



DIGITALISEREN VAN ANALOGE ARCHIEFDOCUMENTEN

1. Woord vooraf

Analoge opslagmedia kunnen archiefdocumenten bevatten 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 onbepert 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. Samen met een operationele beheersomgeving zijn 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 (bijv. vervanging, preservering, raadpleging, enz.) 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 documenten
 - restaureren van de analoge en de digitale documenten
 - verwerken en controleren digitalisering
 - aanmaken en invoeren metadata van de gedigitaliseerde documenten.
 - digitalisering vergt veel technische middelen
- ❑ stel duidelijke criteria op en leg indien nodig prioriteiten en een fasering vast.
- ❑ bepaal op voorhand welke types documenten of materialen al dan niet in huis worden gedigitaliseerd.
- ❑ leg voor elk type document vast welke technische en beschrijvende metadata bij digitalisering worden geregistreerd.
- ❑ bepaal op voorhand of lege pagina's (tekst), pauzes (geluid), enz. al dan niet worden mee gedigitaliseerd, dan wel worden gedocumenteerd door middel van metadata
- ❑ besteed de nodige aandacht aan het beschrijven van de gedigitaliseerde documenten en aan de duurzame koppeling van de gedigitaliseerde archiefdocumenten met hun beschrijving.
- ❑ bepaal de bestemming van de analoge archiefdocumenten na digitalisering. Ga op voorhand na of de gedigitaliseerde documenten de oorspronkelijke documenten kunnen vervangen (substitutie) en of hiervoor een machtiging is vereist.
- ❑ 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, tenzij de archiefbeherende instelling onder een uitzondering van het auteursrecht valt.
- ❑ 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:
 - in staat om de essentiële eigenschappen van de archiefdocumenten in tijd over te brengen
 - gestandaardiseerd: open, gedocumenteerd, stabiel
 - wijdiverspreid en voldoende marktpenetratie
 - uitwisselbaar: onafhankelijk van bepaalde besturingssystemen, netwerkprotocollen en applicaties
 - voorziet een robuust foutopsporings- en verbeteringsmechanisme
 - goed gestructureerde opslag van informatie
 - mogelijkheid tot insluiten van (zelfgedefinieerde) metadatavelden
 - 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 *lossless* (gedecomprimeerde bitstream = originele bitstream) of een *visual lossless* compressiemethode toe: vermijd dat er ten gevolge van compressie (betekenisvolle) informatie en/of kwaliteit verloren gaat
 - pas enkel *lossy* compressie (originele bitstream ≠ 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 wijze waarop de samenhang tussen de digitale objecten en de documenten wordt aangeduid
 - 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 online 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
- ❑ besteed aandacht aan het kleurenmanagement: hanteer bij voorkeur een gestandaardiseerd kleurprofiel. Maak ook elke scandag referentiescans en archiveer deze bij de gedigitaliseerde documenten.
- ❑ 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 op:
 - correcte toepassing van de voorgeschreven bestandsformaten, codecs, kleurprofiel 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:

- ❑ voor meerbladige documenten of documenten waarvan ook de achterkant is beschreven: leg vast hoe op welke wijze meerbladige documenten en de onderlinge volgorde worden opgeslagen, bijv:
 - bundeling van alle bladen in 1 bestand (bijv. multipage TIFF)
 - groeperen van alle bladen in 1 map
 - documenteren door middel van een wrapper of een metadatabestand
 - toepassen van een gemeenschappelijke benaming in de bestandsnaam
- ❑ digitaliseringsparameters en veel gebruikte instellingen zijn:
 - resolutie: 150 PPI, 300 PPI, 400 PPI
 - bitdiepte:
 - kleur afbeeldingen: 24 of 48 bits/pixel
 - grijswaarden afbeelding: 8 of 16 bits /pixel
 - zwart-wit afbeelding: 1 bit/pixel
- ❑ opslagformaat masters: TIFF, JPEG2000
 - TIFF
 - technisch profiel: baseline TIFF (versie 6.0)
 - compressie: geen (uncompressed)
 - voeg metadatavelden in de fileheader in. De baseline TIFF-fileheader voorziet de volgende velden:
 - 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 'NULL' 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.).

- JPEG2000:
 - technisch profiel: bewaar de JPEG2000 codestream in het JPEG2000 (.jp2) formaat zoals gedefinieerd in part 1 van de JPEG2000-norm. Gebruik de uitbreiding voorzien in deel 2 van de JPEG2000 standaard (.jpx formaat) alleen wanneer een kleurprofiel noodzakelijk is dat niet wordt ondersteund door .jp2. Let wel op: het .jpx-formaat kent een mindere ondersteuning dan het .jp2-formaat. Bewaar de afbeeldingen niet als codestream (.jpc).
 - compressie: geen (*uncompressed*) of *visual lossless*. Mathematische *lossless* JPEG2000-bestanden zijn niet altijd even gemakkelijk te creëren. Vanwege deze reden en om een hogere compressiefactor (kleinere bestandsomvang) te behalen, wordt soms voor *visual lossless* gekozen. Deze laatste optie is eigenlijk een *lossy* compressiemethode, maar houdt geen visueel waarneembaar kwaliteits- of informatieverlies in.
 - gebruik het sRGB, grijswaarden of een beperkt ICC kleurenprofiel (i.c. gamma matric style ICC). Deze kleurprofielen worden rechtstreeks ondersteund door deel 1 van de JPEG2000 standaard.
 - pas tiling, verschillende lagen, levels en resoluties toe zodat de scans op een efficiënte wijze kunnen worden geraadpleegd. Bepaal de tiling, de lagen, de levels en de resoluties in functie van de kenmerken van het analoge document
 - metadata:
 - documenteer expliciet in welke resolutie (x pixels/meter) de digitale opname werd gemaakt. Registreer deze metadata in de JP2 header box
 - bed de metadata in XML-vorm in het JPEG2000 bestand in. Voeg hiervoor een XML-box aan het JPEG2000 bestand toe.
 - **Opmerking:** JPEG2000 kent nog steeds geen brede ondersteuning. Software voor het samenstellen en het bekijken van JPEG2000-bestanden is (nog) niet weidverspreid.

4. Digitaliseren van afbeeldingen

Kwaliteitsvereisten moederkopie:

- ❑ digitaliseringsparameters en veel gebruikte instellingen:
 - resolutie: 300 PPI, 400 PPI, 600 PPI, 3000 PPI
 - bitdiepte:
 - kleur afbeeldingen: 24 of 48 bits/pixel
 - grijswaarden afbeelding: 8 of 16 bits /pixel
 - zwart-wit afbeelding: 1 bit/pixel
 - kleurprofiel:
 - afbeeldingen in kleur: sRGB (IEC 61966-2-1)
 - afbeeldingen zonder kleuren: grijswaarden
- ❑ opslagformaat masters: TIFF, JPEG2000
 - TIFF: zie 3. digitaliseren van tekst
 - JPEG2000: zie 3. Digitaliseren van tekst

5. Digitaliseren van audio

Kwaliteitsvereisten moederkopie:

- ❑ digitaliseringsparameters en veel gebruikte instellingen:
 - sample-frequentie: 44.1 KHZ, 48 KHZ, 96 KHZ, 192 KHZ
 - sample-resolutie: 16 bits, 24 bits
 - aantal kanalen: 2 (stereo)
- ❑ opslagformaat masters: WAVE, AIFF
 - WAVE
 - technisch profiel: pas de RIFF-bestandsstructuur toe (*Resource Interchange File Format*)
 - codec: lineaire PCM (*pulse code modulation*)
 - compressie: geen
 - pas 'WAVE_FORMAT_EXTENSIBLE' toe wanneer:
 - meer dan 2 kanalen nodig zijn
 - een expliciete mapping van de kanalen met speakers is vereist
 - een hoge audioresolutie (> 16 bits/sample) wordt gebruikt
 - pas BWF/RF64 toe wanneer:
 - meer dan 2 kanalen nodig zijn
 - de omvang groter is dan 4 GB.
 - metadata: een RIFF-header voorziet de volgende metadatavelden
 - 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
- AIFF:
 - codec: lineaire PCM (pulse code modulation)
 - compressie: geen
 - metadata: de AIFF-fileheader voorziet de volgende metadatavelden
 - Common chunk - "comm":
 - numChannels
 - numSampleFrames
 - sampleSize
 - sampleRate

6. Digitaliseren van video

Kwaliteitsvereisten moederkopie:

- 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 of 2
- opslagformaat masters: MXF, MJPEG2000
 - MXF:
 - technisch profiel: MXF OP1A (SMPTE 378M-2004: Operational Pattern 1a (Single Item, Single Package)).
 - compressie: geen (CCIR 601-ITU-R 601 standaard) of *lossless* compressie (bijv. Motion-JPEG2000)
 - codec:
 - video: *lossless* of *visual lossless* Motion-JPEG2000
 - analoge masters:
 - beeld: uncompressed YUV 4:2:2 met hoge bitdiepte (8 of 10)
 - geluid: uncompressed PCM (48 KHz en 24 bits bitdiepte)
 - digitale masters:
 - video: uncompressed YCbCr, 4:2:2 of 4:4:4 met hoge bitdiepte (8 of 10)
 - geluid: uncompressed PCM (48 KHz en 24 bits bitdiepte)

- metadata: 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.
- MJPEG2000:
 - technisch profiel: MJPEG2000 Simple Profile:
 - 1 videotrack: max. 30 frames/seconde
 - max 1 geluidtrack: max. 48Khz, 8 of 16 bits, 1 of 2 kanalen, PCM-codec
 - geen externe bronnen of referenties: alle gegevens zijn in het bestand opgeslagen
 - compressie: *lossless* of *visual lossless*
 - opslag: in een MXF-wrapper, of rechtstreeks in MJPEG2000:
 - als 1 digitaal object die alle metadata, video- en audio informatie bevat
 - als meerdere digitale objecten: het MJPEG2000 metadata-bestand verwijst naar de afbeeldingen (bijv. JPEG2000) en het audiobestand (bijv. WAV).