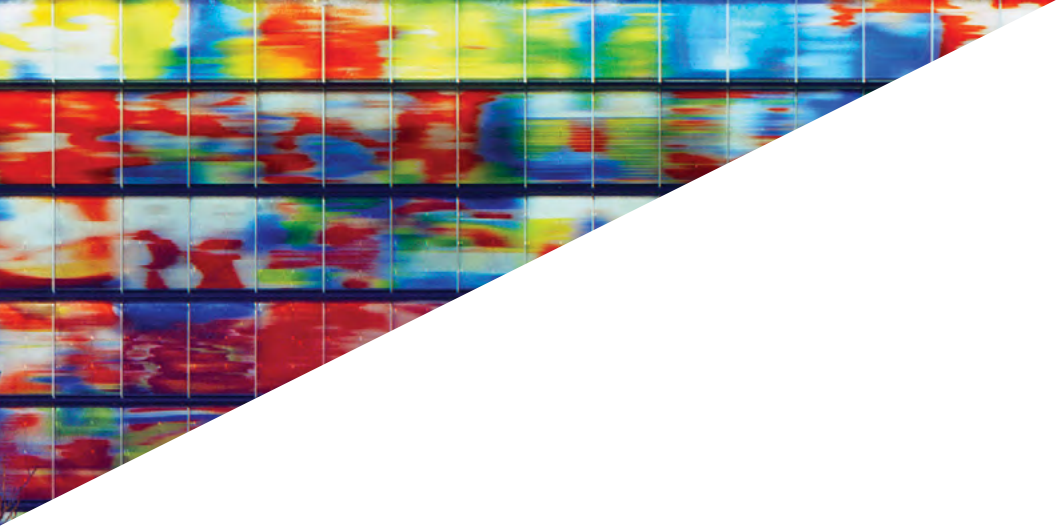


STORAGE BINNEN OAIS NORMATIEF MODEL EN GAP ANALYSIS VOOR BEELD EN GELUID

Versie 1.0

BEELD EN GELUID

Hilversum, Juni 2013
Versie: 1.0
Auteur: Daniel Steinmeier

ENGLISH SUMMARY

One of the organisational goals of the Netherlands institute of Sound and Vision is to become an OAIS-compliant trustworthy digital archive. This document focuses on the Storage-functions within the OAIS-model and what measures and strategies need to be in place in order to fulfill the guidelines related to storage as described in ISO-16363. This document also surveys the different storage systems that are part of the infrastructure at Sound and Vision and presents a gap analysis between the normative functions and guidelines related to storage on the one hand and the technical possibilities and limitations of our storage infrastructure on the other hand.

There are important differences between storage management solutions and the storage function within the OAIS-model. Most storage-related technological solutions have ways of ensuring files won't get corrupted or objects getting properly backed up which is comparable to the goals of digital preservation. On the other hand these systems often present an abstraction layer that hides the technical details from the user whereas being able to prove objects are handled correctly, is one of the most important points of digital preservation.
When these concerns aren't properly adressed risks that might endanger the preservational goals are corruption of the digital objects, loss of trustworthiness or degradation of services.
Our main storage management system is DivArchive. This system handles ingest of the digital objects (MXF, DPX and WAV), manages tape groups and enforces a storage plan so backup copies are created. The files are stored within the open AXF-format that also provides room for the following technical metadata:

- Provenance Collection
  - Provenancedata (repeatable)
    - Application
      - Application Name
      - Version
      - Description
      - Licensor
      - Licensee
      - Serial Number
    - Source
      - Manufacturer
      - Make
      - Model
      - Firmware
      - Description
      - UUID
      - Label
      - OS
      - Root path
      - Location
    - Destination (same fields as source)
    - Object Owner
      - Name
      - Facility
      - Description
      - Operator
- File tree (directorystructuur)
  - File (herhaalbaar)
    - Filename
    - File ID
    - Size
    - Position
    - Checksum

These fields are not a complete set of preservation metadata but they can be used to fill some of the fields in our metadata dictionary. DivArchive is also able to calculate checksums on ingest and verify checksums that are delivered with the object. Internally stored checksums are also used to verify copies have been made correctly. Some event-data in the lifecycle of the digital object is stored in the AXF.

In conclusion we can state that our storage management system has OAIS-compliant ways of ensuring integity on ingest and when creating copies. On the other hand the system can not be used to monitor bitrot other than during migrations and copy-events which is currently also the only feasible way of monitoring bitrot considering the volume of storage data that needs to be checked. Preservation metadata stored in DivArchive needs to be extracted and used to create a complete set of preservation metadata. To achieve this data from DivArchive needs to be combined with some of the events and technical metadata as provided by our new MAM-system. When this system is implemented this document also needs to be reviewed and updated.

INHOUD

INLEIDING

Een van de doestellingen in het meerjarenplan 2012-2015 van Beeld en Geluid is het bereiken van overeenstemming met het OAIS-model[1] met als doel het verkrijgen van de status van 'Trusted Digital Repository'. Het TDR-project is opgezet ten einde de randvoorwaarden om dit te bereiken in kaart te brengen. Binnen het OAIS-referentiemodel is een basisopzet gedefinieerd die aangeeft hoe een 'trusted' digitaal archief ingericht moet zijn om te kunnen zorgen dat een digitaal object van ingest tot aan access op een betrouwbare manier verwerkt, opgeslagen en uitgeleverd kan worden.

Het OAIS-referentiemodel verdeelt de verschillende onderdelen die een rol spelen binnen de lifecycle van het object op in functionele entiteiten. De entiteit 'Storage' is de functionele entiteit die verantwoordelijk is voor het opslaan en veilighouden van de AIP of Archival Information Package.

SCOPE VAN DIT DOCUMENT

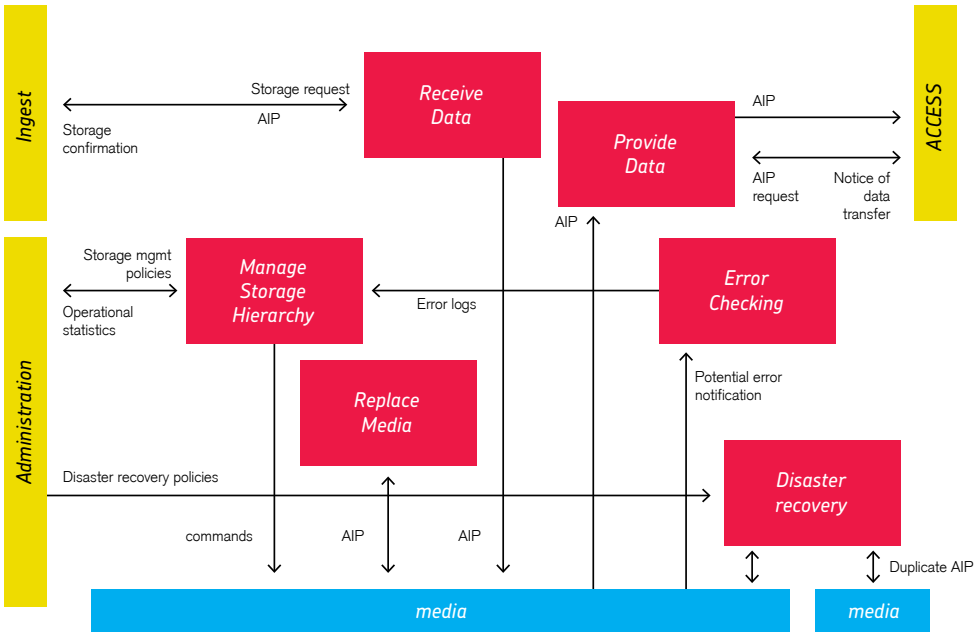
Dit document bevat een beschrijving van het normatieve model van de entiteit Storage zoals deze beschreven is in het OAIS functioneel model. Het normatieve deel bevat tevens een opsomming van de kwaliteitseisen zoals deze aan de verschillende functies binnen deze entiteit worden gesteld vanuit het document Kwaliteitseisen Digitaal Archief[2]. Tenslotte wordt het normatieve deel afgesloten met een korte beschrijving van de relatie tussen storage en preservering.

Het document bevat verder een beschrijving van de storage-infrastructuur zoals deze momenteel bij Beeld en Geluid in gebruik is. Vervolgens wordt bekeken in hoeverre de huidige infrastructuur de functies vanuit het OAIS-model vervult. Dit deel van het document vormt feitelijk een 'gap analysis'. Er wordt met name gekeken in hoeverre de huidige workflow met DIVArchive 6 voldoet aan de eisen die gesteld worden vanuit OAIS. Andere storageoplossingen worden minder uitgebreid onderzocht aangezien DIVArchive uiteindelijk ook de storage moet gaan verzorgen voor objecten die nu nog onder andere Storage Management-oplossingen vallen.

In bijlage A is een onderzoek te vinden dat laat zien in hoeverre de nieuwe versie van DIVArchive, versie 7, voldoet aan de eisen vanuit het OAIS-model. Tenslotte is in bijlage B ook een verslag opgenomen van een gesprek met Matthew Addis dat voor een groot deel ook relevante informatie biedt rondom storage en preservering.

[1] *ISO 14721:2003. The Open Archival Information System Reference Model.*
[2] *Kwaliteitseisen Digitaal Archief v1.0*

DE FUNCTIES VAN 'STORAGE' BINNEN OAIS



Figuur 1: Storage binnen OAIS, bron: OAIS functioneel model

Hieronder wordt uiteengezet welke functies precies worden vervuld binnen de entiteit 'Storage' zoals deze is geformuleerd in het functioneel model ISO 14721. Er wordt tevens bekeken hoe deze functies samenhangen met de eisen zoals deze zijn opgesteld in het document Kwaliteitseisen Digitaal Archief v1.0, gebaseerd op ISO-16363.

De *'Archival Storage'*-entiteit bevat de functie die nodig zijn voor opslag, behoud en uitlevering van de Archival Information Packages oftewel de definitieve archiefobjecten die gepreserveerd moeten worden.
Functies die vervult worden binnen deze entiteit zijn:

- het ontvangen en permanent opslaan van AIP's,
- het managen van de storage hiërarchie,
- het vervangen van dragers,
- het uitvoeren van error checking,
- het verzorgen van Disaster recovery-mogelijkheden
- het uitleveren van archiefobjecten die besteld worden.

De *"Receive Data"*-functie is verantwoordelijk voor het aannemen van het verzoek voor storage zoals deze wordt gedaan aan het eind van de ingest-fase.

- De functie moet zorgen voor het opslaan van de AIP op zijn definitieve storagelocatie en hiervan een confirmatie afgeven.
- Het verzoek tot storage kan ook een indicatie bevatten van het verwachte gebruik van het bestand zodat de storage-functie hiermee rekening kan houden bij de selectie van een opslagmedium.

Relatie met Kwaliteitseisen Digitaal Archief

- Deze functie vervult kwaliteitseis 1.8 waarin gesteld wordt dat het archief moet kunnen aantonen wanneer preserveringsverantwoordelijkheid voor de opgeslagen objecten precies is geaccepteerd[3]
- Deze functie vervult verder ook delen van eisen 2.9 en 2.11 doordat de actie van het opslaan van de AIP ook een controlemechanisme dient te bevatten dat kan garanderen dat de AIP correct is opgeslagen en doordat de actie van het opslaan zelf ook beschikbaar is als preserveringsmetadata[4]

De ***"Manage Storage Hierarchy"***-functie is verantwoordelijk voor

- het verplaatsen van AIP's tussen verschillende storageniveau's op basis van policies
- het monitoren van de statistieken van de storage om zeker te kunnen stellen dat AIP's niet corrupt raken tussen verplaatsingen
- het bewaken dat er voldoende ruimte is op de storage en dat de performance voldoet aan de gestelde eisen
- het controleren van de rapporten zoals deze zijn aangeleverd in het kader van de functie "Error Checking" en de resultaten hiervan te rapporteren aan de functionele entiteit Administratie.

Relatie met Kwaliteitseisen Digitaal Archief

- Deze functie vervult eis 1.9 waarin gesteld wordt dat alle incidenten gedocumenteerd en gerapporteerd moeten worden evenals de stappen die zijn ondernomen om de problemen te verhelpen.[5]

De ***"Replace Media"***-functie is verantwoordelijk voor het reproduceren van de AIP's in de loop van de tijd. Deze functie vervult de rol van verschillende migratieacties zolang deze niet resulteren in data verlies of compressie. Het betekent in de praktijk dus dat objecten op een nieuwe drager worden gezet om dat de oude drager verouderd is geraakt of versleten. Ook kunnen objecten eventueel in een nieuw containerformaat worden opgeslagen wanneer dit geen invloed heeft op de inhoud van de data maar alleen op de structuur, zoals TAR omzetten naar AXF.

- Dit betekent in de praktijk dat deze functie verschillende migratiestrategieën uitvoert zoals 'replication', 'refreshment' en 'repackaging' voor zover deze niet resulteert in verlies van informatie.

Relatie met Kwaliteitseisen Digitaal Archief

- Deze functie maakt daarmee een onderdeel uit van eis B4.2 waarin gesteld wordt dat het archief strategieën moet hebben om migratie mogelijk te maken, waaronder tools.[6]
- Het is hierbij belangrijk dat acties en resultaten daarvan terug te traceren zijn in logs of door middel van preserveringsmetadata om te kunnen bewijzen dat gedane acties overeenstemmen met het vooraf bepaalde beleid.

[3] *Kwaliteitseisen Digitaal Archief V1.0, blz. 7*
[4] *"B2.9: Beeld en Geluid checkt iedere AIP op het moment van creatie op compleetheid en correctheid.", "B2.11: Beeld en Geluid houdt up-to-date documentatie bij van alle acties en administratieve processen die relevant zijn voor de AIP creatie.", Kwaliteitseisen Digitaal Archief V1.0, blz. 9*
[5] *Kwaliteitseisen Digitaal Archief V1.0, blz. 13-14*
[6] *Kwaliteitseisen Digitaal Archief V1.0, blz. 10*

De ***"Error Checking"***-functie verzorgt uitdraaien van statistieken die garanties moeten bieden dat de AIP's niet corrupt raken in de opslag of tussen migraties. Dit betekent dat deze functie moet zorgen dat:

- alle hardware en software die betrokken is bij het verplaatsen en opslaan van AIP's foutmeldingen genereert die in een rapport te tonen zijn. Het controleren van de rapporten door personeel gebeurt binnen eerdergenoemde ***"Manage Storage Hierarchy"***-functie.
- de metadata niet corrupt raakt. Het is dus belangrijk om zeker te zijn dat ook de database van bijvoorbeeld een Storage management-applicatie niet corrupt raakt aangezien deze belangrijke informatie bevat over welke bestanden op welke locatie staan. Ook andere metadata zoals beschrijvende metadata is uiteraard op een vergelijkbare manier van belang.

Relatie met Kwaliteitseisen Digitaal Archief
Onder verantwoordelijkheid van deze functie hoort de vervulling van de eisen B4.4 en C1.8 waarin gesteld wordt dat het archief:

- de integriteit van de objecten actief moet monitoren
- de benodigde middelen moet hebben om bitcorruptie te kunnen detecteren.[7]

De ***"Disaster Recovery"***-functie is verantwoordelijk voor het repliceren van de AIP's naar een off-site locatie als voorziening voor noodgevallen.

- Binnen deze functie is het belangrijk dat ook voor de metadata rekening is gehouden met maatregelen voor noodgevallen zodat belangrijke informatie over bijvoorbeeld de locatie en de geschiedenis van de objecten niet verloren gaat.

Relatie met Kwaliteitseisen Digitaal Archief
Vanuit de kwaliteitseisen wordt gesteld dat deze functie in staat moet zijn om:

- backupfunctionaliteit[8] te verzorgen op een manier zodat er controle is over de hoeveelheid backup's die er worden gemaakt.
- te garanderen dat de creatie van de backup geslaagd is door middel van een integriteitstest en dat verschillende backup's met elkaar gesynchroniseerd zijn.[9]
- te zorgen dat van alle objecten en metadata een offsite backup is gemaakt, dat er een disaster recovery plan ligt om de data in noodgevallen weer terug te kunnen halen en dat van het plan ook een offsite-backup beschikbaar is.[10]

[7] *Kwaliteitseisen Digitaal Archief V1.0, blz. 10-11, 13*
[8] *Backup kan hier begrepen worden als reservekopie, niet als een volledige dump van het hele systeem.*
[9] *Eisen C1.7 en C1.16, Kwaliteitseisen Digitaal Archief V1.0, blz. 13-15*
[10] *Eis 2.4, Kwaliteitseisen Digitaal Archief V1.0, blz. 15*

De ***"Provide Data"***-functie is verantwoordelijk voor het kopiëren van de AIP vanuit de definitieve opslaglocatie naar een tijdelijke locatie om daar verder verwerkt te worden in het geval van een access-verzoek. De functie verzendt een notificatie na aflevering van de bestelling.

Relatie met Kwaliteitseisen Digitaal Archief
Deze functie is verantwoordelijk voor het vervullen van eisen B2.5 en C1.15 waarin gesteld wordt:

- dat het systeem een linking service moet hebben om objecten te vinden zonder dat dit hard gekoppeld is aan de fysieke locatie waar deze zich bevinden
- dat het systeem alle objecten, inclusief kopieën, moet kunnen localiseren en ophalen.

DATA STORAGE EN DIGITALE PRESERVERING

HET VERSCHIL TUSSEN STORAGE EN PRESERVERING

Het is belangrijk om duidelijk te maken wat het verschil is tussen storage en preservering omdat beide begrippen bepaalde gezamenlijke doelen delen en daardoor in de praktijk soms als synoniemen gebruikt worden. Echter zijn er op enkele essentiële onderdelen belangrijke verschillen te benoemen tussen deze twee begrippen.

Storage als proces draait in de basis om het schrijven van digitale bestanden op een medium zodat deze bewaard kan worden en later teruggelezen kan worden. Geavanceerdere storageoplossingen bieden tevens functionaliteit voor de controle van integriteit van de bestanden, door middel van checksuminformatie en disc scrubbing, replicatie van files, bijvoorbeeld in een RAID-oplossing en optimalisatie door middel van Hiërarchisch Storage Management.[11] Hierbij kan bijvoorbeeld gedacht worden aan het plaatsen van populaire objecten op schijf en het plaatsen van minder gebruikte bestanden op tape om snelheid van uitlevering te kunnen garanderen. Ook bevatten geavanceerdere storage-oplossingen functionaliteit voor de controle van het medium zelf, bijvoorbeeld door bij te houden of er niet te veel schrijf- en leesfouten ontstaan. Hiermee kan een schijf vervangen worden en de objecten op de schijf gemigreerd om dataverlies te voorkomen.

Digitale preservering deelt een aantal uitgangspunten met storage bijvoorbeeld het voorkomen van schade aan objecten door verschillende controlemechanismes in te bouwen. Een strategie om dit te bereiken kan bijvoorbeeld zijn om objecten van verouderde media tijdig naar nieuwe dragers te migreren. Preservering gaat echter in tegenstelling tot storage niet alleen om behoud van de data maar ook om de begrijpelijkheid ervan. Goede storage zorgt voor het behoud van de bits maar binnen preservering is het ook noodzakelijk dat de data begrepen kan blijven worden door de gebruikers van het archief op de lange termijn.[12] Ook is belangrijk binnen preservering dat de bits niet alleen behouden blijven, maar dat hiervan ook bewijs geleverd kan worden. Storage-oplossingen kunnen dus een rol spelen binnen preservering maar kunnen niet alle taken vervullen die nodig zijn voor preservering.

[11] *Phillips, Service-Oriented Models for Audiovisual Content Storage, https://prestoprimews.ina.fr/public/deliverables/PP_WP2_D2.3.1_SOAforAV_R1_v1.01.pdf, blz. 10*
[12] *Ibid. blz. 6.*

RISICO'S BIJ STORAGE

Zoals ook blijkt uit verschillende publicaties, waaronder het artikel 'Threats to data integrity from use of large-scale data management environments' van Matthew Addis, kunnen storage-infrastructuren dus niet standaard alle functionaliteit bieden zoals deze nodig zijn voor digitale preservering.[13] Bewuste aanvullende maatregelen moeten genomen worden vanuit de instelling die gebruik maakt van deze infrastructuren. Welke maatregelen noodzakelijk zijn hangt af van de specifieke infrastructuur en de risico's die hieraan verbonden zijn. Een risicoanalyse kan uitwijzen welke gevaren te voorzien zijn binnen de gekozen storage-oplossing, wat de kosten zijn van eventueel dataverlies, wat het kost om dit risico te vermijden en wat de opbrengst is van een dergelijke actie.

Addis benoemt een aantal risicocategorieën in relatie tot storage:[14]

- *Verlies van authenticiteit:* wanneer het archief niet in staat is om acties die uitgevoerd zijn op digitale objecten terug te traceren tot het origineel.
- *Corruptie/degradatie:* wanneer onvoorziene en onbedoelde veranderingen aan het digitale object plaatsvinden, bijvoorbeeld door fouten in de storagetechniek of door menselijke fouten.
- *Uitval van service:* bijvoorbeeld wanneer processen die de integriteit van de objecten bewaken uitvallen
- *Verlies van integriteit door verwachtingen van buitenaf:* wanneer het archief niet in staat is om verwachtingen van producers of de designated communities waar te maken vanwege beperkingen in het systeem, bijvoorbeeld wanneer producers bewijs van integriteit door middel van checksum noodzakelijk achten terwijl het archief dit niet kan leveren.[15]

Hoewel fabrikanten van storageoplossingen zich bewust zijn van de beperkingen van hardware en maatregelen inbouwen om dataverlies en corruptie te voorkomen, zijn deze maatregelen voor het vervullen van garanties rondom integriteit, authenticiteit en begrijpelijkheid, zoals deze centraal staan binnen digitale preservering, vaak niet voldoende.

Een van de redenen is dat fabrikanten dergelijke veiligheidsmaatregelen vaak 'onder water' laten draaien zodat de gebruiker van het systeem in de praktijk niet weet welke precieze veiligheidsmaatregelen er getroffen zijn om bijvoorbeeld een geslaagde kopieeractie te garanderen. In de praktijk gaat men er van uit dat gebruikers niet willen weten wat er tot in detail binnen een systeem gebeurt zolang het maar goed gaat.

Voor preserveringsdoeleinden is het echter juist uiterst belangrijk dat getroffen maatregelen bewezen kunnen worden, ook als er geen fouten optreden. Wanneer de kopieeractie wordt opgeslagen met timestamp en resultaat, gekoppeld aan het object, biedt dit aantoonbaar bewijs van authenticiteit.

---

[13] *Addis, Threats to data integrity from use of large-scale data management environments, https://prestoprimews.ina.fr/public/deliverables/PP_WP3_ID3.2.1_ThreatsMassStorage_R0_v1.00.pdf, blz. 4.*
[14] *Ibid.*
[15] *Ibid, blz. 5*

Een andere reden is dat fouten in software ook een rol spelen; storageoplossingen bestaan niet alleen uit hardware maar ook uit software aangezien de software in storage-management-systemen in de praktijk acties coördineert en locaties van objecten bijhoudt. De software zal er daarom ook voor moeten zorgen dat fouten gedetecteerd worden om het ontstaan van latente fouten te voorkomen. Een fabrikant kan echter niet garanderen dat er nooit fouten in de software zullen ontstaan en gekozen oplossingen zullen hoe dan ook nadelen met zich mee kunnen brengen. Voor preserveringsdoeleinden zijn garanties van de leverancier alleen daarom niet voldoende.

Tenslotte speelt de menselijke factor ook een rol; er kunnen fouten gemaakt worden bij het bedienen van de software, bijvoorbeeld wanneer logs niet worden nagekeken of er niet meteen actie wordt ondernomen bij foutmeldingen. Ook beleid speelt hier een belangrijke rol. Wanneer niet in beleid bepaald wordt dat integriteit een essentieel onderdeel van veilige storage is kan het bijvoorbeeld gebeuren dat bepaalde veiligheidsmaatregelen worden uitgezet in de configuratie omdat dit de doorloop sneller maakt. Het precies in kaart brengen van welke acties wel en niet uitgevoerd worden op de digitale objecten en de rollen en verantwoordelijkheden daarbij van het bedienend personeel, is dus noodzakelijk voor preservering omdat alleen dan bekeken kan worden of de gehele workflow in overeenstemming is met het preserveringsbeleid van de instelling.

Zeker wanneer er meerdere storageoplossingen naast elkaar gebruikt worden is het voor preservering belangrijk dat zaken als authenticiteit en integriteit van digitale objecten door de gehele keten heen gecontroleerd kunnen worden. Elke storageoplossing biedt wel een bepaalde vorm van controle tegen corruptie bijvoorbeeld door intern gebruik van Cyclic Redundancy Checks of een ander checksum-mechanisme maar dit kan niet aangetoond worden wanneer er geen checksuminformatie beschikbaar gemaakt kan worden vanaf het begin van de keten tot aan het eind van de keten, dus vanaf ingest tot aan access. Een ‘extern’ checksummechanisme dat niet gekoppeld is aan een specifieke software- of hardwarecomponent zorgt bovendien voor een grotere onafhankelijkheid van een bepaalde ict-oplossing zodat migratie naar een andere oplossing in de toekomst makkelijker is en het risico van verlies van integriteit vermindert.

Voor authenticiteit geldt hetzelfde; binnen het archief worden acties wellicht uitgevoerd volgens plan, maar zonder ‘audit trail’ van gedane acties per object kan dit achteraf niet meer bewezen worden. Ook de begrijpelijkheid moet meegenomen worden in het kader van storage; alleen monitoring van de Designated communities, de key-users van het archief, kan uitwijzen of de objecten nog voldoen in dit opzicht.

Wanneer storageoplossingen deze functionaliteit niet standaard ondersteunen zal de instelling dus extra maatregelen moeten nemen om dergelijke functionaliteit binnen de Enterprisearchitectuur in te voeren om te kunnen voldoen aan de eisen die worden gesteld aan de entiteit Storage binnen het OAIS-model.

STORAGE BINNEN BEELD EN GELUID

Om als referentie te kunnen dienen voor de rest van dit product zal eerst kort beschreven worden welke verschillende storagecomponenten momenteel bij Beeld en Geluid gebruikt worden.

Bij Beeld en Geluid zijn momenteel meerdere storagecomponenten gelijktijdig in gebruik die gedeeltelijk van elkaar verschillen in functionaliteit en in de manier waarop ze worden ingezet binnen de gehele architectuur. De storageoplossingen bedienen momenteel vooral verschillende typen materiaal en verschillende typen functionaliteit richting de eindgebruikers.

MXF- en *WAV*-bestanden worden opgeslagen binnen DIVArchive. Dit systeem is niet de storage zelf maar het managementsysteem dat opslag naar verschillende storagelocaties regelt. MXF en WAV bestanden worden de eerste twee weken opgeslagen op een disk cache om snelle uitlevering van recent materiaal te kunnen waarborgen. Dit gebeurt op een Isilon cluster dat bestaat uit een RAID 60 opstelling van harde schijven. Na twee weken zorgt DIVArchive voor migratie van de bestanden naar de StorageTek tapelibrary bij Ericsson en voor replicatie naar eenzelfde Tapelibrary op locatie bij Beeld en Geluid. De Tapelibrary bij Beeld en Geluid dient als een fail-over wanneer de Tapelibrary bij Ericsson zou uitvallen. DIVArchive zorgt tevens voor uitlevering naar Extranet zodat alle bestanden die binnen deze constructie worden opgeslagen ook direct besteld kunnen worden.

MPG1- en *MP3*-browsingkopieën worden opgeslagen op een vergelijkbaar Isilon cluster als eerdergenoemde cluster en worden gerepliceerd naar eerdergenoemde Isiloncluster voor fail-over-doeleinden. Een backup naar tape wordt ook gedaan voor alle browsingkopieën. Ondanks het feit dat deze bestanden lage-resolutie afgeleiden zijn van de masterbestanden wordt er toch voorzien in verschillende backupmaatregelen vanwege de investeringen in tijd en kosten die gedaan zijn om deze afgeleiden te maken. De software van Symantec netbackup wordt gebruikt voor de backup van MPG1-bestanden naar tape. Browsebestanden kunnen niet besteld worden maar alleen bekeken via Extranet.

DPX-bestanden worden opgeslagen door de software van de Stornext Hierarchical Storage Manager. Deze oplossing bevat een online disk cache voor tijdelijke opslag en zorgt voor archivering van de DPX-bestanden naar de nearline StorageTek-tapelibrary bij Beeld en Geluid evenals replicatie van de bestanden op offline tape die bewaard wordt op locatie in Rijswijk. DPX-bestanden kunnen niet rechtstreeks besteld worden. Van de DPX-bestanden wordt een 'mezzanine'-formaat gemaakt in de vorm van MXF die via extranet besteld kan worden. Deze vallen onder de 'MXF'-oplossing van DIVArchive.

TIFF-, *JPG*-, *STL*- en *PDF*-bestanden worden opgeslagen op de Sun storage. Dit is een cluster van disks in RAID-opstelling. Voor deze bestanden wordt geen replicatie verzorgd.

In de nabije toekomst zullen de DPX-bestanden die nu als TAR-archiefbestand opgeslagen zijn binnen Stornext gemigreerd worden naar DIVArchive 7 waarbij de TAR-indeling omgezet wordt naar het AXF-formaat dat DIVArchive 7 'out-of-the-box' ondersteunt. Zie bijlage A voor de details van wat dit betekent voor preservering.

Ook is het de bedoeling dat de sun-storage niet meer gebruikt wordt voor opslag van masterbestanden zoals de STL-bestanden maar dat deze gemigreerd worden naar het Isiloncluster waar ook de browsingkopieën staan.

GAP ANALYSIS: OAIS-FUNCTIES EN KWALITEITSEISEN BINNEN BEELD EN GELUID

In onderstaande deel wordt de gap analysis gepresenteerd. De huidige workflows binnen Beeld en Geluid worden vergeleken met de functionele eisen van het OAIS model en de Kwaliteitseisen Digitaal archief. De gap analysis wordt afgesloten met een tabel waaruit precies blijkt aan welke eisen voldaan moet worden binnen de toekomstige Enterprise-architectuur en aan welke eisen nu al wordt voldaan.

Receive data
Binnen Beeld en Geluid wordt de "Receive data"-functie voornamelijk uitgevoerd door de DDV voor uitzendmateriaal en door de DFI voor erfgoedcollecties en filmscanning.

Binnen Beeld en Geluid wordt prioriteit niet door middel van een indicatie per file aangegeven maar door middel van een policy waarbij alle nieuwe instroom van tv-uitzendingen twee weken op een snellere disk cache bewaard wordt alvorens naar tape gearchiveerd te worden omdat materiaal binnen twee weken na uitzending het meest wordt opgevraagd.

Vanaf het moment dat het bestand opgeslagen wordt op disk valt deze al onder de gemanagede storage van DIVArchive 6 en wordt er een checksum voor het bestand uitgerekend die binnen het systeem ook wordt gebruikt om te controleren of een kopieeractie geslaagd is. De tijd van ingest wordt hierbij opgeslagen. Via de database van DIVArchive is deze informatie weer op te halen.

Ericsson kopieert elke dag de metadata uit de DIVArchive database naar een eigen database zodat de informatie beschikbaar blijft aangezien de DIVArchive database maar een beperkt aantal events kan opslaan.

In de huidige workflow wordt een bestand bij ingest niet gecontroleerd tegen een meegeleverde checksum. Wel wordt de file standaard geanalyseerd om te zien of deze voldoet aan de specificaties van het file formaat. Hierbij wordt gecontroleerd of het bestand een correcte header, footer en correcte data-compartimenten heeft. Als dit niet het geval is wordt de instroom afgebroken.

Ook wordt er een tweede check gedaan in het geval van MXF waarbij gekeken wordt of het bestand aan de afgesproken specificaties vanuit de Ketenafspraken voldoet.
De uitslag van dit onderzoek kan de instroom niet tegenhouden maar het rapport wordt wel opgeslagen en bewaard bij Ericsson. Wanneer het bestand is opgeslagen wordt hiervan geen confirmatie afgegeven aan de producer.

Conclusie

- Resultaten van checks, errors en ingest-events worden dus wel gegenereerd en opgeslagen in verschillende systemen maar worden niet beschikbaar gesteld als preserveringsmetadata. Hierdoor is er eigenlijk veel informatie beschikbaar zonder dat hier gebruik van wordt gemaakt voor preserveringsdoeleinden. Hierdoor wordt het moment waarop verantwoordelijkheid voor preservering is aangenomen ook niet gecommuniceerd.
- Er zijn controlemechanismes voor garantie van correcte aflevering, echter niet door middel van checksum
- Er wordt geen confirmatie aan de producer afgegeven van correcte opslag.

Manage storage hierarchy
In het geval van Beeld en Geluid wordt deze functie grotendeels vervuld door de twee Storage managementsystemen Stornext en Diva.

Binnen Stornext wordt bijvoorbeeld in het systeem gelogd of files zijn gearchiveerd naar twee verschillende tapes voordat de file van de disk cache wordt verwijderd. Diva kent de policy om materiaal na twee weken naar tape te schrijven op twee locaties en het materiaal hierna van de disk cache te verwijderen.

Binnen DIVArchive 7 worden alle kopieeracties gecontroleerd op fixity[16] en wordt de tijd van de verplaatsing opgeslagen als provenance data. Doordat de bestaande provenancedata hierbij ook wordt meegenomen blijft de oorspronkelijke ingesttijd ook bewaard.

Aangezien deze functie ook verantwoordelijk is voor het nakijken van binnengekomen error-rapportages is het echter niet mogelijk om deze functie geheel door een systeem te laten uitvoeren omdat uiteindelijk mensen de logs moeten nalopen en de benodigde acties uitvoeren op basis van workflows vastgelegd in beleid.

In de huidige praktijk worden errors gecontroleerd en verholpen door Ericsson, bijvoorbeeld wanneer een tape niet goed blijkt te zijn. DIVArchive kan bijvoorbeeld constateren dat er teveel leesfouten optreden op een bepaalde tape of dat een bepaald bestand helemaal niet leesbaar is. Ook kan het zijn dat de tape fysiek stuk gaat. In dat geval worden de benodigde bestanden vanaf de backup-tape weer naar een nieuwe tape gekopieerd.

Bij kopieeracties van disk naar tape en bij het maken van een offsite backup worden in de huidige workflow door DIVArchive de intern aangemaakte checksums gebruikt om de fixity te controleren. De checksum van de reservekopie wordt dus uitgerekend. Wanneer deze overeenstemt met de originale checksum is de kopieeractie geslaagd.
De verplaatsingsactie en de nieuwe locatie worden gelogd. Deze informatie wordt door Ericsson in een eigen database opgeslagen.

Conclusie

- Op basis van policies wordt materiaal op verschillende storage media opgeslagen, dit beleid is echter niet gedocumenteerd.
- Verplaatsingen tussen storage media worden gecontroleerd om zeker te stellen dat het bestand niet corrupt is geraakt.
- Errormeldingen worden binnen deze functie bekeken en afgehandeld.
- Er zijn echter geen procedures hiervoor gedocumenteerd.
- Ook zijn er geen vaste procedures gedocumenteerd voor het opstellen van rapportages over errors en performance van het systeem.

[16] *Hierbij wordt een checksum uitgerekend van het nieuwe bestand die wordt vergeleken met de checksum van het oude bestand om zeker te kunnen zijn dat het bestand niet veranderd is tijdens de kopieeractie.*

Replace media
Binnen Beeld en Geluid zou deze functie bijvoorbeeld verantwoordelijk zijn voor het repliceren van archiefobjecten van de ene taperobot naar de andere taperobot of voor het migreren van objecten van LTO-4 naar LTO-5.

Ook deze functie wordt momenteel binnen Beeld en Geluid grotendeels vervuld door de twee Storage managementsystemen. In DIVArchive worden intern checksums gebruikt om te controleren of kopieeracties geslaagd zijn. Deze checksums zijn opgeslagen binnen de data management-structuur van Diva. Kopieeracties die via Diva verlopen worden ook geëxporteerd uit de logs en opgeslagen binnen een eigen databasesysteem van Ericsson. Bij de recentste migratie van LTO-4 naar LTO-5 zijn de interne checksums *echter niet gebruikt*.

Conclusie

- Er zijn voor deze functie dus tools beschikbaar om migratieacties uit te kunnen voeren.
- Er is van voorgaande migratieactie echter niet direct provenance-data beschikbaar hoewel deze herleid kan worden door te kijken naar de aanmaakdatum van de bestanden op de nieuwe tapes.
- Opgeslagen events uit deze functie zijn momenteel voor Beeld en Geluid niet toegankelijk als preserveringsmetadata.

Error checking
Deze functie wordt vervuld door de logging-functionaliteit van de twee storage-management-systemen DIVArchive en Stornext.

Binnen DIVArchive wordt bjivoorbeeld gelogd wanneer er een kopieerfout optreedt. Deze informatie wordt momenteel ook geëxporteerd uit de database van Diva en opgeslagen in een database bij Ericsson zodat deze informatie bewaard blijft.

Conclusie

- Er worden binnen DIVArchive een aantal zaken gelogd, zoals kopieerfouten.
- Deze data is echter niet direct beschikbaar voor Beeld en Geluid en is ook niet opgeslagen als preserveringsmetadata.
- Ook kan de huidige oplossing niet geheel zorgen voor de benodigde monitoring van bit-integriteit aangezien deze alleen gecheckt kan worden tijdens een kopieeractie.
- Logs worden opgeslagen, checksums worden aangemaakt na ingest maar niet gemanaged als essentiële metadata die bij het object hoort.

Disaster recovery
De werkwijze zoals deze gebeurt in het geval van de DPX-bestanden bij Beeld en Geluid, waarbij bestanden worden gedupliceerd op offline tape bewaard in een off-site-locatie, is een invulling van deze functie. De MXF en WAV bestanden worden ook gedupliceerd van de locatie bij Ericsson naar de locatie bij Beeld en Geluid.

Conclusie

- Binnen de huidige infrastructuur is er dus controle over het aantal backups en is een offsite backup geregeld voor alle master-AV-bestanden dus voor de DPX-, de MXF- en de WAV-bestande,
- Dit geldt echter niet voor alle additionele bestanden zoals de STL-bestanden.
- De backups worden bij creatie gecontroleerd op integriteit door middel van checksum binnen DIVArchive.
- Deze informatie is echter niet beschikbaar als preserveringsmetadata.
- Er is geen recovery-plan, op locatie of off-site, beschikbaar.

Access data
Binnen Beeld en Geluid wordt deze functie vervuld door DIVA.

Op basis van ID het juiste mediabestand is opgezocht op tape en deze kopieert naar het Isilon-cluster voor verdere verwerking, bijvoorbeeld upload naar een ftp-locatie.

Conclusie

- Na aflevering volgt er een vorm van terugmelding waarbij een ander proces vervolgens zorgt voor doorvoer naar eventuele FTP-locaties.
- Ook backup-versies van de bestanden zijn op te halen en zijn in het verleden ook gebruikt om beschadigde bestanden te restoren.
- De access-requests zijn echter niet als preserveringsmetadata beschikbaar.

ALGEMEEN CONCLUSIES EN TEKORTKOMINGEN

Om de functies van de entiteit storage te kunnen vervullen op een manier die voldoet aan de eisen van het OAIS-model, zijn een aantal acties noodzakelijk die nu niet gebeuren of waar geen bewijs van getoond kan worden.

Dit vereist in de toekomst aanvullend werk om deze mogelijk te maken binnen de Enterprise architectuur, zowel in de vorm van beleid en vastgestelde workflows als ook door middel van technische maatregelen en functionaliteit.

Onderstaande tabel geeft weer welke functies en eisen vanuit OAIS momenteel wel en niet gebeuren in de huidige workflow, ook wordt er aangegeven of de eis vervuld kan worden door DIVArchive6, DIVArchive 7 of een eventuele andere tool.

OAIS functies en eisen	Mogelijk in DIVArchive 6	Mogelijk in DIVArchive 7	Andere tool	Gap in huidige workflow	Opmerking
Receive data					
Fixitycheck met meegeleverde checksum	ja	ja		ja	
Message digest calculation na ingest	ja	ja		nee	
Formatcontrole voor ingest	nee	nee	"Fast check" Ericsson	nee	
Compatibiliteitscontrole voor ingest	nee	nee	"MXF check" Ericsson	nee	
Controlerapport opslaan	nee	nee	database Ericsson	nee	
Tijd van ingest loggen	ja	ja	database Ericsson	nee	
Confirmatie geslaagde ingest aan producer	nee	nee	MAM?	ja	
Controle, fixity en events opslaan als preserveringsmetadata	nee	niet volledig	MAM?	ja	
Manage storage hierarchy					
Policies gedocumenteerd	nvt	nvt		ja	Niet als beleidsdocument beschikbaar
HW/SW garanties/support gedocumenteerd	nvt	nvt		nee	Beschikbaar in Topdesk, moet gecontroleerd worden hoe volledig
AIP-verplaatsing op basis van policy	ja	ja		nee	
Fixity check tussen verplaatsingen	ja	ja		nee	Wordt net als de verplaatsing zelf niet opgeslagen als preserveringsmetadata
Verplaatsing loggen	ja	ja		nee	
Verplaatsing opslaan	niet permanent	ja	database Ericsson	nee	
Verplaatsing als preserveringsmetadata	nee	niet volledig	MAM?	ja	
Check voldoende ruimte	ja	ja		nee	
Oplossen kopieerfouten	nvt	nvt		nee	Wordt opgelost in workflow

Rapportage afhandeling error checking	nvt	nvt		ja	Moet opgelost in workflow
Replace media					
Migratieactie uitvoeren	ja	ja		nee	
Actie loggen	ja	ja		nee	
Actie opslaan	niet permanent	ja		ja	data te herleiden op basis van ingestdatum
Fixity check tijdens migratie	ja	ja		ja	Is alleen niet gebruikt bij laatste migratieactie
Actie opslaan als preserveringsmetadata	nee	niet volledig	MAM?	ja	
Error checking					
Foutmelding bij kopieerfout	ja	ja		nee	Standaardfunctionaliteit
Controle bit integriteit tijdens storage	nee	nee	MAM?	ja	
Controle tapekwaliteit	ja*	ja		ja	*Functionaliteit vereist aanschaf extra pakket, nu niet ingebouwd in de workflow
Controle leesfouten	ja	ja		nee	
Controle metadata	nvt	nvt		ja	
Rapportage tonen voor controle	ja	ja		nee	
Rapportage opslaan	niet permanent	?	database Ericsson	nee	
Errors opslaan als preserveringsmetadata	nee	nee		ja	
Disaster Recovery					
Backup beleid gedocumenteerd	nvt	nvt		ja	Niet als beleidsdocument beschikbaar
Controle hoeveelheid backups	ja	ja		nee	
Controle fixity backup-creatie	ja	ja		nee	Gebeurt binnen DIVArchive
Controle synchronisatie	nee	nee		ja	
Off-site backup	ja	ja		nee	
Recovery plan	nvt	nvt		ja	Bestaat nu niet
Recovery plan off-site backup	nvt	nvt		ja	Bestaat nu niet
Recovery test	nvt	nvt		ja	Gebeurt nu niet
Access data					
Objecten localiseren	ja	ja		nee	
Objecten ophalen	ja	ja		nee	
Backups ophalen	ja	ja		nee	
Confirmatie van aflevering	ja	ja		nee	
Access gelogd	ja	ja		nee	
Access opgeslagen	niet permanent	niet permanent	database Ericsson	nee	
Access opgeslagen als preserveringsmetadata	nee	nee		ja	

TEKORTKOMINGEN

Uit bovenstaande tabel blijken enkele belangrijke tekortkomingen:

Preserveringsmetadata
Er wordt momenteel veel provenance-informatie gelogd en opgeslagen binnen verschillende systemen.

- Deze data is echter niet centraal toegankelijk gemaakt als preserveringsmetadata
- Deze data is niet gemapt naar het formaat zoals voorgesteld binnen de Preservation Metadata Dictionary
- Het belang van deze data voor preserveringsdoeleinden is vanuit het archiefbeleid niet kenbaar gemaakt aan partijen die deze data genereren en opslaan.
- Ditzelfde geldt voor error-meldingen en technischeanalyserapporten vanuit ingest.

Er is dus eigenlijk veel informatie beschikbaar die heel waardevol is voor preserveringsdoeleinden maar doordat deze niet gemapt is naar preserveringsmetadata en niet centraal opgeslagen en doorzoekbaar is gemaakt kan deze nog niet gebruikt worden voor authenticeitsgaranties naar producers en gebruikers toe.

Fixity
Binnen de verschillende workflows worden momenteel checksums gegenereerd en ook gebruikt om kopieeracties te verifiëren voor zover deze kopieeracties binnen het systeem plaatsvinden.

- Er is momenteel echter nog geen end-to-end-oplossing waarbij producers een checksum meeleveren voor ingest en deze checksum door de gehele keten wordt gebruikt om aan te tonen dat het bestand niet veranderd is sinds de aanlevering.
- Ook is het momenteel niet mogelijk om fixity te controleren tijdens opslag, alleen bij kopieeracties.

Beleid
Om te kunnen bewijzen dat workflows binnen het archief gestandaardiseerd en consistent zijn is het nodig om documentatie te hebben die ook daarwerkelijk gevolgd wordt binnen de verschillende workflows rondom storage.

- Uit de Gap-analysis blijkt duidelijk dat er momenteel geen Disaster recovery plan is.
- Ook beleid rondom afhandeling en rapportage van errors, storage hiërarchie en het maken van backups is niet gedocumenteerd.

BIJLAGEN

BIJLAGE A: DIVARCHIVE 7 BINNEN EEN OAIS-WORKFLOW[17]

Het Front Porch DIVArchive 6.0 Content Storage Managementsysteem vormt een bestaand onderdeel van de huidige IT Enterprise architectuur van Beeld en Geluid. In het kader van het Next Archive-programma loopt momenteel een requirementstraject waarin de vereisten voor een nieuw MAM-systeem worden geformuleerd. Dit MAM-systeem gaat onderdeel uitmaken van de systeemarchitectuur. Tegelijkertijd worden er vanuit het TDR-project een normatief informatiemodel en workflow gedefinieerd, gebaseerd op OAIS-compliant-businessprocessen. Overeenstemming met de eisen van OAIS, met het oogmerk de status van TDR (Trusted Digital Repository) te bereiken, vormt een van de doelstellingen uit het Meerjarenplan 2012-2015 van Beeld en Geluid. TDR-businessprocessen zullen moeten worden ondersteund door meerdere applicaties binnen de Enterprise architectuur.

De requirements die uit het TDR project zijn gekomen in het MAM RFI-traject hebben vragen opgeleverd omtrent de mate waarin DIVArchive de vereiste OAIS compliant workflow ondersteunt. Front Porch zelf claimt dat zijn DIVArchive en de AXF formats inderdaad OAIS-compliant zijn : “Onze AXF V1 implementatie bevat zgn. “Provenance” en “Fixity” velden. Provenance wordt voor iedere verplaatsing/beweging binnen het archief bijgehouden door DIVArchive, d.w.z. door de DIVArchive Actor, en bevat een lijst records met de geschiedenis van de verplaatsingen van de betreffende asset . De inhoud van dit AXF veld kan worden bekeken met het AXF reader programma.”[18] DIVArchive vormt ook na aanschaf van het nieuwe MAM-systeem een vast gegeven binnen de Enterprise architectuur van Beeld en Geluid. De storageapplicatie moet dus aansluiten bij de OAIS-vereisten. Om deze reden is aan medewerkers van het TDR projectteam Information Packages en Preservation Metadata (IPS/PMD) gevraagd om vooronderzoek te doen naar de vraag in hoeverre DIVArchive 7 inderdaad is in te zetten is binnen een OAIS-workflow.

Functionaliteit

Vanuit de requirements die opgesteld zijn vanuit TDR in het kader van het RFI is een aantal vragen gesteld aan medewerkers van Front Porch met als doel te achterhalen in hoeverre DIVArchive 7 de vereiste functionaliteit kan bieden[19]. Onderstaande informatie is het voorlopige resultaat hiervan.

Integriteit

- DIVArchive 7 is in staat om bestanden bij ingest te controleren op integriteit aan de hand van meegeleverde checksums mits deze voldoen aan bepaalde door DIVArchive 7 gestelde eisen.
- Als een checksum niet is meegeleverd wordt deze door DIVArchive uitgerekend en opgeslagen in de AXF als metadata behorende bij de file. Deze kan vervolgens geëxporteerd worden via de API.
- Bij elke kopieeractie wordt de checksum gebruikt om te controleren of de kopieeractie geslaagd is.
- Buiten kopieeracties is er geen fixity checking, alleen de mogelijkheid om handmatig de tape te laten controleren, echter zonder dat de uitkomst hiervan wordt opgeslagen in de metadata.

---

[17] *Dit onderzoek was 25 Maart 2013 door Daniel Steinmeier afgerond.*
[18] *Email Phillip Maher, 6 December 2012.*
[19] *DIVArchive v.7 omdat we ervan uitgaan dat Beeld en Geluid de huidige versie zal upgraden.*

Preserveringsmetadata

- Wat binnen DIVArchive wordt beschouwd als preserveringsmetadata is een zeer beperkte set aan metadata die aangeduid wordt als 'provenance data'.
- Provenance data wordt alleen gegenereerd bij een schrijfactie en behelst dus niet de complete lijst aan events zoals deze binnen de Preservation Metadata Dictionary beschreven is.
- De velden binnen het metadataschema zijn niet zo uitgebreid als binnen een formaat als PREMIS vereist is voor provenancemetadata.
- Bij elke kopieeractie wordt de provenancedata van het object bijgewerkt. Dit betekent dus dat de kopie zowel de provenancedata van het origineel als van de kopie bevat.
- Behalve bij schrijfacties wordt de provenancedata niet bijgewerkt.
- De metadata is te exporteren door middel van de API op DIVArchive 7.
- Fouten bij kopiëren of uitleveren zijn alleen te zien en te exporteren vanuit de monitoringtool en worden niet opgeslagen in de metadata.
- De locatie van alle kopieën wordt opgeslagen en de kopieën zijn individueel op te vragen.
- Alle objecten krijgen een eigen persistente ID (UUID).
- De preserveringsmetadata zoals deze standaard binnen het systeem wordt toegevoegd aan de AXF kan als volgt weergegeven worden:
    - Provenance Collection
        - Provenancedata (herhaalbaar)
            - Application
                - Application Name
                - Version
                - Description
                - Licensor
                - Licensee
                - Serial Number
            - Source
                - Manufacturer
                - Make
                - Model
                - Firmware
                - Description
                - UUID
                - Label
                - OS
                - Root path
                - Location
            - Destination (zelfde velden als source)
            - Object Owner
                - Name
                - Facility
                - Description
                - Operator
    - File tree (directorystructuur)
        - File (herhaalbaar)
            - Filename
            - File ID
            - Size
            - Position
            - Checksum

Conclusie
Uit bovenstaande blijkt dat, hoewel DIVArchive 7 ingezet kan worden om aan een aantal eisen binnen de OAIS-workflow te voldoen, het geen complete oplossing kan bieden voor de gehele workflow.

Punten waar DIVArchive overeenkomt met de in het RFI gestelde eisen:

- Het is mogelijk om te controleren of een bestand goed is ge-ingest door middel van een meegeleverde checksum.
- Het is ook mogelijk om een checksum te genereren en te exporteren wanneer een checksum niet is meegeleverd.
- DIVArchive 7 kan gebruikt worden voor het maken van kopieën.
- Kopiën zijn individueel opvraagbaar en uit de metadata zijn locatiegegevens en persistent ID's te halen.
- Er wordt gecontroleerd of een backup zonder fouten is aangemaakt.

Punten waar DIVArchive niet overeenkomt met de in het RFI gestelde eisen:

- De informatie over de controle van de backup wordt niet als Event opgeslagen in de metadata. Dit is nodig om te bewijzen dat deze controle heeft plaatsgevonden. Wellicht is deze informatie nog wel te exporteren uit de monitoringtool.
- De eis om regelmatig te controleren op bitcorruptie is moeilijk te vervullen aangezien DIVArchive 7 geen tussentijdse check kan doen van losse objecten. Wellicht zou vanuit OAIS-standpunt het checken na migratie alleen, en niet tussendoor, voldoende kunnen zijn wanneer dit goed gemotiveerd is, bijvoorbeeld omdat dit extra slijtage voorkomt.
- De eis om de oorspronkelijke file te restoren in geval van bit corruptie kan daarmee ook niet vervuld worden.
- Het opslaan van alle gevallen van bitcorruptie gekoppeld aan het object is niet direct mogelijk, dat wil zeggen de koppeling wordt wellicht inde monitoringtool gemaakt bijvoorbeeld als de ID van het object vermeld staat bij een eventuele foutmelding.
- Errors die optreden bij access en kopiëren kunnen niet direct gekoppeld aan het object worden opgeslagen.
- Versiebeheer is binnen DIVArchive 7 niet aan de orde doordat het alleen mogelijk is om een AXF te updaten door een nieuwe AXF apart op te slaan of door de oud versie te verwijderen en te vervangen voor een nieuwe versie.
- Tenslotte lijkt het niet realistisch om te verwachten dat preserveringsmetadata binnen de AXF opgeslagen kan worden. Aangezien metadata alleen bijgewerkt kan worden door de hele AXF opnieuw te schrijven is dit geen oplossing voor preserveringsmetadata die vaak geupdate moet kunnen worden om de lijst met events bij te werken. Dezelfde overweging maakt ook dat het waarschijnlijk niet werkbaar is om hiërarchische structuren (zoals schone inlassen van een journaal) binnen dezelfde AXF op te slaan wanneer deze niet altijd tegelijkertijd worden aangeleverd.

Mogelijke oplossingen
Voor de punten waar DIVArchive niet aan kan voldoen zal een oplossing gezocht moeten worden binnen de Enterprise architectuur.

Om genoeg preserveringsmetadata gelinkt aan het object te kunnen opslaan zal data uit de monitoringstool en uit de provenancedata van DIVArchive geëxporteerd moeten worden en opgeslagen worden op een andere locatie, bijvoorbeeld als XML in een XML-database, zodat de metadata ook doorzoekbaar is. Het is daarbij nog wel de vraag hoeveel informatie uit de monitoringstool te halen is.

De link tussen metadata en object zal dan bestaan uit de identifier die DIVArchive 7 aan het object meegeeft en niet uit het opslaan van metadata en essence binnen dezelfde container. Wanneer gekozen wordt voor een metadataformaat dat dit ondersteunt, zoals METS, kunnen hiërarchische relaties en versieinformatie ook opgeslagen worden in de metadata.

Ook zal dit het mogelijk maken om errors en uitzonderingen blijvend op te slaan als preserveringsmetadata.
Er zal dan een proces ingericht moeten worden dat verschillende soorten data bij elkaar kan verzamelen en deze kan opslaan als preserveringsmetadata.

Vervolgonderzoek
De voorlopige conclusies uit dit onderzoek maken duidelijk dat er nog een aantal vragen open staan, met name omtrent de informatie die uit de monitoringstool te halen is. Voor de benodigde preserveringsmetadata zoals beschreven in de Preservation Metadata Dictionary 1.0 zou precies bepaald moeten worden welke data uit Diva gehaald zou kunnen worden en welke data door andere applicaties, zoals bijvoorbeeld het MAM of aparte file analyser software, geleverd moet worden. In de aanbeveling voor metadata-standaarden zou vervolgens moeten worden uitgewerkt hoe deze informatie op te slaan en te structureren.

In een vervolgonderzoek zouden tevens aanvullende vragen beantwoord kunnen worden die een completer beeld van de inzet van DIVArchive binnen een OAIS-workflow zouden kunnen geven. Vragen die nog gesteld kunnen worden zijn onder andere:

Wat betekent technisch gezien de actie 'verify tape'?
Dit om te achterhalen of het hier een controle van alle objecten op de tape betreft op slechts een check van de tape zelf en om te bepalen of dit voldoende zekerheid zou geven over de integriteit van de objecten zoals bedoeld binnen OAIS.

Hoe is provenanceinformatie nu opgeslagen in versie 6.0?
Dit is de versie die momenteel nog draait bij Beeld en Geluid en die voor zover ik weet nog geen ondersteuning biedt voor AXF. Het zou interessant zijn om te weten of er desondanks bij de afgelopen tapemigratie ook al provenanceinformatie is bijgehouden die meegenomen zou kunnen worden in de toekomst en of zaken als UUID's en locaties ook momenteel al te exporteren zouden zijn van alle huidige objecten binnen DIVArchive.

Hoe gebruiken andere archieven die DIVArchive 7 hebben afgenomen, zoals Det Danske Filminstitut en VRT, AXF binnen hun preserveringsworkflow?
Het zou bijvoorbeeld interessant zijn om te bekijken welke bestanden andere archieven opslaan in de AXF, of ze ook preserveringsmetadata in de AXF opslaan en of ze nog aanvullende technieken gebruiken om extra provenanceinformatie en technische metadata te verzamelen.

Hoe lang wordt monitoringinformatie bewaard in de database, wat wordt er gelogged en is dit per object op te vragen?
Dit om te bepalen of de monitoringtool op zichzelf voldoende biedt om de eis vanuit Trac te ondersteunen dat alle errors en access-events gekoppeld aan het object worden opgeslagen en om verder te onderzoeken hoe deze informatie is te exporteren en te mappen naar preserveringsmetadata.

Q&A DIVArchive 7-OAIS-onderzoek: Bijlage A

Deze bijlage is als volgt gestructureerd: in het kader is de eis uit het RFI weergegeven (links) en de TRAC-eis (rechts) waar deze op gebaseerd is en die betrekking heeft op functionaliteit die DIVArchive zou moeten vervullen. Vervolgens is in cursief een of meerdere vragen weergegeven die gesteld zijn aan de medewerkers van Front Porch en daaropvolgend de antwoorden die we van hen hebben ontvangen.

Integriteit

11. The system must be able to check the integrity of a file by comparing a newly generated checksum against an externally computed checksum delivered with the file.	4.1.5 The repository shall have an ingest process which verifies each SIP for completeness and correctness. Supporting Text: This is necessary in order to detect and correct errors in the SIP when created and potential transmission errors between the depositor and the repository.

Is DIVArchive 7 able to do fixity checks on externally computed checksums? For instance if a checksum is delivered along with the file on ingest?
Yes we support checksum manifest or Genuine Checksum (A Genuine Checksum is a checksum retrieved by the Actor from a DIVArchive Source). In forthcoming version 7.1, we support multiple modes to get the checksum calculated externally.

A Genuine Checksum Source must be configured in order for the system to read the Checksum from the external source (e.g. SAMMAsolo, external MAM system) providing the file. This initiates Actor on-the-fly checksum calculation to compare the checksums provided by the external source and checksums calculated.

DIVArchive V7.1 supports now three different of Genuine Checksum modes:

- 'MDF_XML' allows DIVArchive to compare checksum provided in an external XML file
- 'TEXT' allows DIVArchive to archive all files and subfolders in a specified folder while comparing their checksum values against known values stored in an external checksum text file.
- 'AXF' allows DIVArchive to validate the integrity of every file within an AXF package. The Actor compares calculated checksums with checksum values (metadata) present in every AXF package.

Fixity checking en preserveringsmetadata

12. The system must be able to generate and save checksums as preservation metadata for ingested files.	"4.4.1.2 The repository shall actively monitor the integrity of AIPs." Examples: "Fixity information (e.g., checksums) for each ingested digital object/AIP; logs of fixity checks; documentation of how AIPs and Fixity information are kept separate;."

Is fixity checking configurable? If so, how often could this approximately be done per file considering the scale of our collection?
Automatic/scheduled re-checking of object/tapes is a DIVArchive roadmap item.
For the time being, you can perform a "Verify Tape" from the control GUI but this is an operator-driven operation.

Is checksum always calculated on ingest if not delivered with the file?
If the application that deals with AXF objects has been configured to create/validate checksums of course. For DIVArchive, yes.

Is the time of ingestion logged as provenance?
Yes. Each provenance record is associated to a timestamp. So the first provenance record is the one that describes how the first object instance has been created

Are the checksums for all files exportable?
Yes checksums can be retrieved via the API (C++ and Java) and soon through the Web Services API.
Tools can be developed to export them using the API call.

Is technical metadata extracted and saved within the AXF?
Metadata stored by DIVArchive in AXF are:

- Object name, object category
- Object UUID
- A set of Provenance records (see above)
- A set of (File name, file size, file checksum) for each file stored within the AXF container as well as the directory structure of the object (you can have nested directories within an AXF object)

Backup-functionaliteit

24. The system must be able to retrieve all copies.	"5.1.2 The repository shall manage the number and location of copies of all digital objects."

Are backup copies individually retrievable?
Yes. With DIVArchive you can retrieve any object instance.

Do they have individual or shared provenance data?
Each object instance has its own set of provenance records. For example, imagine you archive a set of digitized files within DIVArchive. This will create a first disk object instance (ID =0) with one provenance record.
Then your SPM may automatically copy that disk instance to a tape => that will create a second object instance (ID=1) with two provenance records (the initial one + the new provenance record created by the copy operation.

Is location information included, for example if the file is on tape or on disk?
Yes. See above.

How is the provenance data updated when the data is stored on tape?
Provenance records are created/added only when we create a first tape instance (i.e Archive to tape) or when we create additional tape instances (i.e. Copy to tape)

25. The system must be able to provide evidence of the integrity of the backups by documenting all integrity test outcomes in the preservation metadata.	"5.1.2.1 The repository shall have mechanisms in place to ensure any/multiple copies of digital objects are synchronized."

Is the creating of backup copies logged?
Yes. Each time a new AXF object instance is created (e.g. you copy an object from a DIVArchive AXF managed disk to a DIVArchive AXF tape group), this will add a new provenance record => you will the full history within the object itself.

Uit de presentatie:
Each time a data mover (e.g. DIVArchive Actor) moves or replicates an AXF object, AXF Object checksum values (one per file and one per structure) are validated!

Bit corruption

26. The system must be able to detect bit corruption.	"4.4.1.2 The repository shall actively monitor the integrity of AIPs." "5.1.1.3 The repository shall have effective mechanisms to detect bit corruption or loss."

Is checking for corruption logged?
Checking (only) for corruption is possible within DIVArchive (you have a "Verify Tape" command) but we do not log this checking WITHIN the AXF object itself when this verification occurred. It would be REALLY tough from a resource point of view (tape drives, Actors) to do that since we would have to delete the current instance and re-write this instance (with the updated AXF metadata – object instance checked timestamp)

Is fixity checking configurable? If so, how often could this approximately be done per file considering the scale of our collection?
Automatic/scheduled re-checking of object/tapes is a DIVArchive roadmap item.
For the time being, you can perform a "Verify Tape" from the control GUI but this is an operator-driven operation.

27. The system must provide mechanisms for restoring the original file in the case of detected file corruption.	"5.1.1.3.1 The repository shall record and report to its administration all incidents of data corruption or loss, and steps shall be taken to repair/replace corrupt or lost data."

Logging

28. The system must be able to generate status reports on all incidents of bit corruption and associated restore actions and save these as Events in the preservation metadata.	"5.1.1.3.1 The repository shall record and report to its administration all incidents of data corruption or loss, and steps shall be taken to repair/replace corrupt or lost data."

Can logged events be exported as preservation metadata?
The AXF provenance field is updated with all actions impacting the object (original source, tape IDs, disk instances, …).
[…] Otherwise DIVArchive stores all system events in the database and these can be exported.

What information is logged for each event?

- Which application has created this instance (e.g. DIVArchive V7.1)
- Where those files have been pulled from (e.g. DIVArchive managed disk path)
- Where this object has been stored to (e.g. Tape barcode) and which storage device has written this tape (Tape drive serial number)

29. The system must be able to generate status overviews of all access errors and exceptions occurring within the system at all times.	"4.6.1.1 The repository shall log and review all access management failures and anomalies"

Are read-actions to the file logged?
No. As mentioned above It would be REALLY tough from a resource point of view (tape drives, Actors) to do that since we would have to delete the current instance and re-write this instance (with the updated AXF metadata – object instance read timestamp)

Are access and storage errors logged and linked to the file?
Same answer. Those information are stored at the application dealing with AXF objects (e.g. DIVArchive) but not logged in the AXF objects (e.g. if we have problem accessing a storage where AXF object are stored to, how to log this information within the objects?)

If so, can this information be exported and for how long is this information available?
See above. Information not logged => not exported. But of course you have this information within DIVArchive database (e.g. DIVAprotect metrics) and those metrics can be exported.

Hiërarchische relaties en versiebeheer

41. The system must be able to support hierarchical relations between files, for instance linking an MXF with an additional subtitle file.	"4.2.5 The repository shall have access to necessary tools and resources to provide authoritative Representation Information for all of the digital objects it contains." Discussion […] Hierarchical schemes of description can allow some descriptive elements to be associated with many items.

Are different files that belong to the same intellectual entity stored in the same AXF, for example a news broadcast with additional clean footage?
This is entirely dependent on the application dealing with AXF objects. For DIVArchive, this is the upper layer application (e.g. MAM, newsroom application) that defines which files will be part of an object. Some applications store only A/V material, some others store as well meta-data (e.g. XML) within the object…

Can a file be added to the AXF later on? If so, is versioning information part of the metadata?
This is entirely dependent on the application dealing with AXF objects. For DIVArchive, you cannot add a file to an existing object. You need to restore that object, add the file (e.g. add a new audio track) and re-archive this new object (either with another name/category… or if you want to keep the same name/category you need to delete first the existing object before re-archiving it).

BIJLAGE B: GESPREKSVERSLAG MATTHEW ADDIS 14-03-2013[20]

Achtergrond
In de besprekingen rondom de TDR-eisen binnen het RFI-traject voor een nieuw MAM-systeem kwamen een aantal vragen naar boven die betrekking hadden op de noodzaak van sommige eisen en de precieze praktische uitwerking hiervan. Hieruit ontstond de behoefte deze vragen voor te leggen aan een externe partij met ervaring op het gebied van OAIS-workflows. De vragen zijn opgenomen in het document RFI-TDR v4.2. Deze zijn uiteindelijk voorgelegd aan Matthew Addis in een skype-sessie d.d. 14-03-2013. Matthew Addis is CTO van Arkivum, een bedrijf dat een cloud-oplossing aanbiedt voor storage van digitale objecten met 100% data-integriteit-garanties. Als zodanig is hij bekend met de TRAC-eisen en het bedrijf is gecertificeerd voor ISO 27001 dat de basis vormde voor een aantal eisen uit de ISO 16363-standaard 'Audit and certification for Trusted Digital Repositories'.

Risk Assessment
Het eerste punt dat Matthew behandelt betreft Risk Assessment. Hij waarschuwt ervoor in het kader van het MAM-traject om niet te afhankelijk te worden van externe partijen.
Beeld en Geluid als geheel moet kunnen voldoen aan de eisen van TRAC, daarvoor moet je niet afhankelijk zijn van een specifieke vendor. Het gaat erom dat wij continuity of service kunnen garanderen.

De aanpak van Arkivum is om eerst een Risk Assessment te doen zodat je weet wat je grootste/belangrijkste risico's zijn, wat je zelf moet doen en wat de leverancier moet doen om deze te voorkomen of op te lossen. Als je weet welke risico's je wil voorkomen of beperken kun je kijken of bepaalde standaarden (ISO 27001, TRAC) voorwaarden noemen die je zou kunnen implementeren. Details van implementatie zijn dan duidelijker in te vullen, want je weet precies welk risico je aanpakt.

Metadata bij object of in apart systeem?
Op basis van eis 10/38 uit het RFI-document, namelijk dat het systeem technische informatie moet kunnen extraheren en opslaan als een vastgestelde set preserveringsmetadata, kwam de vraag naar voren of het in overeenstemming is met TRAC om verschillende delen van de preserveringsmetadata op te slaan in verschillende systemen.

Dit om te bepalen of informatie door het MAM bij elkaar verzameld moet worden uit verschillende systemen, waaronder DIVA, of dat deze informatie ook verspreid over meerdere systemen opgeslagen mag zijn, zolang als het maar ergens opgeslagen is.

Arkivum gebruikt voor opslag van de lifecycle-metadata, de 'audit trail', een Cassandra Database[21], gedistribueerd over meerdere locaties en dubbel uitgevoerd. De digitale objecten worden op data tape opgeslagen. Objecten en metadata staan dus niet op dezelfde storage.

Ze bewaren tevens een escrow-kopie op een derde locatie. Hier gaan data en metadata samen op tape inclusief open source tools om het van tape te halen en te verifiëren zodat de klant te allen tijde zijn materiaal terug kan krijgen. De metadata van deze escrow-versie wordt alleen bijgewerkt bij migratie.

[20] *Aanwezig in het gesprek vanuit B&G waren: Ernst van Velzen, Daniel Steinmeier en Maaike de Bie; verslag opgesteld door Daniel Steinmeier en Maaike de Bie.*
[21] *en.wikipedia.org/wiki/Apache_Cassandra*

Van updaten van beschrijvende metadata of objecten is geen sprake. Bij Arkivum komt materiaal binnen en daarna gaat alles op slot. Een object toevoegen kan alleen door een nieuw object aan te maken.

Arkivum biedt vooral een storagedienst, geen asset managementsysteem. Alle informatie die opgeslagen moet worden, wordt bij ingest meegegeven. De preserveringsmetadata wordt wel bijgewerkt, maar verder niets tenzij klanten dat hebben aangegeven.

Voor onze catalogus is deze situatie anders, hier moet beschrijvende metadata wel altijd aangepast kunnen worden, dus voor ons is het niet handig om metadata op tape op te slaan.

Vanuit eis 42 'het systeem moet alle prerveringsmetadata kunnen bewaren, bewerken en exporteren' kwam tevens de vraag of het echt noodzakelijk is om al deze informatie bij elkaar te hebben en of het niet ook gewoon door middel van queries bij elkaar gezocht kan worden wanneer de klant er om vraagt.

Het advies van Matthew is om de informatie wel op een plek bij elkaar te houden en om het ook doorzoekbaar te maken via een aparte database. Bij Arkivum wordt provenancedata gegenereerd op het moment dat een event plaatsvindt en er een resultaat is. Ze staan klanten niet toe zelf een query te draaien, maar hebben alle data inzichtelijk voor de klanten. De audit trails bieden deze inzichtelijkheid doordat hier precies in vermeld staat wanneer een file ontvangen is, dat deze niet corrupt is, wanneer er migraties zijn uitgevoerd, etc. Audit trails zijn op zich niet belangrijk voor hun ISO27001-certificering, maar wel voor de klanten uit de medische sector die deze nodig hebben. Voor ISO27001 is het vooral van belang dat ze kunnen laten zien dat ze de juiste processen hebben om integriteit te garanderen.

Zijn advies is tevens om een extra kopie van de metadata te maken (bijvoorbeeld door middel van export naar xml zoals ook aangegeven als eis 43 in het RFI-document) die periodiek geupdate wordt, zodat je niet afhankelijk bent van de database van een specifieke leverancier. Mocht er dan iets misgaan met deze database dan kan een nieuwe versie opgebouwd worden op basis van de extern opgeslagen metadata.

Error reporting

Op basis van eis 15 over het weergeven van alle events en errors die in de levenscyclus van een object hebben plaatsgevonden kwam de vraag naar voren hoe uitgebreid de error-reporting moet zijn en in hoeverre deze in verschillende systemen opgeslagen mag zijn.

Dit om te weten te komen of het MAM integratie moet leveren met de error-reporting zoals deze geleverd wordt door bijvoorbeeld DIVA en zo ja, hoe gedetailleerd deze informatie dan moet zijn.

Binnen Arkivum worden alle fouten die betrekking hebben op objecten gelogd als metadata binnen de audit trail. Hiermee zijn deze dus ook gekoppeld aan het object. Ook alle access-acties en eventuele niet-toegestane access-pogingen worden gelogd op deze manier.[22]
Welke vorm van reporting voldoende zou zijn is volgens Matthew afhankelijk van de systemen die klanten gebruiken. Bij Arkivum staat in het contract met klanten dat als data verloren gaat, dit gerapporteerd wordt. Het gaat hierbij dus niet om een automatische rapportering. Verder is de werkwijze zo dat als het echt aan de kopie ligt, de hele service wordt stilgelegd totdat het probleem is opgelost om andere kopien niet bloot te stellen aan eenzelfde gevaar. Pas als de bestanden op basis van de backup in de oorspronkelijke staat zijn hersteld, worden de objecten weer toegankelijk gemaakt. Voor Arkivum is authenticity belangrijker dan availability!

[22] *http://www.arkivum.com/content/Digital-Archive-Solutions-for-Life-Sciences*

Dit geeft ook meteen antwoord op de vraag bij eis 28 over het opslaan van alle gevallen van bit corruptie in de preserveringsmetadata. De vraag luidde: is het nodig dat gevallen van corruptie worden opgeslagen als events in de preserveringsmetadata of is een ander type logging genoeg?

Het antwoord van Matthew maakt duidelijk dat hij het noodzakelijk acht deze data persistent op te slaan als preserverinsmetadata, gekoppeld aan het object.

Ingest en checksums
Op basis van eis 16, waarin gesteld wordt dat het systeem informatie over de status van ingest moet kunnen leveren via mail en webinterface, kwam de vraag naar voren in hoeverre het vereist is dat dit proces geautomatiseerd wordt. Moet van het MAM gevraagd worden dat mails geautomatiseerd verstuurd worden op bepaalde momenten van de workflow?

Bij Arkivum gaat het ingestproces als volgt:

1) Er wordt een checksum gegenereerd bij binnenkomst.
2) Deze wordt indien door de klant gewenst opgestuurd naar de klant zodat zij kunnen bevestigen dat deze checksum klopt met de checksum van de oorspronkelijke versie.
3) Het object wordt gedupliceerd naar 2 verschillende datacenters opgeslagen.
4) Als er een complete tape geschreven is, wordt die uit de drive gehaald en naar een andere drive gestuurd om daar de geschreven data te lezen en nieuwe checksums te genereren die weer overeen moeten komen met de originele checksums.

Volgens Matthew kan niet vertrouwd worden op het uitrekenen van de checksum tijdens het schrijven maar moet deze geverifieerd worden door deze na de schrijfactie terug te lezen.
Op de vraag of het veiliger is om de tape daarna niet meer te lezen/controleren om slijtage te voorkomen stelt Matthew dat zij ervoor gekozen hebben alle objecten toch één keer per jaar te controleren. De controleacties worden afgewisseld over de verschillende backups zodat niet één kopie sneller slijt dan de andere (staggered checks). Ook worden de checksums nog eens uitgerekend bij uitlevering om er zeker van te zijn dat er alleen correcte bestanden worden uitgeleverd.
Alle leesacties ten behoeve van de controle worden als provenance metadata opgeslagen in de audit trail, dus niet op dezelfde tape als het object. De escrow-kopie bevat wel provenance data opgeslagen op tape samen met het object maar alleen tot aan de ingest in de escrow. Er wordt wel geverifieerd dat de escrow-tape goed geschreven is, maar daarna wordt deze niet meer bijgewerkt tot aan de migratie, ongeveer elke 3 jaar. Bij de migratie wordt de audit trail van de escrow-kopie ook weer bijgewerkt met de metadata uit de database.

Concluderend is het bij Arkivum dus zo geregeld dat uiteindelijk de klant bepaalt welke vorm van terugmelding gewenst is, of de klant dus een checksum wil ontvangen ter verificatie of niet en of de checksums digitaal gesigneerd moeten zijn of niet.[23]

Bi-directional linking.
Eis 19 stelt dat beschrijvende metadata altijd gekoppeld moet zijn aan een object en vice versa. Dit riep de vraag op of dit eigenlijk wel mogelijk is wanneer de metadata en de objecten op verschillende locaties staan en zo ja, hoe die link er dan uit zou moeten zien.

Zoals uit bovenstaande al blijkt is de metadata bij Arkivum ook niet rechtstreeks bij het object opgeslagen, behalve in het geval van de escrow-kopie. De connectie tussen metadata en object wordt gevormd door middel van de unieke ID die elke file krijgt binnen het archief. Deze

[23] *http://www.arkivum.com/content/benefits-archiving-and-seven-questions-you-should-always-ask*

maakt dat op basis van de ID de metadata opgezocht kan worden en dat het object opgezocht kan worden doordat de ID in de metadata vermeld staat. Dit is volgens Matthew de enige manier om aan deze eis te kunnen voldoen. De andere optie, namelijk om metadata en object bij elkaar op te slaan beschouwt hij als ondoenlijk.

Vanuit het DIVA-onderzoek blijkt dat alle objecten binnen DIVA ook een unieke ID meekrijgen (een UUID), die gebruikt zou kunnen worden om aan deze eis te voldoen.

Versiebeheer
Eis 21 stelt dat het systeem versiebeheer moet ondersteunen om waar nodig terug te kunnen gaan naar een vorige versie van een object. De vraag is hoe dit bij Arkivum is geregeld.

Binnen het systeem van Arkivum is versiebeheer niet aan de orde. Een bestand gaat na ingest op slot en er vinden geen updates meer plaats. Als een klant een nieuwe versie van een bestand wil uploaden wordt dit feitelijk een hele nieuwe file die als zodanig in het archief verwerkt wordt. Klanten moeten dan zelf het versiebeheer bijhouden. De metadata wordt wel geupdate voor de audit trail maar dit leidt niet tot verschillende versies.

Files deleten
Een andere vraag in dit kader vanuit het RFI-traject was, kan een origineel bestand verwijderd worden wanneer de producer daar akkoord voor heeft gegeven? Bijvoorbeeld doordat er een formaatmigratie heeft plaatsgevonden en het oorspronkelijke object eigenlijk beschikbaar is in een nieuwe versie waarbij de oude versie overbodig is geworden.

Bij Arkivum is het deleten van data bewust heel moeilijk gemaakt. Deleten gebeurt altijd alleen maar handmatig en op verzoek van de producer. Matthew waarschuwt dat het niet genoeg is om dit recht te koppelen aan een admin-account maar dat het ook echt gekoppeld moet zijn aan een geautoriseerd persoon zodat altijd duidelijk is wie welke actie op welk moment heeft gedaan. Dit moet zijns inziens als vraag ook uitgezet worden bij een MAM-leverancier, namelijk of ze een dergelijk authenticatie geregeld hebben binnen hun systeem.

Deleten is dus mogelijk op verzoek van de producer mits de functionaliteit van het deleten goed genoeg is afgeschermd tegen fouten en misbruik.

Access
Eis 29 stelt dat het systeem een statusoverzicht moet kunnen leveren van alle access-errors en uitzonderingen. De vraag hierbij is, wat voor soort error reporting is voldoende?

Bij Arkivum dient de audit trail als invulling van deze eis. Binnen de audit trail zijn alle raadplegingen van het object gelogd inclusief eventuele pogingen van niet-geautoriseerde personen om het object te raadplegen.

Preservation planning
Eis 30 stelt dat preserveringsacties gelinkt moeten kunnen worden aan preserveringsplannen. De vraag hierbij is, hoe moet dit praktisch uitgewerkt worden? Is het voldoende om een link aan te brengen naar een word-bestand of moet de data actionable zijn?

Binnen Arkivum is formaatmigratie niet direct aan de orde omdat ze beloven dat de data beschikbaar blijft zoals deze is. Wel vindt er migratie plaats van de opslagmedia.

Representation information
Eis 34 stelt dat het systeem in staat moet zijn om digitale objecten te koppelen aan externe algemene informatie over file formaten. De vraag hierbij is hoe dit praktisch te verwezenlijken is. Moet deze informatie opgenomen zijn in de metadata van alle objecten?

Zoals hierboven ook al aangegeven biedt Arkivum vooral opslag van objecten met de garantie dat deze hetzelfde blijven. Als zodanig houden ze zich niet bezig met het formaat van de bestanden. Ze weten bij binnenkomst niet wat voor bestand iets is, of het bijvoorbeeld video is of pdf. Ze extraheren momenteel ook geen technische metadata van de bestanden hoewel dit in de toekomst misschien wel gaat gebeuren omdat klanten in deze functionaliteit wel interesse hebben getoond. Arkivum houdt om deze reden dus ook geen algemene informatie bij over file formaten.

Wrappers
Eis 41 stelt dat het systeem hiërarchische relaties tussen objecten moet kunnen ondersteunen. De vraag vanuit het RFI-traject was of aan deze eis ook voldaan wordt indien bestanden niet op dezelfde locatie worden opgeslagen.

Achtergrond van deze vraag is of het noodzakelijk is dat bestanden die bij elkaar horen altijd binnen eenzelfde wrapper opgeslagen moeten worden of dat een ander type ordening ook voldoende kan zijn.

Binnen het systeem van Arkivum worden geen wrappers gebruikt. Alle objecten worden op tape opgeslagen binnen een LTFS-bestandsindeling. Objecten staan in een folder en er is een metadatabestand met gegevens over wat er in de folder staat. Ze gebruiken dus geen formaat zoals bag-it om alles te wrappen.
Arkivum heeft niet echt de notie van een AIP, er is ook geen relatie tussen bestanden onderling, het zijn losstaande binary files.
Arkivum heeft verder het escrow manifest, waarin staat wat er allemaal moet zijn, alle ids en checksums. Deze gebruiken ze om te controleren of alles volledig is. Ze gebruiken het manifest om te controleren of alle bestanden ook echt in de folder staan en kijken tevens of er geen extra dingen op tape staan die niet in het manifest voorkomen.

Hoewel de situatie van Arkivum niet helemaal vergelijkbaar is met de situatie van Beeld en Geluid is wel duidelijk dat het niet noodzakelijk is objecten in een wrapper op te slaan om hiërarchische relaties te ondersteunen. Verbinding tussen bestanden kan in de metadata gelegd worden.

ISO/Audits
Matthew heeft ook kort toegelicht hoe een audit proces werkt.

Een audit gaat als volgt: er worden een paar gebieden van aandacht genoteerd en wat voor problemen (non-conformities) er zijn. 6 maanden later moet alles zijn gefixed, moeten er processen zijn verbeterd (periodieke reviews van alles). Dan is er weer een audit en 6 maanden later weer audit met al die stappen etc. Dit is dus een continue proces, geen momentopname.
ISO27001 is heel rigoreus, heel erg op details en geven veel non-conformities aan.
ISO16363 is wel een standaard voor TDR zelf, maar er is momenteel nog geen goedgekeurde ISO-standaard voor auditors – er is dus niemand die echt in staat is om de eisen te beoordelen en er is momenteel daardoor nog verschil in interpretatie van de eisen.

Conclusies

1) Solve problems on a lower level, not in information systems. Als je een low tech benadering hebt, is het makkelijker te herstellen als er iets mis gaat. Ook is het voor certificering zo makkelijker aan te tonen hoe je met je risico's omgaat en ben je minder afhankelijk van gekochte systemen en leveranciers.

2) MAM moet mogelijkheid hebben voor exporteren van metadata in xml-structuur voor audit trails en dergelijke informatie. Team TDR zal dus niet gaan kijken naar verschillende wrappers (wordt evt AXF), maar wel naar bijvoorbeeld een METS xml-structuur. In het RFI-document staat al als eis dat alle preserveringsmetadata geëxporteerd moet kunnen worden als xml.

Er zijn afspraken gemaakt om dit gesprek een vervolg te geven waarbij Matthew zich desgewenst nog meer kan voorbereiden op specifieke onderwerpen.