

Digitale containers voor het digitaal archiefdepot

Filip Boudrez
Expertisecentrum DAVID
Antwerpen, 2005

0. INHOUD

1. Inleiding.....	1
2. De DAVID bewaarstrategie.....	2
3. Ingekapselde digitale archiefdocumenten	4
3.1 Inkapseling als opslagmethode.....	4
3.2 Archiefdocumenten als Archival Information Packages	6
3.3 AIP's als containerbestanden.....	7
3.3 Van metadata voorziene archiefdocumenten.....	7
4. Een model AIP.....	10
4.1 De AIP-structuur.....	10
4.2 XML als bestandsformaat.....	11
5. XML Schema's.....	12
5.1 AIP.....	13
5.2 Metadataschema voor digitale archiefdocumenten.....	14
5.3 Archiefbeschrijving.....	14
5.4 E-mail, agenda's en databanken.....	14
6. Samenstellen van de XML-AIP's.....	15
7. Besluit.....	16

1. INLEIDING

Nu digitaal documentbeheer en digitale archivering alsmaar meer geïmplementeerd geraken, komt de dag dichterbij waarop digitale archiefdocumenten aan archiefdiensten of -instellingen worden overgedragen. Deze digitale archiefdocumenten worden idealiter in speciaal ingerichte digitale archiefdepots bewaard.

Momenteel lopen diverse onderzoeken en pilootprojecten met digitale depots als onderwerp¹. Vanuit het bijzonder karakter van digitale archiefdocumenten - die zoveel meer zijn dan louter digitale objecten² - gelden speciale vereisten voor de organisatie, de functionaliteiten en de workflow van een digitaal archiefdepot. Een belangrijk aandachtspunt dat men zeker niet uit het oog mag verliezen, is de digitale duurzaamheid van het digitaal depot zelf. Elk informatiesysteem, en dus ook een digitaal depot, is immers onderhevig aan technologische veroudering. Idealiter wordt een duurzame opslagmethode toegepast die niet dient aangepast te worden telkens een hardware of softwarecomponent van het digitaal depot wordt vervangen. De infrastructuur van een digitaal depot moet hierop voorzien zijn.

¹ Voorbeelden van initiatieven door nationale archiefdiensten zijn ondermeer: National Archives of Australia (AtoR-project), National Archives and Records Administration, Schweizerisches Bundesarchiv (Arela-project), enz. Een project dichterbij huis is het e-Depot van het Gemeentearchief Rotterdam en de Archiefschool.

² Zie hiervoor: K. THIBODEAU, *Boundaries and transformations: an object oriented strategy for the preservation of electronic records*, in: *Proceedings of the DLM-Forum on electronic records*, Brussel, 1996, p. 161-167; F. BOUDREZ, *Inleiding. Het digitaal archiefdocument*, in: F. BOUDREZ EN H. DEKEYSER, *Digitaal archiefbeheer in de praktijk. Handboek*, Antwerpen-Leuven, 2004.

De complexiteit van deze laatste kwestie verhoudt zich rechtstreeks met de wijze waarop de inhoud van het digitaal depot is georganiseerd en samengesteld. De archiefbeheerder kan op dit probleem anticiperen door bij de keuze van een opslagwijze voor de digitale archiefdocumenten en hun metadata hier al rekening mee te houden. Een goedgekozen opslagwijze kan het beheer van de digitale archiefdocumenten op intellectueel en technologisch vlak eenvoudig houden, hen ook beter wapenen tegen calamiteiten en risico's tot een minimum beperken.

In deze bijdrage wordt de digitale opslagmethode toegelicht die eDAVID ontwikkelde en die door de stad Antwerpen zal worden toegepast. Deze opslagmethode bouwt verder op de digitale bewaarstrategie voor digitale archiefdocumenten die het DAVID-project ontwikkelde. Dit artikel start dan ook met een beschrijving van deze bewaarstrategie. Zonder een duidelijk zicht op de te volgen bewaarstrategie is het quasi onmogelijk om de inhoud van het digitaal depot op een gestructureerde wijze te organiseren. De hoofdaandacht gaat in het tweede deel van dit artikel niet alleen uit naar een opslagwijze die bij deze digitale bewaarstrategie aansluit, maar die ook een oplossing biedt voor de archivering van digitale archiefdocumenten in relatie met hun metadata. Hierbij wordt de invalshoek van het Open Archival Information System (OAIS)³ gevolgd. In het bijzonder staat de samenstelling en het beheer van *Archival Information Packages* (AIP) centraal. In het derde deel van deze bijdrage wordt een modelstructuur en bijhorend formaat voor een AIP voorgesteld. Aansluitend worden de XML Schema's voorgesteld zoals zij door eDAVID zijn ontworpen voor de stad Antwerpen. XML zal immers volop worden gebruikt als archiverings-, metadata- en inkapselingsformaat in het digitaal archiefdepot van de stad Antwerpen. Deze XML Schema's gelden als implementatievoorbeeld en worden in het voorlaatste deel van deze bijdrage gedocumenteerd. Tot slot wordt beschreven hoe de voorgestelde opslagmethode praktisch kan worden toegepast.

2. DE DAVID BEWAARSTRATEGIE

Alvorens de archiefbeheerder kan onderzoeken hoe de digitale informatie in het digitaal depot wordt georganiseerd, dient hij zicht te hebben op de wijze waarop men in de toekomst op basis van de gearchiveerde bits en bytes opnieuw digitale archiefdocumenten op scherm reconstrueert. De digitale bewaarstrategie die een organisatie toepast, zal immers bepalen welke informatie wordt gearchiveerd en hoe die in het digitaal archiefdepot wordt beheerd.

Het DAVID-project⁴ evalueerde de bestaande digitale bewaarstrategieën voor digitale objecten en onderzocht in welke mate zij geschikt zijn om digitale archiefdocumenten op lange termijn raadpleegbaar te houden. Dit hield ondermeer in dat niet alleen werd bekeken hoe digitale archiefdocumenten op termijn raadpleegbaar blijven, maar dat ook werd nagegaan hoe de authenticiteit en interpreteerbaarheid van de digitale archiefdocumenten wordt gewaarborgd.

Dit onderzoek wees uit dat op dit ogenblik migratie en emulatie de potentieel meest geschikte digitale bewaarstrategieën zijn⁵. Over de voor- en nadelen van beide strategieën is al heel wat inkt gevloeid⁶, maar in essentie sluiten migratie en emulatie elkaar niet uit. Beide strategieën zijn zelfs complementair: migratie en emulatie zijn elk het meest geschikt voor bepaalde types archiefdocumenten. Voor tekstuele documenten zonder enige functionaliteit is migratie wellicht de eenvoudigste bewaarstrategie, terwijl voor dynamische archiefdocumenten met een bepaald gedrag emulatie de meest aangewezen optie is. Tijdens de levensloop van een archiefdocument kunnen een

³ ISO-14721(2003): *Space data and information transfer systems. Open archival information system. Reference model*

⁴ <http://www.antwerpen.be/david>

⁵ F. BOUDREZ, *Bewaarstrategieën*, in: F. BOUDREZ EN H. DEKEYSER, *Digitaal archiefbeheer in de praktijk. Handboek*, Antwerpen-Leuven, 2004.

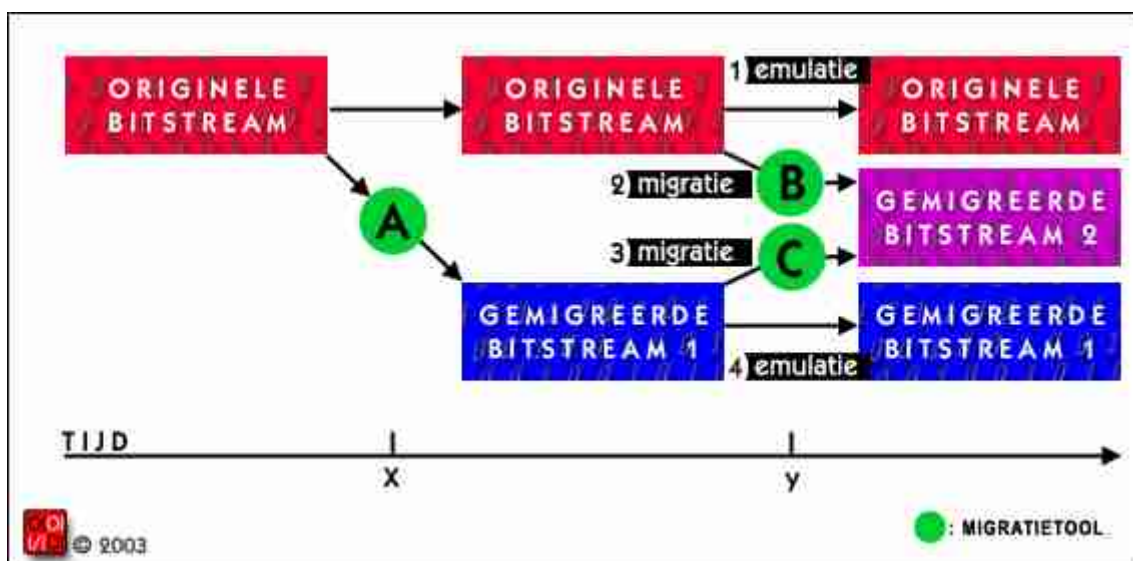
⁶ Zie bijv. J. ROTHENBERG, *Ensuring the longevity of digital information*, 1999 (<http://www.clir.org/pubs/archives/ensuring.pdf>); D. BEARMAN, *Reality and chimeras in the preservation of electronic records*, in: *D-Lib Magazine*, april 1999 (<http://www.dlib.org/dlib/april99/bearman/04bearman.html>)

migratie- en emulatiefase elkaar ook opvolgen. Emulatie heeft immers de beste kans op slagen bij archiefdocumenten in een open, gedocumenteerd en gestandaardiseerd bestandsformaat zodat bij emulatie van een gesloten formaat migratie naar een open bestandsformaat wellicht een eerste en essentiële stap is.

Om zoveel mogelijk leesbaarheidsgaranties te bieden en om zowel de migratie- als emulatiemogelijkheid open te houden, worden de archiefdocumenten in zowel hun origineel bestandsformaat als hun archiveringsformaat in het digitaal depot opgenomen. Elk formaat kan beschouwd worden als een andere representatie van hetzelfde archiefdocument. In de praktijk zullen echter niet altijd exact twee digitale objecten voor één digitaal archiefdocument worden gearcheveerd. In de gevallen waarbij het origineel formaat onmiddellijk ook het geschikt archiveringsformaat voor dat documenttype is, zal slechts één digitaal object voor het archiefdocument in het archief worden opgenomen. Anderzijds is het ook mogelijk dat voor een bepaald type digitaal document meerdere archiveringsformaten worden gebruikt, zodat bijvoorbeeld drie digitale objecten worden gearcheveerd.

Wanneer in de toekomst de vereiste hard- en software voor het raadplegen van het origineel formaat en het archiveringsformaat ontbreken, heeft men met de DAVID-bewaarstrategie op zijn minst vier opties om zijn archiefdocumenten op scherm te reconstrueren:

- emulatie van het origineel formaat
- migratie van het origineel formaat
- migratie van een geschikt archiveringsformaat
- emulatie van een geschikt archiveringsformaat



Door zowel de originele als gemigreerde bitstream in het digitaal depot op te nemen, anticipeert men ook op de technologische evolutie in de toekomst. De IT blijft snel evolueren en het valt momenteel niet te voorspellen wat in de toekomst allemaal mogelijk zal zijn. Zo bestaat de kans dat op termijn digitale documenten raadpleegbaar zijn met een technologie die op het moment van opname in het archief nog niet beschikbaar is.

Deze bewaarstrategie biedt naast zoveel mogelijk leesbaarheidsgaranties ook nog enkele andere voordelen. Ten eerste kunnen archiefgebruikers afhankelijk van hun voorkeur of van de software waarover ze beschikken een digitaal archiefdocument in zowel de originele bitstream als een gemigreerde bitstream raadplegen. Ten tweede blijft door het archiveren van de originele bitstream authenticatie mogelijk op basis van technologieën die gebaseerd zijn op de oorspronkelijke bitstreams.

Geavanceerde digitale handtekeningen zijn hier een voorbeeld van. Voorwaarde is wel dat alle elementen van de 'validation chain' beschikbaar zijn⁷. Ten derde kunnen archiefdocumenten in hun originele en gemigreerde bitstream worden vergeleken.

3. INGEKAPSELDE DIGITALE ARCHIEFDOCUMENTEN

De DAVID-bewaarstrategie houdt in dat één of meerdere representaties van eenzelfde digitaal archiefdocument in het digitaal archief worden bijgehouden. Naast de representaties dienen nog de metadata van het archiefdocument opgeslagen te worden. Voor de opslag van de verschillende representaties en de metadata van een digitaal archiefdocument in een digitaal archiefdepot heeft de archiefbeheerder diverse opties. De organisatie van het digitaal depot hangt sterk af van wat de archivaris als basiseenheid binnen het digitaal depot verkiest en van de wijze waarop de archiefdocumenten worden geïdentificeerd en hun metadata worden opgeslagen.

3.1 Inkapseling als opslagmethode

Een uitermate belangrijk aandachtspunt bij de keuze van een opslagmethode is het identificeren en relateren van de verschillende componenten van een digitaal archiefdocument. De componenten van een archiefdocument vormen immers een logisch geheel die na overbrenging naar het archief nooit meer verloren mag gaan.

In de meeste opslagmethoden zullen de verschillende componenten van een digitaal archiefdocument geen fysieke eenheid vormen, maar op gescheiden plaatsen (een databank, een bestandssysteem of een combinatie van beide⁸) en als verschillende digitale objecten worden opgeslagen. Hun onderlinge band worden met behulp van links, databankrelaties, pointers en bestandsnamen aangegeven. Het archiveren van die relaties op (middel)lange termijn is geen evidentie. De snel evoluerende informatietechnologie vraagt dat de band tussen de digitale objecten op een duidelijke en duurzame wijze wordt vastgelegd. Dit is geen onoverkomelijk probleem, maar is wel een belangrijk aandachtspunt en kan op termijn wel een uitdaging inhouden. Bovendien bestaat het gevaar altijd dat de relatie verloren geraakt.

Het gescheiden bewaren van de componenten van een digitaal archiefdocument houdt altijd een risico in. Van zodra de onderlinge relaties verbroken en niet meer reconstrueerbaar zijn, dient het archiefdocument als verloren beschouwd te worden. Metadata zijn immers essentieel voor de archiefstatus en de langetermijnleesbaarheid van het archiefdocument. De archiefbeheerder kan dit risico vermijden door metadata mee op te nemen in de computerbestanden die de documenten bevatten. Door beide componenten in één fysiek object samen te brengen wordt vermeden dat de relatie tussen het archiefdocument en zijn metadata verloren geraakt.

Het toevoegen van metadata aan digitale objecten wordt inkapseling ('encapsulation') of inbedding genoemd. Inkapseling wordt soms in één adem met migratie en emulatie als digitale bewaarstrategie vermeld⁹, maar hoort strikt genomen niet in dit rijtje thuis. Inkapseling is eigenlijk niets meer dan een opslagtechniek waarbij metadata aan een digitaal object wordt toegevoegd en/of meerdere documenten in één digitaal object worden gegroepeerd. Inkapseling is geen methode die voorschrijft

⁷ Zie F. BOUDREZ, *Digitale handtekeningen en archiefdocumenten*, Antwerpen, 2005.

⁸ Mogelijkheden zijn:

- metadata: databank / representaties: databank (BLOB's)
- metadata: databank / representaties: bestandssysteem
- metadata: bestandssysteem / representaties: bestandssysteem.

⁹ Bijv. <http://www.nla.gov.au/padi/topics/18.html>; TESTBED DIGITALE BEWARING, *XML en digitale bewaring*, Den Haag, 2002.

hoe digitale documenten in de toekomst op scherm worden gereconstrueerd of hoe de toegankelijkheid wordt bewaard.

Het idee om metadata in de digitale objecten in te kapselen is niet nieuw. Inkapseling is één van de basisprincipes van object georiënteerd programmeren. Het belang van inkapseling voor de digitale archivering werd in 1996/97 sterk gepromoot door David Bearman¹⁰. Ondertussen wordt inkapseling ondermeer toegepast in de *Persistent Object Preservation* van het NARA, de VERS-archiveringsstrategie van de Public Record Office van Victoria en het AtoR-project van de National Archives van Australië¹¹.

Het verrijken van archiefdocumenten met metadata is geen absolute voorwaarde voor duurzame digitale archivering, maar is meer dan het overwegen waard aangezien dit belangrijke voordelen oplevert:

- de metadata vormen een onderdeel van het gearchiveerd digitaal object en worden niet op een externe locatie opgeslagen. De metadata zijn onlosmakelijk verbonden met het archiefdocument. Men hoeft zich geen zorgen te maken om links of pointers tussen digitale objecten en hun metadata. Inkapseling vergemakkelijkt ook het beheer op (middel)lange termijn.
- alle componenten van een digitaal archiefdocument kunnen gemakkelijk samen worden getransfereerd en gemigreerd
- de digitale archiefdocumenten zijn zelfbeschrijvend en autonoom: ze identificeren en documenteren zichzelf
- de ingekapselde metadata kunnen ten allen tijde geëxtraheerd en centraal worden opgeslagen
- de digitale objecten in het digitaal depot hebben de status van archiefdocument zonder dat ze hiervoor externe informatie nodig hebben. Digitale archiefdocumenten en niet digitale objecten vormen de basiseenheden van het archiefdepot.
- bij geval van ramp (bijv. crashen databank met de metadata) zijn de gevolgen minder erg (risk assessment):
 - het digitaal archiefdepot bevat nog steeds archiefdocumenten
 - de metadata zijn snel reconstrueerbaar.

De archiefbeheerder kan op diverse wijzen inkapseling als opslagmethode toepassen. Een eerste methode is metadata opnemen in de computerbestanden die één bepaalde representatie van het archiefdocument bevatten (bijv. invullen van de metadatatags in de header van een TIFF-bestand). Deze inkapselingsmethode helpt de archiefbeheerder niet veel verder. Bij archiefdocumenten met meerdere representaties worden dezelfde metadata dan meermaals bijgehouden en de verschillende representaties dienen nog steeds op één of andere manier aan elkaar te worden gerelateerd. Ook op technisch vlak zijn er enkele minpunten aan deze inkapselingsmethode. Het inkapselen van metadata levert bij pure tekstbestanden niet al te veel problemen op, maar bij binaire bestandsformaten is dit veel minder vanzelfsprekend. De meeste binaire bestandsformaten voorzien wel standaard een aantal velden voor het insluiten van metadata, maar deze voldoen niet aan alle archiefnoden. In de voorziene metadatavelden is doorgaans te weinig plaats voorzien en het uitbreiden van de velden kan uitwisselings- en leesbaarheidsproblemen opleveren. Bovendien vraagt het toevoegen van metadata aan binaire bestanden voor elk formaat een afzonderlijke module of softwaretool, want een dergelijke functionaliteit wordt lang niet altijd door courante computerprogramma's ondersteund¹².

¹⁰ Zie bijvoorbeeld: D. BEARMAN, *Item level control and electronic recordkeeping*, in: *Archives and museum informatics*, vol. 10 (3), p. 195-245; MARK D. GIGUERE, *Metadata-Enhanced Electronic Records*, Philadelphia, 1997.

¹¹ K. THIBODEAU, R. MOORE en C. BARU, *Persistent Object Preservation: Advanced computing infrastructure for digital preservation*, in: *Proceedings of the DLM-Forum on electronic records*, Brussel, 2000, p. 113-118; http://www.prov.vic.gov.au/vers/standard/advice_11; <http://www.naa.gov.au/recordkeeping/er/guidelines/10-preservation.html#pres6>.

¹² Met Adobe Photoshop kunnen bijvoorbeeld niet langer alle administratieve TIFF-tags worden ingevuld. Adobe trekt volop de kaart van zijn eigen eXtentible Metadata Platform en biedt in recente softwareversies niet de functionaliteit om de administratieve metadatatags van de TIFF-standaard te editeren.

Om deze nadelen op te vangen past de archivaris beter een andere inkapselingsmethode toe: het inkapselen van de verschillende representaties van hetzelfde archiefdocument in één computerbestand. Hierdoor moeten de verschillende representaties niet meer aan elkaar gelinkt te worden, want hun band wordt vastgelegd door ze in hetzelfde fysiek object op te nemen. Deze inkapselingsmethode biedt ook het voordeel dat de archiefbeschrijvende metadata slechts éénmaal hoeven opgenomen te worden. Bovendien hoeven de metadata niet meer in de headervelden van de binaire bestandsformaten ingekapseld te worden, maar kunnen ze gewoon in het computerbestand opgenomen worden. Hierdoor combineert de archivaris de twee inkapselingsmogelijkheden: enerzijds worden metadata toegevoegd terwijl anderzijds verschillende digitale objecten in hetzelfde computerbestand worden opgenomen. Het eindresultaat van deze inkapselingsmethode zijn zogenaamde containerbestanden¹³. Eén containerbestand bevat alle componenten van één digitaal archiefdocument.

Inkapseling als opslagmethode houdt ondermeer in dat digitale archiefdocumenten voor opname in het digitaal archief eerst verpakt worden in containerbestanden en dat ze pas geconsulteerd kunnen worden nadat ze worden uitgepakt. Het samenstellen van containerbestanden vraagt dus minstens twee extra stappen in de workflow van het digitaal depot.

3.2 Archiefdocumenten als Archival Information Packages

Voor het uitwerken van een opslagmethode voor archiefdocumenten in een digitaal archiefdepot is de OAIS-norm een geschikt referentiekader. In het bijzonder is het informatiemodel van deze algemeen aanvaarde ISO-norm een interessante leidraad. Binnen het OAIS-model vormen informatiepakketten de basiseenheid van het digitaal archief.

OAIS beschrijft de functies, de processen en de informatiestroom van een digitaal archief. Hierin staan de informatiepakketten die in het digitaal archief worden opgenomen, beheerd en geraadpleegd centraal¹⁴. In het OAIS-model is elk *Information Package* een conceptuele container die uit twee types informatie bestaat: *Content Information* en *Preservation Description Information*. De *Content Information* bevat naast de bitstream van de eigenlijke digitale informatiebron (het *Data Object*) ook zijn *Representation Information*. Met behulp van die *Representation Information* wordt een *Data Object* tot een begrijpbaar *Information Object* getransformeerd. De informatiepakketten bevat dus naast de eigenlijke bitstream van het gearchiveerd digitaal object diverse metadata:

- de *Representation Information*: alle informatie die nodig is om het digitale object te vertalen naar een interpreteerbaar concept (bijv. broncode of de gecompileerde installatiebestanden voor een viewer)
- de *Preservation Description Information*:
 - unieke identifiers (“reference information”)
 - informatie over de context (“context information”)
 - informatie over de herkomst (“provenance information”)
 - informatie voor de integriteit of de validatie van de inhoud (“fixity information”)¹⁵

De archiefbeheerder dient er wel rekening mee te houden dat in het OAIS-model een informatiepakket niet noodzakelijk samenvalt met één archiefdocument. OAIS richt zich op de langetermijnarchivering van digitale informatie in het algemeen. Eén AIP kan overeenkomen met één archiefdocument, maar

¹³ De fysieke of logische structuur waarbinnen de bitstreams van een document en de metadata worden samengebracht, worden doorgaans met de termen “containers” of “wrappers” aangeduid.

¹⁴ Naast de AIP's onderscheidt het OAIS-model ook nog twee andere Information Packages: de Submission Information Packages (SIP's) en de Dissemination Information Packages (DIP's). De SIP's worden door de archiefvormer aan de archiefbeheerder overgedragen die ze tot AIP's transformeert. Op basis van de AIP's worden DIP's onder de gebruikers verspreid.

¹⁵ ISO-14721 (2003): *Space data and information transfer systems -- Open archival information system -- Reference model*.

ook met één component van een archiefdocument of met meerdere archiefdocumenten. Binnen het OAIS-model kunnen bijgevolg diverse soorten AIP's worden samengesteld en de archivaris kan verschillende aggregatieniveaus kiezen. Samen met de relevante metadata kan een AIP als inhoud hebben:

- één component van een archiefdocument: bijv. een afbeelding ingevoegd in een tekstdocument
- alle componenten van één representatie van een archiefdocument: bijv. een tekstdocument (incl. de afbeelding(en) en de tekst)
- alle representaties van één archiefdocument: bijv. een tekstdocument in zijn origineel en zijn archiveringsformaat
- alle onderdelen van een archiefbestanddeel: bijv. alle archiefdocumenten die deel uitmaken van een dossier- of onderwerpsmap¹⁶.

Aangezien de archivering van digitale archiefdocumenten het uitgangspunt is, ligt het voor de hand dat de archivaris alle representaties van één archiefdocument als aggregatieniveau voor een AIP verkiest. Naast de bitstreams van de digitale objecten wordt ook metadata in de AIP opgenomen: technische metadata over elke representatie en archiefbeheersmetadata die betrekking heeft op het aggregatieniveau, i.c. het archiefdocument.

3.3 AIP's als containerbestanden

Binnen het OAIS-model worden informatiepakketten als een conceptuele container beschouwd: één informatiepakket kan verspreid worden over meerdere digitale objecten en dit geheel vormt een logische eenheid. Wil de archivaris de inkapselingsvoordelen ten volle benutten, dan is het aangewezen om de AIP's als één digitaal object of één fysiek containerbestand te bewaren. Een containerbestand functioneert op die wijze als een volledige AIP en bevat naast de metadata ook alle representaties van een digitaal archiefdocument.

Deze AIP's of van metadata voorziene archiefdocumenten vormen de basiseenheid van het digitaal depot. De optie om alle representaties en de essentiële metadata van één archiefdocument in één AIP en één computerbestand samen te brengen, leunt het best aan bij het concept dat een digitaal archiefdocument verschillende verschijningsvormen kan hebben en biedt het voordeel dat de metadata die betrekking heeft op het logisch archiefdocument slechts één keer hoeven vermeld te worden.

3.3 Van metadata voorziene archiefdocumenten

Digitale archiefdocumenten met metadata verrijken heeft tot doel ze zo zelfverklarend en zo autonoom mogelijk te maken. Metadata zijn onmisbaar voor het beheer van digitale archiefdocumenten en ondersteunen diverse archiefbeheerstaken: administratief beheer, beschrijven van archief, aantonen van authenticiteit, langetermijnraadpleging, opzoeking en raadpleging, enz.

Een algemene metadatanorm voor digitale archiefdocumenten is momenteel nog niet voorhanden. De ISAD(G)-norm voor het beschrijven van archieven voorziet wel enkele velden waarin metadata specifiek voor digitaal archief kunnen worden geregistreerd, maar is hoofdzakelijk gericht op het intellectueel beheer, ontsluiting en toegankelijk maken van archieven. ISAD(G) is hierdoor te eng om

¹⁶ De *Persistent Object Preservation*-benadering van het NARA voorziet deze mogelijkheid. Op deze wijze wordt de archivalische band tussen de stukken van een dossier op een fysieke wijze geregistreerd. Deze aanpak heeft als nadeel dat de containers al snel heel groot worden (> 150 MB) en niet meer zo gemakkelijk verwerkbaar zijn. Een alternatief voor het fysiek vastleggen van het onderling verband is een referentie in de AIP voor de digitale archiefdocumenten naar de dossiers waarvan ze deel uitmaken.

als algemene metadataset voor digitale archiefdocumenten te gebruiken, want ISAD(G) dekt niet alle metadatafuncties voor digitale archiefdocumenten. ISAD(G) is uiteraard wel bruikbaar voor de beschrijving en de ontsluiting van digitale archiefdocumenten, maar voor de andere metadatafuncties zijn nog andere metadatavelden nodig. Immers, digitale archiefdocumenten vragen als digitaal object en als archiefdocument bijkomende metadata voor hun beheer.

Als archiefdocument heeft een digitaal archiefdocument in de eerste plaats metadata inzake records management nodig: metadata over de context waarin de archiefdocumenten worden opgemaakt en ontvangen en metadata over de records management procedures waarbinnen ze worden beheerd. De essentiële metadata moeten minstens de identiteit en het unieke karakter van de archiefdocumenten aangeven zodat ze van andere (archief)documenten kunnen onderscheiden worden. Deze metadata hebben niet alleen een identificerende functie, maar zijn eveneens belangrijk om de authenticiteit van de archiefdocumenten aan te tonen zodat ze als bewijs kunnen dienen. Deze metadata relateren de archiefdocumenten ook aan werkprocessen en andere documenten zodat ze interpreteerbaar zijn voor archiefgebruikers. In courante bestandssystemen of documentbeheersapplicaties worden ze doorgaans gescheiden van de digitale archiefdocumenten bewaard, terwijl ze essentieel zijn voor de status en functie van de archiefdocumenten¹⁷. Momenteel wordt als verdere verfijning van de records management standaard (ISO-15489) een norm voor records management metadata uitgewerkt (ISO-23081). In afwachting van het vastleggen van deze norm kunnen bestaande records managements metadatasets (Australië, UK, enz.) als inspiratiebron dienen.

Naast de archiefbeheersmetadata komen ook een aantal technische gegevens over de digitale representatie(s) van de digitale archiefdocumenten voor inkapseling in aanmerking. Deze technische metadata documenteren de gearcheveerde bitstreams waaruit de archiefdocumenten op scherm dienen gereconstrueerd te worden. Zonder enige technische documentatie wordt het moeilijk, zo niet onmogelijk, om op termijn digitale archiefdocumenten te raadplegen. De technische metadata ondersteunen de reconstructie van de gearcheveerde bits en bytes tot een menselijk leesbaar document op scherm. Welke technische gegevens essentieel zijn, is afhankelijk van het type archiefdocument en de bewaarstrategie die wordt gevolgd.

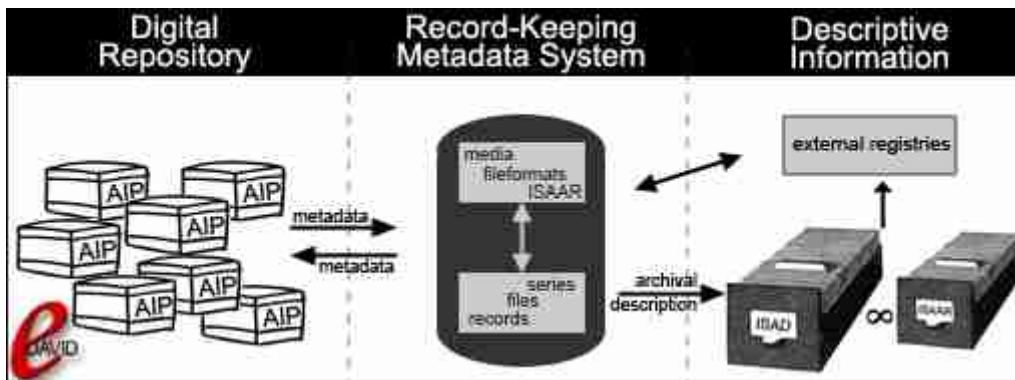
Inzake technische metadata is het interessant om over de grenzen van het archiefveld naar initiatieven zoals OAIS en Premis te kijken. In het OAIS-referentiemodel wordt voorgeschreven welke types metadata nodig zijn voor de archivering van digitale informatie, maar de OAIS-norm geeft niet aan welke metadatavelden precies nodig zijn. Dit vindt de archivaris wel terug in de metadataset voor de bewaring van digitale objecten die de Premis Working Group uitwerkte. Verder bouwende op het OAIS-informatiemodel heeft de Premis Working Group de metadata geïdentificeerd die nodig zijn voor de archivering van digitale objecten¹⁸.

Een cruciale kwestie bij het toepassen van inkapseling als opslagmethode is de vraag welke metadata in de containerbestanden worden opgenomen. Technisch is het mogelijk om alle relevante metadata in te kapselen, maar dit is niet altijd efficiënt of opportuun. Men dient ten eerste rekening te houden met het incrementeel karakter van het metadataproces. Om te vermijden dat de ingekapselde metadata frequent moeten aangepast worden, is het beter om enkel de essentiële metadata op te nemen. Alle metadata inkapselen leidt ten tweede ook tot overkill van de containerbestanden en tot grote redundantie binnen het digitaal archiefdepot. Het heeft bijvoorbeeld weinig zin om de specificatie of viewers voor bepaalde bestandsformaten of documentatie over de archiefvormer in elke container op te nemen. Dergelijke gemeenschappelijke metadata hoeven slechts éénmaal bijgehouden te worden.

¹⁷ Het belang van metadata voor de 'recordness' van digitale archiefdocumenten en van de inkapseling van metadata werd in 1996 o.a. al door David Bearman onderstreept. Het project 'Functional Requirements for Recordkeeping' beschouwt een archiefdocument als een object met ingekapselde metadata en vergeleek een archiefdocument met een document bewaard in een envelop die voorzien is van de nodige metadata. (D. Bearman, *Virtual archives*, op: *ICA Meeting*, Peking, september 1996 (<http://www.archimuse.com/papers/nhprc/prog6.html>).

¹⁸ PREMIS, *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*, mei 2005 (<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>)

Inkapseling van metadata betekent immers niet dat geen metadata centraal of extern meer worden bewaard, wel integendeel.



Het centrale metadatasysteem (RKMS) kan diverse metadata bevatten: bijv. beschrijvingen van archiefvormers (ISAAR), documentatie over bestandsformaten, beheersinformatie over de opslagmedia, archiefbeschrijvende metadata op een hoger niveau dan dat van het archiefdocument (bijv. dossiers of series), enz. Een aantal metadata kunnen ontleend of gekoppeld worden aan externe informatiebronnen. Zelfs de ingekapselde metadata kunnen nogmaals opgenomen worden in het centrale metadatasysteem. Dit is niet alleen een extra veiligheidsmaatregel, maar maakt ook snellere zoekacties mogelijk. Hierdoor is er wel nood aan synchronisatie tussen de AIP's en het centrale metadatasysteem. Voor raadpleging door archiefgebruikers of voor uitwisselingsdoeleinden, kunnen op basis van de metadata in het algemeen metadatasysteem idealiter ook ISAD(G)-conforme archiefbeschrijvingen (automatisch) worden samengesteld. Dit veronderstelt dat de archiefbeschrijvende metadata uit het algemeen metadatasysteem gemapt worden aan de ISAD(G)-elementen.

De archivaris maakt bijgevolg best een selectie van de metadata die worden ingekapseld. Bij de selectie van de essentiële metadata die mee worden ingekapseld dient de archivaris een evenwicht te zoeken tussen enerzijds de autonomie en anderzijds het beheersbaar houden van de containerbestanden. Criteria bij deze selectie zijn de keuze voor:

- metadata op documentniveau
- metadata die essentieel zijn voor de archieffunctie en de interpretatie door mens en machine.

Voorbeelden van metadata die voor inkapseling in aanmerking komen zijn ondermeer:

Metadata als archiefdocument	Metadata als digitaal object
<ul style="list-style-type: none"> ■ het uniek ID ■ de titel ■ het onderwerp ■ de archiefvormer (auteur, afzender, geadresseerde) ■ de datum ■ de relaties met andere archiefdocumenten ■ het werkproces waarbinnen de archiefdocumenten werden ontvangen, opgemaakt of beheerd 	<ul style="list-style-type: none"> ■ de identificatie van het bestandsformaat ■ het versienummer van het bestandsformaat ■ het toegepaste profiel van het bestandsformaat ■ de gebruikte codetabel ■ (een verwijzing naar) de '<i>Representation Information</i>'.

4. EEN MODEL AIP

Een *Archival Information Package* die naast alle representaties ook de essentiële metadata van een archiefdocument bevat, kan op diverse wijzen gemodelleerd worden. Hieronder wordt een hiërarchische modelstructuur gepresenteerd en wordt toegelicht hoe een volledige AIP in één computerbestand kan worden opgeslagen.

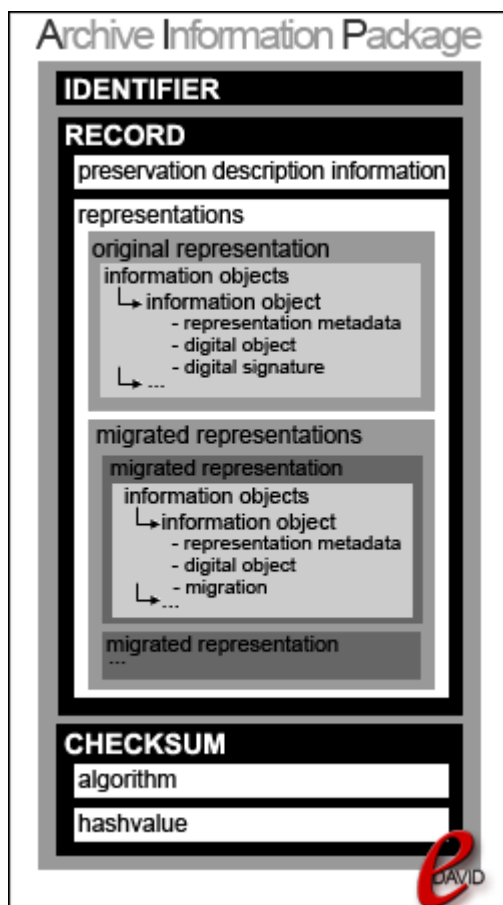
4.1 De AIP-structuur

De hoofdstructuur van een AIP bestaat uit drie delen:

- de identifier voor de AIP
- alle representaties en de essentiële metadata van het archiefdocument
- de checksum.

De identifier en de checksum dienen hoofdzakelijk voor het beheer van de AIP's. De identifier bevat de unieke ID van het computerbestand met de AIP als inhoud en is de referentie naar de AIP. Dit is bij voorkeur een duurzame ID zodat met deze identifier de AIP op lange termijn kan geïdentificeerd worden. De checksum functioneert als 'fixity information'¹⁹ en kan ook als (onderdeel van de) AIP-identifier gebruikt worden. Met een checksum kan achteraf de geldigheid van de AIP's gecontroleerd worden door de ingekapselde en de herberekende hashwaarde met elkaar te vergelijken. Deze controle kan volledig automatisch en steekproefgewijs verlopen. Van zodra de ingekapselde hashwaarde niet gelijk is aan de herberekende hashwaarde dient een alarmfunctie in werking te treden (bijv. veiligheidskopie terugplaatsen). Van de checksum wordt niet alleen de hashwaarde bijgehouden, maar ook het toegepaste hashingalgoritme.

Het tweede onderdeel in deze AIP-structuur bevat alle componenten van het archiefdocument en is verder opgesplitst in enkele subelementen. De archiefbeschrijvende metadata en de records management metadata worden in het subelement 'preservation description information' opgenomen. Deze metadata hebben betrekking op elke verschijningsvorm van het digitaal archiefdocument en hoeven bijgevolg slechts éénmaal vermeld te worden. Het tweede subelement ('representations') bevat alle verschijningsvormen en hun technische metadata van het archiefdocument. Na de originele verschijningsvorm is in AIP-structuur plaats voorzien om één of meerdere gemigreerde representaties op te nemen. Bij iedere representatie is de mogelijkheid voorzien om één of meerdere computerbestanden op te nemen. Er is immers een één-op-één relatie of een één-op-veel relatie mogelijk tussen een bepaalde representatie van het archiefdocument en de computerbestanden. Bij de originele verschijningsvorm is ook plaats voorzien om een optionele digitale handtekening en alle bijhorende metadata op te nemen. Dit element is



¹⁹ In de OAIS-norm wordt 'fixity information' als volgt gedefinieerd: the information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information object has not been altered in an undocumented manner.

voorzien voor de archivering van digitaal ondertekende archiefdocumenten en zal in principe enkel als subelement van het origineel document voorkomen. In dit element kunnen naast de digitale handtekening en zijn metadata ook alle essentiële elementen van de validation chain worden opgenomen²⁰. In de structuur is plaats voorzien voor één of meerdere archiveringsformaten. Een archiefdocument kan meer dan één geschikt archiveringsformaat hebben of in de toekomst kunnen nieuwe migraties nodig zijn. Naast de technische metadata en het eigenlijke digitale object is er ook ruimte om de migratieoperatie te documenteren ('migration').

4.2 XML als bestandsformaat

De containers met de AIP's kunnen op diverse wijzen als één digitaal object worden opgeslagen. Als bestandsformaat voor de containerbestanden is XML te verkiezen boven gecomprimeerde 'archiverings'-formaten zoals tar, zip, gzip, jar, gz, enz²¹. Deze laatste formaten zijn te vermijden vanwege de extra reconstructieschakel die het gebruik van compressie met zich meebrengt. Digitale bestanden die uitsluitend uit tekstkarakters bestaan, kunnen direct in XML opgenomen worden. Binaire bestanden dienen wel eerst via Base64 naar tekstkarakters omgezet te worden. Het in- en uitpakken via Base64 kan eveneens als een extra reconstructieschakel aanzien worden, maar Base64 is heel goed gedocumenteerd, wijdverspreid op diverse platformen en heel eenvoudig. Binaire bestanden via Base64 omzetten heeft inzake bestandsomvang wel het omgekeerd effect dan compressie. De bestandsomvang neemt namelijk met ongeveer één derde toe.

Een tweede reden om geen gecomprimeerde 'archiverings'-formaten te gebruiken, schuilt in de beperkte mogelijkheid van deze formaten om metadata in te sluiten²². Aangezien XML uitbreidbaar is, kan men in functie van de eigen behoeften metadatavelden voorzien in de XML-containers. Ongeacht het formaat kunnen de metadata in een XML-container op een uniforme wijze met één softwarecomponent worden verwerkt. Het insluiten van metadata in de XML-containers biedt het voordeel dat men zich niet meer hoeft te bekommeren om een persistente link tussen de digitale objecten enerzijds en de metadata anderzijds wanneer die gescheiden van elkaar worden opgeslagen.

Het inkapselen van metadata in de XML-containers biedt niet alleen het voordeel dat de metadata onlosmakelijk verbonden zijn met het archiefdocument, maar zorgt er ook voor dat de metadata op een gestructureerde en statische wijze zijn vastgelegd zodat ze achteraf gemakkelijk kunnen worden herbruikt. Door de metadata in XML bij te houden, zorgt men er ook voor dat deze gegevens op een digitaal duurzame wijze zijn vastgelegd en gemakkelijk door computers verwerkt kunnen worden. Tenslotte staat het flexibele karakter van XML toe, dat de elementen binnen de XML-containerbestanden ten allen tijde kunnen worden uitgebreid of getransformeerd. Dit is dan alweer een pluspunt met het oog op het incrementeel beschrijven van archiefdocumenten of latere migraties.

De keuze voor XML als bestandsformaat voor de AIP's wordt niet alleen verantwoord door technische redenen of de langetermijnraadpleegbaarheid, maar is ook gebaseerd op de nood aan het documenteren van de AIP's. De informatie en kennis over de samenstelling en de inhoud van de AIP's moet in tijd worden overgebracht. Toekomstige generaties archiefbeheerders moeten kunnen begrijpen hoe de representaties en de metadata van een digitaal archiefdocument in een container verpakt zijn. De semantiek van de AIP-onderdelen en hun onderlinge relaties moet daarom op een duidelijke wijze worden gedocumenteerd. Met XML kan de samenstelling van de AIP's op een overzichtelijke en goed gestructureerde wijze gecommuniceerd worden, zonder dat men hiervoor

²⁰ Zie F. BOUDREZ, *Digitale handtekeningen en archiefdocumenten*, Antwerpen, 2005.

²¹ F. BOUDREZ, *<XML> en digitaal archiveren*, Antwerpen, 2002.

²² Een oplossing voor deze beperking is het customiseren van de metadatabestanden binnen de compressieformaten. Een voorbeeld van een dergelijk initiatief is het onderzoek naar het uitbreiden van JAR-bestanden zodat metadata kunnen worden ingesloten (W.E. UNDERWOOD, *A java JAR implementation of an archival information package*, Consultative committee on space data systems, XML Workshop, NASA Goddard, 20 August 2001).

externe informatie moet raadplegen. In OAIS-termen zal het XML Schema voor de AIP functioneren als *Packaging Information* (de informatie die de onderdelen van een informatiepakket met elkaar verbindt). Het XML Schema voor de AIP identificeert en relateert de componenten van de *Content Information* en de *Preservation Description Information*. Deze Packaging Information wordt grotendeels in de AIP's mee inkapseld in de vorm van geneste XML-elementnamen. De AIP's zijn hierdoor zelfbeschrijvend. XML gebruiken voor de AIP's levert dezelfde voordelen op als het inkapselen van metadata in de archiefdocumenten. Voorwaarde is wel dat het documentmodel aan de basis van de AIP logisch en gestructureerd is opgebouwd en dat semantische XML-elementnamen worden gebruikt.

Bij het gebruik van XML als AIP-formaat kan de archivaris nog overwegen om niet zomaar 'normale' XML maar XML in combinatie met het Resource Description Framework (RDF)-model toe te passen. RDF is net zoals XML een W3C Recommendation²³ en is voornamelijk gericht op het semantische web. RDF verhoogt de machineleesbaarheid en de interoperabiliteit van de XML-documenten. Daartegenover staat dat het semantisch web niet de hoofdbekommernis van archieven is en dat door het toepassen van de RDF-syntax de XML structuur complexer wordt. RDF vraagt immers een andere manier van informatiemodellering. Dit laatste gaat ten koste van de vlotte leesbaarheid door de mens, want de interpretatie van de structuur en het distilleren van de semantiek verloopt moeizamer.

5. XML SCHEMA'S

Door XML als bestandsformaat voor de containerbestanden te gebruiken, kan elke organisatie in functie van haar eigen noden en aanpak een op maat gemaakt documentmodel voor de AIP's uitwerken.

Voor de implementatie van de hierboven beschreven opslagwijze in de vorm van XML containerbestanden heeft eDAVID voor de stad Antwerpen diverse XML Schema's ontwikkeld. Deze XML Schema's definiëren het formeel model voor de XML-documenten. Er is een XML Schema voor:

- het XML-containerbestand of de AIP
- een algemeen metadataset voor het beheer van digitale archiefdocumenten
- een archiefbeschrijving conform ISAD(G)
- de documenttypes waarvoor XML als archiveringsformaat wordt gebruikt: e-mails, agenda's en databanken.

Deze strategische keuze inzake het gebruik van XML heeft een gecombineerde toepassing van XML als gevolg. XML wordt ten eerste gebruikt als taal waarin alle onderdelen van een AIP's als digitaal archiefdocumenten worden verpakt. XML wordt hier gebruikt als inkapselingsformaat. XML wordt ten tweede ook gebruikt als geschikt archiveringsformaat voor een aantal documenttypes. Ten derde is XML ook het metadataformaat voor de essentiële metadata. Deze metadata worden immers rechtstreeks in XML opgeslagen.

Bij het uitwerken van de XML Schema's heeft eDAVID niet geopteerd voor het toepassen van RDF, maar voor 'gewone' XML. Naast de hierboven vermelde nadelen, werd het niet wenselijk geacht om een tijdsgebonden technologie als RDF in te bedden in de basiseenheden van het digitaal archief. Om de overzichtelijkheid te bewaren bij het combineren van diverse XML Schema's werden verschillende namespaces in de XML Schema's toegepast. Alle XML-elementen die in het AIP XML Schema zijn gedefinieerd, hebben de prefix 'aip'. De metadata uit het algemeen metadataschema kregen de prefix 'rkms' mee. De elementen uit de ISAD(G) beschrijvingsstandaard hebben als prefix 'isad', terwijl de XML Schema's voor de documenttypes waarbij XML als archiveringsformaat geldt, de prefix naar het documenttype verwijst (i.c. email, calendar en database).

²³ W3C, *Resource Description Framework (RDF) Model and Syntax Specification*, 22 februari 1999 (<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>)

De ontwerpschema's waren het onderwerp van een 'request for comments' die op 30 mei 2005 werd uitgestuurd naar diverse (inter)nationale archiefinstellingen en collega's. Hun reacties werden ingezameld tegen 30 juni 2005 zodat hun opmerkingen in juli en augustus konden verwerkt worden.

5.1 AIP

Het XML Schema voor de AIP is het algemene documentmodel voor de inkapseling van de essentiële metadata en alle representaties van een digitaal archiefdocument. Binnen dit XML Schema worden de andere XML Schema's als subschema's gebruikt.

Het eerste subelement van de AIP bevat de unieke identifier waarmee de AIP wordt aangeduid. De metadata en de verschillende representaties van het archiefdocument vormen de inhoud van het tweede subelement ('record'). De metadata die betrekking hebben op het archiefdocument hoeven slechts éénmaal vermeld te worden en wordt opgeslagen in het element *preservation description information*. In het XML Schema is de mogelijkheid voorzien om deze metadata te mappen aan het algemeen metadataschema voor digitale archiefdocumenten, of bij ontbreken hiervan aan ISAD(G). Vervolgens worden de verschillende representaties van het archiefdocument opgenomen. Naast de oorspronkelijke representatie is in het XML Schema de mogelijkheid voorzien om nul, één of meerdere gemigreerde bitstreams op te nemen. De oorspronkelijke representatie hoeft niet voor te komen: in een aantal gevallen is het bijvoorbeeld niet mogelijk om deze representatie op te nemen (bijv. bij databanken, gedigitaliseerd materiaal, enz.) Bij elke representatie wordt de relevante technische metadata vermeldt (*representation metadata*). De digitale objecten kunnen zowel binaire bestanden of XML-documenten bevatten. In het geval XML als archiveringsformaat wordt gebruikt dan wordt voor deze documentmodellen een afzonderlijk XML Schema gebruikt (zie verder).

Het derde childelement van het rootelement 'AIP' is voorbehouden voor het registreren van een checksum²⁴. Deze checksum wordt berekend op het volledige eerste childelement 'record' van de 'AIP' met inbegrip van alle XML-tags van de subelementen. De functie van de checksum is het aantonen van de bitintegriteit van de AIP. Elke bitwijziging in de AIP levert immers een andere checksum op en zal bij het vergelijken van de ingekapselde en herberekende checksum aan het licht komen. Hoewel in het OAIS-referentiemodel de *fixity information* enkel bedoeld is voor de *content information* van de AIP, wordt bij de implementatie voor de stad Antwerpen de checksum op het volledige 'record'-element berekend. Hier zijn twee redenen voor. Ten eerste kan de checksum hierdoor de bitintegriteit van de ingekapselde metadata en packaging information mee controleren. Zo wordt de functie van de fixity information uitgebreid. Ten tweede vraagt dit minder programmeerwerk en is deze werkwijze gemakkelijker te implementeren.

Bij toekomstige transformaties (bijv. wijzigen XML-structuur of toevoegen van een nieuwe gemigreerde versie) dient de checksum opnieuw berekend te worden. Indien nodig kan men op dat tijdstip een sterker hashingalgoritme gebruiken, want algoritmes worden kwetsbaarder met het toenemen van de computerkracht.

Het XML Schema voor een AIP is beschikbaar op: <http://www.edavid.be/xmlschemas/aip.xsd>

²⁴ Het inkapselen van de checksum is technisch niet de gemakkelijkste oplossing: men dient er immers over te waken dat de checksum op de correcte bitstream wordt berekend en dat componenten zoals de XML-epiloog, de XML-tags voor het <AIP>-element en het volledige element <checksum> niet worden meegerekend. Een gemakkelijker oplossing is de MD5-checksum extern bewaren of als bestandsnaam voor de AIP-container gebruiken. In dit laatste geval gaat de semantiek van de bestandsnaam verloren, wat op het opzoeken van documenten zonder toegang bijna onmogelijk maakt.

5.2 Metadataschema voor digitale archiefdocumenten

Voor het beheer van de digitale archiefdocumenten beschikt elke organisatie idealiter over een algemeen metadataschema. Bij gebrek aan een algemene norm zijn er echter nog maar weinig archiefbeherende instelling die over een dergelijk algemeen metadataschema beschikken.

Een dergelijk metadataschema hoeft niet volledig van nul opgebouwd te worden. De archiefbeschrijvende metadata worden bij voorkeur op ISAD(G) gebaseerd, terwijl de metadataset van de Premis-werkgroep voor de technische metadata een prima aanzet is.

Dit algemeen metadataschema voor het langetermijnbeheer van digitale archiefdocumenten is nog in ontwikkeling. In het metadataschema zijn alvast metadatasets voorzien voor een digitaal archiefdocument, een dossier- of onderwerpsmap, een serie, een bestandsformaat en de opslagmedia. De archiefbeschrijvende metadata-elementen zijn gelinkt aan de corresponderende ISAD(G)-velden door als attribuut de ISAD(G)-referentie op te nemen.

Het algemeen metadataschema is beschikbaar op: <http://www.edavid.be/xmlschemas/rkms.xsd>

5.3 Archiefbeschrijving

In afwachting van de uitwerking van een algemeen metadataschema voor digitale archiefdocumenten kan een organisatie de archiefbeschrijvende metadata rechtstreeks conform ISAD(G) in de XML-AIP opnemen. Aangezien deze betrekking hebben op alle mogelijke verschijningsvormen van hetzelfde archiefdocument, hoeft deze beschrijving slechts één keer opgenomen te worden.

Het XML Schema voor ISAD(G) is zo ontworpen dat men de keuze heeft tussen het samenstellen van een volledige ISAD(G)-beschrijving en tussen het gebruik van individuele beschrijvingselementen van de ISAD(G)-norm. In het eerste scenario dienen minimaal de 6 verplichte ISAD(G)-velden in het beschrijvingselement voor te komen. Aangezien dit niet altijd mogelijk of nodig is, kan men in plaats van een volledige ISAD(G)-beschrijvingsfiche ook individuele ISAD(G)-velden in dit beschrijvingselement opnemen. Veel hangt af van de mate waarin men automatisch de beschrijvingsvelden kan samenstellen.

In het XML Schema voor ISAD(G) hebben alle elementen als attribuut een verwijzing naar het paragraafnummer in de ISAD(G)-norm. Hoewel de ISAD(G)-norm voorschrijft dat men deze nummers niet *moet* gebruiken om beschrijvingselementen aan te duiden, is dit toch wel aanbevolen om redenen van duidelijkheid en uitwisseling.

Het XML Schema voor ISAD(G) is beschikbaar op: <http://www.edavid.be/xmlschemas/isad.xsd>

5.4 E-mail, agenda's en databanken

Het stadsarchief Antwerpen maakt gebruik van XML als archiveringsformaat voor een aantal types digitale archiefdocumenten:

- e-mail
- Outlook agenda's
- databanken

Voor elk documenttype is een XML Schema uitgewerkt. Deze XML Schema's kunnen op twee verschillende wijzen worden gebruikt: als archiveringsformaat of binnen de XML-containerbestanden.

Het rootelement voor elk documenttype is “<document>” aangezien de essentiële archiefbeschrijvende en contextualiserende metadata in de XML Schema's voor de documenttypes ontbreken. Deze informatie maakt immers deel uit van de AIP-container.

De XML Schema's zijn bereikbaar op:

- e-mail: <http://www.antwerpen.be/david/website/eng/rfc/email.xsd>
- Outlook agenda's: <http://www.antwerpen.be/david/website/eng/rfc/calendar.xsd>
- databanken: <http://www.antwerpen.be/david/website/eng/rfc/databank.xsd>

6. SAMENSTELLEN VAN DE XML-AIP'S

Alvorens de digitale archiefdocumenten in het digitaal depot worden opgenomen, dienen de digitale archiefdocumenten naar AIP's getransformeerd te worden. Afhankelijk van de interne afspraken en verantwoordelijkheden kan deze transformatie door de archiefvormer en/of de archivaris worden uitgevoerd. Het samenstellen van AIP's houdt immers een aantal acties in:

- migratie van de originele formaten naar geschikte archiveringsformaten
- inkapselen van de originele en gemigreerde bitstreams in XML
- registreren en inkapselen van de essentiële technische en archiefbeschrijvende metadata
- genereren van een checksum voor het controleren van de bitintegriteit
- controleren van de kwaliteit van de XML-AIP's.

De herkomst van de ingekapselde metadata kan divers zijn. Bij voorkeur worden de metadata van bij de creatie of ontvangst van de archiefdocumenten zoveel mogelijk op een geautomatiseerde wijze vastgelegd. De metadata kunnen automatisch geëxtraheerd worden uit het informatiesysteem waarbinnen de digitale archiefdocumenten worden beheerd of kunnen uit de digitale archiefdocumenten zelf worden gehaald. Voorwaarde hiervoor is natuurlijk dat deze metadata op een gestructureerde wijze zijn opgeslagen. Anderzijds kunnen de metadata ook door de archiefvormer of archivaris via de migratie- of inkapselingstool worden toegekend.

Voor de praktische implementatie van het samenstellen van AIP's zijn tal van scenario's mogelijk. Een belangrijk gegeven waarmee rekening moet worden gehouden is de beschikbaarheid van software. Migratie van archiefdocumenten naar een geschikt archiveringsformaat gebeurt best alvorens de originele software in onbruik geraakt. Inkapseling vindt ten laatste voor opname in het digitaal archiefdepot plaats. Beide acties kunnen dus in tijd gespreid worden. Zo kunnen archiefvormer en archivaris een taakverdeling uitwerken. Een mogelijkheid is dat de archiefvormer de metadata registreert en de archiefdocumenten migreert, terwijl de archivaris voor de inkapselingsoperatie zorgt en eventueel metadata toevoegt. Een andere mogelijkheid is dat alle acties op hetzelfde tijdstip worden uitgevoerd. Wanneer de archiefvormer de containerbestanden samenstelt en in die vorm overdraagt aan de archivaris dan functioneren de XML-containers niet alleen als AIP's, maar eigenlijk ook als Submission Information Packages (SIP's). Men kan dit laatste trouwens gemakkelijk automatiseren wanneer men hiervoor over een migratie- of inkapselingstool beschikt of de RMA voorziet van een AIP-exportmodule²⁵. Wanneer een organisatie bijvoorbeeld naar een hogere versie of een andere RMA overschakelt, zullen waarschijnlijk niet alle archiefdocumenten mee overgezet worden. In dat scenario kunnen dan XML-AIP's uit de RMA in afbouw geëxporteerd worden.

Het migratie- en inkapselingstijdstip kan dus samenvallen, maar dit hoeft niet altijd het geval te zijn. Het stadsarchief Antwerpen ontwikkelde een stand alone migratie- en inkapselingstool voor het samenstellen van XML-AIP's. Met deze tool worden momenteel e-mails en tekstverwerkingsdocumenten volledig automatisch gemigreerd en ingekapseld. De metadata worden zoveel mogelijk automatisch vastgelegd door ze uit het informatiesysteem of uit de documenten zelf op

²⁵ Een dergelijke AIP-exportmodule kan niet alleen archiveringsdoelinden dienden, maar kan ook gebruikt worden bij het overzetten van de digitale archiefdocumenten van de ene RMA naar de andere.

te halen. Voor de archivering van e-mails worden XML-API's samengesteld die naast de metadata de bitstream van het MS Outlook messageformaat (origineel formaat) en van de XML-versie (geschikt archiveringsformaat) bevatten. Bij tekstverwerkingsdocumenten wordt naast het MS Wordformaat de multipage TIFF-versie ingekapseld. Binnenkort wordt deze tool uitgebreid met een module voor de archivering van MS Accessdatabanken.

7. BESLUIT

Verder bouwende op de digitale archiveringsstrategie aanbevolen door het DAVID-project, werd een geschikte digitale opslagmethode voor digitale archiefdocumenten gezocht. Een oplossing werd gevonden in een combinatie van enerzijds het toepassen van het OAIS-informatiemodel en anderzijds het fysiek inkapselen van alle componenten van een digitaal archiefarchiefdocument in één containerbestand of AIP.

Het inkapselen van de verschillende verschijningsvormen en de essentiële metadata in containerbestand is voor tal van redenen een interessante archiveringsmethode. De digitale archiefdocumenten worden als het ware samen met hun metadata in containers verpakt met het oog op transport doorheen de tijd. Vooral het voordeel dat de essentiële metadata voortaan onlosmakelijk verbonden zijn met de archiefdocumenten is heel belangrijk. Hierdoor worden niet alleen nu, maar ook in de toekomst risico's vermeden. De digitale objecten in het digitaal archiefdepot winnen hierdoor aan autonomie en hebben onmiddellijk de status van archiefdocument zonder dat ze hiervoor van externe informatie afhankelijk zijn. De archieffunctie of de archiefstatus van de archiefdocumenten blijft dus ten allen tijde bewaard.

Het archiveren van met metadata verrijkte digitale archiefdocumenten biedt ook voordelen voor het digitaal depot zelf. Digitale archiefdocumenten vormen de basiseenheden van het digitaal archiefdepot en zijn beter gewapend tegen calamiteiten of toekomstige migraties en transformaties. Het digitaal archiefdepot vraagt voor het beheer van de digitale archiefdocumenten weinig of geen intelligentie, waardoor de archivaris minder zware eisen aan de software-infrastructuur voor het digitaal depot kan stellen. In principe kan elk robuust opslagsysteem als digitaal depot volstaan. Alle essentiële informatie steekt immers in de XML-AIP's. Op die manier heeft men ook weinig te vrezen van de technologische veroudering van het informatiesysteem dat men als digitaal depot gebruikt. Eén van de weinige functionaliteiten van het digitaal depot is het steekproefgewijs controleren van de bitintegriteit van de gearchiveerde digitale objecten en de bijhorende foutafhandeling wanneer een probleem opduikt.

De geautomatiseerde archieftoegangen kunnen deel uitmaken van het digitaal depot, maar men kan die evengoed beheren in gerelateerde informatiesystemen die functioneren als inventaris. In de archieftoegangen worden de archiefbeschrijvingen aan de ID's van de archiefdocumenten binnen het digitaal depot gekoppeld. De archiefbeschrijvingen kunnen gedeeltelijk automatisch worden samengesteld door metadata te extraheren op het tijdstip van opname en/of beschrijving. De (ingekapselde) metadata kunnen ook worden geïndexeerd.

Bij de implementatie van deze archiveringsmethode koos de stad Antwerpen ervoor om alle verschijningsvormen (originele en gemigreerde formaten) en de essentiële metadata in één XML-container in te kapselen. Deze methode vraagt wel dat digitale documenten bij opname in het digitaal archief worden getransformeerd tot containerbestanden. Deze containerbestanden zijn in grote mate zelfbeschrijvend, wat hun interpretatie ten goede komt. Een goede interpretatie veronderstelt wel dat de archiefbeheerder vertrouwd is met de OAIS-norm, want de architectuur en de semantiek van de AIP's zijn gebaseerd op het OAIS-informatiemodel.