

Een digitaal archief in 10 stappen

FILIP BOUDREZ
EXPERTISECENTRUM DAVID VZW
ANTWERPEN, 2009

0. INHOUD

1. Inleiding.....	1
2. Overzicht.....	2
3. De 10 stappen.....	3
3.1 Gebruik een betrouwbaar en beveiligd opslagsysteem en pas een adequaat beheer toe.....	3
3.2 Bewaar de metadata op een digitaal duurzame wijze.....	5
3.3 Voorzie een duurzame band tussen het document en zijn metadata.....	6
3.4 Registreer voor elk digitaal object metadata	7
3.5 Definieer en documenteer de essentiële eigenschappen van het originele document.....	8
3.6 Bewaar de leesbaarheid van de digitale documenten.....	9
3.7 Archiveer alle componenten nodig voor een getrouwe reconstructie van de documenten.....	10
3.8 Documenteer het beheer van de digitale documenten.....	11
3.9 Beschrijf de digitale documenten.....	12
3.10 Documenteer het digitale archief.....	13
4. Besluit.....	14
5. Meer informatie.....	14

1. INLEIDING

Dit document wil een stappenplan voor het opzetten en het beheren van een betrouwbaar en duurzaam digitaal archief aanbieden. De stappen zijn gebaseerd op de functionele vereisten van digitale archivering en op de bijhorende maatregelen voor risicobeheer. Organisatorische of procedurele kwesties zijn niet de primaire focus van dit document¹, maar komen hier en daar wel zijdelings aan bod. De concrete implementatie van iedere stap zal immers verschillen in functie van de organisatorische context waarbinnen het digitale archief operationeel is. De stappen worden algemeen geformuleerd zodat ze te realiseren zijn door zowel particulieren als kleine of middelgrote archiefbeherende instellingen.

Voor dit stappenplan wordt het uitgangspunt gehanteerd dat digitaal archiveren een risicovol en kwetsbaar reconstructieproces is². Dit uitgangspunt leidt tot 5 terugkerende basisprincipes bij het bouwen en het beheren van een digitaal archief:

1. het vermijden van (externe) afhankelijkheden
2. het vermijden van overbodige reconstructiestappen
3. het toepassen van risicospreiding en -beheer
4. het toepassen van normen en standaarden
5. het documenteren van de implementatie en het beheer.

Deze 5 basisprincipes vormen de basis voor de acties die voor elke stap moeten worden uitgevoerd. Van elke stap wordt eerst een kleine beschrijving en situering binnen het functioneel proces van digitale archivering gegeven. Vervolgens worden voor elke stap de volgende deelaspecten beschreven:

¹ Zie hiervoor: S. SCHAULE, *Organisatorische aspecten bij het bouwen en het beheren van een digitaal depot*, Antwerpen, 2009.

² Deze visie wordt meer gedetailleerd uitgewerkt in: F. BOUDREZ, *Digitaal archiveren en digitale duurzaamheid*, Antwerpen, 2009.

- **WAT**: de acties of handelingen die moeten worden uitgevoerd
- **WANNEER**: op welk moment in de levenscyclus van het digitale document dienen deze acties of handelingen te worden uitgevoerd
- **HOE**: voorbeelden van de wijze waarop de voorgeschreven acties of handelingen kunnen worden uitgevoerd.

Het stappenplan richt zich op de realisatie van een digitaal archief waarin bruikbare digitale documenten worden opgeslagen. Naast het regelen van de opslag en het handhaven van de leesbaarheid, betekent dit ook dat het digitale archief betrouwbare en begrijpbare digitale documenten moet kunnen afleveren. De stappen richten zich dan ook op twee niveaus: de functionele stappen die vereist zijn om digitale documenten te archiveren en op het digitale archief als archiveringssysteem.

2. OVERZICHT

In termen van het OAIS-referentiemodel³ hebben de stappen vooral betrekking op de hoofdprocessen 'opname' en 'beheer'⁴. Inzake beschikbaarstellen zijn er immers geen fundamentele verschillen tussen de analoge en de papieren wereld. Wel dienen een aantal elementen in functie van de raadpleging van digitale documenten van bij de opname en tijdens het beheer worden geregeld. Deze elementen worden beschreven bij de respectievelijke stappen die betrekking hebben op de opname en het beheer van digitale documenten.

De 10 stappen kunnen in drie groepen worden onderverdeeld. De eerste drie stappen hebben betrekking op de keuze en de inrichting van de hard- en softwarearchitectuur van het digitale archief. Stap 4 tot en met 9 richten zich op de functionaliteiten die nodig zijn om digitale documenten leesbaar, authentiek en integer te houden. In de laatste stap worden alle onderdelen van het digitale archiveringssysteem gedocumenteerd.

De 10 stappen worden hieronder in een logische volgorde opgesomd en beschreven. Hun onderlinge volgorde is echter indicatief, en zeker niet leidend. Bepaalde stappen kunnen omgewisseld worden en/of kunnen met elkaar worden gecombineerd. Hun naleving is belangrijker dan de volgorde waarin ze worden gerealiseerd.

De 10 stappen zijn:

- hard- en softwarearchitectuur van het digitale archief:
 - stap 1: gebruik een betrouwbaar en beveiligd opslagsysteem en pas een adequaat beheer toe
 - stap 2: bewaar de metadata op een digitaal duurzame wijze
 - stap 3: voorzie een duurzame band tussen het document en zijn metadata
- functionaliteiten van het digitale archief:
 - stap 4: registreer voor elk digitaal object metadata
 - stap 5: definieer en documenteer de essentiële eigenschappen van het originele document
 - stap 6: bewaar de leesbaarheid van de digitale documenten
 - stap 7: archiveer alle componenten nodig voor een getrouwe reconstructie van de documenten
 - stap 8: documenteer het beheer van de digitale documenten

³ ISO 14721:2003, Space data and information transfer systems -- Open archival information system -- Reference model. OAIS is een algemeen aanvaard referentiemodel voor een digitaal archiveringssysteem. Om OAIS-compliant te zijn dienen de verantwoordelijkheden en het informatiemodel zoals beschreven in de norm, nageleefd te worden. OAIS bevat impliciet ook een functioneel model voor digitale archivering, maar dit is niet normatief.

⁴ Zie ook: F. BOUDREZ, *Basisprocessen voor een digitaal archiefdepot*, Antwerpen, 2006.

- stap 9: beschrijf de digitale documenten
- documentatie van het digitale archief:
- stap 10: documenteer het digitale archief

3. DE 10 STAPPEN

3.1 **Gebruik een betrouwbaar en beveiligd opslagsysteem en pas een adequaat beheer toe**

Voor (middel-)langetermijnarchivering geldt het algemene uitgangspunt dat duurzame gegevensdragers aangewezen zijn. Samen met een goed materieel beheer en aangepaste manipulatie moet dit de leesbaarheid en de bewaring van de documenten garanderen. Althans, deze redenering geldt voor analoge archieven. Voor deze archieven garanderen beide aandachtspunten een leesbaarheid van tientallen, zo niet honderden, jaren. Het ligt dan ook voor de hand dat ook voor duurzame digitale archivering gegevensdragers zouden worden gekozen die voor langetermijnarchivering geschikt zijn. Alleen stelt zich de vraag of dat soort digitale gegevensdragers wel bestaan. In de digitale wereld gaat het immers niet enkel om de duurzaamheid van de gegevensdrager, maar ook om de beschikbaarheid van de overeenstemmende leestecnologie. Die technologie is sterk onderhevig aan veroudering, zodat digitale objecten toch periodiek naar andere gegevensdragers moeten worden overgezet. Bovendien is digitale opslag kwetsbaarder. Digitale objecten en hun gegevensdragers dienen bijgevolg het onderwerp van een actief preserveringsbeleid te zijn.

Net zoals bij alle andere facetten van digitale archivering zullen de principes van risicospreiding,-preventie en -herstel hier worden toegepast. In die zin is de discussie of optische, magnetische dan wel op harde schijven gebaseerde opslag het best is, weinig relevant. Belangrijkere criteria in de keuze van een opslagsysteem zijn het volume (de totale bestandsomvang én het aantal digitale objecten), de verwachte groei, de responstijden bij beschikbaarstelling en de arbeidsintensiteit van de overeenstemmende beheersprocedure.

Dit laatste criterium is niet onbelangrijk. Elk type opslag vraagt immers een vorm van actief beheer waarbij de integriteit van de digitale objecten en de kwaliteit van de gegevensdragers wordt bewaakt. De geavanceerde opslagsystemen zoals SAN en NAS zijn standaard uitgerust met technologie voor het geautomatiseerd opsporen en verbeteren van fouten in de bitopslag. Geen enkel digitaal opslagsysteem is echter feilloos, zodat extra controles ook bij deze meer geavanceerde opslagsystemen geen overbodige luxe zijn. Aangezien de digitale objecten online bereikbaar zijn, kunnen deze controles geautomatiseerd worden uitgevoerd. Voor meer eenvoudige opslagoplossingen zijn ook opsporings- en verbeteringsacties mogelijk⁵, maar deze vragen doorgaans manuele handelingen, waardoor deze opslagoplossingen minder geschikt zijn voor de opslag van grote volumes digitale objecten.

In de keuze van een opslagsysteem voor de digitale objecten kan ook worden geopteerd voor een opslag als 'files' of voor een opslag als 'blobs' in een database. In dit laatste geval worden de computerbestanden ingeladen in een database en als één databaseveld opgeslagen. Dit gebeurt veelal omwille van redenen van performantie en snelle beschikbaarstelling. Nadelen van deze keuze zijn echter:

- afhankelijkheid van een databasemanagementsysteem

⁵ Voor voorbeelden zie:

- Digitaal ArchiVeren: rlichtlijn & aDvies, nr. 2: *Duurzame CD's*
- Digitaal ArchiVeren: rlichtlijn & aDvies, nr. 6: *Duurzame magnetische dragers*

- inladen en uitlezen van de computerbestanden zijn extra stappen in het opname- en beschikbaarstellingsproces
- bij grote digitale archieven (> 10 miljoen computerbestanden) kan deze aanpak leiden tot een lage performantie van de database
- de opslag van digitale documenten die uit veel digitale objecten bestaan (bijv. een database, een website, CAD, enz.) is vrij complex
- duurdere opslag
- digitale documenten en hun metadata worden in 1 systeem opgeslagen, wat extra risico's kan inhouden

WAT:

- kies genormeerde gegevensdragers: vermijd zoveel mogelijk afhankelijkheid van één bepaalde technologie of één bepaalde producent
- pas genormeerde bestandssystemen toe
- zorg voor back-ups en/of veiligheidskopieën
- voer kwaliteitscontroles uit. Controleer hierbij permanent of periodiek of:
 - alle digitale objecten nog aanwezig zijn
 - de bitintegriteit van de digitale objecten in orde is
 - de kwaliteit en de performantie van het opslagsysteem en/of de opslagmedia nog intact zijn
 - de digitale objecten virusvrij zijn
- organiseer toegangscontrole zodat enkel gemachtigden toegang hebben tot de digitale objecten. Maak hierbij een onderscheid tussen lees- en wijzigingsrechten.
- kies voor een uitbreidbaar opslagsysteem
- kies een opslagsysteem gescheiden van de beheersapplicatie en dat meerdere interfaces ondersteunt

WANNEER:

- bij implementatie van het digitale depot
- bij uitbreiding van de opslagcapaciteit
- kwaliteitscontroles:
 - idealiter: permanent en systematisch. Bij voorkeur verlopen de controles volledig geautomatiseerd.
 - alternatief: steekproefgewijs op een representatief staal.
- plaats de digitale objecten tijdig over naar een nieuw opslagsysteem of nieuwe opslagmedia van zodra:
 - de vereiste leestecnologie in onbruik geraakt en/of niet langer wordt ondersteund: houd rekening met de compatibiliteit van de fysieke dragers, van de logische indeling (bestandssysteem van de dragers), van de aansluitingen, enz.
 - het aantal fouten op een drager opmerkelijk snel stijgt
 - de gegevensdrager degradeert

HOE:

- spreid het risico door bijv.:
 - meerdere kopieën van hetzelfde digitaal object bij te houden
 - verschillende opslagsystemen of -media voor moeder- en veiligheidskopie te gebruiken
 - verschillende bestandssystemen voor moeder- en veiligheidskopie te gebruiken (bijv. bij harde schijven: NTFS voor Windows, ext3 voor Linux, HFS (Plus) voor Mac OS, enz.)
 - opslagmedia van verschillende producenten te gebruiken
 - bewaar moeder- en veiligheidskopie op verschillende locaties
- controleer de bitintegriteit van de digitale objecten door de geregistreerde checksums of CRC's te vergelijken met de herberekende checksums of CRC's (zie ook stap 3.4)

- bij opslag op harde schijven:
 - pas een RAID-configuratie (min. niveau 5 of 5+) toe
 - zorg voor een mirroring van de digitale objecten in twee (of drie) systemen
 - monitor de performantie van de harde schijven
- bij optische schijven en magnetische banden: controleer de schijf of het tapepack visueel op anomalieën
- bij opslag op optische schijven:
 - gebruik schijven van verschillende merken voor moeder- en veiligheidskopie
- controleer de digitale objecten op virussen alvorens ze in het digitale archief op te nemen

3.2 Bewaar de metadata op een digitaal duurzame wijze

De digitale duurzaamheid van de documenten is belangrijk, maar heeft maar weinig nut als de metadata van en over de digitale documenten niet digitaal duurzaam worden beheerd. Het begrijpbaar en betrouwbaar archiveren van digitale documenten is niet mogelijk zonder metadata. Die metadata worden doorgaans in het formaat en/of de database(s) van de archiefbeherende applicaties opgeslagen. Deze formaten of database(-modellen) zijn lang niet altijd open en/of gedocumenteerd. Dit houdt een even groot risico in als de opslag van digitale documenten in een applicatieafhankelijk bestandsformaat.

De digitale duurzaamheid van de metadata is een belangrijk aandachtspunt bij de keuze en bij de inrichting van het informatiesysteem waarmee het (digitale) archief wordt beheerd. Digitale duurzaamheid betekent in eerste instantie systeemafhankelijke opslag. Idealiter zijn de metadata over de digitale documenten dan ook autonoom van de archiefbeherende applicatie opgeslagen, zodat de metadata niet bij elke versieslag van de archiefbeherende applicatie dient gemanipuleerd, geconverteerd of gemigreerd te worden. Bij de keuze van een bestaand opensource of commercieel informatiesysteem als archiefbeheersapplicatie dient men immers rekening te houden met ontwikkelingscyclus van softwarepakketten, wat neerkomt op een nieuwe versie elke 3 à 4 jaar.

WAT:

- vermijd een vendor lock-in voor de archiefbeherende applicatie(s)
- bewaar de metadata niet exclusief in een applicatieafhankelijke formaten en/of database
- ontwikkel een archiveringsstrategie voor de metadata
- archiveer documentatie over het datamodel van de database

WANNEER:

- bij implementatie van het digitale archief
- bij versiebeheer van het digitale archief

HOE:

- archiveer metadata rechtstreeks in het digitale archiveringssysteem, door bijv. metadata in XML-formaat in het digitale depot te bewaren. Gebruik hiervoor geen systeemafhankelijke of weinig semantische XML, maar ontwerp XML Schema's die duidelijk zijn voor de archiefbeheerder(s)⁶.

⁶ eDAVID ontwikkelde diverse systeemafhankelijke XML Schemas voor digitale archivering. Er zijn XML Schemas voor de archivering van bepaalde documenttypes (bijv. e-mails, databases, adresboeken, SMS en MMS, enz.), van metadata en voor de inkapseling van (archief-)documenten en hun metadata in XML-containers. Deze XML Schemas zijn beschikbaar op: www.edavid.be/xmlschemas

3.3 Voorzie een duurzame band tussen het document en zijn metadata

Archiefdocumenten hebben maar betekenis en zijn slechts (her)bruikbaar als informatie over hun ontstaans- en beheerscontext beschikbaar is⁷. Die metadata wordt binnen de meeste archiveringssystemen apart van de documenten opgeslagen (bijv. in de database van het documentmanagement-, archiefbeheers- of collectieregistratiesysteem). Dit houdt een risico in want de band tussen een document en zijn metadata kan verbroken geraken waardoor het document zijn functie(s) verliest. Mogelijke oorzaken hiervan zijn niet geldige URI's, URL's, pointers of protocollen die in onbruik geraken.

Mogelijke oplossingen zijn het gebruik van persistente links of het samen bewaren van de documenten en hun metadata. Voor dit laatste zijn inbedding en/of inkapseling mogelijke opties⁸. Hierdoor worden de documenten en hun metadata die logisch één geheel vormen, ook fysiek samen bewaard.

WAT:

- inbedden van metadata in de header van de bestanden
- inkapseling van het document zijn digitale objecten en metadata in een containerformaat
- opslag van de (essentiële) metadata bij de digitale objecten in het opslagsysteem
- voorzien van een duurzame en/of reconstrueerbare link tussen een document en zijn metadata

WANNEER:

- bij classificatie, klasseren, inventarisatie of registratie
- bij opname in het digitale archief
- bij migraties en transformaties

HOE:

- gebruik maken van de beschikbare metadatavelden in de headers van bestandsformaten (inbedding):
 - invullen van de standaard beschikbare metadatavelden die het bestandsformaat voorziet
 - de metadata in de vorm van XML-isles bewaren in één metadataveld van een bestandsheader bijv.: alle metadata conform een eigen XML Schema in 1 metadataveld van een TIFF-bestand bewaren⁹
- XML metadata streams aan de digitale objecten toevoegen, bijv. XMP-metadata in een PDF/A-bestand (inbedding)
- de metadata als XML-documenten bij de digitale documenten bewaren, bijv. aan een digitaal dossier of een digitaal inventarisnummer een XML-document met de metadata toevoegen
- de metadata en de digitale objecten van een document in één containerbestand bewaren (inkapseling), zoals bijv. in de eDAVID-opslagwijze voor digitale archiefdocumenten¹⁰

⁷ Voor meer info over digitale archieven en hun metadata, zie: H. HOFMAN, *Een uitdijend heelal? Context van archiefbescheiden*, in: P.J. HORSMAN, F.C.J. KETELAAR en T. THOMASSEN, *Context. Interpretatiekaders in de archivalieken*, 's-Gravenhage, 2000; F. BOUDREZ, *Beschrijven van digitaal archief*, in: M. Van Der Eycken en E. Houtman, *LACH. Liber amicorum Coppens Herman*, Brussel, 2007, p. 69-87.

⁸ Voor een uitgebreidere bespreking en vergelijking van inbedding en inkapseling, zie: F. BOUDREZ, *Digitaal archiveren en digitale duurzaamheid*, Antwerpen, 2009.

⁹ S. HEUSCHER, *Persistent and integer lineage for digital objects. The SIMPLE approach*, Bern, 2006

¹⁰ F. BOUDREZ, *Basisprocessen voor een digitaal archiefdepot*, Antwerpen, 2006.

3.4 Registreer voor elk digitaal object metadata

Een efficiënt en duurzaam beheer van een grote hoeveelheid digitale documenten vereist dat een aantal metadata over elk digitaal object wordt geregistreerd¹¹. Elk digitaal object moet uniek identificeerbaar en lokaliseerbaar zijn. Om periodiek of systematisch de bitintegriteit van de digitale objecten te kunnen controleren dient bij opname in het digitale archief voor elk digitaal object een checksum of hashvalue ('fixity information') te worden berekend. Door herberekening van deze checksum of hashvalue en het herberekend resultaat te vergelijken met het vastgelegde resultaat, kunnen wijzigingen in de digitale objecten worden opgespoord¹².

Elk digitaal object wordt ook in een bepaald bestandsformaat opgeslagen ('representation information'). Om het digitale object achteraf te kunnen lezen dient bekend te zijn in welk formaat het is opgeslagen en welke codecs werden gebruikt¹³. Identificerende informatie zoals MIME-types en/of bestandsextensies zijn in veel gevallen onvoldoende, zeker daar waar verschillende versies van het formaat bestaan of waar zogenaamde wrapperformaten meerdere profielen toelaten (TIFF, PDF, AVI, WAVE, enz.). Het is dan ook beter om de identificerende formaat- en profielinformatie van elk digitaal object expliciet te registreren.

Naast deze technische metadata worden bij voorkeur nog een aantal andere elementen expliciet gedocumenteerd. Een (digitaal) document kan meerdere representaties hebben¹⁴ en elke representatie kan uit één of meerdere digitale objecten bestaan. Deze relaties worden best goed geregistreerd zodat de samenhang tussen het document, zijn representatie(s) en de bijhorende digitale objecten duidelijk is.

WAT:

Het registreren van:

- de unieke identificatiekenmerken waarmee het digitale object binnen het digitale archief wordt geïdentificeerd
- de lokatie ('reference information') binnen het opslagsysteem van het digitale archief
- de CRC/checksum/Reed-Solomon coding ('fixity information') van het digitale object
- de identificatie van het bestandsformaat en toegepast profiel ('representation information') van het digitale object (incl. codecs en streams)
- de relatie met de representatie en het digitale document

WANNEER:

- bij opname in het digitale archief
- bij het uitvoeren van conserveringsacties die een impact hebben op de digitale objecten (transformaties, normalisaties, migraties, conversies, enz.)
- bij het verversen van de gegevensdragers

HOE:

- noteer de ID, de lokatie, de checksum en informatie over het bestandsformaat /-profiel van elk digitaal object:

¹¹ Een digitaal document wordt niet noodzakelijk als 1 digitaal object opgeslagen. Er kunnen verschillende soorten relaties mogelijk zijn tussen digitale documenten en digitale objecten:

- 1 op 1: 1 brief opgeslagen als 1 tekstverwerkingsbestand
- 1 op veel: 1 (ingescand) register opgeslagen als 450 afbeeldingen
- veel op 1: 10 jaarrekeningen opgeslagen in 1 spreadsheet

¹² Voorbeelden van tools om checksums te berekenen zijn: MD5summer, MD5/SHA hash tool, MD5 Calculator, md5, FastSum, MD5sums, Advanced Checksum Verifier, SHA1 Hash Generator, enz.

¹³ Een video opgeslagen in AVI-formaat kan bijv. uit verschillende video- en audiostreams bestaan. Naast technische informatie over het AVI-bestand dient ook elke stream gedocumenteerd te zijn.

¹⁴ Bijv. een ingescand register kan een digitale moederkopie en raadplegingskopie hebben of een e-mail gearhiveerd als een Outlookitem en als een XML-document

- registreer deze info in een datatabel of voeg aan elk inventarisnummer, object, boek een XML- of tekstbestand toe met deze informatie
- gebruik voor de extractie en registratie van de formaatinformatie tools zoals:
 - JHOVE¹⁵
 - DROID¹⁶
 - FITS¹⁷
 - National Library of New Zealand Metadata Extractor¹⁸
 - Apache Tika¹⁹
 - MediaInfo²⁰
- verwijst naar file format registries zoals het Unified Digital Formats Registry (UDFR²¹; samengaan van Pronom en Global Digital Formats Registry) voor meer gedetailleerde specificaties van bestandsformaten
- noteer in één of meerdere datatabellen de relatie tussen digitale documenten en hun digitale objecten
- combineer de ID met de 'reference information' van een digitaal object door een relatief path als ID te gebruiken. Evt. kan hierin ook de checksum worden verwerkt.

3.5 Definieer en documenteer de essentiële eigenschappen van het originele document

Het bewaren van de leesbaarheid van digitale documenten is ten gevolge van technologische veroudering en snelle evoluties niet vanzelfsprekend. Zonder een adequate bewaarstrategie is de leesbaarheid van digitale informatie gedoemd om te verdwijnen. Er zijn verschillende benaderingen om de leesbaarheid te bewaren (zie stap 3.6), maar het toetsingspunt binnen elke strategie is de archivering van authentieke en integere documenten.

Het belangrijkste element hierbij is de controle op de correcte en ongewijzigde archivering van de 'essentie' of de 'essentiële eigenschappen' van de digitale documenten. Dit zijn de eigenschappen die ongewijzigd in tijd moeten worden overgebracht zodat de documenten toegankelijk en betekenisvol blijven²². In de digitale wereld komt dit in grote mate neer op het bepalen van het 'origineel' van een digitaal document. Deze visie vindt zijn oorsprong in de benadering van een digitaal document als een logische entiteit die door meerdere fysieke entiteiten (digitale objecten) kan worden gerepresenteerd.

Afhankelijk van zijn ontstaanscontext en zijn archiefwaarde kan de essentie van een digitaal document heel verscheiden zijn: een rol of functie in een werkproces (archief), een gevoel, een visie, een betekenis of een boodschap (bibliotheek, museum), een artefact uit het verleden (erfgoed), enz. Bij het definiëren van de essentie van het document wordt een onderscheid gemaakt tussen de elementen die voortvloeien uit de reden waarom het document werd gecreëerd en/of beheerd en de eigenschappen die veeleer incidenteel zijn. Deze laatste groep eigenschappen hebben geen specifieke betekenis of functie en maken eerder toevallig of onbewust deel uit van het digitale document of zijn digitale objecten. Essentiële eigenschappen dienen ongewijzigd in tijd worden overgebracht. Incidentele eigenschappen mogen wijzigingen of zelfs verloren gaan, zonder dat

¹⁵ <http://hul.harvard.edu/jhove>

¹⁶ <http://droid.sourceforge.net>

¹⁷ <http://fits.googlecode.com>

¹⁸ <http://meta-extractor.sourceforge.net>

¹⁹ <http://lucene.apache.org/tika>

²⁰ <http://mediainfo.sourceforge.net/nl>

²¹ <http://www.udfr.org>

²² In de Engelstalige literatuur worden voor 'essentiële eigenschappen' de volgende termen gebruikt: 'significant properties', 'significant characteristics' en 'essential characteristics'.

hiervoor het digitale document aan authenticiteit en/of integriteit inboet²³.

WAT:

- documenteer welke elementen van het digitale document essentiële eigenschappen zijn voor het behoud van zijn authenticiteit en zijn integriteit
- baseer keuzes inzake archiveringsformaten, omzettings- en emulatiertools op het behoud van de essentiële eigenschappen van de digitale documenten

WANNEER:

- registreer bij het bepalen van de archiefwaarde welke elementen essentieel zijn

HOE:

- analyseer de ontstaanscontext van de digitale documenten (bijv. functionele analyse)
- bepaal welke inhoudelijke (interne) en formele (externe) componenten van het document of de verzameling documenten essentieel zijn

3.6 Bewaar de leesbaarheid van de digitale documenten

De bits en bytes van digitale objecten zijn conform een bepaald bestandsformaat en -profiel gestructureerd. Of die digitale objecten ook leesbaar op scherm worden gepresenteerd, hangt onder meer af van de ondersteuning door de aanwezige applicatiesoftware. Afhankelijkheden ten aanzien van een specifiek softwarepakket of -versie en/of leverancier zijn absoluut te vermijden. Ook de toepassing van specifieke codecs en compressiemethoden kunnen in een onleesbaar digitaal document resulteren.

De oplossing om de digitale duurzaamheid van digitale documenten op (middel-)lange termijn te garanderen, wordt aan beide zijden van het probleem gezocht. De conversie- of migratiestrategie zet het bestandsformaat van de digitale documenten om naar meer compatibele of duurzamere formaten. Deze benadering gaat uit van de aanname dat digitale documenten aanpassing behoeven om in tijd te worden overgedragen. Voorbeelden hiervan zijn de archivering van e-mails als XML-documenten of de omzetting van een MS Word-document naar Open Document Format. Dergelijke aanpassingen kunnen een rechtstreekse impact hebben op de authenticiteit en integriteit van de digitale documenten. Daarom is het uitermate belangrijk dat deze operaties uiterst zorgvuldig worden gepland en op een gecontroleerde wijze worden uitgevoerd.

Aan de andere kant probeert de emulatiestrategie het leesbaarheidsprobleem op te lossen door de vereiste leestecnologie te archiveren. In deze strategie wordt één van de vereiste hard- en/of softwarecomponenten nagebouwd en worden de documenten gereconstrueerd op basis van de ongewijzigde digitale objecten. Een emulatievoorbeeld voor digitale archivering is de Dioscuri²⁴, een software emulator voor een X.86 Intel machine met DOS als besturingssysteem.

Beide strategieën hebben hun voor- en nadelen en houden een aantal risico's in²⁵. Beide strategieën zijn ook complementair met elkaar. Een archiefbeherende instelling kan voor verschillende documenttypes een andere strategie hanteren. Een combinatie van beide strategieën is ook mogelijk in de levenscyclus van hetzelfde archiefdocument.

In beide gevallen is het belangrijk dat de implementatie en toepassing goed wordt gedocumenteerd. Ter ondersteuning van beide aanpakken is het ook belangrijk dat de technische kenmerken van de bestandsformaten worden gedocumenteerd.

²³ InterPARES 1, *How to preserve authentic records*, 2005.

²⁴ <http://dioscuri.sourceforge.net>

²⁵ Zie voor meer informatie hierover: F. BOUDREZ, *Digitaal archiveren en digitale duurzaamheid*, Antwerpen, 2009.

WAT:

- bewaar de digitale documenten in hun oorspronkelijk en een duurzaam archiveringsformaat
- controleer of de digitale objecten vrij zijn van specifieke reconstructieschakels zoals encryptie en wachtwoorden
- documenteer welke representaties van een digitaal document in het archief aanwezig zijn
- controleer bij opname in het digitale archief of:
 - de vereiste representaties van de documenten aanwezig zijn
 - de representaties voldoen aan het voorgeschreven formaatprofiel
 - de digitale documenten niet corrupt of beschadigd zijn en wel kunnen worden geopend
 - de technische metadata over het formaatprofiel aanwezig zijn
 - de vereiste software-ondersteuning voor het leesbaar maken van de documenten aanwezig is
- volg de technologische evolutie op: bewaak de evolutie van de vereiste ondersteuning

WANNEER:

- bij klasseren
- bij opname in het digitale archief
- bij verdwijnen van ondersteuning van bestandsformaten
- bij raadpleging van digitale documenten

HOE:

- kies archiveringsformaten en codecs die:
 - genormeerd en open zijn
 - voldoende gedocumenteerd zijn
 - de essentiële eigenschappen van het origineel/authentiek document bewaren
 - geen significant informatie- en/of kwaliteitsverlies met zich meebrengen
 - voldoen aan de behoeften van de gebruikersgroep(en)
- vermijd het gebruik van producent- of applicatieafhankelijke compressie en codecs
- gebruik tools zoals JHOVE en/of DROID voor de identificatie en validatie van bestandsformaten
- gebruik de PLATO-tool²⁶ voor:
 - het kiezen van een bewaarstrategie en de tools die worden gebruikt
 - het plannen en het voorbereiden van uitvoering
 - het documenteren van de conserveringsacties
- open alle documenten als controle voor hun leesbaarheid, door bijv. alle documenten geautomatiseerd te laten openen in hun bijhorende applicatie.

3.7 Archiveer alle componenten nodig voor een getrouwe reconstructie van de documenten

Digitaal archiveren beschouwen als een reconstructieproces betekent dat alle componenten die nodig zijn voor hun reconstructie beschikbaar moeten zijn bij raadpleging. Dit is veelal meer dan enkel de digitale documenten zelf, zodat ook extra componenten dienen te worden gearchiveerd.

Welke componenten nodig zijn is onder meer afhankelijk van de bewaarstrategie die wordt toegepast en de wijze waarop de digitale documenten worden gearchiveerd. Bij toepassing van emulatie als bewaarstrategie is het onder meer belangrijk dat naast de emulator zelf ook de vereiste applicatiesoftware en de bijhorende documentatie wordt gearchiveerd en onderhouden.

²⁶ <http://www.ifs.tuwien.ac.at/dp/plato>

Het archiveren van digitale documenten hoeft niet automatisch te betekenen dat elk document als statisch object en/of in documentvorm hoeft te worden gearchiveerd. Digitale documenten kunnen ook worden gearchiveerd door hun brongegevens en alle logica voor hun reconstructie op een duurzame en systeemafhankelijke wijze te archiveren. In plaats van 100000 facturen als PDF/A-document te archiveren, kan men bijvoorbeeld ervoor kiezen om de brongegevens in XML te archiveren. Op basis van bijv. XSL(T)-stylesheets kan men achteraf 1 of meerdere facturen (her-)genereren. Ook in de gevallen waarin de archiefbeheerder opteert om alles in documentvorm te archiveren dient men erover te waken dat alle componenten (bijv. templates, formulieren, stylesheets, enz.) mee worden gearchiveerd.

WAT:

- identificeer alle componenten die nodig zijn voor een getrouwe reconstructie van de documenten
- neem alle vereiste componenten in het digitale archief op en zorg ervoor dat er geen afhankelijkheden zijn van externe bronnen die buiten het beheer van de archiefbeheerder vallen
- documenteer de componenten, hun afhankelijkheden en hun onderlinge relatie

WANNEER:

- bij archiefwaardering
- bij opname in het digitale archief
- bij uitvoering van preserveringstaken zoals conversies, migraties en transformaties

HOE:

- identificeer de vereiste componenten door de documenten te openen op een computer met een minimale hard- en softwareconfiguratie en zonder netwerkverbinding
- bij emulatie: archiveer de applicatiesoftware in de vorm van ghosts of images

3.8 Documenteer het beheer van de digitale documenten

Om de betrouwbaarheid van digitale documenten te garanderen, verschuift de klemtoon van het ongewijzigd preserveren van het object naar de metadata van het digitale document. In plaats van het authentieke en originele object te preserveren, wordt de authenticiteit van de documenten bewaard. Metadata moet de geloofwaardigheid van de digitale documenten onderbouwen. In die metadata wordt onder meer het beheer binnen de archiefvormende en de archiefbeherende applicaties geregistreerd.

Het registreren van beheersmetadata is tijdens de levensloop van een document een incrementeel proces. De beheersmetadata van/over digitale documenten worden telkens aangevuld. Dit proces start bij de aanmaak en het beheer binnen de archiefvormende applicatie en loopt verder door, ook na opname in het digitale archief. Elementen zoals uitgevoerde preservingsacties dienen in de vorm van metadata te worden gedocumenteerd.

Met behulp van de geregistreeerde beheersmetadata moet verklaard worden hoe een digitaal document is geëvolueerd vanaf zijn creatie en/of opname in het archiveringssysteem tot het moment van raadpleging. Met andere woorden, op basis van de metadata moet de levenscyclus van een digitaal document reconstrueerbaar zijn.

WAT:

- documenteren van de creatie, het gebruik en het beheer van de documenten

WANNEER:

- archiveren van beheersmetadata van/over de documenten bij opname in het digitale archief
- registreren van metadata bij het uitvoeren van preserveringsacties

HOE:

- bepaal van welke beheersacties welke metadata worden geregistreerd
- documenteer elke bewerking of preserveringsactie door minimaal bij te houden
 - welke actie wanneer werd uitgevoerd?
 - welke tools/instrumenten werden gebruikt?
 - wie voerde de actie uit?

3.9 Beschrijf de digitale documenten

Om de gearchiveerde documenten te identificeren, terug te vinden en te kunnen begrijpen dienen ze voorzien te worden van een beschrijving. In een beschrijving wordt hun inhoud, hun betekenis en/of hun context gedocumenteerd. Deze beschrijvingen zijn een onderdeel van de metadata van het digitale document. Welke beschrijvingen nodig zijn, is onder meer afhankelijk van het type materiaal en de gebruikersgroep(en).

Als onderdeel van de contextuele beschrijving dienen ook de diverse relaties van de digitale documenten op een expliciete wijze worden geregistreerd. In de analoge wereld is dit niet of veel minder het geval aangezien deze logische relaties blijken uit hun fysieke plaats in het archief. In de digitale wereld is van fysieke plaats eigenlijk geen sprake meer zodat deze relaties expliciet dienen te worden gemaakt. De relatie van een document met zijn reeks, zijn (project-)dossier of zijn onderwerpsmap moet bijgevolg blijken uit zijn metadata. Dit geldt eveneens voor de onderlinge relaties tussen documenten, al kan dit ook worden afgeleid uit gemeenschappelijke metadata zoals een band met een specifiek dossier.

WAT:

- identificeren van de documenten
- contextualiseren van de documenten: beschrijven van de archiefvormer, beschrijven van het werkproces waarbinnen het document een rol/functie vervulde, documenteren van de relatie met archiefbestanddelen en andere documenten
- faciliteren van retrieval

WANNEER:

- bij klasseren
- bij opname in het digitale archief

HOE:

- redactie van archiefbeschrijvingen (ISAD(G), ISAAR(cpf), ISDF), bibliografische referenties (ISBD, Marc21, enz.), objectregistratie (CDWA, enz.)
- toekennen van trefwoorden, tagging, koppeling aan een taxonomie
- scheid de inhoudelijke en/of contextuele beschrijving van de metadata van of over de representaties, maar voorzie wel een koppeling
- samenstellen van dossierlijsten
- faciliteer retrieval door de gearchiveerde documenten te indexeren

3.10 Documenteer het digitale archief

Tenslotte is het belangrijk dat alle onderdelen van het digitale archief worden gedocumenteerd. Deze documentatie zal ten eerste mee helpen om de levenscyclus van een digitaal archiefdocument vanaf zijn (pre-)opname in het digitale archief te reconstrueren. De documentatie zal op een algemeen niveau mee helpen verklaren hoe een digitaal document is geëvolueerd vanaf zijn opname tot het moment van raadpleging (zie ook stap 3.8). Bepaalde meta-informatie hoeft immers niet op het niveau van de digitale documenten te worden bijgehouden, aangezien ze van toepassing is op de volledige inhoud van het digitale archief. Dit zou anders leiden tot een gigantische redundantie.

Een tweede belangrijke reden voor het documenteren van het digitale archief is het beheer ervan. In geval van versiebeheer en probleemopvolging, is goede documentatie van onschatbare waarde. Men mag er immers niet van uitgaan dat de hard- en software van een digitaal archiveringsstelsel een lange levensduur heeft of per definitie bugvrij is. In dit opzicht is een digitaal archiveringsstelsel identiek aan elk ander digitaal informatiesysteem. Een goede documentatie wapent ten derde een archiefbeherende instelling of persoon ook beter tegen een mogelijke vendor lock-in. De documentatie moet ten vierde ook vermijden dat het digitale archief een black box is voor de archiefbeheerder(s) en dat bepaalde onderdelen niet kunnen worden ondersteund. Een laatste reden voor het bijhouden van goede documentatie is het afleggen van verantwoording over de werking en het beheer van het digitale archief.

Het is verstandig om deze documentatie niet exclusief in het eigen digitale archief op te slaan. In geval van calamiteiten, systeemcrashes, enz. schiet men niet veel op wanneer de documentatie in het systeem zelf is opgeslagen. Beter is om deze documentatie ook in een ander systeem of digitaal archief te bewaren.

WAT:

- documenteer het digitale archiveringsbeleid, de digitale bewaarstrategie, de geïmplementeerde normen en standaarden, enz.
- documenteer de databasemodellen en de broncode
- documenteer de procedures en de workflows
- documenteer het opslagsysteem en evt. logica dat bij plaatsing van digitale documenten wordt gehanteerd

WANNEER:

- bij ontwikkeling en implementatie van het digitale archief
- bij versiebeheer op elk onderdeel van het digitale archief

HOE:

- archiveer de (policy-)documenten waarin het archiveringsbeleid, de digitale bewaarstrategie, de geïmplementeerde normen en standaarden, enz. worden beschreven
- archiveer functionele en technische analyses, waaronder:
 - workflows en UML-diagrammen
 - use cases met beschrijving van de betrokken 'agents' en hun rollen
 - schema's met de informatie- en de systeemarchitectuur van het digitale archief
- laat databasemodellen en broncode documenteren en deponeren (bijv. escrowregelingen, deponering bij notaris, overdracht aan opdrachtgever / archiefbeherende instelling)
- documenteer de procedures door o.a. volgende documenten te archiveren:
 - handleidingen, handelingenlijsten en werkwijzen
 - richtlijnen
 - checklists

4. BESLUIT

Door bovenvermelde stappen in de praktijk te brengen moet iedere persoon, organisatie of instelling in staat zijn om een betrouwbaar en digitaal duurzaam archief uit te bouwen. Zoals eerder aangegeven, zal de concrete implementatie sterk afhankelijk zijn van de organisatorische context van het digitale archief. Eén van die elementen zijn de financiële middelen, maar zelfs met een beperkt budget zijn deze 10 stappen in de praktijk te brengen.

Aangezien digitaal archiveren een kwetsbaar en risicovol proces is, zijn veel functionele handelingen gericht op maatregelen voor risicopreventie en -herstel. Op die wijze vormt een calamiteitenplan een integraal onderdeel van het digitale archief. Samen met het bijhouden van de vereiste documentatie zal dit in grote mate bijdragen tot de betrouwbaarheid van het digitale archiveringssysteem.

5. MEER INFORMATIE

- F. BOUDREZ, *Digitale containers voor het digitaal archiefdepot*, Antwerpen, 2005.
- F. BOUDREZ, *Basisprocessen voor een digitaal archiefdepot*, Antwerpen, 2006.
- F. BOUDREZ, *Digitaal archiveren en digitale duurzaamheid*, Antwerpen, 2009.
- ED³ Eisen duurzaam digitaal depot, 2008.
- *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, 2007.
- A. WILSON, *Significant properties report*, 2007.