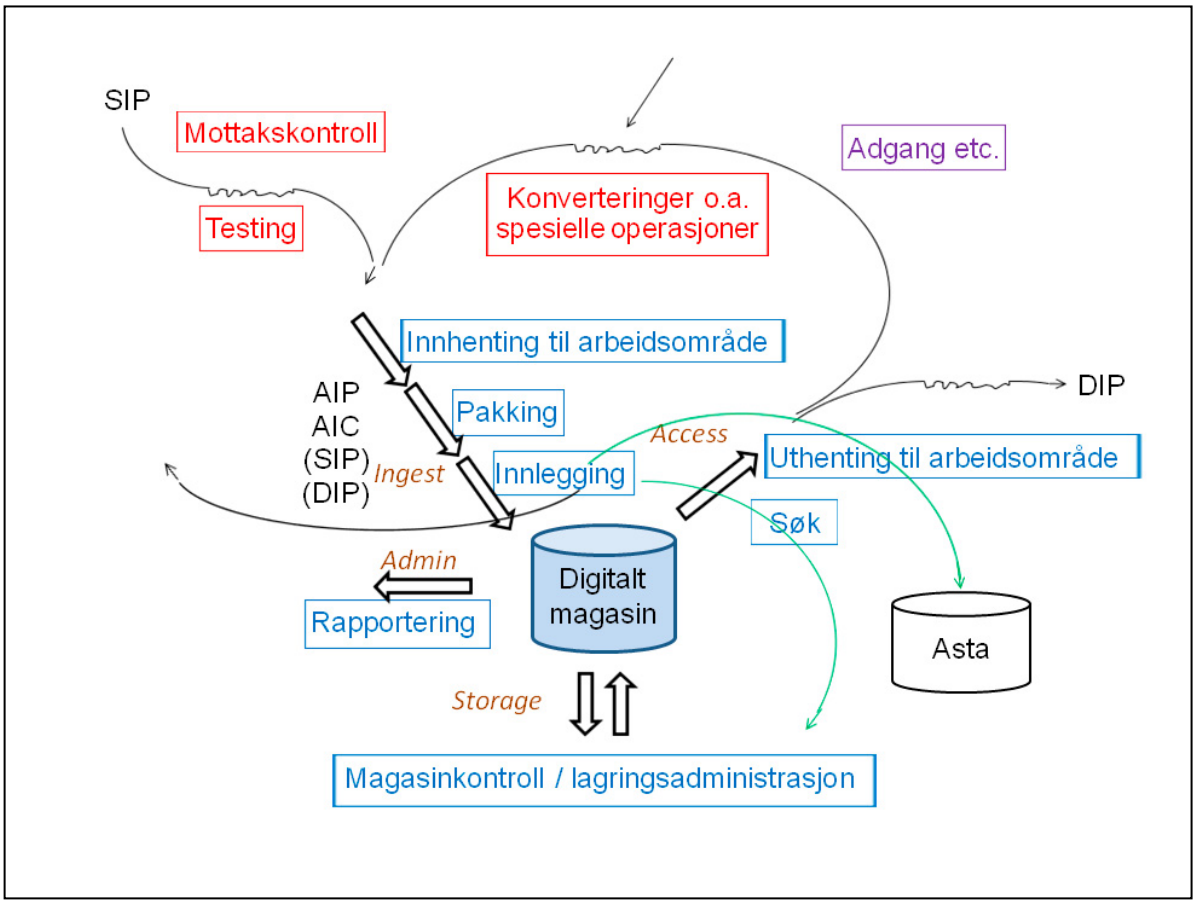
Tilleggskrav til lagringssystem (opsjon i de tilfeller lagringsadministrasjon inngår)

18. Filsystemets konsistens må kunne verifiseres.
19. Det må finnes mekanismer som sikrer synkronisert lagring av versjoner (kopier) av samme objekt på ulike medier (typisk disk og tape-robot).
20. Alle (identiske) kopier av de lagrede objektene må kunne lokaliseres.
21. Objekter må kunne migreres til nye medier, og medie-migrering må kunne verifiseres.
22. Alle former for korrumpering eller tap av data må kunne detekteres umiddelbart, også feil ved migrering eller synkronisering av kopier.
23. Det må finnes mekanismer for å monitorere bestanden av lagret informasjon, og bekrefte at den faktisk omfatter de objekter som skal finnes der – og bare disse.

24. De lagrede objektenes integritet må kunne monitoreres ved kontroll av sjekksummer.
25. Mekanismer for validering og logging av tilgang bør kunne implementeres på objektnivå slik at de omfatter individuelle arkivpakker, jf. også punkt 15 og 16.
26. All kopiering av objekter ut av digitalt depot må kunne logges.

Del 2: Krav til forvaltningssystemet i de enkelte håndteringsfaser

De ulike aktivitetsfasene ved håndtering av DIAS-arkivpakker i et digitalt depot er illustrert i figuren nedenfor:



Spesielt om aktiviteter i røde bokser (”Mottakskontroll”, ”Testing” og ”Spesielle operasjoner”): Det forventes ikke at forvaltningssystemet skal støtte disse funksjonene, men det skal kunne lagre resultater av operasjonene.

A. Mottakskontroll

Funksjoner og rutiner

En avleveringspakke (SIP) skal håndteres i et kontrollert miljø. Den skal integritetssikres ved mottak for å hindre en senere uautorisert endring av innhold, og for å kunne verifisere at innholdet er bevart uendret fra og med mottak. Det skal derfor genereres en samlet sjekksum ved mottak av en SIP. Sjekksum(mer) som følger med en SIP, skal verifiseres, enten som ledd i mottakskontrollen eller i tilknytning til etterfølgende testing, jf. punkt B. I mottakskontrollen inngår også følgende øvrige oppgaver: Materialet skal sjekkes for virus – etter en definert periode med karantene. Det skal kontrolleres at tilfredsstillende dokumentasjon er vedlagt, og foretas en initiell kontroll av at informasjonsinnholdet er i samsvar med bestemmelser og inngåtte avtaler. Mottakskontrollen må ivareta krav til konfidensialitetssikring. Utførte aktiviteter ved kontrollen skal dokumenteres og autentiseres.

Innledende virussjekk forutsettes å bli foretatt på frittstående utstyr.

Krav til forvaltningssystemet

A.1 Mottatte objekter (SIP) skal kunne hentes inn til kontrollert arbeids- og lagringsområde. Alternativt skal det kunne opprettes et separat område med dedikerte tilgangsrettigheter for å innhente og lagre en SIP med dokumentasjon fra eksternt utført mottakskontroll.

A.2 Det skal kunne genereres en samlet sjekksum for en mottatt SIP.

A.3 Det skal være mulig å verifisere sjekksummer som følger med en SIP (knyttet til samlet innhold og/eller til selve transporten).

A.4 Mottatte objekter skal kunne pakkes ut.

A.5 Utførte aktiviteter ved mottakskontrollen skal kunne logges (registreres) i forvaltnings-systemet og/eller lagres som en egen dokumentasjonsfil med tilknyttet sjekksum.

A.6 Henting/kopiering av objekter til og fra kontrollert arbeidsområde skal logges automatisk av systemet. Loggen skal også vise hvem som har utført operasjonene.

B. Testing

Funksjoner og rutiner

Prosedyrer skal utføres for å verifisere om innholdet i en SIP er teknisk konsistent, korrekt og komplett, om det oppfyller definerte struktur- og formatkrav, og om det er tilknyttet de tekniske og logiske metadata som kreves for å bevare informasjonen med opprettholdt lesbarhet og autentisitet. Sjekksummer for enkeltfiler i SIP-en skal verifiseres (eventuelt også samlet SIP-sjekksum, jf. punkt A). Testingen skal utføres innenfor et kontrollert og beskyttet område, og ivareta krav til konfidensialitetssikring. De utførte operasjonene og resultatene av testingen skal dokumenteres og autentiseres.

Krav til forvaltningssystemet

B.1 En SIP med tilhørende objekter og sjekksummer skal kunne hentes inn til et eget (dedikert) kontrollområde for testing. Både eksport av objekter til testområdet og tilbakehenting av objekter til sentralt kontrollområde skal logges automatisk av systemet.

B.2 Sjekksummer for enkeltfiler skal kunne verifiseres ved å generere nye sjekksummer for sammenligning med de mottatte.

B.3 Utførte aktiviteter ved testingen skal kunne registreres i en logg.

B.4 Operasjonslogger og dokumentasjon av testresultater skal kunne lagres integritetssikret med sjekksummer.

Ingest

C. Innhenting

Funksjoner og rutiner

Genereringen av en bevaringspakke (AIP) må tilrettelegges. Første versjon (generasjon) av en AIP skal være den opprinnelige SIP-en slik den ble mottatt. Dokumentasjon av operasjonene ved mottak og testing skal lagres i AIP versjon 2 dersom det også blir generert og lagret en slik versjon ved lagringen av SIP-en. Dersom det ikke blir generert en AIP nr. 2 ved lagringen av SIP-en, skal operasjonene ved mottak og testing av vedkommende SIP lagres i en separat AIU som er tilknyttet SIP-en via overordnet AIC.

Objektene som skal inngå i AIP-er, AIC-er og AIU-er, skal hentes fra dedikerte områder til et sentralt, beskyttet kontrollområde (arbeids- og lagringsområde) som første trinn i arbeidet med å generere en AIP. En eksisterende AIP og den overordnede AIC (arkivsamlepakke) som den er tilknyttet, skal også hentes ut fra digitalt depot til det sentrale kontrollområdet for å kunne oppdateres, jf. punkt D og E. Fra kontrollområdet skal det tilrettelegges for senere eksport av opplysninger i eller om arkivpakker, herunder til et sentralt arkiv-informasjonssystem (vanligvis Asta), jf. punkt E.

Krav til forvaltningssystemet

C.1 En SIP med sjekksummer og tilhørende operasjonslogger og dokumentasjon fra mottaks-kontroll og testing skal kunne hentes inn til sentralt kontrollområde for å tilrettelegges som en AIP.

C.2 En innhentet SIP skal kunne verifiseres mot SIP-innholdet på testområdet med logg-hendelser.

C.3 Innhentet dokumentasjon av operasjoner ved mottak og testing skal kunne verifiseres.

C.4 Metadata som skal finnes som egne objekter i en AIP og en AIC, skal kunne hentes fra SIP og fra logger/dokumentasjon av operasjonene ved mottak og testing, og lagres som egne filer. Nødvendige tilleggsmetadata skal kunne registreres og lagres som egne filer.

C.5 Nygenererte filer skal kunne integritetssikres med sjekksummer.

C.6 Data skal kunne hentes ut for å tilrettelegges for senere eksport til arkivinformasjons-system (ASTA) samt til forvaltningssystemets egen database.

C.7 Utførte aktiviteter i fasen med innhenting skal kunne registreres i en logg.

C.8 Trafikk (henting/kopiering av objekter) inn/ut av dedikerte arbeidsområder - også mellom dem - skal styres av forvaltningssystemet, jf. punkt A.6.

D. Pakking

Funksjoner og rutiner

En ny arkivpakke skal genereres som en AIP eller DIP (bruksversjon) i henhold til reglene for organisering og pakking i DIAS XML-skjema. For en ny AIP skal det også genereres en ny, overordnet AIC med peker til vedkommende AIP. For hver arkivpakke skal det genereres en samlet sjekksum. En AIP skal også pakkes som en tar-fil, og pakkens sjekksum skal i dette tilfellet være knyttet til tar-filen. Samlet sjekksum for en AIP skal ligge i tilhørende AIC (AIC-ens samlede sjekksum skal lagres i forvaltningssystemet, eventuelt også eksporteres til SAN-systemet og til arkivinformasjonssystemet, jf. punkt E). Oppdatering av en eksisterende AIP eller AIU skal skje ved at det pakkes og genereres en ny, fullstendig AIP- eller AIU-versjon med utgangspunkt i den foregående. Dette krever i seg selv en oppdatering av tilhørende AIC – med ny pakking og generering av ny samlet sjekksum.

Krav til forvaltningssystemet

D.1 En SIP, AIP, AIC og DIP skal kunne genereres etter reglene i DIAS XML-skjema, tildeles en unik ID (URN), og organiseres som en METS-fil med PREMIS innbygd.

D.2 En SIP, AIP og DIP skal i tillegg kunne pakkes som en tar-fil.

D.3 Samlet sjekksum for en SIP, AIP og DIP skal kunne genereres etter tar-pakking.

D.4 En SIP, AIP, AIC og DIP skal kunne valideres i forhold til DIAS XML-skjema.

D.5 Utførte aktiviteter i fasen med pakking skal kunne registreres i en logg og lagres i forvaltningssystemet.

E. Innlegging

Funksjoner og rutiner
Nye og oppdaterte arkivpakker skal overføres til digitalt depot og innlemmes. Informasjon om arkivpakkene skal dessuten kommuniseres til arkivinformasjonssystem (Asta), forvaltningssystemets database og et eventuelt eget system for lagringsadministrasjon (SAN-system). En oppdatert versjon av en AIP skal innlemmes i digitalt depot som en ny versjon i tillegg til den foregående. Den krever at det også innlemmes en tilhørende overordnet AIC i oppdatert versjon. En oppdatert AIC skal imidlertid erstatte foregående AIC-versjon. Istedenfor å oppdateres kan en eksisterende AIP tilknyttes en enkel, frittstående AIU, bestående f.eks. av en oppdatert ADDML-fil. Teknisk skal en slik AIU alltid tilknyttes en AIP via dens overordnede AIC. Lagring av en AIU i digitalt depot krever følgelig også oppdatering og ny innlegging av vedkommende AIC.

Krav til forvaltningssystemet
E.1 Nygenererte og oppdaterte arkivpakker skal kunne lagres i digitalt depot. Systemet skal bare tillate lagring av en ny eller oppdatert AIP, AIU eller DIP når den lagres sammen med en tilknyttet ny eller oppdatert AIC.

E.2 En oppdatert versjon av en AIP skal lagres i tillegg til den foregående, og ikke kunne overskrive den. En oppdatert versjon av en AIC skal erstatte den foregående. En oppdatert DIP skal også erstatte den foregående. En lagret DIP skal dessuten kunne slettes.

E.3 En AIU (med tilknytning til overordnet AIC) skal kunne lagres i digitalt depot som et eget objekt. En AIU skal miste flagg for ”aktiv” i forvaltningssystemets database eller alternativt kunne slettes automatisert ved innlegging av en ny, oppdatert AIP-versjon under vedkommende AIC.

E.4 Tilleggsinformasjon om en innlemmet arkivpakke skal kunne overføres til et eventuelt separat system for lagringsadministrasjon (SAN-system), herunder sjekksum for overordnet AIC.

E.5 Lagringsbekreftelse og Magasin-ID for arkivpakke skal kunne mottas fra eget lagringsadministrasjonssystem – om dette finnes.

E.6 Informasjon om den enkelte arkivpakke – herunder pakke-ID, (evt.) magasin-ID, status for AIP og for tilknyttede AIU-er (”flagg” for status = aktiv/ikke aktiv), tilgangsbestemmelser, forekommende formattyper, sikkerhetsmerke/markør, type klausul (restriksjon) og varighet for klausul – skal kunne overføres til forvaltningssystemet og lagres i dets database.

E.7 Tilrettelagt arkiv- og aktørbeskrivelse skal kunne eksporteres til bruk for sentralt arkivinformasjonssystem (ASTA), jf. punkt C.6.

E.8 Innlegging av objekter i digitalt depot skal logges automatisk av systemet.

E.9 Utførte aktiviteter i fasen med pakking skal kunne registreres i en logg og lagres i forvaltningssystemet.

Aksess

F. Uthenting

Funksjoner og rutiner
En arkivpakke (AIP + AIC) skal kunne hentes fra digitalt depot til sentralt arbeidsområde for å oppdateres eller for å tilpasses som en DIP. Teknisk skal en slik uthenting alltid skje som en kopiering av objekter fra digitalt depot. En AIP som hentes ut for oppdatering, skal kopieres

(og senere oppdateres) sammen med dens tilhørende AIC. Det samme gjelder for en eksisterende DIP som hentes ut for oppdatering - med mindre en oppdatert DIP-versjon ikke skal lagres i digitalt depot. Forvaltningssystemet skal gi oversikt over objekter som er kopiert fra digitalt depot til sentralt arbeidsområde. Systemet skal også gi oversikt over uthentede objekter fra digitalt depot som er kopiert (eksportert) ut av sentralt arbeidsområde

Krav til forvaltningssystemet

F.1 Objekter (AIP, AIU, AIC og DIP) skal kunne hentes fra digitalt depot ved å kopieres til kontrollert og beskyttet lagringsområde.

F.2 Systemet skal kunne sperre for uthenting/kopiering av objekter med en bestemt sikkerhetsmerking eller klausultype.

F.3 Systemet skal kunne låse for (samtidig) uthenting av mer enn én kopi av en AIC.

F.4 Uthenting (kopiering) av objekter fra digitalt depot skal kunne logges automatisk av systemet, likeledes eksport (kopiering) av uthentede objekter til områder utenfor sentralt arbeidsområde.

F.5 Det skal kunne produseres en samlet oversikt over uthentede objekter fra digitalt depot til sentralt arbeidsområde som viser tidspunktet for kopiering, tidspunktet for sletting og hvem som har foretatt operasjonene.

F.6 Det skal kunne produseres en tilsvarende oversikt over uthentede objekter som er eksportert (kopiert) til områder utenfor sentralt arbeidsområde.

G. Søk

Funksjoner og rutiner

På grunnlag av informasjonen som er registrert i forvaltningssystemet, skal det være mulig å fremsøke en oversikt over lagrede arkivpakker i depot, og fremfinne arkivpakker for uthenting (kopiering) fra depot.

Krav til forvaltningssystemet

G.1 Det skal kunne fremsøkes en samlet oversikt over arkivpakker og avgrensede oversikter over pakker etter valgte kriterier på grunnlag av opplysningene i forvaltningssystemet.

G.2 Arkivpakker skal enkeltvis kunne fremsøkes for uthenting (kopiering) fra depot.

H. Spesielle operasjoner

Funksjoner og rutiner

En sentral funksjon ved testing av arkivpakker vil være formatkontroll. En viktig funksjon i tilknytning til vedlikehold – eventuelt også ved mottak – vil være formatkonverteringer. Forvaltningssystemet forutsettes ikke å kunne støtte utførelsen av slike funksjoner, men det må kunne dokumentere operasjonene, resultatene og verifiseringen av resultatene .

I. Rapportering

Funksjoner og rutiner

Det skal være mulig å produsere ulike rapporter om arkivobjektene i digitalt depot på grunnlag av informasjonen som forvaltningssystemet har registrert om dem.

Krav til forvaltningssystemet

I.1 Det skal kunne produseres rapporter med oversikt over:
- alle lagrede pakker i depot,
- pakker med gradert materiale,
- pakker med sensitive personopplysninger,
- forekommende dokumentformater i arkivpakker,
- utførte operasjoner ved generering og oppdatering av pakker
- utført verifisering av sjekksummer
- statistikk

J. Magasinkontroll

Funksjoner og rutiner

Det må finnes systemer for å identifisere alle lagrede arkivobjekter i digitalt depot. Objektene må kunne aksesseres for å utføre monitoreringsfunksjoner. Det må finnes funksjoner for å overvåke den samlede arkivbestanden, og bekrefte at den faktisk omfatter de arkivpakker som skal finnes der – og bare disse. Arkivpakkenes opprettholdte integritet må være gjenstand for en aktiv monitorering. Det må eksistere en logg for kontrollen av sjekksummer. Det må også finnes tilgjengelige mekanismer for monitorering som varsler om fare når dokumentformater og tekniske metadata er i ferd med å bli teknologisk forgjengelige.

Krav til forvaltningssystemet

J.1 Det må finnes mekanismer for å identifisere og aksessere alle lagrede arkivobjekter.

J.2 Det må finnes mekanismer for å sikre at lagrede arkivobjekter bare kan endres eller slettes ved spesielle prosedyrer og med spesiell autorisasjon.

J.3 Det må finnes mekanismer for å styre hvilke operasjoner som er lovlige og mulige i forhold til de enkelte arkivpakker - basert på forvaltningssystemets opplysninger om klausuler, sikkerhetsmerker og andre tilgangsbestemmelser.

J.4 Basert på forvaltningssystemets opplysninger om forekommende dokumentformater i arkivpakker skal det kunne implementeres opplegg for en monitorering av formater.

J.5 Jf. for øvrig Del 1, punkt 18-26 om krav til lagringssystem.

K. Adgangskontroll

Funksjoner og rutiner

Det må finnes særskilte mekanismer for å styre tilgang til arkivobjekter i depot. For å ivareta behovet for konfidensialitetssikring må tilgangsstyringen også omfatte adgangen til å kopiere materiale ut av kontrollerte områder.

Krav til forvaltningssystemet

K.1 Systemet skal gi mulighet for å differensiere bruker- og tilgangsrettigheter

K.2 Mekanismer for validering og logging av tilgang skal kunne implementeres på objekt-nivå, og være knyttet til den enkelte arkivpakke.

K.3 Kopiering av informasjonsinnhold ut av kontrollerte områder skal kunne logges særskilt.