

# Beschrijven van digitaal archief

Filip Boudrez  
Expertisecentrum DAVID vzw  
Antwerpen, 2007

## 0. INHOUD

---

1. Inleiding.....	1
2. Metadata.....	1
3. Beschrijven vanaf de creatie.....	3
3.1 Bij implementatie van informatiesystemen.....	4
3.2 Bij creatie en opname.....	6
3.3 Bij gebruik en beheer.....	8
4. Transfert naar het archief.....	8
4.1 Digitale overdrachtslijsten.....	9
4.2 Verwerking tot AIP's.....	9
5. Duurzame archivering.....	11
5.1 Authority files over bestandsformaten.....	11
5.2 Documenteren van het beheer.....	12
6. Besluit.....	12

## 1. INLEIDING

---

Iedere archiefbeheerder die digitale archiefdocumenten wil archiveren, heeft oplossingen nodig voor het metadataverhaal dat digitale archivering altijd vergezelt. Metadata spelen immers een essentiële rol in het proces van digitaal archiveren. Zonder metadata kunnen gearchiveerde bits en bytes niet tot digitale archiefdocumenten worden gereconstrueerd, zijn de documenten niet herbruikbaar en kan hun betrouwbaarheid niet worden aangetoond. Kortom, van digitale archiefdocumenten zijn hun metadata minstens even belangrijk als het document zelf.

Deze bijdrage gaat in op het beschrijvingsproces voor digitale archiefdocumenten. Het incrementeel karakter en het registreren van specifieke metadata tijdens de verschillende fasen van de levenscyclus van het digitaal archiefdocument staan hierbij centraal. Voor elke fase wordt aangegeven waarom en welke metadata worden vastgelegd. De aandacht gaat uit naar de archiefbeschrijvende, de technische en de administratieve metadata die nodig zijn om digitaal archief te beheren.

Bij elke fase wordt eveneens aangegeven hoe de metadata kunnen worden geregistreerd en georganiseerd. Deze aspecten zijn hoofdzakelijk een kwestie van implementatie binnen de archiefvormende en archiefbeherende instellingen en kunnen bijgevolg op uiteenlopende wijzen worden uitgewerkt. Om de algemene uitgangspunten concreet te maken, wordt hierbij één en ander geïllustreerd aan de hand van een casus bij de administratie of het archief van de stad Antwerpen. Diverse elementen van het beschrijvingsproces voor digitaal archief werden er de voorbije jaren gaandeweg geïmplementeerd.

## 2. METADATA

---

Het containerbegrip 'metadata' dekt meerdere ladingen en wordt doorgaans met digitaal archief geassocieerd. Dit laatste hoeft geen verwondering te wekken, want het begrip 'metadata' is afkomstig uit de computerwetenschap, zodat de link met digitaal archief voor de hand ligt. Het

concept van metadata in zijn algemene betekenis van 'informatie over informatie'<sup>1</sup> is echter niet exclusief verbonden aan digitaal archief, maar is evenzeer van toepassing op papieren archief. Dit betekent echter niet dat het concept metadata voor papieren en voor digitaal archief dezelfde invulling krijgt. In vergelijking met papieren archief, kennen de metadata van digitaal archief een ruimere invulling, wat kan verklaard worden door de nood aan meer metadata in de digitale wereld, maar ook door het feit dat de metadata voor digitale archiefdocumenten meer functies hebben. Concreet betekent dit onder meer dat over elk digitaal archiefdocument metadata dienen te worden geregistreerd en gearhiveerd, terwijl dit voor papieren archiefdocumenten lang niet altijd het geval is.

Papieren en digitale archiefdocumenten hebben met elkaar gemeen dat ze metadata nodig hebben voor hun archiefbeschrijvingen en toegankelijkheid. Toch vallen hier al de eerste verschillen te noteren. De verschillen op het vlak van beschrijvende metadata situeren zich niet zozeer op het niveau van het beschrijven van de archiefcontext of het archiefdocument als logische eenheid, dan wel op het niveau van de representaties van het archiefdocument. Voor het beschrijven van de context waarin archief wordt gecreëerd, ontvangen, gebruikt en beheerd maakt het niet uit of een archiefdocument in papieren of digitale vorm bestaat. Ditzelfde geldt ook voor het beschrijven van een archiefdocument als een intellectuele eenheid. In een adequaat archiefbeschrijvingssysteem wordt voor de beschrijving van het archiefdocument immers best abstractie gemaakt van zijn verschillende representaties. Binnen beide aspecten van het intellectueel archiefbeheer wordt bij voorkeur een integratie van de beschrijving van papieren en digitaal archief nagestreefd. Niet alleen is dit de aangewezen plaats om de beschrijvingen van papier en digitaal archief te integreren, dit zal ook leiden tot een efficiënter beschrijvingsproces waarbij dubbele invoer en dubbel beheer van archiefbeschrijvingen worden vermeden.

Het onderscheid tussen papier en digitaal archief is wel belangrijk voor het beschrijven van de representaties. De digitale representaties van archiefdocumenten dienen altijd van metadata voorzien te zijn. Van digitale archiefdocumenten dienen immers meerdere types relaties op een expliciete wijze te worden gedocumenteerd:

- het digitale archiefdocument en zijn archiefcontext: de (hiërarchische) band tussen twee verschillende beschrijvingsniveaus of tussen archiefbestanddelen van hetzelfde beschrijvingsniveau blijkt niet meer uit de fysieke locatie van de stukken. Van een digitaal archiefdocument dient minimaal zijn relatie met een werkproces te worden aangeduid. Dit is dan meteen ook een impliciete aanduiding van zijn relatie met de andere archiefdocumenten die binnen hetzelfde werkproces ontstonden<sup>2</sup>.
- de digitale archiefdocumenten en hun representaties: hetzelfde digitaal archiefdocument kan verschillende representaties hebben. De beschrijving van het digitaal archiefdocument dient aan zijn representatie(s) te zijn gekoppeld.
- de representaties en hun digitale objecten: één representatie kan uit één of meerdere digitale objecten bestaan. De beschrijving van de representatie dient aan de ID's van de digitale objecten zijn gekoppeld.
- de digitale objecten en een identificatie van hun bestandsformaten (inclusief versie nummers en formaatprofielen)

<sup>1</sup> Het begrip 'metadata' wordt doorgaans letterlijk vertaald als 'gegevens over gegevens'. Gegevens op zich hebben echter geen betekenis en kunnen dus niet bijdragen tot kennis. Vanwege die reden is het beter om 'metadata' te definiëren als 'informatie over informatie'.

Van de begripsbepaling 'metadata' binnen de archiefwetenschap zijn veel definities in omloop. De ISO-15489 standaard voor Records Management definieert metadata als: "*data describing the context, content and structure of records and their management through time*".

Over de metadata van digitale archiefdocumenten zijn talloze referentiedocumenten beschikbaar. Goede startpunten zijn: A. GILLILAND-SWETLAND, *Defining metadata*, Los Angeles, 2000; A. CUNNINGHAM, *Dynamic descriptions: Australian strategies for the intellectual control of records and recordkeeping systems*, in: P.J. HORSMAN, F.C.J. KETELAAR en T. THOMASSEN, *Naar een nieuw paradigma in de archivalieken*, 's-Gravenhage, 1999, p. 133-142; H. HOFMAN, *Een uitdijend heela? Context van archiefbescheiden*, in: P.J. HORSMAN, F.C.J. KETELAAR en T. THOMASSEN, *Context. Interpretatiekaders in de archivalieken*, 's-Gravenhage, 2000,

<sup>2</sup> Het reconstrueren van deze archivalische band kan eenvoudig gerealiseerd worden door middel van bevraging van het databanksysteem waarin deze metadata zijn opgeslagen.

Metadata van digitale archiefdocumenten zijn niet alleen belangrijk voor het toegankelijk maken en het administratief beheer. Metadata van digitale archiefdocumenten zijn ook essentieel voor het intact bewaren van hun digitale objecten, het verzekeren van hun leesbaarheid op (middel)lange termijn, het archiveren van authentieke en integere archiefdocumenten, en het ondersteunen van hun geloofwaardigheid. De metadata over digitale archiefdocumenten hebben bijgevolg betrekking op diverse soorten informatie die elk een specifieke functie hebben. Om intacte digitale objecten te bewaren worden checksums en pariteitsinformatie bijgehouden. Het leesbaar houden van digitale archiefdocumenten vraagt de toepassing van een digitale bewaarstrategie. Voor de uitvoering van een digitale bewaarstrategie zijn technische metadata nodig zoals specificaties van bestandsformaten en/of software(-modules), maar anderzijds dient ook de toepassing van digitale preservingsacties zoals conversie, migratie, emulatie of inkapseling in de vorm van metadata te worden geregistreerd. Deze laatste metadata maken deel uit van de historiek van digitale archiefdocumenten. In de historiek van het archiefdocument worden ook de wijzigingen na vastlegging (wie, wat en wanneer) bijgehouden. Samen met gegevens over de identiteit is de gedocumenteerde historiek belangrijk voor het bepalen of het inschatten van de mate van authenticiteit van de digitale archiefdocumenten.

Metadata zijn ook nodig om integere archiefdocumenten te archiveren. Essentiële gegevens die ontbreken in een bepaalde representatie van een digitaal archiefdocument dienen in de vorm van metadata te worden vastgelegd. Deze metadata zorgen er dan voor dat complete archiefdocumenten kunnen worden gearhiveerd. In de digitale omgeving is het immers eerder de uitzondering dan de regel dat alle essentiële gegevenselementen van de digitale documenten deel uitmaken van de standaard representatie van het document zelf. Doorgaans zijn veel, zo niet alle, gegevens wel aanwezig, maar ze worden maar zelden allemaal in het document zelf weergegeven. Heel dikwijls zijn een aantal gegevens ook voortdurend onderhevig aan wijzigingen en worden ze buiten de documenten beheerd. Soms worden een aantal gegevens ook uit andere informatiesystemen opgehaald en bevat het informatiesysteem waarbinnen de archiefdocumenten worden beheerd enkel pointers of links naar deze externe info. Essentiële componenten van het archiefdocument die niet in een gefixeerde vorm duurzaam archiveerbaar zijn, dienen in de vorm van metadata te worden beschreven. Wil men integere en accurate documenten archiveren, dan dienen alle essentiële gegevens op een statische wijze te worden vastgelegd en/of op een expliciete wijze aan de documenten te worden gekoppeld.

In de metadata van digitale archiefdocumenten worden bijgevolg niet alleen beschrijvingen over het digitaal archiefdocument bijgehouden, maar worden ook componenten van het digitaal document zelf beschreven. Beide soorten metadata dienen meerdere meesters, want ze hebben verschillende functies. Die functies van dezelfde metadata kunnen tijdens de verdere levensloop nog wijzigen. Belangrijk is dan ook dat deze metadata op een zodanige wijze worden vastgelegd en georganiseerd, dat ze gemakkelijk herbruikbaar en uitwisselbaar zijn.

---

### 3. BESCHRIJVEN VANAF DE CREATIE

---

De diverse functies van metadata leiden er toe dat met het beschrijven van digitale documenten niet kan worden gewacht tot dat digitale documenten naar het archief zijn overgebracht, maar dat nog voor de creatie van digitale documenten functionaliteiten voor het registreren van metadata moeten worden voorzien. Dit is een belangrijk aandachtspunt bij de implementatie van informatiesystemen. Dan dienen de functionaliteiten ingebouwd te worden zodat bij de creatie en opname, maar ook bij het verder gebruik en beheer van digitale archiefdocumenten de vereiste metadata zo efficiënt mogelijk worden geregistreerd en bijgehouden.

### 3.1 Bij implementatie van informatiesystemen

Om een doeltreffende digitale archivering te realiseren, is het belangrijk dat nog vóór de ingebruikname van het informatiesysteem de nodige maatregelen worden getroffen zodat de digitale archiefdocumenten en hun metadata worden vastgelegd. Bij creatie of opname van nieuwe digitale archiefdocumenten is het immers van belang dat gelijktijdig een aantal metadata worden geregistreerd.

Samen met de nodige functionaliteiten voor het fixeren van de inhoud van de archiefdocumenten en het uitvoeren van het selectieproces vormt dit de basis voor de records management functionaliteiten die in het informatiesysteem moeten worden voorzien. Het spreekt voor zich dat dit enkel nodig is voor die systemen waarbinnen archiefdocumenten worden gecreëerd en/of beheerd. Anderzijds blijft dit idealiter niet beperkt tot die informatiesystemen waarbinnen archiefwaardige informatie wordt beheerd, maar is dit een aandachtspunt voor elk informatiesysteem dat digitale archiefdocumenten bevat. Ook al hebben die digitale archiefdocumenten slechts een middellange bewaartermijn en worden ze op termijn niet naar het archief overgebracht. De levensduur van een doorsnee informatiesysteem is in de regel korter dan die van de meeste archiefdocumenten<sup>3</sup>, zodat digitale archiefdocumenten maar zelden binnen hun oorspronkelijk informatiesysteem blijven beheerd. De digitale archiefdocumenten met een middellange bewaartermijn zullen in de regel dan wel niet naar het digitaal depot van de archiefbeherende instelling worden overgebracht, maar ze zullen op termijn wel naar andere informatiesystemen moeten worden getransfereerd. Deze digitale archiefdocumenten dienen voor de organisatie evengoed herbruikbaar, en dus authentiek en integer, te zijn. Met andere woorden, elk processpecifiek informatiesysteem dat archiefdocumenten beheert, is idealiter ook een archiverings- of metadatasysteem<sup>4</sup>.

Voor archivariissen gaat de prioriteit uit naar de informatiesystemen waarbinnen archiefwaardige informatie wordt beheerd. Dit veronderstelt wel dat de archivaris zo snel mogelijk wordt betrokken bij het proces om een nieuw informatiesysteem te kopen of te maken, en dat men hiermee niet wacht tot enkele archiveringsvragen zich na verloop van tijd opdringen ten gevolge van bijvoorbeeld performantie-, overzettings- of capaciteitsproblemen. Om dit mogelijk te maken is in de modelaanpak voor ICT-projecten bij de stad Antwerpen voorgeschreven dat het stadsarchief door de projectleiders wordt gecontacteerd van zodra een pakket wordt gemaakt, gekocht of het onderwerp is van een grote versieverandering. In heel belangrijke projecten waarbij op voorhand al duidelijk is dat het om archiefwaardige digitale informatie gaat, is het stadsarchief Antwerpen in de stuurgroep van het project vertegenwoordigd.

De eerste opdracht voor de archivaris is het waarderen van de informatie die binnen het informatiesysteem wordt beheerd. Op basis van een analyse van de archiefvorming binnen de werkprocessen die door het informatiesysteem worden aangestuurd, worden de archiefreeksen, de archiefdocumenten en hun essentiële componenten geïdentificeerd. Vervolgens worden hun bestemmingen en bewaartermijnen vastgelegd. De analyse richt zich op de werkprocessen, de archiefvormers en de archiefreeksen en vormt de basis van de records management functionaliteiten die vanuit archiefstandpunt in het informatiesysteem dienen te worden voorzien. Dit is meteen ook de start van het beschrijvingsproces. Op basis van de uitgevoerde analyse moet duidelijk zijn welke essentiële componenten van de digitale archiefdocumenten en welke procesgerelateerde informatie in

<sup>3</sup> Major releases van courante pakketten hebben een levenscyclus van gemiddeld 3 tot 5 jaar. Dit is ook de periode waarvoor leveranciers ondersteuning voor een bepaalde versie garanderen. Na afloop van deze periode dient in de regel naar een nieuwe versie te worden overgeschakeld. Dergelijke versieveranderingen houden doorgaans conversies of migraties van de beheerde informatie in.

<sup>4</sup> Dit laatste hoeft niet altijd te betekenen dat de archiefdocumenten en hun metadata altijd binnen het processpecifiek informatiesysteem worden opgeslagen. Ze kunnen even goed rechtstreeks naar het digitaal depot van het archief worden overgebracht, terwijl raadplegingskopieën binnen het archiefvormende informatiesysteem achterblijven. Vanzelfsprekend is dit enkel mogelijk voor de archiefdocumenten die na vastlegging onherroepelijk definitief zijn en waarover geen aanvullende records management metadata meer nodig zijn.

de vorm van metadata worden bijgehouden. Deze metadata zullen doorgaans in het processpecifiek informatiesysteem worden geregistreerd en opgeslagen. Hierbij wordt bij voorkeur meteen ook de nodige aandacht besteed aan de wijze waarop de metadata worden opgeslagen. De metadata van digitale archiefdocumenten zullen immers overwegend in digitale vorm worden bijgehouden en hebben een bewaartermijn die minstens even lang is als die van de archiefdocumenten waarop ze betrekking hebben. De metadata van digitale archiefdocumenten worden bijgevolg best op een expliciete, statische, gefixeerde en gestructureerde wijze opgeslagen. Ook de relatie met de digitale archiefdocumenten waarbij ze horen, dient op een duurzame wijze te worden aangegeven.

Het voorzien van de nodige functionaliteiten voor het registreren en bewaren van metadata in processpecifieke applicaties is in het beste geval een kwestie van parametrisatie of configuratie van het informatiesysteem. Records management functionaliteiten zoals het registreren van metadata mee in het lastenboek laten opnemen, betekent een serieus pluspunt want dit zal tijd en middelen besparen. Door de eisen in het lastenboek te specificeren, kunnen ze als criterium tijdens de onderhandelingen en bij de uiteindelijke keuze worden gehanteerd. De ervaring van het stadsarchief Antwerpen leert dat deze eisen niet van doorslaggevende aard zijn in de finale keuze van een pakket, maar doordat ze deel uitmaken van het lastenboek zullen ze ook gemakkelijker in de projectplanning worden opgenomen wanneer een informatiesysteem wordt gekozen dat niet aan deze eisen voldoet. Zijn de eisen niet in het lastenboek vermeld, dan zal het achteraf moeilijker zijn om ze in de projectplanning te laten opnemen. In het minder gunstige geval waarin de functionaliteiten voor het registreren van metadata nog aan het informatiesysteem moeten worden toegevoegd, dient men wel oog te hebben voor een aantal consequenties: het implementatietraject duurt langer, het programmeerwerk op maat brengt extra kosten met zich mee, verregaande customisaties brengen duurdere onderhoudscontracten en complexere upgrades met zich mee, enz.

De Antwerpse praktijkervaring leert ook dat na selectie van het nieuwe pakket de implementatie best verder op de voet wordt opgevolgd. Tijdens de inrichting van het systeem is het aangewezen om meer gedetailleerd het informatiesysteem te analyseren. Deze analyse past binnen de acceptatietesten die door eindgebruiker en archivaris worden uitgevoerd. Hierbij wordt nagegaan in welke mate de gevraagde records management functionaliteiten correct worden uitgevoerd. Deze extra - en doorgaans ook meer diepgaande - controle is geen overbodige luxe, want in weerwil van voorschriften in lastenboeken en de diverse onderhandelingen met kandidaat-leveranciers blijkt op dit moment soms nog dat enkele functionele vereisten niet altijd even correct worden uitgevoerd. Als dit snel aan het licht komt tijdens het implementatietraject dan kunnen nog de nodige bijstellingen gebeuren. Een grondige analyse van het geïmplementeerde systeem op dit tijdstip is ook nog vanwege een tweede reden belangrijk. Dit is namelijk ook het geschikte moment om de archiveringsprocedure voor de digitale archiefdocumenten en hun metadata concreet uit te werken en uit te testen.

De archiefwaardering en meer bepaald de analyse die hiervoor werd uitgevoerd, kan ook in een tweede opzicht de start van het beschrijvingsproces betekenen. Doordat bij de analyse niet alleen de archiefreeksen, maar ook de werkprocessen en de archiefvormer(s) in beeld komen, kan de archivaris op dit tijdstip al een groot deel van het archief en zijn context beschrijven. Vooral voor het beschrijven van de archiefcontext is het aangewezen om nu al met beschrijven te starten. Nooit zal meer informatie en documentatie over het werkproces en de functie van de archiefdocumenten daarin beschikbaar zijn. Stelt de archivaris dit uit naar een later tijdstip (bijv. tot na de transfert van de archiefdocumenten naar het archief), dan zal het retroactief verzamelen van de nodige informatie niet alleen tijdrovend zijn, het valt ook te betwijfelen of deze hetzelfde kwaliteitsniveau zal halen. Voor de betrokken archiefvormer(s) ligt het voor de hand dat meteen ISAAR(cpf)-fiches worden aangemaakt of aangevuld. Voor het beschrijven van de werkprocessen mag worden verwacht dat de aankomende ISAF-norm een aanzet zal vormen<sup>5</sup>. Ook de archiefbeschrijvingen voor de hoogste niveaus kunnen eigenlijk best al op dit moment in de vorm van ISAD(G)-conforme beschrijvingen worden aangemaakt. De beschrijvingen van het archief en zijn context zijn op dit tijdstip nog niet definitief. Belangrijk is dat

<sup>5</sup> ISAF (International Standard on Activities/Functions of Corporate Bodies) zal pas in 2008 definitief worden vastgelegd en gepresenteerd.

deze metadata zodanig georganiseerd worden dat ze in de toekomst nog kunnen aangroeien en dat er tussen de drie beschrijvingsnormen onderling kan worden verwezen. Deze beschrijvingen van het archief en zijn context wordt binnen het archiefbeheerssysteem van de archiefbeherende instelling opgeslagen. Vermits de archiefdocumenten nog dienen gemaakt of ontvangen te worden, en zeker nog niet in het beheer zijn van de archivaris, is het opportuun om deze beschrijvingen nog niet te publiceren.

Tot de context waarbinnen de digitale archiefdocumenten tot stand komen en worden gebruikt, behoort ook het informatiesysteem waarin ze worden beheerd. Van deze informatiesystemen wordt dan ook best beschrijvende informatie bijgehouden. Deze informatie kan de archiefwaardering en het uitwerken van een archiveringsprocedure ondersteunen, maar zal op termijn hoofdzakelijk nodig zijn om te documenteren binnen welk informatiesysteem de archiefvormer zijn archiefdocumenten beheerde. Op basis van die informatie kan bijvoorbeeld de betrouwbaarheid van de archiefdocumenten vóór hun overbrenging naar het archief worden ingeschat<sup>6</sup>. Belangrijk is dat deze informatie op termijn deel uitmaakt van de beschrijving van de digitale archiefdocumenten. De informatie over de archiefvormende informatiesystemen kan op diverse wijzen worden bijgehouden. De archivaris kan voor het registreren en bewaren van deze informatie over het informatiesysteem een nieuw instrument opzetten of dit onmiddellijk integreren in zijn beschrijvingssysteem. Het stadsarchief Antwerpen registreert deze informatie in de portfolio van informatiesystemen die de IT-partner van de stad Antwerpen bijhoudt. Bij overbrenging van de digitale archiefdocumenten naar het archief zal de relevante informatie uit de portfolio aan het systeem voor het intellectueel archiefbeheer worden toegevoegd.

### 3.2 Bij creatie en opname

De hierboven beschreven werkzaamheden zijn er voornamelijk opgericht om de nodige voorbereidingen te treffen zodat archiefdocumenten en hun metadata worden vastgelegd op het tijdstip dat het document zijn status van archiefdocument heeft. Op dat moment in de levenscyclus dient namelijk de 'capture' van het archiefdocument te gebeuren. De 'capture' richt zich niet alleen op het document zelf, maar houdt ook in dat op dit ogenblik de vereiste metadata worden geregistreerd.

Bij de opname van digitale archiefdocumenten in het archiveringssysteem van de archiefvormer is het belangrijk dat gelijktijdig enkele metadata worden geregistreerd. Tijdens deze fase in de levenscyclus dienen in de regel de volgende elementen worden beschreven:

- de identiteit van het archiefdocument
- de essentiële componenten van het archiefdocument die ontbreken in de standaard representatie of dreigen te verdwijnen (bijv. omdat ze op externe plaatsen worden bijgehouden en/of voortdurend het onderwerp van wijzigingen zijn, of omdat ze niet op een duurzame wijze kunnen worden vastgelegd)
- de archivalische band: binnen welk werkproces werd het archiefdocument gecreëerd, ontvangen en/of gebruikt, de relatie met de andere archiefdocumenten die binnen hetzelfde werkproces tot stand kwamen<sup>7</sup>.

<sup>6</sup> F. BOUDREZ, *Archief onder controle? De beheersinventaris als instrument bij de archivering van digitale documenten*, Antwerpen, 2002.

<sup>7</sup> Voor Luciana Duranti is de archivalische band in de eerste plaats de relatie van het archiefdocument met de andere archiefdocumenten die binnen hetzelfde werkproces ontstonden (L. DURANTI, *The archival bond*, in: *Archives and museum informatics*, 1997, nrs. 3-4, p. 213-218.) . Vertrekkend vanuit het begrip archief als 'procesgebonden informatie' ligt het voor de hand dat de primaire archivalische band van het archiefdocument die met het werkproces is. Hieruit kan de relatie met andere archiefdocumenten worden afgeleid, zodat de relaties tussen de documenten onderling niet altijd op een expliciete wijze moeten worden geregistreerd.

- de processpecifieke informatie: routing, dossierafhandeling, dossiergegevens, enz.

Het registreren van de metadata kan op diverse wijzen verlopen: volledig geautomatiseerd, invoer door de eindgebruiker of een combinatie van beiden. Het spreekt voor zich dat het automatisch registreren van de nodige metadata de voorkeur geniet. Automatiseren verhoogt de gebruiksvriendelijkheid voor de eindgebruiker en de accuraatheid van de metadata. Toch zal de 'capture' maar zelden volledig geautomatiseerd kunnen verlopen. Zeker wanneer geen work-flow in het informatiesysteem is ingebed, zal bijvoorbeeld de situering binnen zijn ontstaanscontext enkel met tussenkomst van een eindgebruiker kunnen gebeuren. Automatisering kan eveneens deze actie ondersteunen, maar de eindgebruiker zal moeten aangeven op welke zaak of onderwerp een document betrekking heeft. Een praktijkvoorbeeld van een semi-geautomatiseerde 'capture' van metadata is het klasseren van e-mails en bijlage(n) bij de administratie van de stad Antwerpen. Van de afzender of ontvanger van een e-mailbericht wordt verwacht dat hij/zij de e-mails en bijlage(n) die een archiefdocument zijn voor de stad Antwerpen in het gemeenschappelijk digitaal klassement van zijn dienst of afdeling klasseert. De manuele handeling die de afzender of de ontvanger uitvoert, is het aanduiden op welke zaak of onderwerp e-mails en/of bijlage(n) betrekking hebben. Het klasseren en registreren van de essentiële metadata verloopt vanaf dan volledig automatisch<sup>8</sup>. Op het moment van klasseren worden verschillende metadata-elementen vastgelegd. Deze metadata worden geregistreerd omdat ze essentieel zijn voor de archiefstatus, in de Outlook-messagerepresentatie van de e-mail ontbreken of hun langetermijnarchivering ondersteunen. Deze metadata zijn:

- beschrijvende metadata:
  - de naam van het dossier of het onderwerp waarop de e-mail betrekking heeft
  - de plaats van de dossier- of onderwerpsmap in het digitaal klassement van de organisatie
  - de relatie tussen e-mail en bijlage(n), indien van toepassing
- integriteitsmetadata:
  - het e-mailadres van de afzender en eventueel diens gemachtigde
  - de datum en het tijdstip van verzending en ontvangst
  - de samenstelling van de verzendlijst, indien van toepassing
- technische metadata:
  - het versienummer van het e-mailclientprogramma waarmee de e-mail werd geklasseerd, ter identificatie van het bestandsformaat van de geklasseerde e-mail.

Al deze informatie worden ingebed in de geklasseerde e-mail zodat het archiefdocument vanaf dan altijd wordt vergezeld door zijn metadata. Deze metadata zijn niet zichtbaar of editeerbaar door de eindgebruiker. Om geklasseerde e-mails en bijlagen achteraf gemakkelijker terug te vinden worden ze op het klasseermoment ten tweede ook verrijkt met enkele uitwendige metadata: de naam van de eindgebruiker die de e-mail klasseerde, de oorspronkelijke onderwerpsaanduiding van de e-mail, de datum en het tijdstip van ontvangst en ten slotte ook de relatie tussen e-mails en bijlage(n).

In deze casus worden de metadata op het tijdstip van het klasseren van de e-mails en bijlagen geregistreerd. In theorie is het beter om de metadata op het moment van creatie, verzending of ontvangst te registreren, want dan is de kans het grootst dat de metadata 100% accuraat zijn<sup>9</sup>. In de praktijk is dit evenwel niet altijd mogelijk. In het geval van het archiveren van de content-items die binnen het Content Management Systeem van de stad Antwerpen worden aangemaakt en vervolgens

<sup>8</sup> De volledige archiveringsstrategie voor e-mail bij de stad Antwerpen wordt beschreven in: F. BOUDREZ, *E-mails: hoe klasseren en goed archiveren?*, Antwerpen, 2006.

<sup>9</sup> In het geval van e-mails is het mogelijke tijdsverschil tussen de verzending of ontvangst enerzijds en het klasseren anderzijds enkel maar relevant voor de samenstelling van verzendlijsten die in de tussentijd kan veranderen. Klasseert de eindgebruiker de e-mail na verzending of ontvangst niet onmiddellijk, maar bewaart hij/zij de e-mail toch nog in MS Outlook, dan wordt de samenstelling van de verzendlijst op dat ogenblik al vastgelegd en in de e-mail binnen MS Outlook ingebed. Wordt de e-mail later toch nog geklasseerd, dan wordt de ingebede samenstelling van de verzendlijst uit de e-mail opgehaald. Om achteraf te kunnen inschatten in welke mate de vastgestelde samenstelling van de verzendlijst accuraat is, wordt ook de datum en het tijdstip van het vastleggen van de verzendlijst geregistreerd.

via diverse kanalen worden gepubliceerd kon dit wel en dit werd dan ook zo ingericht. Hier gebeurt het registreren van de metadata ook volledig automatisch zonder dat de eindgebruiker enige actie moet ondernemen. Simultaan met het publiceren van een content-item wordt achter de schermen een archiveringsactie opgestart waarbij zowel het content-item als zijn metadata naar een XML-bestand in een archiefmap wordt weggeschreven. Dit XML-bestand bevat de gefixeerde inhoud en de metadata van het content-item.

### **3.3 Bij gebruik en beheer**

Na het vastleggen van de essentiële metadata op het 'capture'-moment worden de metadata over het digitaal archiefdocument verder aangevuld met gegevens over het beheer van de documenten en de werkprocessen waarbinnen ze verder worden gebruikt. Digitale documenten hebben immers het grote voordeel dat ze gemakkelijk herbruikbaar zijn, maar hierdoor zijn ze ook kwetsbaarder en vluchtiger. Alleen nog maar omwille van hun betrouwbaarheid dient hun beheer doorheen de tijd te worden gedocumenteerd. Ook de toegestane wijzigingen na de vastlegging (wie? wat? wanneer?) dienen te worden geregistreerd.

Opnieuw betekent dit dat het informatiesysteem waarbinnen de archiefdocumenten worden beheerd met enkele records management functionaliteiten dienen te zijn uitgerust. Die functionaliteiten zijn erop gericht om van welbepaalde acties een aantal gegevens te registreren. Bij de inrichting van het informatiesysteem moet worden gedefinieerd welke acties en wat van deze acties dient te worden geregistreerd. De courante open systemen (e-mailsystemen, gemeenschappelijke serverschrijven, enz.) beschikken standaard niet over de nodige functionaliteiten om de nodige metadata aan te leggen, zodat DMS/RMA-pakketten of andere processpecifieke applicaties met records management functionaliteiten vereist zijn.

De metadata kunnen in de metadatafiches van de digitale archiefdocumenten, in een logbestand of een combinatie van beide worden bijgehouden. Welke optie wordt toegepast, zal niet alleen afhangen van het pakket en de implementatie, maar ook van de behoeften die vanuit het werkproces ten aanzien van de digitale archiefdocumenten worden gesteld. In het geval met metadata-fiches per digitaal archiefdocument wordt gewerkt, dan zal het datamodel van de beheersapplicatie hiervoor de nodige tabellen en velden moeten voorzien. Dit is niet nodig wanneer de metadata van/over de digitale archiefdocumenten in een logbestand worden bijgehouden. Deze logbestanden mogen niet worden verward met de standaard logbestanden die vanuit het standpunt van algemeen systeembeheer (performantie, disaster recovery, enz.) worden bijgehouden. Deze laatste bevatten immers te veel irrelevante informatie en worden snel (te) groot.

Op basis van de metadatafiches en/of de logbestanden moet een constructie van de audit-trail van een digitaal archiefdocument mogelijk zijn. Tot die audit-trail behoren in principe ook een aantal gegevens over de informatiesystemen waarin de digitale archiefdocumenten worden beheerd. De mate van betrouwbaarheid van de digitale archiefdocumenten moet immers niet alleen blijken uit de geschiedenis van het document zelf, maar hangt ook af van het informatiesysteem waarbinnen de documenten werden beheerd.

## **4. TRANSFERT NAAR HET ARCHIEF**

Na uitvoering van de selectiebeslissingen worden de digitale archiefdocumenten met archiefwaarde en hun metadata naar het archief overgebracht. Samen vormen ze in de terminologie van de Open



Archival Information System-norm<sup>10</sup> een Submission Information Package (SIP). Vervolgens dient de archivaris de digitale archiefdocumenten te verwerken tot duurzame Archive Information Packages (AIP) en verder te beschrijven .

## 4.1 Digitale overdrachtslijsten

Voor de overbrenging van digitale archiefdocumenten en hun metadata naar het archief zijn diverse mogelijkheden. De digitale archiefdocumenten en hun metadata dienen uit het informatiesysteem van de archiefvormer te worden geëxporteerd op een zodanige wijze dat hun band behouden blijft en dat de ontvangende archiefdienst in staat is om ze op te nemen.

Voor de transfert van de digitale archiefdocumenten en hun metadata van de administratie van de stad Antwerpen naar het stadsarchief wordt met digitale overdrachtslijsten in XML gewerkt. De basis van zo'n overdrachtslijst voor digitale archieven is een hiërarchisch gestructureerd overzicht waarin de relaties tussen digitale archiefbestanddelen, digitale archiefdocumenten en hun digitale objecten op een expliciete wijze wordt vastgelegd<sup>11</sup>. In de regel wordt slechts één representatie van het digitaal archiefdocument overgedragen, maar het XML Schema voor deze overdrachtslijsten voorziet de mogelijkheid dat ook meerdere representaties van hetzelfde archiefdocument kunnen worden neergelegd. Van deze optie zal voornamelijk gebruik worden gemaakt wanneer het stadsarchief Antwerpen niet zelf, maar de archiefvormer de migratie naar geschikte archiveringsformaten uitvoert. Bijgevolg draagt de archiefvormer het digitaal archiefdocument in zowel zijn oorspronkelijk bestandsformaat als zijn geschikt archiveringsformaat over.

Het samenstellen van zo'n digitale overdrachtslijst verloopt volledig automatisch. Voor processpecifieke pakketten met archiefwaardige informatie is dit één van de vereiste records management functionaliteiten. Voor het genereren van een XML-overdrachtslijst voor digitale archiefdocumenten uit open systemen zoals digitale klasseringen op gemeenschappelijke serverschijven, werd een eigen tool geprogrammeerd. Bij het samenstellen van de overdrachtslijst wordt voor elk over te dragen digitaal object een checksum berekend en in de overdrachtslijst geregistreerd zodat de archivaris bij ontvangst de bitintegriteit van de getransfereerde digitale objecten kan controleren. De checksums functioneren hierbij als *fixity information* en zijn technische metadata waarmee wordt gecontroleerd of de digitale objecten intact zijn. Net zoals de controle op volledigheid van de transfert verloopt deze operatie volledig automatisch.

Belangrijker voor het beschrijven van digitale archieven is echter de opname van records management metadata in deze overdrachtslijsten. Het XML Schema voorziet de nodige elementen om op het niveau van archiefbestanddelen en archiefdocumenten records management metadata op te nemen. Bij het genereren van de XML-overdrachtslijst dienen de metadata uit de metadatafiches of de logbestanden van het archiefvormende informatiesysteem uitgelezen te worden en vervolgens aan de overeenstemmende elementen in de overdrachtslijst te worden toegevoegd. Hierdoor worden de records management metadata onmiddellijk in een geschikt uitwisselingsformaat aan de archiefdienst aangeleverd.

## 4.2 Verwerking tot AIP's

<sup>10</sup> ISO 14721 (2003): Space data and information transfer system – Open Archival Information System – Reference model.

<sup>11</sup> eDAVID ontwikkelde het XML Schema voor deze digitale overdrachtslijsten. Het XML Schema is beschikbaar op: [www.edavid.be/xmlschemas/sip.xsd](http://www.edavid.be/xmlschemas/sip.xsd).

Bij ontvangst van de overdracht van digitale archieven worden diverse kwaliteitscontroles uitgevoerd. De XML-overdrachtslijsten vervullen hierbij de functie van referentiedocument. Op basis van de informatie in de XML-overdrachtslijst wordt gecontroleerd of de overdracht volledig is. Naast de hierboven beschreven controles wordt hierbij ook nagegaan of de vereiste metadata over de digitale archiefdocumenten aanwezig zijn.

Na validatie van de overdracht worden de digitale archiefdocumenten en hun metadata verwerkt tot AIP's. Voor de metadata houdt dit in dat ze eerst uit de XML-overdrachtslijsten worden uitgelezen. De metadata worden ofwel gemapt naar de overeenstemmende elementen in het beschrijvingsmodel van het archief ofwel als een gerelateerd XML-metadatabestand in het digitaal depot bijgehouden. In de praktijk worden de metadata die betrekking hebben op reeksen doorgaans aan de overeenstemmende ISAD(G)-fiches toegevoegd. In de regel zal de archivaris op dit ogenblik de ISAD(G)-fiche die na de archiefwaardering werden aangemaakt, verder aanvullen. Voor de metadata die betrekking hebben op het niveau van de digitale archiefdocumenten wordt meestal de tweede optie toegepast. Hiervoor zijn twee redenen. Ten eerste worden op documentniveau lang niet altijd ISAD(G)-fiches aangemaakt of bijgehouden. Ten tweede zijn er op documentniveau doorgaans veel meer specifieke records management metadata aanwezig dan op reeksniveau. Vermits de ISAD(G)-norm geen specifieke elementen voor records management metadata voorziet en het uitwerken van een generiek datamodel voor applicatie- of processpecifieke metadata niet evident is, is de optie van de metadata als een gerelateerd digitaal object bij te houden het gemakkelijkst te implementeren. Deze keuze biedt ook naar de toekomst toe de meeste flexibiliteit. Het ligt voor de hand dat deze metadata dan als XML-documenten worden gearchiveerd.

Dit is echter niet de enige actie die met de metadata uit de overdrachtslijst wordt uitgevoerd. Bij de inkapseling van de verschillende representaties van het digitaal archiefdocument in één AIP-container worden ook een aantal metadata in het containerbestand opgenomen (*preservation description information*)<sup>12</sup>. Deze ingekapselde metadata hebben betrekking op het document als intellectuele eenheid en gelden bijgevolg voor alle representaties van het document. De metadata die hiervoor in aanmerking komen, zijn die metadata die essentieel zijn voor de status van archiefdocument: de link met het werkproces, de ID van het dossier of de onderwerpsmap waarvan ze deel uitmaken, de locatie binnen het classificatiesysteem, enz.

In de AIP-container - waarvoor opnieuw XML wordt gebruikt, maar nu als inkapselingsformaat - worden de relaties tussen het digitaal archiefdocument, de verschillende representaties en hun digitale objecten op een expliciete en semantische wijze aangeduid. Op het niveau van de digitale objecten worden best ook een aantal identificerende technische metadata bijgehouden: de naam en het versienummer van het bestandsformaat of de applicatiesoftware, het toegepaste formaatprofiel, de encoding, enz. Er mogen dan wel softwaretoepassingen voor formaatherkenning of -identificatie beschikbaar zijn<sup>13</sup>, het expliciet beschrijven van deze technische metadata is gemakkelijker voor het beheer en toekomstige preserveringsacties<sup>14</sup>. Vanwege dezelfde redenen is het geen overbodige luxe om deze technische beschrijvingen niet alleen in de AIP's te bewaren, maar ook binnen de beheersapplicatie van het digitaal depot bij te houden,

De verwerking van een digitale overdracht tot AIP's houdt in dat diverse kwaliteitscontroles en preserveringsacties worden uitgevoerd. Het is eveneens belangrijk dat deze acties in de vorm van metadata worden gedocumenteerd. Op het niveau van de overdracht kan worden bijgehouden wie welke acties uitvoerde, welke software hierbij werd gebruikt en wat het resultaat van elke actie was. Ook deze informatie zal naderhand deel uitmaken van de audit-trail van de digitale archiefdocumenten.

<sup>12</sup> Voor meer informatie over deze opslagwijze, zie: F. BOUDREZ, *Digitale containers voor het archief*, Antwerpen, 2005. Het eDAVID XML Schema voor een AIP is beschikbaar op: <http://www.edavid.be/xmlschemas/aip.xsd>.

<sup>13</sup> Open sourcevoorbeelden van dergelijke tools zijn onder andere JHOVE (<http://hul.harvard.edu/jhove>) en DROID (<http://droid.sourceforge.net>)

<sup>14</sup> Vermits binaire objecten via Base64-encoding in de AIP's zijn ingekapseld, is automatische formaatherkenning pas mogelijk na Base64-decoding.

De administratieve gegevens over de overdracht die in de XML-overdrachtslijst worden vermeld, kunnen opnieuw als basis dienen en integraal worden overgenomen.

## 5. DUURZAME ARCHIVERING

---

De verwerking van de digitale overdrachten tot AIP's betekent niet het einde van het beschrijvingsverhaal voor digitale documenten. Digitale archivering vraagt immers een actief conserveringsbeleid waarbij steeds nieuwe oplossingen voor het leesbaarheidsprobleem van digitale informatie nodig zijn. Samen met de andere conserveringsacties zal het beheer van de digitale archiefdocumenten verder moeten worden gedocumenteerd. In functie hiervan zal het beschrijvingsproces van digitaal archief nog worden voortgezet.

### 5.1 Authority files over bestandsformaten

Op termijn kunnen de bewaarde bits en bytes ten gevolge van de technologische veroudering niet meer tot een leesbaar document worden gereconstrueerd. Voor dit leesbaarheidsprobleem worden diverse oplossingen uitgewerkt. Een aantal oplossingen richt zich op het bestandsformaat waarin digitale informatie is opgeslagen (conversie, migratie). Andere zoeken oplossingen voor de software die nodig is om toegang te verschaffen tot de inhoud van het document (viewers, emulatie). Er zijn ook enkele digitale bewaarstrategieën die beide aanpakken combineren.

Ongeacht de toegepaste digitale bewaarstrategie is er steeds behoefte aan technische informatie over bestandsformaten. Met behulp van deze informatie zullen migratiepaden worden bepaald en/of softwareoplossingen voor raadpleging worden ontwikkeld. Bijgevolg is het aangewezen dat de archiefbeherende instelling technische documentatie bijhoudt over de bestandsformaten die voor langetermijnarchivering worden ondersteund. Het is aangewezen om voor elk bestandsformaat technische informatie in de vorm van authority fiches bij te houden, want deze informatie hoeft slechts op één plaats aanwezig te zijn. Op het niveau van de gearchiveerde digitale objecten zijn in principe enkel identificerende metadata over het formaat, de formaatversie en eventueel toegepaste compressie, encoding en formaatprofielen nodig. Deze metadata-elementen maken ook deel uit de PREMIS 1.0 -metadataset voor objecten als semantische eenheid<sup>15</sup>.

Het bijhouden van de authority files over bestandsformaten is eigenlijk een onderdeel van het digitaal depot. Alhoewel momenteel diverse registers over bestandsformaten worden aangelegd (bijv. PRONOM<sup>16</sup> of het Global Digital Format Registry<sup>17</sup>) loont het toch de moeite om als archiefbeherende instelling zelf zijn authority files bij te houden. Op die manier is deze belangrijke informatie ten allen tijde in huis aanwezig, wat opnieuw belangrijk is voor de autonomie van het digitaal archief. Bovendien richten de grote bestaande registers zich momenteel hoofdzakelijk op het verzamelen en ter beschikking stellen van identificerende technische informatie. De specificaties van de bestandsformaten zelf worden in deze registers niet opgeslagen, terwijl deze informatie toch essentieel is.

Het spreekt voor zich dat men voor het aanleggen van de technische fiches niet van nul dient te starten. De archivaris kan zijn technische fiche koppelen aan fiches in bestaande registers, of kan de informatie uit deze registers overnemen.

---

<sup>15</sup> De PREMIS-metadataset schrijft deze metadata-velden voor de eenheden 'file' en 'bitstream' voor. (<http://www.oclc.org/research/projects/pmwg/>)

<sup>16</sup> <http://www.nationalarchives.gov.uk/pronom>

<sup>17</sup> <http://hul.harvard.edu/gdfr>

## 5.2 Documenteren van het beheer

Om de geloofwaardigheid en de herbruikbaarheid van de digitale archiefdocumenten te handhaven, is het belangrijk dat het beheer door de archiefbeherende instelling verder wordt gedocumenteerd. De metadata inzake de beheerscontext worden verder aangevuld. Zaken zoals conserveringsacties, betrouwbaarheidswaarborgen, beheer van het digitaal depot, enz. worden idealiter goed beschreven zodat ten allen tijde een audit-trail van de digitale archiefdocumenten kan worden samengesteld. Een dergelijke audit-trail bevat dan de relevante gegevens van het documentbeheer door de archiefvormer én door de archiefbeheerder. Hiermee kan de archiefgebruiker inzicht in de ontstaans- en beheerscontext van het digitaal archiefdocument verwerven.

De beheerscontext zal in de praktijk op diverse plaatsen doorheen de tijd verder worden aangeduid. Gemeenschappelijke informatie (bijv. over het beheersregime van het digitaal depot) kan éénmalig en centraal worden bijgehouden. Deze informatie is immers van toepassing op alle digitale archiefdocumenten die binnen het digitaal depot worden beheerd. De uitgevoerde conserveringsacties zullen eerder op het niveau van digitale objecten worden geregistreerd en bijgehouden. In de AIP's van het stadsarchief Antwerpen zijn hiervoor de nodige velden voorzien (o.a. de identificatie van de uitvoerder, de naam van de gebruikte software, de datum en het tijdstip van de actie).

Het documenteren van het beheer van de digitale archiefdocumenten is één van de voornaamste processen die binnen de digitaal depotarchitectuur van de archiefbeherende instelling worden uitgevoerd. Gelet op het belang van het documenteren van het beheersregime voor de betrouwbaarheid van de digitale archiefdocumenten, is het aangewezen om dit op een zodanige wijze te organiseren dat deze metadata betrouwbaar zijn.

## 6. BESLUIT

Digitale archiefdocumenten hebben veel metadata nodig. Die metadata dienen niet alleen om het digitaal archief toegankelijk te maken, maar hebben diverse functies.

Uit de verschillende functies van metadata van of over digitale archiefdocumenten vloeit voort dat het beschrijven van digitale archiefdocumenten een incrementeel proces is. Dit proces start nog voor de creatie van de eigenlijke archiefdocumenten met het bepalen van de metadata eisen als onderdeel van de records management functionaliteiten die in processspecifieke informatiesystemen of RMA's dienen te worden voorzien. Het incrementeel karakter van het beschrijvingsproces brengt met zich mee de metadata van digitale archiefdocumenten dynamisch zijn. Tijdens elke fase van de levenscyclus worden nieuwe metadata toegevoegd. In de (semi-)dynamische fase ligt de klemtoon op het registreren van records management metadata, terwijl vanaf de statische fase de focus hoofdzakelijk uitgaat naar het expliciteren van de archiefcontext van de documenten, het continueren van de leesbaarheid en het documenteren van het beheer op lange termijn.

Het efficiënt beschrijven van digitaal archief veronderstelt dat de archivaris nauw betrokken is bij het archiefvormingsproces. Vanuit de archiefwaardering zal hij mee gestalte geven aan de records management functionaliteiten van de archiefvormende informatiesystemen. Een groot deel van deze functionaliteiten zijn gericht op het vastleggen van beschrijvende informatie over het ontstaan en het beheer van de digitale archiefdocumenten. Hier ligt ook een grote kans voor de archivaris, want zo kan hij/zij de archiverings- en metadatasystemen van de archiefvormer meteen afstemmen met het



metadataverhaal van het digitaal depot en het archiefbeheersysteem. Dit zal leiden tot een efficiënt archiveringsproces.