



DIGITALISEREN VAN ANALOGIE ARCHIEFDOCUMENTEN

1. Woord vooraf

Analoge opslagmedia kunnen archiefdocumenten dragen die voor lange termijnarchivering in aanmerking komen. Deze informatiedragers zijn echter onderhevig aan gebruiksslijtage, degeneratie en technologische veroudering waardoor ze niet geschikt zijn als lange termijn drager voor bijvoorbeeld geluids- en/of videomateriaal.

Het tijdig omzetten van de archiefdocumenten naar nieuwe analoge dragers is geen goede oplossing, want bij elke analoge omzetting gaat informatie en kwaliteit verloren. Dit verlies is onvermijdelijk en wordt doorgaans met de term 'generatieverlies' aangeduid. Aangezien analoge media en de nodige afspeltechnologie snel verouderen, is er met analoge archivering aan (te) hoog tempo generatieverlies.

Een betere oplossing is de digitalisering van het archiefmateriaal dat op analoge dragers is opgeslagen. Mits digitalisering is het mogelijk om in de toekomst de bitstreams onbeperkt te kopiëren en naar andere dragers over te brengen zonder dat dit met informatie- en kwaliteitsverlies hoeft samen te gaan. Met digitalisering moet men echter ook oppassen, want er zijn tal van factoren waarmee men rekening dient te houden. Door te digitaliseren komt men op het terrein van digitale archivering en er zijn een aantal valkuilen die absoluut moeten vermeden worden.

Immers, door te digitaliseren is men in de toekomst niet meer alleen afhankelijk van de nodige afspelapparatuur en de kwaliteit van de dragers maar ook van bestandssystemen en softwareapplicaties om de digitale documenten te raadplegen. Bovendien verouderen software en digitale formaten aan een hoger tempo dan analoge apparatuur. Omzettingen blijven bijgevolg nodig. Bij elke omzetting schuilt het gevaar op informatie- en kwaliteitsverlies om de hoek.

Digitalisering is als lange termijn archiveringsstrategie voor analoge archiefdocumenten pas effectief wanneer men in staat is om op termijn de leesbaarheid en de kwaliteit intact te houden. Dit vraagt van bij het digitaliseringsmoment een digitale bewaarstrategie voor de gedigitaliseerde archiefdocumenten die maximale waarborgen biedt voor de lange termijn leesbaarheid en voor het instandhouden van het kwaliteitsniveau van de archiefdocumenten. Dit is realiseerbaar door van bij de digitalisering de gedigitaliseerde archiefdocumenten onmiddellijk in een geschikt archiveringsformaat te bewaren en door zorgvuldig de digitaliseringsparameters te kiezen. Dit zijn absolute voorwaarden voor elk digitaliseringsinitiatief waarbij men de gedigitaliseerde documenten voor het nageslacht wil bewaren en eventueel de analoge archiefdocumenten wil vernietigen.

In onderstaand overzicht wordt voor elke type document de belangrijkste digitaliseringsparameters, de meest courante instellingen en de geschikte archiveringsformaten opgesomd. De keuze van de geschikte archiveringsformaten is gebaseerd op de lange termijn bewaarstrategie voor digitale archiefdocumenten die het DAVID-project in zijn vademecum voorstelt. Dit overzicht wordt voorafgegaan door een aantal algemene aanbevelingen.

2. Algemene aanbevelingen voor digitalisering

Belangrijke aandachtspunten en vuistregels voor elk digitaliseringsproject zijn:

... bij de planning en de voorbereiding:

- ❑ formuleer op voorhand duidelijke doelstellingen en weeg af of digitalisering de efficiëntste en meest kost effectieve oplossing is.
 - digitalisering is duur en arbeidsintensief. Naast het eigenlijke digitaliseringswerk vergen volgende werkzaamheden ook veel budget en tijd:
 - reinigen, verpakken, nummeren en beschrijven van de archiefdocumenten
 - verwerken en controleren digitaliserings
 - aanmaken en invoeren metadata van de gedigitaliseerde documenten.
 - digitalisering slurpt veel technische middelen op
- ❑ stel duidelijke selectiecriteria op en leg indien nodig prioriteiten en een fasering vast.
- ❑ besteed de nodige aandacht aan het beschrijven van de gedigitaliseerde documenten en aan de duurzame koppeling van de gedigitaliseerde archiefdocumenten met hun beschrijving
- ❑ leg voor elk type document vast welke technische en beschrijvende metadata in het archiveringsformaat worden ingekapseld in de fileheader. Inkapseling van deze metadata biedt het voordeel dat ze
 - onlosmakelijk verbonden zijn met het gedigitaliseerde document
 - geëxtraheerd kunnen worden voor de automatische beschrijving
- ❑ bepaal de bestemming van de analoge archiefdocumenten na digitalisering: voor vernietiging van de originele archiefdocumenten hebben openbare instellingen de toestemming van de Algemene Rijksarchivaris of diens gemachtigde nodig (Archiefwet van 24 juni 1955: art. 5)
- ❑ zorg ervoor dat het digitaliseringsproject het auteursrecht respecteert. Voor auteursrechtelijk beschermde werken is voor digitalisering (= maken van een digitale kopie) de toestemming van de dragers van het auteursrecht vereist.
- ❑ plan een lange termijn digitale bewaarstrategie alvorens met digitalisering te starten:
 - voorzie een oplossing voor het probleem van de technologische veroudering van digitale formaten en dragers die de kwaliteit van de gedigitaliseerde archiefdocumenten niet schaadt.
 - denk ook aan de lange termijnbewaring van de beschrijvingen en de bijhorende ontsluitingsmechanismen. De beschrijvingen hebben minstens dezelfde levensduur als de gedigitaliseerde archiefdocumenten.
 - zorg voor een niet technologie gebonden koppeling tussen gedigitaliseerde archiefdocumenten en hun beschrijving: kies een wijze van relateren die duurzaam is.
- ❑ plan de valorisatie en het gebruik van het gedigitaliseerd materiaal. Zorg ervoor dat de gedigitaliseerde archiefdocumenten toegankelijk zijn en onder de aandacht worden gebracht.
- ❑ laat eerst een staal digitaliseren en controleer de kwaliteit van dit staal zodat bijstellingen nog mogelijk zijn vooraleer alle geselecteerd materiaal wordt gedigitaliseerd.

... bij het kiezen van een digitaal formaat

- ❑ kies voor de digitale moederkopieën een geschikt archiveringsformaat. Bewaar digitale moederkopieën rechtstreeks in dit formaat. Een archiveringsformaat beantwoordt aan volgende criteria:

- gestandaardiseerd: open, gedocumenteerd, stabiel
- wijdverspreid en voldoende marktpenetratie
- uitwisselbaar: onafhankelijk van bepaalde besturingssystemen, netwerkprotocollen en applicaties
- voorziet een robuust foutopsporings- en verbeteringsmechanisme
- goed gestructureerde opslag van informatie
- opslag zonder informatieverlies (geen *lossy* compressie)
- mogelijkheid tot insluiten van (zelfgedefinieerde) metadatavelden
- in staat om de essentiële eigenschappen van de archiefdocumenten in tijd over te brengen
- bewaren van de authenticiteit van de archiefdocumenten
- autonoom en zelfvoorzienig
- niet afhankelijk van specifieke dragers en afspeelapparaten
- in het geval er nog geen geschikt archiveringsformaat beschikbaar is:
 - stel digitalisering zo lang mogelijk uit
 - indien uitstel geen optie is:
 - houdt het analoge origineel bij zodat het eventueel nog als digitaliseringsbasis kan dienen
 - migreer naar een geschikt archiveringsformaat van zodra dit beschikbaar is
 - onderzoek of de analoge of de eerder gedigitaliseerde versie het beste digitaliseringsresultaat geeft
 - beperk informatie- en kwaliteitsverlies tot een absoluut minimum
- in het geval compressie niet kan vermeden worden:
 - gebruik altijd een open, gedocumenteerde en gestandaardiseerde (de)compressiemethode
 - pas een *loss/less* compressiemethode toe (originele bitstream = gedecomprimeerde bitstream): vermijd dat er ten gevolge van compressie informatie en/of kwaliteit verloren gaat
 - pas enkel *lossy* compressie (originele bitstream \neq gedecomprimeerde bitstream) toe wanneer het niet anders kan
- leg voor elk type document volgende parameters vast:
 - het bestandsformaat en de digitaliseringsparameters voor de digitale moederkopie en eventueel de raadplegingskopieën
 - de structuur van de bestandsnaam die aan de bestanden wordt toegekend
 - de softwareapplicaties waarmee de gedigitaliseerde documenten raadpleegbaar moeten zijn en waarmee de gedigitaliseerde documenten worden gecontroleerd

... bij het digitaliseren:

- maak een digitale moederkopie van een zo hoog mogelijke kwaliteit, die:
 - voor meerdere doeleinden kan worden gebruikt (reproducties, onderzoek, samenstellen raadplegingskopieën, enz.)
 - een zo getrouw mogelijke kopie van het analoge origineel is: baseer de digitaliseringsparameters op de kwaliteit van het analoge origineel en bouw bijkomend een veiligheidsmarge in. Houd ook rekening met het digitaliseringsdoel (vervanging analoge originelen of on line terbeschikkingstelling).

- ❑ besteed aandacht aan de kwaliteit van de gedigitaliseerde documenten bij het vastleggen van:
 - de sampling-rate: de frequentie of het interval van de metingen
 - de sampling-precision of de kwantificering: de accuraatheid of de precisie waarmee het meetresultaat in een binaire waarde wordt omgezet
 - de encoding: de omzetting van de data in een software leesbaar formaat
- ❑ bewaar een onbewerkte versie van de digitale moederkopie ('preservation master copy'). Indien nodig, maak een gekuiste of bewerkte versie voor raadplegingsdoeleinden en die ook als basis voor de productie van raadplegingskopieën kan dienen ('production master copy').
- ❑ houd het analoge origineel bij zolang je geen sluitende zekerheid hebt over de kwaliteit en de lange termijnbewaring van de gedigitaliseerde moederkopie
- ❑ gebruik voor de digitalisering afspeelapparatuur, hard- en software van uitstekende kwaliteit die internationale standaarden correct toepast: calibreer de apparaten alvorens met digitalisering te starten
- ❑ documenteer het digitaliseringsproces. Houd documentatie bij over de selectiecriteria, het geselecteerd materiaal, de opslagformaten, de digitaliseringsparameters en -instellingen, de opslagmedia, de procedure, de beschrijving, de eventuele bewerkingen, de bestemming van de analoge archiefdocumenten, enz.

... bij de verwerking en de controle

- ❑ controleer of valideer de gedigitaliseerde archiefdocumenten:
 - correcte toepassing van de voorgeschreven bestandsformaten, codecs en digitaliseringsinstellingen
 - toekenning bestandsnamen conform de afspraken
 - correcte invulling van de metadata (ingekapselde metadata in de archiveringsformaten, metadata op de dragers)
 - volledigheid: alle geselecteerd materiaal gedigitaliseerd? volledigheid van de gedigitaliseerde documenten

... bij de opslag op dragers

- ❑ bewaar de digitale moederkopieën op dragers die geschikt zijn voor lange termijnarchivering. Zie onder '7. Meer informatie' voor verwijzingen naar andere DAVID publicaties over duurzame digitale dragers.

3. Digitaliseren van tekst

Kwaliteitsvereisten moederkopie:

- ❑ Bestandsformaten: TIFF (versie 6.0)
 - extensie Windows, Unix en OS/2: ".TIF"
 - extensie Mac: ".TIFF"
- ❑ Compressie: geen
- ❑ Byte-volgorde: "Intel" TIFF
- ❑ Metadatavelden in de TIFF-header voor archiefmetadata:
 - Compression (tagnr. 259): 1 (=uncompressed)

- `DocumentName` (tagnr. 269): bijv. bestandsnaam
- `ImageDescription` (tagnr. 270): beschrijving van de afbeelding
- `Make` (tagnr. 271): fabrikant van de scanner, video, apparatuur die werd gebruikt voor het maken van de afbeelding
- `Model` (tagnr. 272): gebruikte scanner
- `Software` (tagnr. 305): software naam en versienummer van de software die werd gebruikt voor het maken van de afbeelding
- `DateTime` (tagnr. 306): datum van digitalisering (JJJJ:MM:DD UU:MM:SS)
- `Artist` (tagnr. 315): bijv. naam van de archiefbeherende instelling
- `HostComputer` (tagnr. 316): bijv. hardware, besturingssysteem
- `CopyRight` (tagnr. 33432): houder van het copyright
- ! **Opmerking1:** Deze tags bevatten een ASCII-veld (7 bits ASCII) waarvan de laatste byte 'NUL' moet zijn.
- ! **Opmerking2:** Een TIFF-document is in zekere zin uitbreidbaar. Organisaties kunnen zelfgekozen tags (vanaf tagnr. 32768) aan TIFF-documenten toevoegen. Deze zogenaamde "private tags" worden beter vermeden, want ze kunnen immers leesbaarheids- en uitwisselingsproblemen opleveren en vereisen specifieke software. De "private tags" en hun inhoud kunnen verloren gaan, wanneer de afbeelding achteraf wordt bewerkt en opgeslagen met een applicatie die deze tags niet ondersteunt. Bepaalde applicaties voegen echter altijd één of meerdere "private tags" aan TIFF-documenten toe (bijv. PhotoShop, PageMaker, enz.).
- Digitaliseringsparameters en veel gebruikte instellingen:
 - resolutie: 400 PPI, 300 PPI, 150 PPI
 - bitdiepte:
 - kleur afbeeldingen: 24 of 48 bits/pixel
 - grijswaarden afbeelding: 8 of 16 bits /pixel
 - zwart-wit afbeelding: 1 bit/pixel

4. Digitaliseren van afbeeldingen

Kwaliteitsvereisten moederkopie:

- Bestandsformaten: TIFF (versie 6)
 - extensie Windows, Unix en OS/2: ".TIF"
 - extensie Mac: ".TIFF"
- Compressie: geen
- Byte-volgorde: "Intel" TIFF (wordt ook aangeduid als "PC"- of "IBM"-volgorde)
- Metadatatavelden in de TIFF-header voor archiefmetadata:
 - `Compression` (tagnr. 259): 1 (=uncompressed)
 - `DocumentName` (tagnr. 269): bijv. bestandsnaam
 - `ImageDescription` (tagnr. 270): beschrijving van de afbeelding
 - `Make` (tagnr. 271): fabrikant van de scanner, video, apparatuur die werd gebruikt voor het maken van de afbeelding
 - `Model` (tagnr. 272): gebruikte scanner

- Software (tagnr. 305): software naam en versienummer van de software die werd gebruikt voor het maken van de afbeelding
- DateTime (tagnr. 306): datum van digitalisering (JJJJ:MM:DD UU:MM:SS)
- Artist (tagnr. 315): bijv. naam van de archiefbeherende instelling
- HostComputer (tagnr. 316): bijv. hardware, besturingssysteem
- CopyRight (tagnr. 33432): houder van het copyright
- ! **Opmerking1:** Deze tags bevatten een ASCII-veld (7 bits ASCII) waarvan de laatste byte 'NUL' moet zijn.
- ! **Opmerking2:** Een TIFF-document is in zekere zin uitbreidbaar. Organisaties kunnen zelfgekozen tags (vanaf tagnr. 32768) aan TIFF-documenten toevoegen. Deze zogenaamde "private tags" worden beter vermeden, want ze kunnen immers leesbaarheids- en uitwisselingsproblemen opleveren en vereisen specifieke software. De "private tags" en hun inhoud kunnen verloren gaan, wanneer de afbeelding achteraf wordt bewerkt en opgeslagen met een applicatie die deze tags niet ondersteunt. Bepaalde applicaties voegen echter altijd één of meerdere "private tags" aan TIFF-documenten toe (bijv. PhotoShop, PageMaker, enz.).
- Digitaliseringsparameters en veel gebruikte instellingen:
 - resolutie: 3000 PPI, 600 PPI, 400 PPI, 300 PPI
 - bitdiepte:
 - kleur afbeeldingen: 24 of 48 bits/pixel
 - grijswaarden afbeelding: 8 of 16 bits /pixel
 - zwart-wit afbeelding: 1 bit/pixel
 - modus kleur afbeeldingen:
 - RGB (rood-groen-blauw): sRGB (IEC 61966-2-1)

5. Digitaliseren van audio

Kwaliteitsvereisten moederkopie:

- Bestandsformaat:
 - Windows/Intel: WAVE,(B)WAV met RIFF-bestandsstructuur (*Resource Interchange File Format*)
 - Mac: AIFF
- Codec: PCM (zowel voor (B)WAV(E) als AIFF)
- Compressie: geen
- Voorziene metadatavelden in de RIFF-WAVE-fileheader:
 - List chunk - "info": datatype is "ASCIIZ" (NULL terminated string)
 - IARL - Archival Location: bijv. bestandsnaam of archiefnummer
 - IART - Artist: bijv. naam uitvoerder(s)
 - ICMT - Comments: commentaar, opmerking
 - ICOP - Copyright: copyright informatie
 - ICRD - Creation date: datum opname of digitalisering
 - IENG - Engineer: naam ingenieur
 - IGNR - Genre: genre

- IKEY - Keywords: trefwoorden (gescheiden door “;”)
- IMED - Medium: medium
- INAM - Name/Title: beschrijving/titel archiefdocument
- ISFT - Software: naam digitaliseringssoftware
- ISRC - Source: bron
- ISRF - Source Form: vorm/type origineel document
- ITCH - Technician: naam technicus
- Format chunk - “fmt”:
- Compression code: 1 (uncompressed)
- Number of channels: 1(mono), 2(stereo)
- Sample rate
- Sampler chunk - “smp1”:
- Manufacturer
- Product
- Sample period
- Voorziene metadatavelden in de AIFF-fileheader:
- Common chunk - “comm”:
- numChannels
- numSampleFrames
- sampleSize
- sampleRate
- Digitaliseringsparameters en veel gebruikte instellingen:
- sample-frequentie: 192 KHz, 96 KHz, 48 KHz
- sample-resolutie: 24 bits, 16 bits
- aantal kanalen: 2 (stereo)

6. Digitaliseren van video

Kwaliteitsvereisten moederkopie:

- Bestandsformaat: AAF of MXF
 - Aandachtspunten bij AAF:
 - sla alle ‘essence’ en metadata in het AAF-bestand op: vermijd pointers naar externe informatie
 - sla geen producent- of softwaregebonden ‘essence’ of metadata (bijv. edit lists) in het AAF op, maar gebruik standaarden in de plaats
- Codec: open, gestandaardiseerde codec (bijv. Motion-JPEG2000)
- Compressie: geen (CCIR 601-ITU-R 601 standaard) of lossless compressie (bijv. Motion-JPEG2000)
- Metadatavelden: voor de registratie van metadata wordt XML gebruikt. De gebruiker kan zelf zijn eigen metadatamodel uitwerken - AAF en MXF zijn immers uitbreidbaar - of een bestaand XML metadataschema aanpassen.
- Digitaliseringsparameters en veel gebruikte instellingen:
 - sample-rate:

- 4:4:4: Y:13,5 MHz (NTSC: 858 samples/lijn; PAL: 864 samples/lijn)
R-Y: 13,5 MHz
B-Y: 13,5 MHz
- 4:2:2: Y:13,5 MHz
R-Y: 6,75 MHz
B-Y: 6,75 MHz
- videoformaat: component (beeld bestaat uit meerdere signalen)
- frame-rate: 30 frames/seconde, 25 frames/seconde
- aantal lijnen: 720 lijnen
- bitdiepte: 10 bits
- aantal audio kanalen: 4

7. Meer informatie

- Digitaliseren van beeldmateriaal
 - NARA, *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files - Raster Images*, juni 2004.
- Auteursrecht:
 - H. DEKEYSER, *Digitale archivering: een juridische stand van zaken vanuit Belgisch perspectief. Deel 2: Auteursrecht, technische beschermingsmaatregelen en wettelijk depot*, Antwerpen-Leuven, 2004.
- Archiveringsformaten:
 - F. BOUDREZ, *Videoarchivering: bruggen bouwen op technologisch drijfzand*, Antwerpen, 2004.
 - F. BOUDREZ, *Standaarden voor digitale archiefdocumenten*, Antwerpen, april 2004, p. 32-34.
- Digitale bewaarstrategie:
 - F. BOUDREZ, *Bewaarstrategieën*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2004.
- Compressie:
 - F. BOUDREZ, *Inleiding*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2004.
 - F. BOUDREZ, *Archiveringsstandaarden. Bestandsformaten*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2004.
- Duurzame digitale dragers:
 - algemeen:
 - F. BOUDREZ, *Archiveringsstandaarden. Dragere*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2004.
 - magnetische dragers:
 - Digitaal ArchiVeren. richtlijn & advies, nr 6: Duurzame magnetische dragers.
 - F. BOUDREZ, *Magnetische dragers voor het archief*, Antwerpen, 2002.
 - optische dragers:
 - Digitaal ArchiVeren. richtlijn & advies, nr. 2: Duurzame CD's.
 - F. Boudrez, *CD's voor het archief*, Antwerpen, 2001.