

Videoarchivering: bruggen bouwen op technologisch drijfzand.

Filip Boudrez
Stadsarchief Antwerpen
Antwerpen, 2004

0. INHOUD

1. Inleiding.....	1
2. Van analoog tot digitaal: noodzakelijk drijfzand.....	1
3. Bewaarstrategieën voor digitale documenten.....	2
3.1 Digitaal archiveren?.....	2
3.2 Hard copy, computermusea en conversies.....	3
3.3 Migratie vs. emulatie.....	3
3.4 Besluit: DAVID-bewaarstrategie.....	5
4. Geschikte archiveringsformaten.....	7
5. Archiveringsformaat voor video.....	9
6. Praktische aanbevelingen voor videoarchivering.....	10

1. INLEIDING

Het onderzoeksproject DAVID (Digitaal Archiveren in Instellingen en Diensten¹) onderzocht de problematiek van de lange termijnbewaring van digitale archiefdocumenten. Hiervoor werden de bewaarstrategieën voor digitale documenten geëvalueerd. Het DAVID-project ontwikkelde een digitale bewaarstrategie die zoveel mogelijk leesbaarheids garanties biedt en waarbij in de toekomst zowel migratie als emulatie mogelijk blijven. Deze bewaarstrategie is toepasbaar op allerlei types digitale documenten, waaronder gedigitaliseerde analoge videobanden. In het tweede deel van deze bijdrage wordt een geschikt archiveringsformaat voor video gezocht en worden concrete aanbevelingen voor videoarchivering geformuleerd².

2. VAN ANALOOG TOT DIGITAAL: NOODZAKELIJK DRIJFZAND

Videobanden zijn niet geschikt voor lange termijnarchivering vanwege hun degeneratie en technologische veroudering. Immers, analoge dragers zoals videobanden zijn onderhevig aan gebruiksslijtage en lijden sterk onder de tand des tijds. Bovendien wordt het binnen afzienbare tijd moeilijk om nog de nodige afspeelapparatuur te vinden. Nochtans kunnen videobanden archiefdocumenten bevatten die voor lange termijnarchivering in aanmerking komen. De archivering van de archiefdocumenten opgeslagen op analoge media vraagt een pro-actieve bewaarstrategie om de raadpleegbaarheid op lange termijn te verzekeren.

Het tijdig overzetten van de videobanden naar nieuwe analoge dragers is geen duurzame oplossing. Bij elke analoge overzetting gaat immers informatie verloren en treedt kwaliteitsverlies (vervorming, ruis) op. Dit verschijnsel wordt 'generatieverlies' genoemd en is onvermijdelijk. Bovendien blijft men ook afhankelijk van een specifieke afspeeltechnologie. Een andere oplossing is digitalisering van de analoge documenten. Door het geluids- en beeldmateriaal te digitaliseren kan men in de toekomst de gedigitaliseerde archiefdocumenten naar andere dragers overzetten zonder dat dit met informatie- en kwaliteitsverlies hoeft samen te gaan. In de digitale wereld kan men bitstreams immers onbeperkt kopiëren zonder dat de gearchiveerde documenten aan kwaliteit

1 De website en het vademecum van het DAVID-project zijn beschikbaar op: <http://www.antwerpen.be/david>.

2 Deze tekst is de uitgewerkte abstract van de lezing 'Digitale archivering: bruggen bouwen op technologisch drijfzand', Studiedag: VideolnVraag, SMAK Gent, vrijdag 19 maart 2004.

inboeten. De kwaliteit van digitale documenten blijft ook intact bij veelvuldig gebruik. Digitalisering is met andere woorden vereist voor de preservatie van audio-visueel archiefmateriaal. Door te digitaliseren kan men ook gebruik maken van extra voordelen op het vlak van consultatie, distributie en reproductie.

Digitalisering heeft echter ook zijn schaduwzijden. Aan digitalisering hangt ten eerste een groot prijskaartje vast. Daardoor kan niet alles gedigitaliseerd worden zodat weloverwogen keuzes zich opdringen en goede selectiecriteria nodig zijn. Naast de hoge kostprijs is ten tweede vooral de grotere afhankelijkheid van bepaalde technologie voor de raadpleging van gedigitaliseerde documenten een belangrijk minpunt. Computertechnologieën verouderen snel en geraken in onbruik, doorgaans in een hoger tempo dan analoge afspeeltechnologieën. Digitalisering is dus geen oplossing voor de technologische veroudering van analoge afspeelapparatuur, wel integendeel. Door te digitaliseren wordt men afhankelijk van hard- en software voor de reconstructie van geluid- en beeldopnamen en komt men op het terrein van digitale archivering.

De lange termijn raadpleegbaarheid van digitale documenten mag dan wel een afzonderlijk onderzoeksgebied zijn, het is een belangrijk aandachtspunt bij digitaliseringsinitiatieven. Pas wanneer men vanaf het digitaliseringsmoment met digitale archivering op lange termijn rekening houdt, is het mogelijk om gedigitaliseerde documenten in de toekomst te blijven raadplegen en de kwaliteit ervan intact te houden. Zonder een duidelijke bewaarstrategie voor digitale documenten loopt men het risico meer informatie en kwaliteit te verliezen in vergelijking met het 'generatieverlies' in de analoge wereld, of zelfs het hele document kwijt te raken. Niet meer raadpleegbare gedigitaliseerde audio-visuele opnamen worden immers als verloren beschouwd.

3. BEWAARSTRATEGIEËN VOOR DIGITALE DOCUMENTEN

3.1 Digitaal archiveren?

Digitale archiefdocumenten zijn digitale objecten en kunnen bijgevolg slechts geraadpleegd worden wanneer men over de vereiste hard- en softwareconfiguratie beschikt. Zonder hard- en software kan men de bewaarde bits en bytes niet weergeven in een menselijk begrijpbare vorm. Aangezien hard- en software snel verouderen en veel digitale archiefdocumenten een langere levensduur hebben, moet men ervan uitgaan dat de digitale documenten raadpleegbaar moeten zijn met andere hard- en software dan waarmee ze werden gecreëerd.

Digitaal archiveren is het bewaren van de mogelijkheid tot reconstructie van digitale archiefdocumenten op basis van de gearchiveerde bits en bytes. Op het moment van raadpleging moet een interactie mogelijk zijn tussen de beschikbare computersystemen en gearchiveerde computerbestanden. Die interactie is bij voorkeur van zo min mogelijk factoren afhankelijk. Het ontbreken van één reconstructieschakel resulteert immers in het verlies van het archiefdocument.

De internationale archiefwereld buigt zich al enige tijd over het lange termijn leesbaarheidsprobleem van digitale archiefdocumenten. Verschillende bewaarstrategieën voor digitale documenten werden hiervoor bestudeerd: hard copy, computermuseum, conversie, migratie en emulatie.

3.2 Hard copy, computermusea en conversies

Bij de hard copy strategie worden digitale documenten afgedrukt en/of op microfilm geplaatst³. Hard copy heeft niet alleen als nadeel dat de digitale eigenschappen en voordelen verloren gaan, voor de

3 Een meer exhaustieve bespreking en evaluatie van elke bewaarstrategie is beschikbaar in het vademecum van het DAVID-project: F. Boudrez en H. Dekeyser, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2004. (on line beschikbaar op <http://www.antwerpen.be/david>)

archivering van geluid en beeld is deze bewaarstrategie helemaal geen optie.

Het bewaren van de originele hard- en software wordt de computermuseumstrategie genoemd. Deze strategie gaat uit van de bewaring van de oorspronkelijke IT-configuraties waarmee de digitale documenten werden gecreëerd, zodat de digitale documenten hiermee kunnen geconsulteerd worden. Deze optie is echter niet realistisch op lange termijn. Hard- en software hebben niet alleen een beperkte levensduur, ook de vereiste kennis voor het werken met oude configuraties verdwijnt. Bovendien is deze oplossing weinig praktisch. De museumstrategie kan hoogstens voor noodoperaties ('digitale archeologie') of korte termijnarchivering worden gebruikt.

Bij conversie worden de digitale documenten telkens omgezet naar een hogere versie van het formaat waarin ze werden omgeslagen ((bijv. MS Word97 → MS Word2000 → MS Word2002 → MS Word2003) Aangezien de versieveranderingen van bestandsformaten elkaar snel opvolgen, dienen de conversies frequent herhaald worden. Bij elke conversie is er kans op wijzigingen of verlies. Bovendien blijft men afhankelijk van specifieke software(versies) en/of producenten.

3.3 Migratie vs. emulatie

Aangezien de hard copy-, de museum- en de conversiestrategie algemeen worden afgewezen voor de lange termijnarchivering van digitale documenten werd de leesbaarheidsproblematiek jarenlang toegespitst op de vraag of migratie dan wel emulatie de beste bewaarstrategie is.

3.3.1 MIGRATIE

Migratie is de bewaarstrategie waarbij digitale documenten naar een geschikt archiveringsformaat worden omgezet. Op die manier verzekert men zich ervan dat de digitale documenten kunnen geraadpleegd worden met andere hard- en software dan diegene waarmee ze werden gecreëerd.

Aangezien geschikte archiveringsformaten bij voorkeur standaardformaten zijn, is dit de bewaarstrategie waarbij de digitale documenten naar een standaardformaat worden omgezet. Standaarden zijn in principe gedocumenteerd, stabiel en niet afhankelijk van één producent. Migratie wordt soms ook wel aangeduid met 'transformatie' of 'normalisatie' wanneer standaarden als doelformaat worden gebruikt.

Voordelen van migratie als bewaarstrategie zijn onder meer:

- archiefbehurende instellingen kunnen zelfstandig de migratiestrategie toepassen
- digitale archiefdocumenten worden niet in een producent-, software- of versiegebonden bestandsformaat bewaard
- de specificatie van het bestandsformaat is beschikbaar: op basis van deze documentatie kan ten allen tijd een nieuwe viewer geprogrammeerd worden
- beschikbaarheid van omzettingstools: er zijn niet alleen veel omzettingstools op de markt, migraties zijn ook gemakkelijk te realiseren met behulp van wijdverspreide computerprogramma's

De migratiestrategie heeft ook een aantal nadelen:

- afhankelijkheid van beschikbare standaarden
- kans op informatieverlies (bijv. wanneer het doelformaat bepaalde componenten van het oorspronkelijk formaat niet kan overnemen)

Tot op heden is migratie de meest toegepaste bewaarstrategie voor digitale archiefdocumenten. Archiefinstellingen hebben ondertussen al een vrij veel expertise en ervaring inzake het gebruik van migratie opgebouwd zodat de risico's en de minpunten vrij goed ingeschat kunnen worden.

3.3.2 EMULATIE

Bij emulatie wordt niet de originele hard- en software bewaard, maar wordt het vereiste platform op een toekomstige computerconfiguratie gereconstrueerd zodat de computerbestanden in hun oorspronkelijk formaat raadpleegbaar zijn.

Emulatie kan op diverse niveau's worden toegepast. Men kan computerhardware, besturingssystemen, specifieke software of een combinatie van dit alles nabootsen. Emulatie is mogelijk op basis van configureerbare chips (emulatie door hardware) of op basis van computerprogramma's (emulatie door software).

Inmiddels bestaan verschillende visies op de wijze waarop emulatie voor digitale archivering kan worden toegepast:

- Jeff Rothenberg: Emulation Virtual Machine⁴
- Steve Gilheany: Turing Machine⁵
- Raymond Lorie: Universal Virtual Machine (data preservatie, programma preservatie⁶)
- Cedars & Camileon project: Migration on request⁷

Emulatie heeft een aantal interessante voordelen:

- in theorie kunnen de documenten in hun oorspronkelijk formaat gearchiveerd worden:
 - o alle originele eigenschappen en functionaliteiten blijven behouden
 - o er gaan geen elementen verloren ten gevolge van omzettingen
 - o de authenticiteit van de digitale archiefdocumenten is gemakkelijker te garanderen
- documenten hoeven niet omgezet te worden telkens hun formaat in onbruik raakt
- de kostprijs is niet afhankelijk van het aantal gearchiveerde digitale documenten

Anderzijds zijn er ook een aantal nadelen aan emulatie verbonden:

- emulatie is technisch complex: de nodige know-how en expertise voor ontwikkeling en onderhoud zijn niet intern aanwezig. Archiefbeherende instellingen zijn bijgevolg afhankelijk van externe diensten en partners.
- emulatie heeft hoge ontwikkelings- en onderhoudskosten: kunnen archieven die nu voor deze benadering kiezen de financiële inspanningen in de toekomst blijven leveren?
- de platformen waarop emulatieprogramma's draaien, evolueren, wat op termijn omzettingen van of aanpassingen aan emulatieprogramma's zal vergen
- overkill: bepaalde emulatiebenaderingen gaan uit van de volledige reconstructie van de oorspronkelijke applicaties met alle functionaliteiten, terwijl in principe een viewer volstaat om het (statisch) archiefdocument weer te geven. Emulatie richt zich hoofdzakelijk op de lange termijn bewaring van systemen en software, terwijl de archivaris in eerste plaats de archivering van het digitaal archiefdocument beoogt.
- de bescherming van het auteursrecht op hard- en software houdt restricties in voor reverse engineering en het bouwen van emulatoren
- archiefvormers maken gebruik van tal van verschillende informatiesystemen, waarvan een aantal op maat van de organisatie zijn gesneden of ad hoc ontwikkeld zijn: archieven dienen over tal van emulatoren te beschikken en kunnen bepaalde kosten niet delen met andere archieven
- emulatie van viewers voor gesloten of niet gedocumenteerde bestandsformaten op basis

4 J. ROTHENBERG EN T. BIKSON, *Digital preservation: carrying authentic, understandable and usable digital records through time. report to the dutch national archives and ministry of the interior*, 1999 (http://www.digitaleduurzaamheid.nl/bibliotheek/docs/final-report_4.pdf); J. ROTHENBERG, *An experiment in using emulation to preserve digital publications*, Den Haag, 2000 (<http://www.kb.nl/coop/nedlib/results/emulationpreservationreport.pdf>); J. ROTHENBERG, *Avoiding technological quicksand: finding a viable technical foundation for digital preservation. a report to the council on library and information resources*, Washington, 1999 (<http://www.clir.org/pubs/reports/rothenberg/pub77.pdf>); J. ROTHENBERG, *Ensuring the longevity of digital information*, Santa Monica, 1999 (<http://www.clir.org/pubs/archives/ensuring.pdf>).

5 <http://www.archivebuilders.com/aba010.html>

6 <http://www.rlg.org/preserv/diginews/diginews5-3.html#feature2>

7 <http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2>

van reverse engineering is risicovol. Emulatie van viewers voor gestandaardiseerde en gedocumenteerde formaten is gemakkelijker en veiliger, maar blijft hoe dan ook technisch complex. Wordt emulatie dan toch voorafgegaan door een migratie naar een open, gedocumenteerd en gestandaardiseerd archiveringsformaat?

- gebruikers werken met oude software en kunnen geen gebruik maken van technologische vernieuwingen
- archieven beheren niet alleen digitale archiefdocumenten, maar ook emulatiehardware en -software en bijhorende documentatie
- de praktische haalbaarheid van bepaalde emulatiepistes zal pas in de toekomst blijken

De promotoren van emulatie als digitale bewaarstrategie schuiven voornamelijk het behoud van het oorspronkelijk computerbestand met alle originele eigenschappen als belangrijkste argument naar voor. Men beklemtoont in het bijzonder de mogelijkheid tot bewaring van de 'look and feel' en de functionaliteiten, terwijl deze eigenschappen bij migratie veelal gewijzigd worden of verloren gaan. Ze stellen niet de vraag of alle 'originele' eigenschappen wel bijdragen tot de archiefstatus van een digitaal object en bijgevolg wel dienen gearhiveerd te worden, hoe de originele 'look and feel' kan gedefinieerd worden en of het behouden van de oorspronkelijke functionaliteiten wel noodzakelijk is. Ze beschouwen digitale archiefdocumenten louter als digitale artefacten waarvan alle eigenschappen behouden dienen te worden. Het is niet toevallig dat de grote emulatievoorstanders in de eerste plaats computerwetenschappers zijn.

Niettegenstaande dit alles, blijft emulatie een potentiële strategie die zijn nut voor de archivering van digitale archiefdocumenten kan hebben. In ieder geval heeft men nog maar beperkte ervaringen met emulatie als digitale bewaarstrategie. Er zijn nog maar weinig praktische en grootschalige emulatietoepassingen voor digitale archivering operationeel.

3.4 Besluit: DAVID-bewaarstrategie

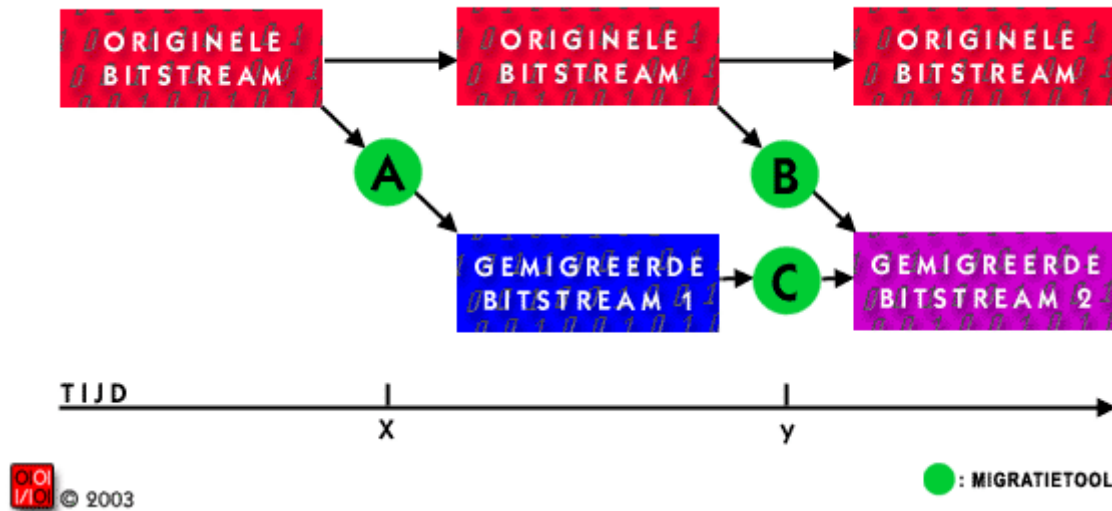
Uit de evaluatie van de mogelijke digitale bewaarstrategieën blijkt dat er momenteel nog geen definitieve oplossing voor het bewaarprobleem van digitale archiefdocumenten is. Migratie en emulatie zijn theoretisch het best geschikt om het digitale duurzaamheidsprobleem op te lossen. Beide oplossingen hebben met elkaar gemeen dat ze een bitstream vertalen naar een leesbaar document. Migratie en emulatie doen dit wel op een verschillend tijdstip. Bij migratie gebeurt dit in het heden, terwijl emulatie deze actie naar de toekomst verschuift. Migratie biedt een oplossing aan documentenzijde, terwijl emulatie voor het leesbaarheidsprobleem aan de hard- en/of softwarezijde een oplossing zoekt.

Inmiddels is het inzicht gegroeid dat beide benaderingen elkaar niet uitsluiten. Beide oplossingen zijn complementair in de levenscyclus van een digitaal archiefdocument of zijn meer geschikt voor een welbepaald type digitaal archiefdocument. In het algemeen is emulatie meer geschikt wanneer de "look and feel" en het gedrag van een document van belang is, terwijl migratie volstaat wanneer inhoud en structuur de essentiële componenten van een archiefdocument zijn. Ondertussen bestaan ook tal van tussenoplossingen die elementen van de migratie- en emulatieoplossing combineren.

De bewaarstrategie die het DAVID-project voorstelt⁸, is een middenweg tussen emulatie en migratie en houdt naar de toekomst toe nog alle opties open. Dit kan door het bewaren van de originele bitstreams te combineren met het creëren en bijhouden van gemigreerde versies in een bestandsformaat die meer garanties inzake leesbaarheid biedt. Digitale archiefdocumenten die niet in een geschikt archiveringsformaat zijn opgeslagen, worden voor opname in het digitaal archief

⁸ Deze zienswijze is geïnspireerd op de "Migration on request"-strategie van het CAMiLEON-project en op de benadering van de Nationale Archiefdienst van Australië (P. MELLOR, P. WHEATLEY EN D. SERGEANT, *Migration on Request - a practical technique for preservation*, <http://www.si.umich.edu/CAMiLEON/reports/mor/index.html>; H. HESLOP, S. DAVIS EN A. WILSON, *National Archives Green Paper: An approach to the preservation of digital records*, Canberra, 2002, http://www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html)

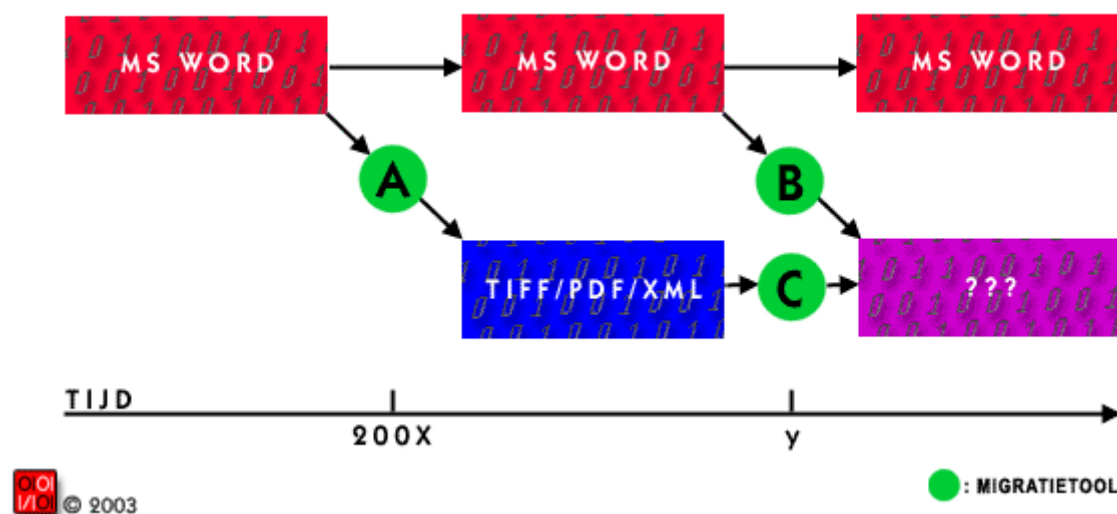
gemigreerd naar een archiveringsformaat. Het archiefdocument in zijn oorspronkelijk bestandsformaat wordt niet vernietigd. Van deze archiefdocumenten worden dus twee bitrepresentaties bijgehouden: één in het oorspronkelijk bestandsformaat, één in het gemigreerd bestandsformaat. Men kan deze bitrepresentaties in afzonderlijke computerbestanden bewaren of in een XML-container inkapselen. Dit biedt het voordeel dat in de toekomst zowel emulatie als migratie van het oorspronkelijk of het gemigreerde bestandsformaat mogelijk zijn. Voor digitale archiefdocumenten die van bij hun creatie in een geschikt archiveringsformaat worden bijgehouden, is geen migratie nodig zodat slechts één bitrepresentatie wordt bewaard.



Toegepast op een tekstdocument dat werd opgeslagen in MS Wordformaat houdt deze bewaarstrategie de volgende stappen in. Ten laatste op het moment van opname in het digitaal archief wordt het tekstdocument in MS Word gemigreerd naar een geschikt archiveringsformaat. MS Word is immers een ongedocumenteerd producent- en applicatiegebonden bestandsformaat met in tijd beperkte ondersteuning⁹, dat helemaal niet geschikt is voor lange termijnarchivering. In het digitaal depot worden het MS Wordbestand en het gemigreerde bestand opgeslagen. Wanneer het archiveringsformaat XML, TIFF of PDF in onbruik dreigt te geraken, heeft men de keuze tussen verschillende opties:

- gebruiken van een emulator voor het MS Wordformaat
- gebruiken van een emulator voor het gemigreerde formaat
- migratie naar een nieuw archiveringsformaat op basis van het MS Wordbestand
- migratie naar een nieuw archiveringsformaat op basis van het gemigreerde formaat

⁹ Microsoft Corporation hanteert vanaf 15 oktober 2002 een formeel Support Life Cycle beleid. Dit beleid bevat richtlijnen voor de beschikbaarheid van productondersteuning. (<http://support.microsoft.com/default.aspx?scid=fn;nl:complifeport>). Zo is er na 30 juni 2008 in principe geen ondersteuning meer voor MS Word 2002.



Zelfs al lijkt emulatie in het geval van MS Word een weinig waarschijnlijke mogelijkheid (auteursrecht, niet gedocumenteerd bestandsformaat), deze bewaarstrategie zou kunnen betekenen dat in het tweede archiveringsformaat meer originele eigenschappen van het archiefdocument aanwezig zijn dan in het eerste archiveringsformaat.

4. GESCHIKTE ARCHIVERINGSFORMATEN

In de DAVID-bewaarstrategie worden de digitale documenten omgezet naar een geschikt archiveringsformaat. Opslag in een archiveringsformaat biedt de grootste raadplegingsgaranties en is het beste vertrekpunt voor migratie en/of emulatie in de toekomst.

Een geschikt archiveringsformaat voldoet aan volgende vereisten:

- gestandaardiseerd: gedocumenteerd, stabiel en niet afhankelijk van één producent
- wijdverspreid en voldoende marktpenetratie
- uitwisselbaar: onafhankelijk van bepaalde besturingssystemen en applicaties
- robuust foutopsporing- en verbeteringsmechanisme: fouten in bitopslag zijn herstelbaar, validatiefunctie
- goed gestructureerde opslag van informatie
- opslag zonder informatieverlies
- mogelijkheid tot insluiten van (zelfgedefinieerde) metadatavelden
- in staat om de essentiële eigenschappen van het archiefdocument in tijd over te brengen
- bewaren van de authenticiteit van de archiefdocumenten
- autonoom en zelfvoorzienig
- drager en apparaat onafhankelijke opslag mogelijk
- eenvoudig en gebruiksvriendelijk.

Tabel: Voorbeelden van geschikte archiveringsformaten en raadplegingskopieën

TYPE DOCUMENT	MOEDERKOPIE	RAADPLEGINGSKOPIE
Tekst	XML(Unicode), TIFF, PDF/A	XML (Unicode), TIFF, PDF/A
Afbeelding	TIFF, CGM	JPEG

TYPE DOCUMENT	MOEDERKOPIE	RAADPLEGINGSKOPIE
Geluid	WAVE (PCM)	(streaming)MP3 RM

Vanuit archiveringsperspectief is het niet alleen van belang dat archiefdocumenten met lange bewaartermijn in een archiveringsformaat worden opgeslagen, maar dat ook een geschikt profiel van het archiveringsformaat wordt toegepast. De meeste archiveringsformaten laten de gebruiker toe om een aantal instellingen en parameters te definiëren. Zo kan men diverse soorten TIFF-, XML-, PDF- en WAV-bestanden produceren, maar niet elk TIFF-, XML-, PDF- of WAV-document is even geschikt om op lange termijn te archiveren. Bij afbeeldingen die als TIFF-bestand worden bewaard, kan JPEG-compressie worden toegepast. MP3-gecomprimeerd audiomateriaal kan verpakt worden in een WAV-bestand. In beide gevallen gaan niet alleen gegevens verloren, maar voor de reconstructie is men afhankelijk van de overeenstemmende decompressie. De kwaliteit van XML-documenten is afhankelijk van de nesting en semantiek van de XML-tags. PDF-documenten bestemd voor lange termijnbewaring zijn bij voorkeur getagd of op zijn minst gestructureerd.

Een goede leidraad bij het toepassen van een geschikt profiel van een bepaald archiveringsformaat is het uitgangspunt om zo autonoom mogelijke digitale archiefdocumenten te archiveren. Voor digitale moederkopieën worden reconstructieafhankelijkheden best tot een absoluut minimum beperkt. Het ontbreken van één noodzakelijke schakel in het reconstructieproces kan immers tot verlies van het archiefdocument leiden. De gearchiveerde documenten zijn bijgevolg best zo zelf voorzienig mogelijk. Omwille deze reden worden externe afhankelijkheden (decompressie, decryptie, paswoorden of andere beveiligingsinstellingen, enz.) zoveel mogelijk vermeden terwijl informatie nodig voor de interpretatie en reconstructie zoveel mogelijk wordt ingekapseld. Deze vereisten gelden hoofdzakelijk voor de digitale moederkopieën en in mindere mate voor de raadplegingskopieën.

Het toepassen van compressie is niet alleen af te raden vanwege de decompressieafhankelijkheid, maar ook vanwege het informatie en kwaliteitsverlies dat compressie in de meeste gevallen met zich mee brengt. De efficiëntste compressiemethoden zijn immers *lossy*, wat inhoudt dat de gedecomprimeerde bitstream niet identiek is aan de oorspronkelijke bitstream. Dit is een gevolg van het wegfilteren van informatie ('bit-rate-reduction') en levert kleinere digitale bestanden op. Verwerking, opslag en uitwisseling van gecomprimeerde bitstreams verloopt sneller en efficiënter. Maar door *lossy* compressie ontstaan ook vervormingen waardoor de kwaliteit van de gearchiveerde documenten daalt. Het meervoudig toepassen van compressie leidt op termijn eveneens tot visueel waarneembaar generatieverlies¹⁰. Dit is het duidelijkst wanneer achtereenvolgens verschillende compressiealgoritmes worden toegepast, maar ook wanneer men meermaals dezelfde compressiemethode gebruikt, treedt opmerkelijk informatie- en kwaliteitsverlies op.

Het vermijden van compressie bij de opslag van digitale moederkopieën is een voorzichtige aanpak die niet alleen vanuit theoretisch en gebruikersonderzoek wordt onderbouwd, maar ook mee gebaseerd is op de expertise en praktische ervaringen die de internationale archiefwereld opdeed bij het digitaal archiveren van tekst, afbeeldingen en geluid. Het ongecomprimeerd bewaren van digitale moederkopieën is bovendien ook veiliger en houdt minder risico's in. Gecomprimeerde digitale documenten zijn immers kwetsbaarder dan ongecomprimeerde documenten. Een fout in een gecomprimeerd bestand leidt sneller tot onherstelbaar verlies.

Overigens mag men niet uit het oog verliezen dat de compressienoodzaak vooral voortvloeit uit de beperkingen die de huidige technologie oplegt en dat ten gevolge van de technologische vooruitgang deze restricties de komende jaren soepeler worden of zelfs helemaal zullen verdwijnen. Dit geldt overigens ook voor de compressiemethoden die continu verbeterd worden.

¹⁰ Gebruikersonderzoek van de RAI wees bijvoorbeeld uit dat Digitale Betacam na 7 generaties al leidt tot visueel waarneembare vervormingen. Na 4 generaties Digitale Betacam wordt al in toenemende mate ruis waargenomen. Digitale Betacam (Sony) past 2.34:1 op DCT-gebaseerde *lossy* compressie toe.

5. ARCHIVERINGSFORMAAT VOOR VIDEO

In tegenstelling tot tekst, afbeeldingen en geluid is er nog geen geschikt digitaal archiveringsformaat voor video beschikbaar. Het digitaal archiveringsformaat voor video voldoet bij voorkeur aan alle kwaliteitsvereisten van de geschikte archiveringsformaten in het algemeen. Naast de algemene kenmerken gelden nog een aantal specifieke vereisten voor digitale video:

- ondersteuning van 10 bits/pixel
- meerdere frame rates/sizes mogelijk
- bevat tijdsaanduiding in het bestand
- opslag van (multichannel) audio en video in één bestand

De kenmerken van een geschikt archiveringsformaat stellen strenge eisen ten aanzien van het formaat waarin digitale video wordt gearchiveerd. Vooral de noodzaak om video en audio data ongecomprimeerd te bewaren, maakt de keuze van een geschikt archiveringsformaat er niet gemakkelijker op. Anno 2004 is ongecomprimeerde opslag van tekst, afbeeldingen en geluid gemakkelijker te realiseren dan het ongecomprimeerd bewaren van bewegend beeld. Ongecomprimeerde opslag van bewegend beeld en geluid is technisch veeleisend en heel duur, maar niet onmogelijk. Voorbeelden van digitale videoformaten die geen compressie toepassen zijn D5 (Panasonic) voor Standard Definition Television en D6/Voodoo (Philips-Toshiba) of D5 HD (Panasonic) voor High Definition Television. Deze formaten zijn echter gesloten, producentgebonden en (nog) niet wijdverspreid en bijgevolg geen geschikte archiveringsformaten voor bewegend beeld. Digitale videoformaten met een grotere marktpenetratie zoals Digitale Betacam (Sony), DVCAM (Sony) en DVCPRO (Panasonic) passen dan weer *lossy* compressie toe¹¹, waardoor ze evenmin als archiveringsformaat bruikbaar zijn. Bovendien zijn deze laatste formaten ook gesloten, producentgebonden en afhankelijk van specifieke dragers. Digitale Betacam is nochtans een veel gebruikt formaat, terwijl het als een “kwetsbaar” formaat wordt geëvalueerd¹². Voorlopig komt geen enkel commercieel digitaal videoformaat in aanmerking om als archiveringsformaat gebruikt te worden.

Vanwege deze lacune lopen momenteel diverse initiatieven en projecten met als doel een geschikt archiveringsformaat voor video te ontwikkelen (bijv. PRESTO, PRESTOSPACE, DHC Digital Video Preservation Reformatting Project, Amicitia). Nader onderzoek moet ook uitwijzen of de bestandsformaten AAF (Advanced Authoring Format¹³) en diens afgeleide MXF (Material eXchange Format¹⁴) eventueel bruikbaar zijn als archiveringscontainers voor digitale video¹⁵.

AAF en MXF zijn in staat om ongecomprimeerde, *loss/less* gecomprimeerde en *lossy* gecomprimeerde digitale videobeelden in tijd over te brengen. Beide formaten voldoen aan de voornaamste criteria van een archiveringsformaat: open, gedocumenteerd, gestandaardiseerd, open source ontwikkeling, uitwisselbaar, mogelijkheid tot ongecomprimeerde opslag van audio en videodata, brede industriële ondersteuning, uitbreidbaar, enz. AAF-documenten kunnen eveneens voor elk programma element een volledige historiek van bronmateriaal tot afgewerkte video bevatten. Als beide formaten een vrij 'natuurlijke' representatie van de videodata (de 'essence') kunnen bevatten en voldoende verspreiding vinden, dan hebben ze een vrij grote kans op slagen. Beide formaten kunnen voor archiveringsdoeleinden ook in combinatie met elkaar worden gebruikt (AAF voor de moederkopie, MXF voor de (streaming) raadplegingskopie). Bij AAF dient men er wel over te waken dat de bestanden volledig zelf voorzienig zijn en niet linken naar externe bronnen.

De Library of Congress en het Information Management Office van de Australische overheid

11 Digitale Betacam: 2.34: 1
DVCAM: 5:1
DVCPRO: 5:1

12 <http://paulmessier.com/Videoid/> → 1985 to present: Digital Betacam. Onder 'vulnerable' verstaat de auteur: “This is a current but highly proprietary format”.

13 AAF is ontwikkeld door de AAF Association (<http://aafassociation.org>)

14 MXF is ontwikkeld door het Professional MPEG Forum (<http://www.pro-mpeg.org>) en de AAF Association.

15 Voor een bespreking van beide formaten voor digitale video, zie: F. BOUDREZ, *Standaarden voor digitale archiefdocumenten*, p. 32-34. (<http://www.antwerpen.be/david>). Zie ook: F. PAVUZA, *Format and file issues for video archiving*, ErpaWorkshop: File formats for digital preservation, 11 mei 2004.

hebben (uncompressed en lossless) MXF inmiddels al opgenomen als geprefereerd formaat voor digitale video¹⁶. Het Amerikaanse Digital Video Preservation Reformatting Project vergeleek AAF en MXF en beveelde tenslotte ook MXF als archiveringsformaat aan¹⁷. Belangrijke argumenten in de keuze voor MXF zijn:

- de zelf-voorzienigheid van MXF-documenten: alle essence en metadata worden in het MXF-bestand zelf bewaard (AAF kan pointers naar externe essence of metadata bevatten)
- MXF is geschikt voor zowel lange termijnarchivering als raadpleging
- AAF is meer gericht op productie en post-productie, terwijl MXF enkel afgewerkt materiaal bevat en dus hoofdzakelijk voor raadpleging dient
- edit lists en metadata in AAF-bestanden kunnen van producent- of softwaregebonden aard zijn

Deze argumenten sluiten AAF nog niet noodzakelijk als archiveringsformaat voor videomateriaal uit. Bij het gebruik van AAF voor videoarchivering dient men er wel over te waken dat de AAF-bestanden: volledig zelfvoorzienig zijn, niet afhankelijk zijn van externe informatie en geen producent- of softwaregebonden items bevatten. Met andere woorden, men kan een bepaald profiel van AAF toepassen zodat AAF wel bruikbaar is voor lange termijnarchivering. Men kan dit vergelijken met het toepassen van een bepaald PDF- of TIFF-profiel voor de archivering van tekstdocumenten.

De keuze van een wrapperformaat voor de archivering van digitale video is één zaak, de keuze van een videocodec en audiocodec een andere. Aangezien lange termijnarchivering hoge eisen stelt tav de beeldkwaliteit zijn *lossy*-codecs absoluut te vermijden (zie hoger). Een geschikte videocodec voor lange termijnarchivering dient bijgevolg geen of *lossless* compressie toe te passen. Ongecomprimeerde opslag en verwerking van videomateriaal zijn echter momenteel nog niet altijd technisch haalbaar, zodat *lossless* compressie zich in de meeste gevallen opdringt. Ten aanzien van de videocodec gelden de meeste vereisten die van toepassing zijn op geschikte archiveringsformaten in het algemeen. Momenteel is M-JPEG2000 (Motion JPEG2000) de videocodec die het best aan al deze vereisten beantwoordt. M-JPEG2000 is gebaseerd op de JPEG2000-standaard voor stilstaande afbeeldingen¹⁸. Met M-JPEG2000 is het mogelijk om digitale video in *lossless* compressie te archiveren. M-JPEG2000 past enkel intra-frame compressie (Discrete Wavelet Transform) toe, en geen inter-frame compressie, waardoor de individuele frames nog zonder kwaliteitsverlies kunnen worden geëditeerd¹⁹. Ongecomprimeerde opslag van het audiosignaal is wel eenvoudig te realiseren. Voor het geluidssignaal geniet bijgevolg ongecomprimeerde WAV de voorkeur.

6. PRAKTISCHE AANBEVELINGEN VOOR VIDEOARCHIVERING

Zolang er nog geen zekerheid over een geschikt archiveringsformaat en codec voor video is, nemen archiefbeherende instellingen eerder een afwachtende en voorzichtige houding aan. Digitale archivering van video staat bijgevolg nog in zijn kinderschoenen. De digitalisering van analoge beeldmateriaal voor preservatiedoeleinden wordt zo lang mogelijk uitgesteld²⁰. De algemene verwachting is dat er nog veel zal bewegen op het vlak van digitale video en televisie.

Wanneer digitalisering zich opdringt en langer wachten geen optie meer is, dan wordt bij voorkeur een digitale moederkopie van een zo hoog mogelijke kwaliteit gecreëerd die voor meerdere doeleinden bruikbaar is. Van de digitale moederkopie worden twee versies bijgehouden: een niet bijgewerkte versie en een 'gekuiste' versie. Op basis van de 'gekuiste' digitale moederkopie worden raadplegingskopieën gemaakt. De digitale moederkopie wordt idealiter ongecomprimeerd opgeslagen. In de gevallen dat compressie onvermijdelijk is, is het vanuit archiveringsstandpunt

16 Library of Congress: http://www.digitalpreservation.gov/formats/content/video_preferences.shtml;
Information Management Office (Australië):
<http://www.agimo.gov.au/practice/delivery/checklists/digitisation>.

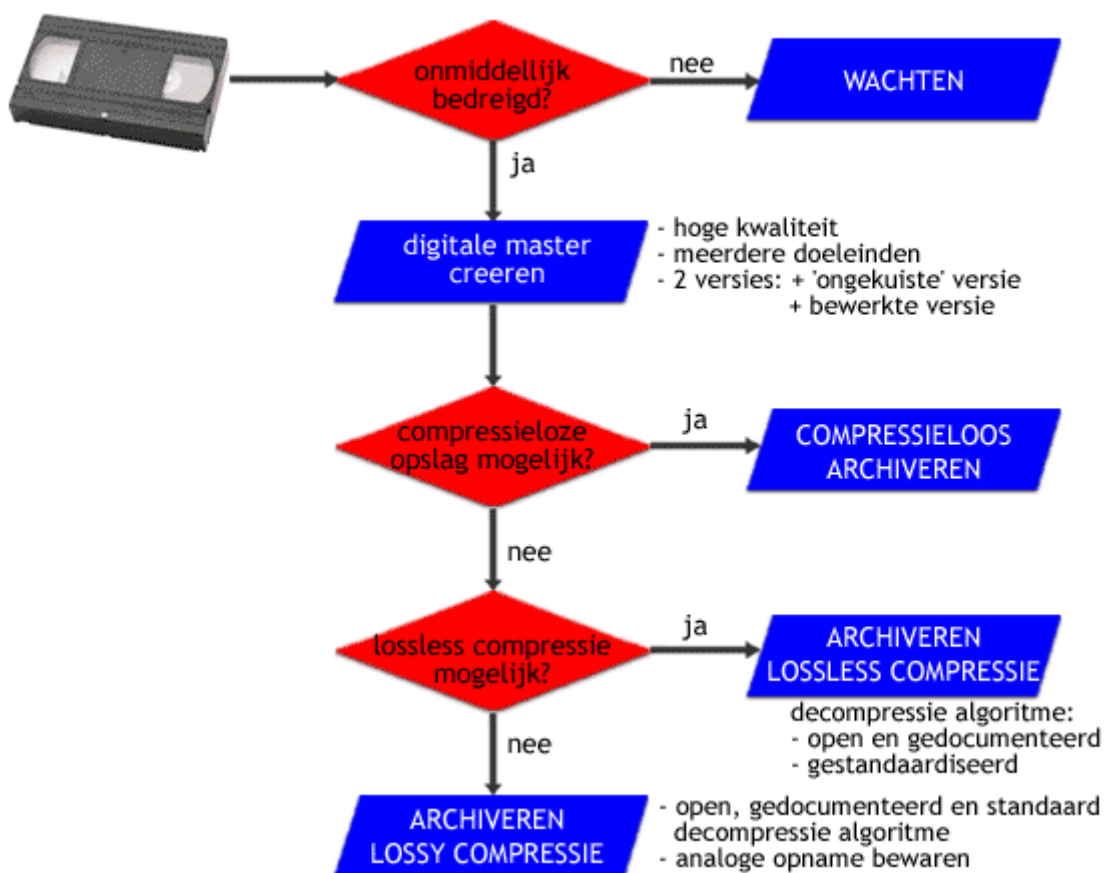
17 DIGITAL VIDEO PRESERVATION REFORMATTING PROJECT, *A Report*, 2004.

18 ISO/IEC 15444-1:2004, Information technology. JPEG 2000 image coding system: Core coding system

19 MPEG past zowel inter-frame als intra-frame compressie toe.

20 Bijv. C. FLEISCHHAUER, *Audio and video reformatting. A Library of Congress perspective*, 2003.

belangrijk om een *lossless* compressiemethode toe te passen. Bij *lossless* compressie is de gedecomprimeerde bitstream identiek aan de oorspronkelijke bitstream. Deze compressiemethode resulteert niet in informatieverlies of vervormingen, maar is doorgaans minder efficiënt dan *lossy* compressie. Het decompressie-algoritme is open, gedocumenteerd en gestandaardiseerd. Van zodra een geschikt archiveringsformaat beschikbaar is, is het aan te bevelen om de digitale moederkopie naar dit formaat om te zetten. Wanneer toch een *lossy* compressiemethode wordt toegepast, blijft de analoge beeldopname bewaard zodat die eventueel nog als bron voor digitalisering kan worden gebruikt wanneer het geschikte archiveringsformaat beschikbaar is. Onderzoek moet dan uitwijzen of de analoge versie dan wel de digitale versie de beste kwaliteit biedt. Vanuit lange termijnarchiveringsstandpunt is het hoe dan ook belangrijk dat elke digitale video naar het geschikte archiveringsformaat wordt omgezet van zodra men over dit formaat beschikt.



Figuur: Aanbevelingen voor het digitaliseren van videomateriaal in afwachting totdat een geschikt archiveringsformaat voor video beschikbaar is.