

Het OAIS-model, een leidraad voor duurzame toegankelijkheid

BARBARA SIERMAN

1 Inleiding

In augustus 2012 publiceerde de International Standard Organisation een bijgewerkte versie van het OAIS-model. In de tien jaren die inmiddels verstreken zijn nadat de eerste versie werd gepubliceerd, heeft het OAIS-model een stevige positie veroverd. Voldoende aanleiding om aan dit model een artikel te wijden in dit Handboek.

1.1 Voorgeschiedenis

Voor iedereen die zich met het toegankelijk houden van digitaal materiaal bezighoudt, is het 'OAIS' een standaardbegrip. Het OAIS, of volledig 'REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM' is ontstaan onder auspiciën van The Consultative Committee for Space Data Systems (CCSDS) in de Verenigde Staten. Dit is een internationale groep ruimtevaartorganisaties. Het resultaat van hun werk is ondergebracht bij de International Standards Organisation (ISO) en bekend als standaard ISO 14721:2003.

Het OAIS Reference Model geeft aanbevelingen bij het inrichten van een archief, dat gericht is op langetermijnbewaring van en toegang geven tot informatie, met name digitale informatie. Het biedt een samenhangend geheel van uitgangspunten en terminologie.

Aan het ontstaan van de standaard ging een jarenlange voorbereiding vooraf, die uitgebreid is beschreven in het proefschrift van Lee (Lee 2005). Daaruit blijkt dat het initiatief tot deze standaard werd genomen door een werkgroep binnen CCSDS. Dankzij het feit dat er in de voorbereiding op ruime schaal publiciteit aan gegeven is en organisaties uitgenodigd werden hun bijdrage te leveren, kwamen er ook reacties uit andere disciplines. Verschillende bibliotheken en archieven droegen hun

steentje bij. Het OAIS-model is dan ook niet alleen geschikt voor één bepaalde discipline, zoals bijvoorbeeld de ruimtevaart, maar zo generiek, dat een grote diversiteit aan organisaties dit kan toepassen. Dat heeft geleid tot een brede acceptatie van de standaard.

ISO kent elke vijf jaar een periodieke herijking van de standaard, waarbij iedereen wijzigingen kan voordragen. Dit geldt ook voor deze standaard. Het internationale commentaar uit het veld heeft geresulteerd in een aangepaste versie, die bekend is onder ISO 14721:2012 en in augustus 2012 gepubliceerd is. Deze versie is het uitgangspunt voor dit artikel.* Inmiddels heeft CCSDS deze standaard gewijzigd van 'Recommended Standard' (de 2003 versie) in 'Recommended Practice' (de 2012 standaard); dit weerspiegelt mooi de status die het OAIS inmiddels heeft.

Belangstellenden die de volledige (Engelse) tekst willen lezen, kunnen de tekst van ISO 14721 bestellen bij het ISO-bureau of downloaden van de CCSDS-website.

1.2 *Gebruikte terminologie*

Ik heb ervoor gekozen in dit artikel de begrippen uit OAIS onvertaald te laten en dus alle Engelse termen in de Nederlandse tekst te gebruiken. Digitale duurzaamheid is immers een internationale activiteit, en in artikelen, lezingen en discussies gebruikt men de Engelse OAIS-terminologie. Niet iedereen maakt deze keuze, in Nederland heeft bijvoorbeeld het Landelijke Overleg Provinciale Archief Inspecteurs ervoor gekozen de termen naar het Nederlands te vertalen in hun Eisen Duurzaam Digitaal Depot (ED3 2008).

1.3 *Wat is een referentiemodel?*

In het algemeen kun je twee soorten standaarden onderscheiden: standaarden gebaseerd op praktijkervaringen en standaarden die ontwikkeld worden voordat er op grote schaal al implementaties en producten zijn. Het OAIS-model behoort tot deze laatste categorie ('anticipatory standards') en verenigt ideeën, concepten en ervaringen uit de (beperkte) praktijk van digitale duurzaamheid tot een referentiemodel. Het model werd ontwikkeld tussen 1995 en 2002, toen er nog niet

* Voor een overzicht van de belangrijkste verschillen, zie <http://digitalpreservation.nl/seeds/standards/oais-2012-update/>

veel lange termijn digitale archieven actief waren. Vandaar ook dat dit referentiemodel een hoge mate van abstractie kent. Het is een leidraad voor het opzetten van een archief, geen blauwdruk, en biedt een kader om op basis van het uitgewerkte model, verdere ontwikkelingen te stimuleren. Elke systeemarchitect zal echter zelf het model vertalen naar een implementatie.*

1.4 *Gebruik van het OAIIS referentie model*

Het OAIIS-model is het uitgangspunt voor iedereen die zich bezighoudt met digitale duurzaamheid. Het gebruik van het model is bij lange na niet beperkt tot de disciplines die het samenstelden: de ruimtevaart, bibliotheken en archieven. Ook de biowetenschappen, aardwetenschappen en medische wetenschappen hanteren dit model voor hun opslagsystemen voor wetenschappelijke data. Commerciële en niet-commerciële bedrijven die duurzame archiefsystemen bouwen, hanteren het OAIIS-model als leidraad, bijvoorbeeld Fedora, Duraspace, Tessella en Ex Libris. Ook wanneer organisaties zelf hun system ontwerpen, is het OAIIS-model vaak de leidraad. De standaarden Premis (preservation metadata) en METS (structurele metadata) conformeren zich aan het OAIIS-model. Verschillende kostenmodellen voor duurzame archieven zoals LIFE en KRDS baseren zich op het OAIIS-model. Al met al kunnen we zeggen, dat in de ruim tien jaar dat het model een ISO-standaard is, de adoptiegraad zeer hoog ligt en dat kennis van het model een basis is voor iedereen die iets met het duurzaam toegankelijk houden van digitaal materiaal te maken heeft.

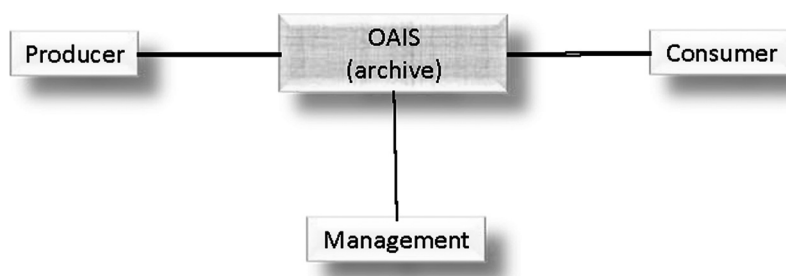
De kracht van OAIIS ligt vooral in het feit, dat door deze standaard iedereen de beschikking heeft over een gemeenschappelijk jargon, een soort Esperanto voor digitale duurzaamheid. De termen zijn gedefinieerd en de standaard zit stevig in elkaar, wat misverstanden voorkomt. Er is ook wel kritiek op verschillende aspecten van het model. Europese projecten als Planets en CASPAR en (inter)nationale organisaties voor digitale duurzaamheid dienden wijzigingsvoorstellen in bij vijfjaarlijks evaluatie van de standaard. Een deel van de kritiek is dan ook verwerkt in de 2012 versie.

* Lee, p.52 e.v.

2 Wat is een OAIS?

Voordat we dieper ingaan op het referentiemodel, is het belangrijk een aantal basisbegrippen toe te lichten. Te beginnen met de vraag: wat is een Open Archival Information System? Met deze omschrijving wordt een 'archief', in de ruimste zin van het woord, aangeduid, dat zich bezighoudt met het voor de lange termijn bewaren en toegankelijk houden van informatie. Het kan een bibliotheek zijn, een museum, een wetenschappelijke instelling of een traditioneel archief, maar ook een archief van een verzekering of een oliemaatschappij. Het kan hierbij zowel om fysieke als digitale informatie gaan. In principe is het OAIS-model op beide toepasbaar. Maar in de praktijk ligt de toepassing en acceptatie van deze standaard toch vooral bij instellingen die zorgdragen voor digitale informatie. Zoals uit de omschrijving blijkt, kun je het bewaren van digitale informatie niet los zien van het toegankelijk houden van deze informatie. Beide aspecten, zowel de opslag als de toegang, zijn onlosmakelijk met elkaar verbonden. Alleen duurzaam opslaan heeft weinig zin, als je de toegang niet kunt garanderen. De toezegging iets blijvend toegankelijk te houden, lukt niet als je de opslag niet duurzaam regelt.

In essentie heeft een duurzaam archief te maken met informatie, die 'ergens' in de omgeving tot stand komt, in het archief wordt bewaard en uiteindelijk weer ter beschikking komt. Dit is het kernplaatje van OAIS. De Producer is degene die de informatie tot stand brengt. De Consumer is de afnemer van de informatie. Zowel de Producer als de Consumer hoeven geen menselijke entiteiten te zijn. Een wetenschappelijk apparaat dat data genereert, kan ook een Producer zijn. En software die informatie uit het archief doorgeeft aan een ander systeem, kan een Consumer van het OAIS-archief zijn.



Wel hanteert de OAIS-standaard een nadere precisering van de Consumer: de Designated Community. Met de Designated Community duiden we de groep gebruikers van het archief aan, die de doelgroep is van het OAIS-archief en die in staat moet zijn deze bewaarde informatie te begrijpen. Uiteraard kent het OAIS-archief zelf een Management, dat zorgt voor het functioneren van het archief. In de rest van dit artikel zullen al deze elementen verder toegelicht worden.

Samenvattend is een OAIS-archief een archief dat de verantwoordelijkheid heeft informatie te bewaren en voor lange termijn toegankelijk te houden voor een speciale groep gebruikers, de Designated Community.

3 Verantwoordelijkheden van een OAIS-Archief

Kan elk digitaal archief, dat digitaal materiaal bewaart en toegankelijk maakt, zich ook een OAIS-archief noemen? Nee, de standaard heeft een aantal eisen geformuleerd, die essentieel zijn voor een archief, dat zich conformeert aan de OAIS-Standaard. De voornaamste daarvan zijn:

1. *Verkrijgen van voldoende informatie van Producer*

De eerste verantwoordelijkheid van het OAIS-archief begint al bij de verwerving van het materiaal. Samen met de 'Producer' overlegt het OAIS-archief welk materiaal en in welke vorm hij aanlevert. Ook bepaalt het archief zelf wat het wel en niet wil accepteren en onder welke criteria. De afspraken met de Producer liggen vast in het 'Submission Agreement'.

2. *Voldoende mandaat*

Om goed beheer te kunnen doen, en zeker voor de lange termijn, is het belangrijk dat het OAIS-archief ook het juridische mandaat heeft om met het materiaal te kunnen wat het archief noodzakelijk vindt in het kader van de duurzame toegankelijkheid. Zoals het uitvoeren van preservation actions (zie paragraaf 7). De huidige wetgeving staat bepaalde acties niet zonder meer toe (zoals het kopiëren van materiaal) en het is belangrijk de expliciete toestemming van de Producer vast te leggen.

3. *Bepaal de Designated Community*

Zoals gezegd, moet het archief een duidelijk beeld hebben wie de beoogde gebruikers, de Designated Community, van het materiaal zijn en over welke kennis zij moeten beschikken om de informatie te kunnen

begrijpen. Dit bepaalt ook welke activiteiten een archief onderneemt om het materiaal voor deze gebruikers zo goed mogelijk toegankelijk te houden. Archieven kunnen meerdere Designated Communities onderscheiden, vaak gerelateerd aan de diversiteit van hun collecties.

4. *Toegankelijkheid van de informatie*

Het OAIS-archief moet ervoor zorgen dat deze gebruikersgroep het materiaal onafhankelijk en zonder extra hulp van buitenaf, bijvoorbeeld experts op het vakgebied, kan gebruiken en begrijpen.

5. *Bewaar de authenticiteit*

De gebruiker van het OAIS-archief zal ook duidelijkheid willen hebben over wat het archief toont: is dit een kopie van het origineel, hebben er wijzigingen op plaatsgevonden? Kortom, hoe authentiek is het object? Authenticiteit is af te leiden, onder andere doordat een OAIS-archief duidelijk rekenschap aflegt van haar handelen (Provenance, zie paragraaf 4.2.1).

6. *Het OAIS-archief heeft een duidelijk beleid, procedures en processen*

Men moet er van op aan kunnen dat het OAIS-archief zijn werk op een betrouwbare manier doet, zonder dat het bewaarde materiaal beschadigt. Dit betekent dat het OAIS-archief duidelijke procedures en beschreven processen heeft, die voorkomen dat er materiaal beschadigd raakt of verdwijnt. De gebruiker moet erop kunnen vertrouwen dat een archief niet op ad hoc-basis informatie uit het archief kan verwijderen.

In het geval het OAIS-archief ophoudt te bestaan, een situatie die zich bij een langertermijnarchief altijd kan voordoen, zal het OAIS-archief maatregelen moeten treffen om de inhoud van het archief veilig te stellen.

Het is dus niet voldoende als een OAIS-archief de juiste terminologie hanteert en zijn systeem heeft gebaseerd op het functionele model dat hieronder nader wordt beschreven. Het voldoen aan de beschreven verantwoordelijkheden hoort er ook bij, wil een archief zich 'OAIS-compliant' kunnen noemen. Deze verantwoordelijkheden betreffen meerdere organisatieonderdelen. Via een procedure van audit en certificering (zie verder) kan een externe instantie dit toetsen.

4 Het OAIS Informatie model

Zoals uit het bovenstaande blijkt, bestaat de inhoud van het archief uit 'informatie'. Informatie is meer dan bits, en ontstaat door een combinatie van data die is opgeslagen (hier het 'Data Object', de bitreeks genoemd) en 'kennis' die tot 'informatie' leidt. Een deel van de benodigde kennis die nodig is, zal bij de gebruiker aanwezig zijn. Binnen OAIS wordt dit de 'Knowledge Base' van de gebruiker genoemd. Maar voor een juiste interpretatie van de bits, kan deze Knowledge Base ontoereikend zijn, omdat er specifieke extra informatie nodig is. Deze informatie moet dan bij het Data Object aanwezig zijn. Dit noemen we de Representation Information.

Representation Information is alle informatie die nodig is om het Data Object om te zetten in een betekenisvol geheel. Het Content Data Object, samen met de Representation Information, leidt tot de Content Information, in het figuur hierna het Information Object genoemd.



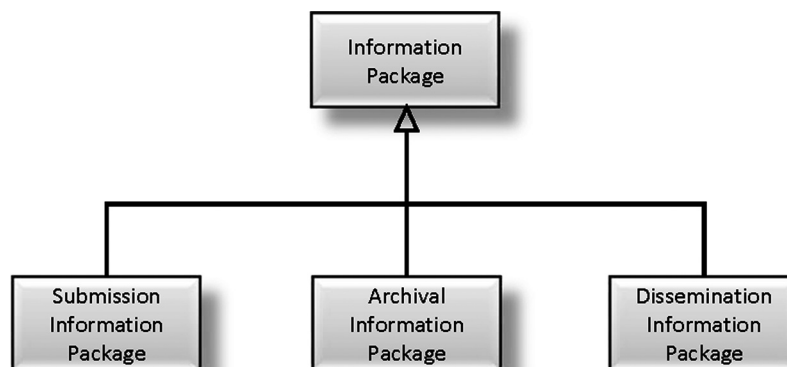
Ter toelichting een simpele uitleg. Alle digitale informatie bestaat uit een reeks nullen en enen, ook het Data Object. Op zich is niet duidelijk wat die reeks nullen en enen betekent, als je niet ook weet volgens welke formaatspecificaties deze nullen en enen geordend zijn, met welke software je dit formaat kunt weergeven en op welke hardware deze software draait (deel van de Representation Information). En als je dit pad gerealiseerd hebt, dan zijn de nullen en enen leesbaar gemaakt, maar de gebruiker moet het dan nog van data kunnen omzetten naar informatie. Daarvoor heeft de gebruiker kennis nodig, die zijn Knowledge Base gebruikt om bijvoorbeeld een Nederlandse tekst van een Engelse tekst te kunnen onderscheiden. Sommige kennis zit in het object zelf, namelijk in de extra informatie die de Producer meegeeft. Zoals de betekenis die bepaalde kolommen in een spreadsheet hebben. Deze informatie kan vaak niet uit het object zelf afgeleid worden.

Het archief verzorgt dus niet alleen nullen en enen van het Data Object maar zorgt er ook voor dat de Representation Information aanwezig is, om van de nullen en enen uit het Data Object een Information Object te maken. Zodat de gebruiker het begrijpt.

Het zal duidelijk zijn dat een OAIS-archief alleen de relevante Representation Information kan verzamelen, als het OAIS-archief ook een duidelijk beeld heeft van zijn gebruikers en hun Knowledge Base. Daarnaast beslist het OAIS-archief wat er, bijvoorbeeld in de vorm van metadata, ook bewaard moet worden om deze kennis uit de Knowledge Base aan te vullen. Niet alleen met het oog op het hier en nu, maar ook met het oog op toekomstige gebruikers. Dat maakt het ingewikkeld, want hoe weten we hoe de kennis zich in de loop der tijd ontwikkelt en wat gebruikers nog wel en niet weten? Gevolg is dat de Representation Information in de loop der tijd kan veranderen en aangepast moet worden. Al is het maar dat het taalgebruik in de Representation Information zelf kan verouderen en toelichting behoeft. Deze recursiviteit noemen we het 'Netwerk van Representation Information'.

4.1 *Information packages*

Het OAIS-archief bewaart de Content Information in zogenaamde Information Packages. Een Information Package is een conceptuele container. De informatie is aanwezig in het archief, maar de onderdelen staan niet noodzakelijkerwijs ook fysiek bij elkaar.



Een Information Package kent drie stadia, die samenhangen met de verwerking in het digitale archief; een Submission Information Package (SIP), een Archival Information Package (AIP) en een Dissemination Information Package (DIP).

1. Submission Information Package

Op het moment dat een Producer digitale objecten aanbiedt aan het archief, doet hij dit als een Submission Information Package. In deze SIP kunnen meerdere objecten zitten, die het O AIS-archief gaat opslaan. In veel gevallen zijn deze 'information packages' nog niet gereed voor permanente opslag. De onderhandelingen, die tussen het O AIS-archief en de producer hebben plaatsgevonden, leiden tot afspraken wat een SIP bevat. Dit is vastgelegd in het Submission Agreement. Een uitgebreide beschrijving wat een Submission Agreement inhoudt, is te vinden in de, aan het O AIS-model gerelateerde standaard: Producer-Archive Interface Methodology Abstract Standard (PAIMAS).

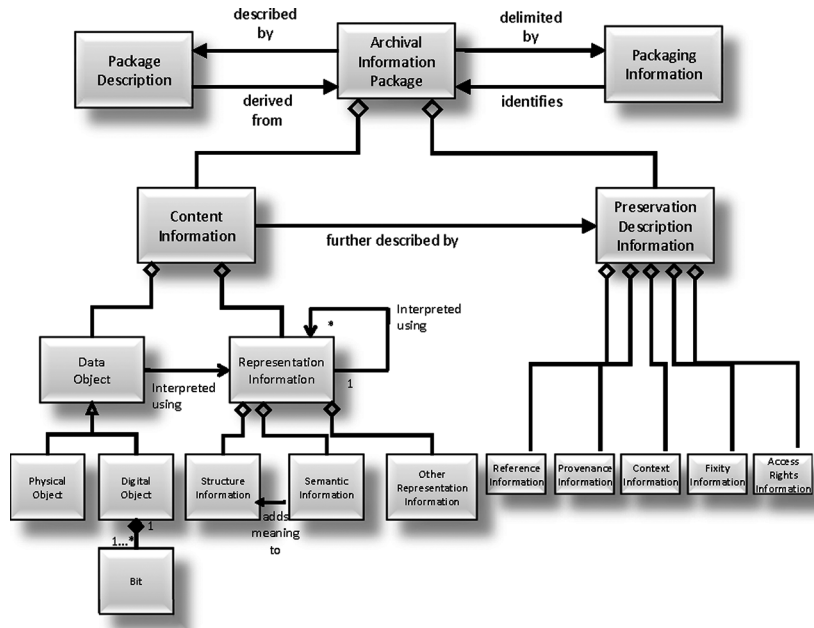
2. Archival Information Package

Na binnenkomst doorloopt de SIP verschillende verwerkingsstappen. Dit is nodig om Information Packages tot stand te brengen die, kwalitatief gezien, gereed zijn voor langetermijn toegankelijkheid. Dit gebeurt in de Ingest-functie. Het pakket dat dan ontstaat, noemen we het 'Archival Information Package'. Een Archival Information Package (AIP) kan ontstaan zijn uit één of meerdere Submission Information Packages.

3. Dissemination Information Package

Wanneer uiteindelijk de gebruiker een object uit het archief opvraagt, krijgt hij (een kopie van) het Archival Information Package, de DIP. Een DIP kan bestaan uit alle informatie in de AIP. Daarnaast kan een archief meerdere vormen van DIPs naast elkaar aanbieden door bijvoorbeeld alleen een deel van de AIP te laten zien (één hoofdstuk van een boek), of alleen de metadata. Via een Order Agreement legt de Consumer vast welke informatie hij wenst uit het archief, dit kan via een formele (gedetailleerde) aanvraag, maar zal in de praktijk vaak een online aanvraag zijn.

4.2 Wat zit er in een Information Package?



De figuur hiervoor toont de meest uitgebreide vorm van een Information Package, en bevat alle informatie die gezien wordt als noodzakelijk voor duurzame toegang. In de praktijk zal de AIP deze informatie bevatten en (nog) niet de SIP.

Op hoofdlijnen bevat een Information Package het Content Object met de bijbehorende Representation Information; samen vormen ze de Content Information. Dit is wat men uiteindelijk wil bewaren, ofwel 'the original target of preservation'. Daarnaast zal een Information Package nog drie onderdelen bevatten: de Preservation Description Information (PDI), de Packaging Information, en de Package Description Information.

4.2.1 Preservation Description Information (PDI)

De PDI kun je pas goed vastleggen, als je ook weet waaruit de Content Information bestaat. Waarom dat zo is, wordt duidelijk als je ziet uit welke

onderdelen de PDI bestaat; alle informatie die hierin staat, slaat immers op onderdelen van het Content Object. Hoewel de PDI hoort bij het AIP, is het niet noodzakelijk dat alle informatie in één enkele file is opgeslagen. Het OAIS-archief bepaalt zelf hoe de informatie binnen het OAIS wordt opgeslagen. De Producer speelt een belangrijke rol bij het samenstellen van de PDI. Voor sommige informatie is hij alleen degene die er weet van heeft, het is niet op een andere manier af te leiden uit de objecten. Vandaar dat het tot de verantwoordelijkheden van het OAIS-archief hoort dit goed te regelen in het Submission Agreement.

Provenance

In de provenance informatie ligt de geschiedenis vast van de Content Information, dus wat er in de loop der tijd mee is gebeurd. Het OAIS-archief legt dit vast vanaf het moment dat de Content Information via de Ingest-functie wordt opgeslagen en houdt dit vanaf dat moment bij. Het kan zijn dat eerdere activiteiten door de Producer ook relevant zijn. In dat geval kan ook de Producer provenance informatie meegeven. Voorbeelden van provenance informatie zijn: wie heeft het Content Object gemaakt, wie heeft het sindsdien beheerd, hebben er migraties op de Content Information plaatsgevonden? Provenance informatie is een belangrijk ingrediënt om de authenticiteit van de Content Information te bepalen. Het is aan het archief te bepalen welke mate van detail de Provenance informatie moet hebben en hoeveel inspanning er verricht wordt om deze te achterhalen.

Context Information

De Context Information plaatst de Content Information in relatie tot zijn omgeving: waar en waarom was het ooit gecreëerd en hoe verhoudt het zich tot andere Content Information. Is het bijvoorbeeld een digitale versie van een oorspronkelijke analoge versie van het object?

Fixity Information

Fixity Informatie legt vast of de Content Information niet onopzettelijk gewijzigd is doordat een bitreeks beschadigd is. Vooral bij het verplaatsen van Information Packages zou de Content Information beschadigd kunnen raken. Bij Ingest wordt gecontroleerd of bijvoorbeeld het Data Object dat de Producer aanleverde nog identiek is aan het Data Object dat is ontvangen bij het OAIS-archief. Voor deze controle maakt men

veelal gebruik van checksums. De bitreeks ondergaat bij de Producer reeds een berekening (checksum algoritme), de uitkomst geeft hij mee bij de levering. Deze berekening wordt bij Ingest opnieuw uitgevoerd. Wanneer de uitkomst identiek is, gaat men ervan uit dat er geen bits verloren zijn geraakt. De uitkomst draagt bij aan het vertrouwen in de authenticiteit van de Content Information.

Reference Information

In de Reference Information is de identificatie, die specifiek aan deze Content Information verbonden is en ondubbelzinnig hiernaar verwijst. De term 'persistent identifier' wordt ook wel gebruikt voor bijvoorbeeld een Digital Object Identifier (DOI) of andere unieke identifier die in de discipline gebruikelijk is. De informatie is meestal ook opgenomen in de Package Description Information om het zoeken en vinden van een bepaald object te vergemakkelijken.

Access Rights Information

Access Rights Information is een veelomvattende term. Het omvat zowel informatie over het wettelijk kader, als over de licentierechten van het OAIS-archief en de toegangsrechten voor de Consumer. Ook de rechten die een OAIS-archief heeft om bijvoorbeeld maatregelen in het kader van de duurzame toegankelijkheid te nemen, als het omzetten naar een ander file format, horen hier thuis.

4.2.2 Package Information

De Package Information geeft aan hoe de onderdelen binnen het Information Package onderling samenhangen. Bijvoorbeeld de filestructuur op een cd-rom.

4.2.3 Package Descriptive Information

De gebruiker moet het object kunnen vinden en daarvoor is beschrijvende (titel-) informatie nodig. Het OAIS-archief voegt deze beschrijvende informatie, hier 'descriptive information' genoemd toe aan het Content Object. Deze toegangsinformatie hoeft niet in het Information Package aanwezig te zijn, hoewel het vaak een afgeleide zal zijn van al meegegeven informatie. Vaak zal de informatie in een catalogus komen.

5 De doelgroep: de Designated Community

De Designated Community is een specialisatie van de ‘Consumer’ en bestaat uit de (toekomstige) gebruikers van het materiaal, dat in het archief is opgeslagen. Deze gebruikers zijn door hun kennis in principe in staat de bewaarde informatie te begrijpen. Het archief moet ervoor zorgen dat zij dit zo zelfstandig mogelijk moeten kunnen. Een OAIS-archief kan meerdere Designated Communities onderscheiden. Het is van belang dat het OAIS-archief een beeld heeft van deze groep(en), om in de gaten te kunnen houden of het materiaal nog zo toegankelijk voor deze potentiële gebruikers is, dat zij ook werkelijk de informatie begrijpen. Dit begrip ligt vooral in de toegevoegde Representation Information en de PDI. Wanneer een specifiek fileformaat bijvoorbeeld binnen een bepaalde gemeenschap in onbruik raakt, dan zal die groep gebruikers op een gegeven moment niet meer de beschikking hebben over de hulpmiddelen (tools) om dit file format te lezen. Als het OAIS-archief dan alleen nog dit format aanlevert, kan het zijn Designated Community niet langer goed van dienst zijn. Kennis van de Designated Community helpt ‘to maximise information preservation’.

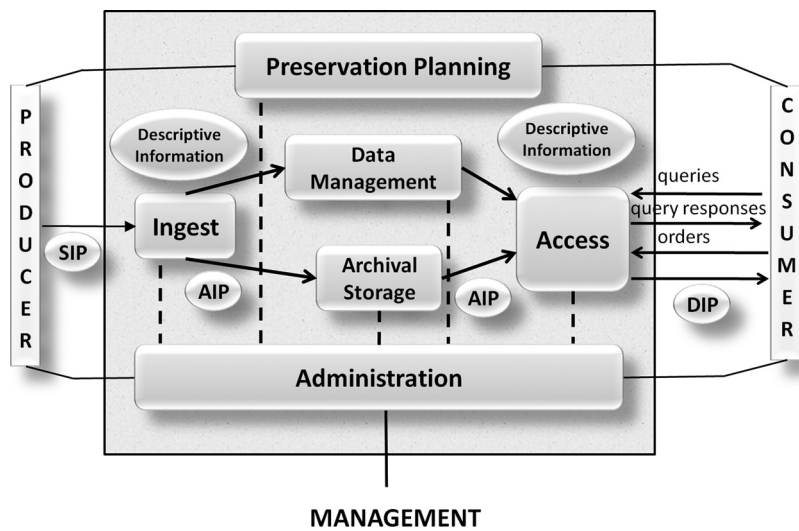
Iemand uit de Designated Community beschikt in principe over een ‘Knowledge Base’, die hem in staat stelt het materiaal te begrijpen. Dat kan bijvoorbeeld betekenen dat die persoon Engels spreekt zodat Engelse informatie begrepen wordt, en kennis van een bepaald vakgebied heeft, zodat de gebruiker een Engelse tekst op dat vakgebied begrijpt. In de loop der tijd kunnen er gebruikers komen die bepaalde kennis niet meer hebben. Periodiek verifieert het OAIS-archief of de aangeboden informatie nog aansluit bij de wensen van deze gebruikers (via de functie Preservation Planning, zie paragraaf 6.5). Is dit niet langer het geval, dan zal het OAIS-archief passende maatregelen nemen. Zoals het wijzigen van het profiel van de Designated Community of door het aandragen van extra informatie die bijdraagt aan het begrip.

Het begrijpen van de Designated Community en de bijbehorende Knowledge Base is voor het archief een voorwaarde om te kunnen bepalen welke Representation Information zij moet vastleggen.

Naast de Designated Community kan een archief ook gebruikers, Consumers, hebben, die de informatie uit het OAIS-archief kunnen opvragen. Dit kunnen zowel personen als systemen zijn. Uiteraard houdt het OAIS-archief de informatie voor deze Consumers toegankelijk.

6 Het functionele model van OAIS

Nu de kernbegrippen van het informatiemodel van OAIS zijn besproken, kijken we naar het functionele model van een OAIS-archief. Een functioneel model beschrijft de functies van een archief en daarbij de onderlinge relaties. Het bijgevoegde plaatje is geeft de kernfuncties weer, gekoppeld aan de rollen die we eerder hebben besproken van Producer, Consumer en Management. Het is misschien wel het meest gepresenteerde overzicht uit het OAIS-model.



De Producer levert zijn materiaal aan het OAIS-archief. De producer kan een externe partij zijn, maar ook een afdeling binnen de organisatie, die bijvoorbeeld een digitaliseringsproject doet. Deze levering noemen we een SIP, een Submission Information Package. De SIP zal meestal nog niet gereed zijn voor opslag in het archief, in de Archival Storage.

Gedurende het Ingest-proces vinden controles plaats op de inhoud van de SIP. Door daarnaast relevante metadata toe te voegen, komt een Information Package tot stand dat voldoet aan de eisen voor een AIP (zie paragraaf 4.2). Deze AIP wordt opgeslagen in Archival Storage. De bijbehorende beschrijvende informatie voor toegang, komt in Data Managementfunctie.

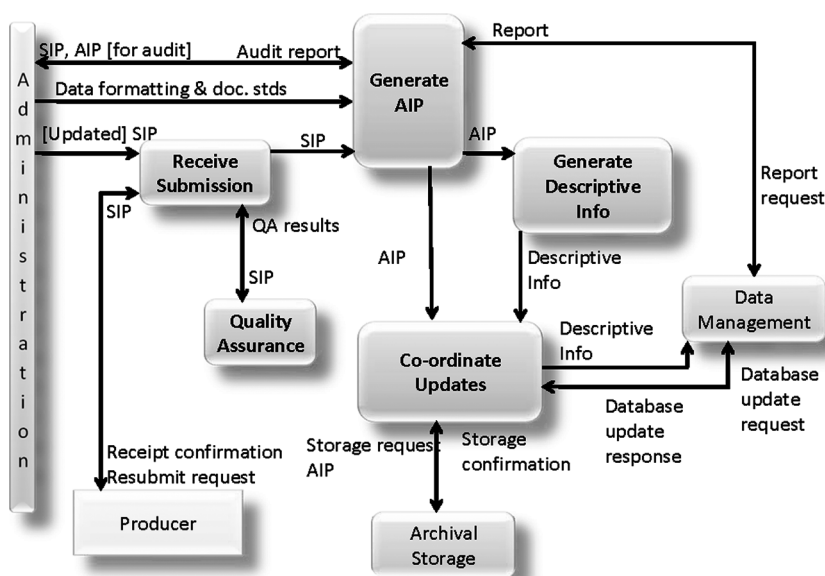
Een Consumer die het object wil raadplegen, dient een verzoek in bij het archief, gebruikmakend van beschrijvende informatie. Als het object aanwezig is, ontvangt hij via de Access-functie een Dissemination Information Package (DIP). Deze DIP kan een kopie van de AIP zijn, maar het is ook mogelijk dat slechts een deel van de AIP getoond wordt of dat er een combinatie van AIPs getoond wordt.

Een belangrijke functie is Administration. Zoals op het plaatje te zien, is deze functie verbonden met alle andere functies. Onder Administration vallen alle activiteiten die samenhangen met het algemene dagelijkse beheer van het archief, in de ruimste zin van het woord.

Preservation Planning omvat alle functies die samenhangen met het monitoren van de omgeving van het OAIS-archief. Het OAIS-archief opereert immers in een wereld in beweging en moet dus voortdurend monitoren of er veranderingen in de omgeving zijn, die consequenties of risico's voor het archief en de opgeslagen informatie opleveren. Samenvattend kent het OAIS-model zes functionele entiteiten, die hierna in detail besproken worden.

1. Ingest
2. Data Management
3. Storage
4. Preservation Planning
5. Administration
6. Dissemination/Access.

6.1 Ingest



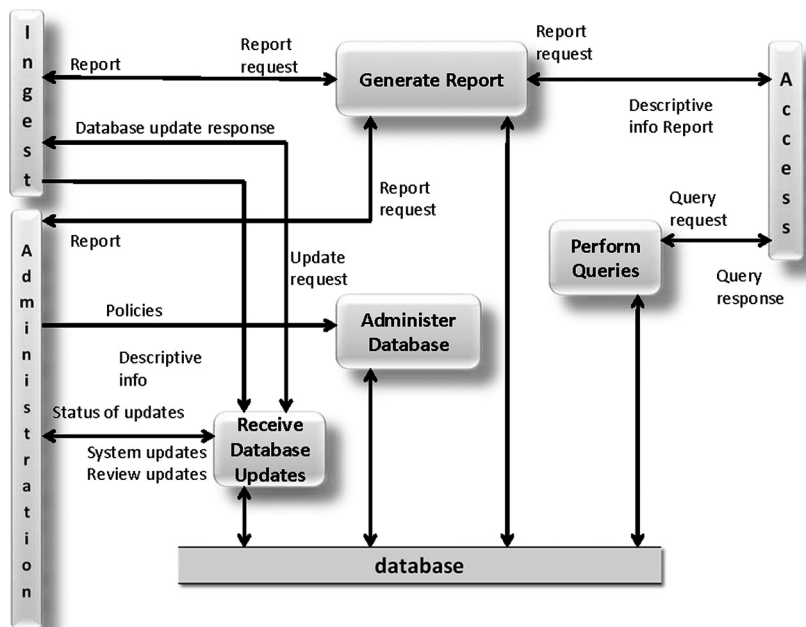
Ingest is het proces van de ontvangst van het Data object, tot en met het gereed maken van de AIP ter archivering. De functie Quality Assurance controleert of de aangeleverde SIPs correct overgebracht zijn van de producer naar de (tijdelijke) opslag van het OAIS-archief. Dit gebeurt meestal via checksums. Daarnaast kunnen er in de Ingest-functie nog meer controles plaatsvinden, als viruscontrole, of identificatie van voorkomende fileformaten. Tevens vindt er controle plaats of het geleverde overeenkomt met de afspraken uit het Submission Agreement.

Uiteindelijk wordt er, op basis van een of meerdere SIPs, een AIP aangemaakt, die voldoet aan de AIP specificatie van dat OAIS-archief, eventueel na het toevoegen van extra informatie. In principe is de AIP dan gereed voor invoer in de Archival Storage, maar voordat dit gebeurt, zal Administration de AIP nog op correctheid en volledigheid controleren (Audit Submission).

Naast de AIP, maakt de Ingest functie ook de metadata voor toegang aan, de descriptive information. Dit kan op basis van informatie, die ook in de AIP staat en door de Producer al is meegegeven. Het kan ook nieuwe informatie zijn.

Uiteindelijk regelt de functie Co-ordinate updates de overdracht van de AIPs naar Archival Storage en de Descriptive Information naar Data Management.

6.2 Data Management

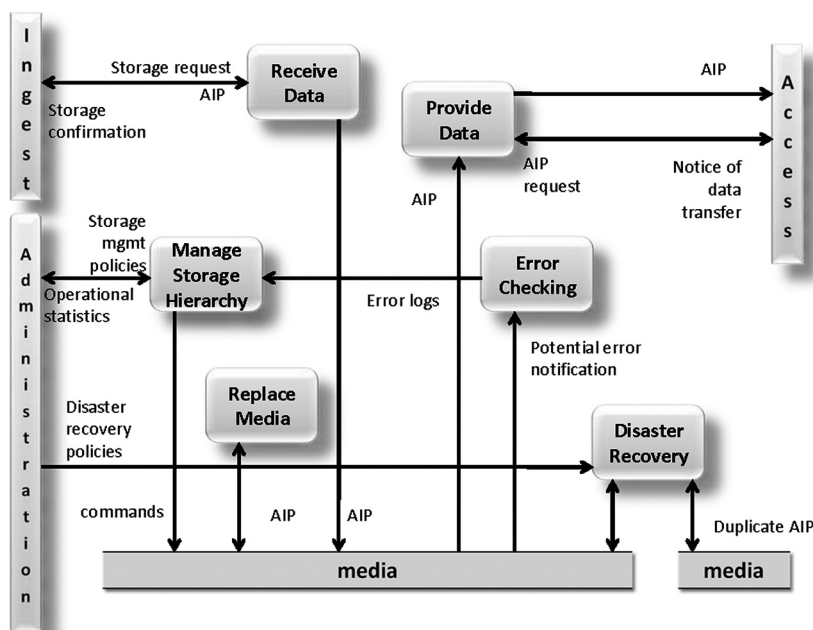


Data management is de plaats in het archief waar diverse informatie over de AIPs opgeslagen wordt, onder meer de toegangsinformatie (Descriptive Information). Voor beheer is het van belang snel te kunnen zoeken op informatie uit de AIP. Dit werkt te omslachtig als er rechtstreeks op Archival Storage gezocht moet worden, omdat dan elke AIP ‘open’ gemaakt moet worden voordat men de juiste informatie heeft. Dus is er vaak een duplicaat van die informatie in Data Management opgeslagen, met eventuele aanvullingen.

Er is dus altijd een onverbreekelijke relatie tussen een AIP en de informatie daarover in Data Management. Data Management houdt ook bij waar (op de hardware) een AIP is opgeslagen.

Verschillende functies in het OAIS-archief vragen via rapportages informatie uit Data Management op (Perform Queries, Generate Report).

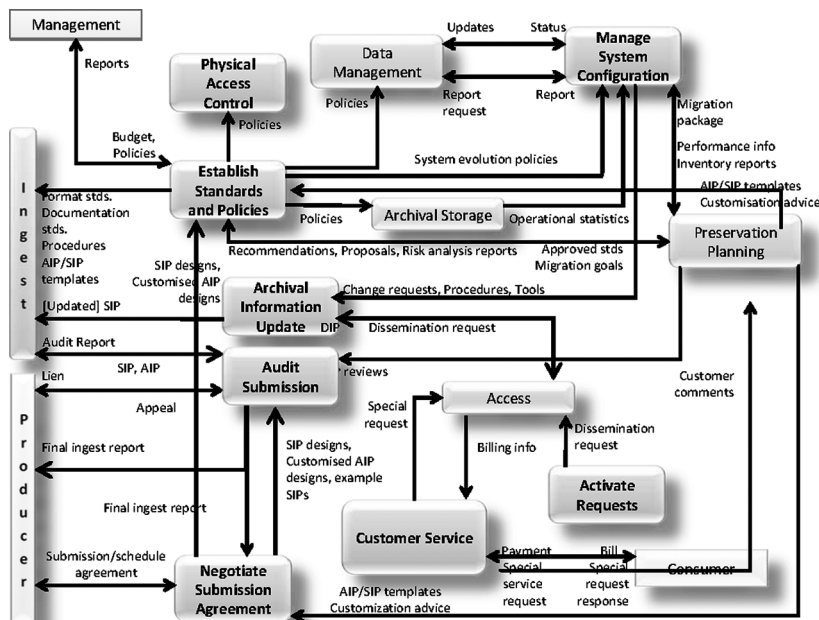
6.3 Archival Storage



Archival Storage regelt de permanente opslag van de AIPs die het via Ingest ontvangt (Receive Data). Tevens behandelt het de verzoeken om duplicaten van AIPs te leveren af (Provide Data), die via de Access-functie komen en daarna als DIP aan de Consumer worden geleverd. Archival Storage regelt ook waar de AIP wordt opgeslagen (Manage Storage Hierarchy). Je kunt je voorstellen, dat er meerdere opslagfaciliteiten zijn en dat er afspraken bestaan waar welke AIPs opgeslagen worden

(dit is overigens in Administration bepaald). Daarnaast omvat Archival Storage functies die zorg dragen voor de integriteit van de opgeslagen data (Error Checking), het opvangen van ernstige storingen (Disaster Recovery, door het aanleggen van back ups) en het periodiek vervangen van de hardware waarop de informatie is opgeslagen(Replace Media). Ook hier is er een nauwe relatie met de Administration-functie, die de randvoorwaarden bepaalt.

6.4 Administration



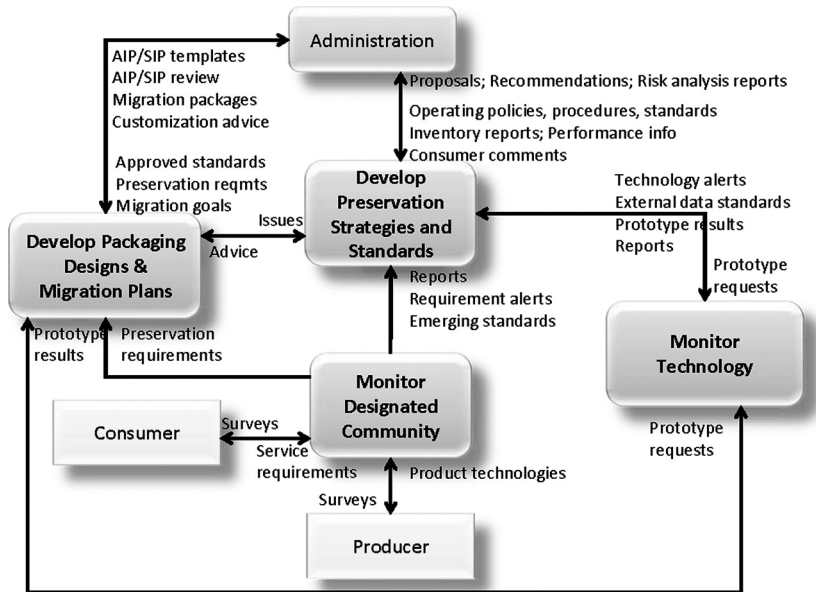
Onder Administration vallen alle diensten/services en functies/ taken die met het dagelijks beheer van alle overige functies samenhangen. Het is een spin in het web van het OAIS-archief. Administration regelt alle afspraken rondom het leveren van materialen door Producers aan het OAIS-archief (via de functie via de Negotiate Submission Agreement). Het Submission Agreement bevat onder meer: de afspraken over

toegangsrechten en preservation rechten, het tijdschema en de wijze van levering en een detailbeschrijving van de structuur van de aan te leveren SIP. Administration stelt vervolgens de bijbehorende procedures op om het materiaal binnen te halen en regelt de terugkoppeling over hetverloop van het proces. Binnen Administration zal ook een test van de SIP en AIP plaatsvinden (Audit Submission).

In de loop der tijd kunnen de eisen aan de samenstelling van SIP en AIP wijzigen. De functie Establish Standards and Policies is verantwoordelijk voor het vastleggen en onderhouden van de standaarden en het beleid dat het OAIS-archief hanteert. En zal dus reageren op wijzigingen. In dit geval zal een voorstel tot wijziging van de SIP en de AIP worden opgesteld. Vervolgens gaat dit ter implementatie naar Ingest. Preservation Planning kan bijvoorbeeld op basis van haar monitoringactiviteiten periodiek adviseren tot het doen van aanpassingen, bijvoorbeeld ten aanzien van toegestane file-formaten. En last but not least, de functie Management levert de belangrijke randvoorwaarden en beleidsuitgangspunten voor het OAIS-archief (missie, budget, etc.).

Daarnaast beheert Administration het systeem (Manage System Configuration) en is verantwoordelijk voor de soft- en hardware en de controle op de toegang tot deze faciliteiten (Physical Access Control). Input zijn daarvoor performance metingen en foutrapportages. Daarnaast informatie uit andere functies, zoals informatie over technische ontwikkelingen die van belang kunnen zijn (via Preservation Planning). Deze input kan leiden tot een noodzakelijke aanpassing van en tot actie bij de functie Archival Information Update, bijvoorbeeld een transformatie (zie paragraaf 7) naar een ander file format.

6.5 Preservation Planning

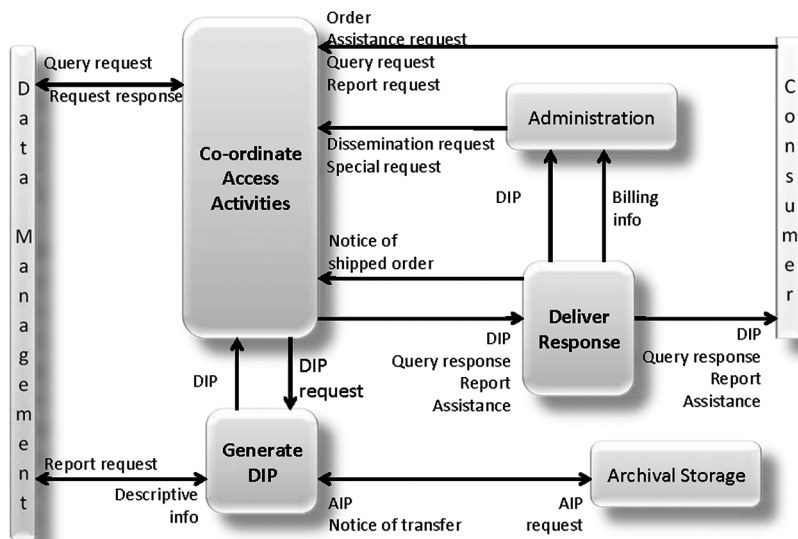


De Preservation Planning-functie monitort de omgeving van het OAIS-archief en waarschuwt, op basis van de verkregen informatie, andere onderdelen van het OAIS-archief maatregelen te nemen. Een belangrijk terrein daarbij is de Designated Community. De wereld van de Consumer en van de Producer van informatie zal tijdens het bestaan van het OAIS-archief dat immers gericht is op de lange termijn, veranderen. Het OAIS-archief moet op deze veranderingen inspelen, anders kan het zijn missie niet vervullen. Dit kan door bijvoorbeeld zijn Consumers een andere DIP aan te bieden (meer gericht op eigentijdse manieren van omgang met digitaal materiaal, zoals een tablet) of door andere file formaten van de Producer te accepteren. Het monitoren kan ook op actieve wijze plaatsvinden: door periodiek vragenlijsten te sturen of gesprekken te voeren met vertegenwoordigers van de Designated Community. Dit kan leiden tot het toepassen van nieuwe of verbeterde standaarden door het OAIS-archief (via de functie Develop Preservation Strategies and

Standards). Het kan ook leiden tot aanpassingen in de Information Packages, die dan de functie Developing Packing Designs aanpast en doorstuurt naar Administration ter verwerking.

Het tweede, brede terrein dat onder de monitoringfunctie valt, is het bijhouden en reageren op technische vernieuwingen. Door de snelle ontwikkelingen in file formaten, software en hardware bestaat het risico dat het OAIS-archief door onkunde en onwetendheid de aansluiting op deze ontwikkelingen mist. Het OAIS-archief kan dan de eigen missie niet meer uitvoeren. Deze functie heeft, net als de vorige monitoringfunctie, een nauwe relatie met de functies Develop Preservation Strategies and Standards en Developing packaging Designs. Daarnaast kan Preservation Planning aanbevelen een migratieplan te ontwikkelen, omdat bijvoorbeeld een file format niet langer meer door de leverancier ondersteund wordt. De Preservation Planning-functie is ook verantwoordelijk om periodiek een risico-inventarisatie te doen en de resultaten daarvan door te sturen naar Administration.

6.6 Access



De Access-functie handelt de informatieverzoeken af uit het OAIS-archief. De belangrijkste klant zal de Consumer zijn of iemand uit de Designated Community, die via een DIP de gevraagde informatie krijgt. Vaak zal een OAIS-archief meerdere varianten DIPs ontwikkelen om aan verschillende behoeften tegemoet te komen. Bijvoorbeeld om alleen de metadata te tonen, of het Content Object in verschillende varianten te tonen (specifiek file format, thumbnail etc.). Via de functie Generate DIP, ontvangt de functie uit Data management een kopie van de beschrijvende metadata, en uit de Archival Storage, een kopie van de AIP. Samen leidt dit tot een DIP voor de Consumer. Uiteraard ziet de Access-functie erop toe dat de rechten ten aanzien van toegang daarbij worden gerespecteerd, zoals afgesproken met de Producer. Waar nodig, regelt de Access-functie ook de financiële afhandeling.

Naast het leveren van één specifiek object via een DIP, ondersteunt de Access-functie ook de aanmaak van rapportages.

7 Activiteiten om de duurzame toegankelijkheid te bewaren

Op termijn is het onvermijdelijk dat een OAIS-archief acties onderneemt om de langetermijntoegang te kunnen waarborgen. Je kunt daarbij denken aan vervanging van hardware (mediamigratie) of het migreren van de objecten naar een ander file format. Deze 'Preservation Actions' zijn kostbare en risicovolle activiteiten. Een van de belangrijkste risico's is dat er informatie verdwijnt door deze handelingen.

7.1 Migratie

Bij migratie kun je onderscheid maken tussen activiteiten waarbij de bitreeks wijzigt en die waarbij de bitreeks ongewijzigd blijft.

De bitreeks blijft ongewijzigd bij 'Refreshment'. Bij 'refreshment' migreert de AIP van een opslagmedium naar een identiek ander opslagmedium. Er verandert dus niets aan de bits. Bijvoorbeeld een bestand op tape dat weggeschreven wordt naar een andere, identieke tape.

Ook bij 'Replication' wijzigt er niets aan de AIP, maar wordt deze op een ander opslagmedium geplaatst. Denk bijvoorbeeld aan een back-up van een bestand, dat op optical disc staat en een back-up daarvan die op tape

staat. Het verschil met refreshment is dat bij replication de verwijzingen (mappings) in de Archival Storage ook wijzigen, ze verwijzen immers naar een andere opslag. Bij refreshment kan deze mapping-structuur gehandhaafd blijven.

De bitreeks van het Content Object kan wijzigen wanneer het OAIS-archief kiest voor andere software voor het digitale archief en de samenstelling van de AIP wijzigt. Dit heet 'repackaging'. In dit geval zal het Data Object niet wijzigen, maar wel de Packaging Information.

De grootste wijzingen treden op als ook het Data Object wijzigt. Dit gebeurt, als bijvoorbeeld het Data object van file format A, omgezet wordt naar file format B. In het spraakgebruik heet dat vaak 'migratie', maar de OAIS-term is 'transformatie': elke migratie waarbij er een wijziging in de content is.

We onderscheiden een transformatie die omkeerbaar is, en één die niet omkeerbaar is. Is de transformatie omkeerbaar, dan kun je het proces weer terugdraaien. Bijvoorbeeld een ASCII-tekst die je omzet naar een UTF-8-tekst: dat is een 1-op-1-transformatie die je ook kunt terugdraaien. Bij een niet-omkeerbare transformatie is er geen 1-op-1-relatie tussen de bitreeks in het origineel en de afgeleide. Bijvoorbeeld een transformatie van een TIFF-file naar een JPEG 2000-file. Deze omzetting kan leiden tot het verlies van bits die niet terugkeren als je de resulterende JPEG 2000-file weer naar TIFF transformeert. Het is daarom belangrijk, van tevoren vast te stellen welke eigenschappen tijdens de transformatie bewaard moeten blijven, zodat bij vooraf testen al gekeken kan worden of bij de transformatie deze eigenschappen bewaard blijven. Deze eigenschappen noemen we Transformation Information Properties, nauw gerelateerd aan wat veelal aangeduid wordt met 'significant properties', de karakteristieke eigenschappen van een object: bijvoorbeeld: kleur en layout. Deze informatie is in hoge mate gerelateerd aan het begrip authenticiteit. Aangezien het bewaren van de authenticiteit een van de doelstellingen van het OAIS-archief zal zijn, zal men proberen deze 'Transformation Information Properties' in kaart te brengen en bij transformatie te bewaren.

Een transformatie is het enige geval van migratie, waarbij een nieuwe AIP wordt gecreëerd. Deze werkwijze zal zijn: van de AIP's die voor transformatie in aanmerking komen, wordt een DIP aangemaakt. Deze

DIP's ondergaan de transformatie en krijgen de bijbehorende (event-) metadata ten behoeve van de Provenance. Vervolgens worden ze weer als SIP aan het OAIS-archief aangeboden en ingelezen. De originele AIP zit dan nog in het archief, de nieuwe AIP is daar een versie van. De gebruiker die de AIP later opvraagt, kan met deze informatie de geschiedenis van het object zien en dus de authenticiteit ervan beoordelen. In de andere genoemde gevallen van migratie ontstaat er dus geen nieuwe AIP.

7.2 *Emulatie*

Het is niet altijd nodig om een AIP te migreren, als de toegankelijkheid in gevaar is. Er zijn ook mogelijkheden om in de Access-functie acties te doen, zodat het object op de originele wijze getoond wordt, ook al wordt het file format niet langer ondersteund door bijpassende toegangsoftware. Denk bijvoorbeeld aan een website, die feitelijk alleen door een contemporaine browser in al zijn functionaliteit ondersteund wordt. Emulatie kan hiervoor een goede strategie zijn. Hierbij wordt de originele software gedraaid op een nagebouwde versie van de originele hardware-omgeving, die op zijn beurt draait op moderne versie van de omgeving. Het oorspronkelijke object blijft onaangetast (de bits hoeven niet gewijzigd te worden, zoals in geval van transformatie) en de gebruiker ziet het object zoals het oorspronkelijk getoond werd. Ook dit draagt bij aan de authenticiteit van het object.

8 **Audit en certificering**

Hoewel het OAIS een raamwerk geeft voor digitale archieven, is dit nog geen garantie dat de implementatie ervan ook resulteert in een betrouwbaar ('trustworthy') OAIS-archief. Een archief waar men ervan verzekerd is dat de opgeslagen data permanent toegankelijk blijven. Daarbij spelen namelijk, behalve het systeem dat gebruikt wordt, meerdere factoren een belangrijke rol, zoals de inrichting van de organisatie, de financiële dekking en de beveiligingsmaatregelen.

De mate van abstractie binnen het OAIS-model leidt tot meerdere interpretaties. Al in een vroeg stadium was er behoefte aan een vorm van

certificering. Dit resulteerde in verschillende initiatieven als Trustworthy Repositories Audit & Certification Criteria en Check-list (TRAC), gepubliceerd in 2007 door het Center for Research Libraries, DRAMBORA (Digital Repository Audit Method Based on Risk Assessment) in 2009 en de Kriteriencatalog vertrauenswürdige digitale Langzeitarchive van Nestor (2008). Op initiatief van CCSDS is TRAC omgewerkt tot de, in 2012 gepubliceerd, ISO-standaard 16363 *Audit and Certification of Trustworthy Digital Repositories*. De bijbehorende concept standaard ISO 16969 *Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories* regelt de benoeming van auditors. Wanneer ook deze standaard vastgesteld is, kan een OAIS-archief een officiële ISO-audit ondergaan en gecertificeerd worden.

9 Tot slot

Zoals reeds in de inleiding is opgemerkt, biedt het OAIS-model voor elke organisatie die digitale informatie voor de lange termijn toegankelijk wil houden, een heldere en samenhangende leidraad voor het opzetten van een digitaal archief. In de loop der tijd zijn er standaarden ontwikkeld, die op basis van het OAIS-model, delen nader hebben uitgewerkt, zoals de Provenance informatie (Premis), de samenstelling van de Submission Agreement (PAIS), het Submission Information Package (PAIMAS). Daarnaast ontstonden er verschillende toetsingsmechanismen (Drambora, RAC, DSA) en zal het binnen afzienbare tijd mogelijk zijn een archief een officiële ISO-audit te laten ondergaan. Daarmee is de cirkel rond: er is een set van uitgangspunten en er is een mogelijkheid tot onafhankelijke, periodieke toetsing.

De digitale wereld is er een van snelle veranderingen en de tijd zal leren of het OAIS-model robuust genoeg blijkt te zijn om deze digitale uitdagingen te verwerken. Het ISO-proces voor periodieke herijking van de standaard biedt iedereen de gelegenheid wijzigingen voor te stellen. Deze publieke toetsing biedt ons allen de kans een praktische en werkbare standaard te onderhouden.

Illustraties: Reprinted with permission of the Consultative Committee for Space Data Systems

Literatuur

- ISO (www.iso.org)
- DRAMBORA Digital Repository Audit Method based on Risk Assessment, <http://www.repositoryaudit.eu/>
- DSA Data Seal of Approval www.datasealofapproval.org
- ED₃ Eisen duurzaam digitaal depot. Toetsingskader voor de beheersomgeving van blijvend te bewaren digitale informatie. Landelijk Overleg Provinciale Archief Inspecteurs. 2008 [http://www.lopai.nl/pdf/ED₃_v1.pdf](http://www.lopai.nl/pdf/ED3_v1.pdf)
- Lee, C.A., *Defining digital preservation work: A case study of the development of the reference model for an open archival information system*. 2005
- OAIS *Reference model for an Open Archival Information System (OAIS)*. Recommended practice CCSDS 650.0-m-2 Magenta Book. Issue 2. Washington, D.C.: CCSDS, June 2012. Idem aan ISO 14721:2012 (<http://public.ccsds.org/publications/archive/650xom2.pdf>)
- PAIMAS *Producer-Archive Interface Methodology Abstract Standard*. Recommendation for Space Data System Standards, CCSDS 651.0-M-1. Magenta Book. Issue 1. Washington, D.C.: CCSDS, May 2004. Idem aan ISO 20652:2006. <http://public.ccsds.org/publications/archive/651xom1.pdf>
- PREMIS Data Dictionary for Preservation Metadata, version 2.2, July 2012 <http://www.loc.gov/standards/premis/>
- TDR, Audit and Certification of trustworthy digital repositories. Recommended practice. CCSDS 652.0-M-1 MAGENTA BOOK. Washington, D.C.: CCSDS September 2011. Idem aan ISO 16363. <http://www.iso16363.org/standards/>
- TRAC Trustworthy Repositories Audit & Certification: Criteria and Checklist. OCLC/CRL, 2007 www.crl.edu/PDF/trac.pdf

