

MIGRATIE NAAR ARCHIVERINGSFORMATEN

1. Woord vooraf

Migratie is tot op heden de meest toegepaste digitale bewaarstrategie¹ voor digitale archiefdocumenten en is een belangrijke stap in de digitale bewaarstrategie die eDAVID voorstelt². Deze bewaarstrategie gaat uit van de archivering van digitale archiefdocumenten in hun oorspronkelijk én hun geschikt archiveringsformaat. Het geschikt archiveringsformaat is doorgaans het resultaat van een migratie operatie en biedt meer leesbaarheids garanties. Door ook het oorspronkelijk applicatieformaat bij te houden, blijven in de toekomst zoveel mogelijk opties open.

Migratie is momenteel de meest praktische oplossing voor de technologische veroudering van hard- en software en het daaruit voortvloeiende leesbaarheidsprobleem van digitale documenten. Door digitale archiefdocumenten te migreren naar een ander bestandsformaat zorgt men ervoor dat deze documenten op scherm presenteerbaar blijven wanneer de software waarmee ze werden gecreëerd, verouderd of in onbruik raakt.

In de praktijk wordt migratie als bewaarstrategie voor digitale archiefdocumenten samen toegepast met:

- ❑ het gebruik van open en gedocumenteerde standaarden als geschikte archiveringsformaten, wat o.a. volgende voordelen oplevert:
 - de mogelijkheid tot ontwikkelen van nieuwe viewers of applicaties op basis van de specificatie en documentatie
 - een migratiecyclus die minder afhankelijk is van IT-producenten en van de achterwaartse compatibiliteit van applicaties
- ❑ het beperken van de archiveringsformaten die door de archiefdienst worden ondersteund, met als voordelen:
 - in de toekomst zijn minder viewers of applicaties voor raadpleging nodig
 - limiteren van de kennis en expertise over geschikte archiveringsformaten
 - in de hand houden van het aantal bestandsformaten die het digitaal archiveringssysteem moet ondersteunen

Migratie als digitale bewaarstrategie wordt echter algemeen als een risicovolle onderneming met kans op informatieverlies beschouwd³. Deze risico's zijn op zich niet groter dan bij andere bewaarstrategieën, maar moeten op voorhand goed worden ingeschat zodat die kunnen worden uitgesloten. Uitvoerige risico-analyses en tests moeten vermijden dat tijdens de migratie ongewenste verliezen, onverwachte neveneffecten of fouten optreden. Kwaliteitscontroles tijdens het migratieproces moeten eventuele anomalieën of problemen melden.

Onderstaand overzicht bevat richtlijnen en aanbevelingen zodat je vermijdt dat migratie een onzeker proces met risico's is. Je doet dit door een migratiepad te definiëren. Zo zorg je er ook

voor dat het hele omzettingsproces op een gecontroleerde, betrouwbare en gedocumenteerde wijze verloopt.

2. Wanneer migreren?

Digitale archiefdocumenten worden gemigreerd wanneer:

- ❑ het applicatieformaat geen ondersteuning meer krijgt of de vereiste software niet meer operationeel is
- ❑ ze niet in een geschikt archiveringsformaat zijn opgeslagen en op (middel)lange termijn leesbaar moeten blijven
- ❑ een geschikt archiveringsformaat verouderd is of geen brede ondersteuning meer geniet
- ❑ een meer geschikt archiveringsformaat beschikbaar wordt.

3. Wie migreert?

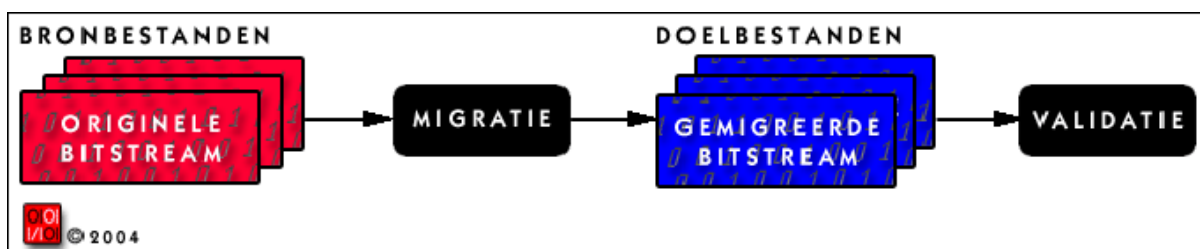
Archiefvormers kunnen beslissen hun digitale archiefdocumenten zelf te migreren, dit uit te besteden of over te laten aan de archiefdienst. Vanwege de noodzaak om de betrouwbaarheid van migraties aan te tonen is het van belang dat:

- ❑ bij migraties door de archiefvormer zelf:
 - de archiefdienst toeziet op het migratiepad bij de archiefvormer
 - het migratiepad een interne/externe audit kan doorstaan.
- ❑ bij migraties door externen:
 - een formele overeenkomst wordt opgesteld waarin duidelijk het migratiepad wordt vastgelegd
 - gekozen wordt voor een betrouwbare externe partner waarvan expertise, ervaring, naleving van normen kunnen worden gecertificeerd.

4. Definiëren van een migratiepad

Een migratiepad bestaat uit de volgende stappen:

1. vaststellen essentiële eigenschappen
2. keuze van het doelformaat
3. keuze van een migratietool
4. tests
5. migratie van de archiefdocumenten
6. validatie van de gemigreerde archiefdocumenten
7. documenteren.



Dit proces kan in ruime mate worden ondersteund door de PLATO-tool te gebruiken. PLATO is een web-based applicatie waarmee een conserveringsplan zoals een migratiepad kan worden vastgesteld en gedocumenteerd. PLATO is beschikbaar op: <http://www.ifs.tuwien.ac.at/dp/plato>

4.1 Vaststellen essentiële eigenschappen

Een migratie van archiefdocumenten naar hun geschikt archiveringsformaat vertrekt vanuit de identificatie van de essentiële eigenschappen van de documenten die behouden dienen te blijven. De basis hiervoor is een identificatie van het archiefstuk en een functionele waardering van zijn componenten. Niet alle componenten van een digitaal document zijn immers essentieel voor de archiefstatus van het document (zie F. BOUDREZ, *Inleiding. Het digitaal archiefdocument*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2003).

- ❑ leg op voorhand vast:
 - welke componenten van de archiefdocumenten essentieel zijn en niet gewijzigd mogen worden
 - welke componenten van de archiefdocumenten incidenteel zijn en bijgevolg gewijzigd mogen worden of niet overgezet hoeven te worden.
- ❑ ken een gewicht toe aan de essentiële componenten zodat hier een hiërarchie of verdeelsleutel in is
- ❑ bepaal in overleg met de archiefvormer de waarde op lange termijn van de componenten van een digitaal archiefdocument:

COMPONENTEN

- de context:
- de inhoud:
- de structuur:
- de 'look and feel':
- het gedrag / de functionaliteiten

AANDACHTSPUNTEN

- bevatten de bronbestanden ingekapselde metadata?
- verschaft de volledige pathaanduiding contextuele informatie?
- hoe wordt gelinkt naar de beschrijving van de archiefdocumenten?
- wordt de inhoud integraal en ongewijzigd overgenomen?
- worden digitale handtekeningen mee gearhiveerd?
- wat is de granulariteit van de structuur in het doelbestand?
- worden structurelementen samengevoegd of opgesplitst?
- hoe bepaal je de originele 'look and feel'?
- is elk gedrag of elke functionaliteit essentieel?
- welke externe bronnen of software is nodig voor het uitvoeren van het geselecteerde gedrag of de vereiste functionaliteiten?

4.2 Keuze van het doelformaat

- ❑ kies een doelformaat dat:
 - aan zoveel mogelijk criteria van een geschikt archiveringsformaat beantwoordt. DAVID formuleerde criteria voor een geschikt archiveringsformaat (zie F. BOUDREZ,

Archiveringsstandaarden. Bestandsformaten, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2003)

- het minste risico's inhoudt:
 - eenvoudige specificatie: gemakkelijker om een viewer te creëren
 - gemakkelijk uit te voeren omzettingsproces: minder kans op fouten en problemen
 - alle essentiële eigenschappen en zoveel mogelijk incidentele eigenschappen van het digitaal archiefdocument overneemt: vergeet hierbij de ingekapselde metadata niet. Houd rekening met de complexiteit van migraties: hoe meer componenten worden overgezet, hoe complexer de omzetting. Vermijd indien nodig de omzetting van overbodige componenten
 - gemakkelijk bruikbaar is of op basis waarvan gemakkelijk een raadplegingskopie wordt samengesteld.
- leg het profiel van het archiveringsformaat vast:
- welke formaatversie wordt toegepast?
 - welke codecs worden gebruikt?
 - welke metadata worden in welke headervelden genoteerd?

4.3 Keuze van een migratietool

- verzamel documentatie en informatie over het bronformaat. De specificaties van open formaten worden doorgaans op het web gepubliceerd of zijn tegen betaling beschikbaar bij de standaardiseringsorganisatie. Producenten van gesloten applicatieformaten verspreiden zelf geen informatie over hun formaat. Op het web vind je over veel gesloten applicatieformaten wel documentatie terug, maar deze is niet altijd even accuraat of up-to-date
- voor de migratie kan gebruik worden gemaakt van web-based of stand alone migratietools:
- commerciële tools
 - open source tools
 - zelf gecreëerde tools
- kies geen migratietool waarvan je niet weet wat er zich binnenin afspeelt (zgn 'black box' migratietool):
- geef de voorkeur aan een open source of een zelf gecreëerde migratietool
 - in geval van de keuze voor een commerciële tool: kies een tool
 - waarvan de bekende problemen en gebreken zijn gedocumenteerd (bijv. meldingen op de website van de producent) en waarvoor eventuele 'workarounds' beschikbaar zijn
 - die heel uitgebreid werd getest
 - ga na of de migratietool de bronbestanden zonder fouten inleest. In geval van gesloten applicatieformaten maak je hiervoor wellicht best gebruik van de producentgebonden applicatie zelf.
- ga na welke componenten van het bronbestand ongewijzigd worden overgenomen en wat de verschillen zijn tussen bron- en doelbestand
- zorg ervoor dat de migratietool aan volgende functionele vereisten voldoet:
- ongewijzigd laten van het bronbestand (enkel openen en inlezen)
 - foutloze omzetting van de essentiële componenten van het digitaal archiefdocument
 - bulkomzettingen

- uitgebreide error-handling: registratie en melding van de fouten/problemen bij het omzetten
- produceren van een doelbestand conform de bestandsformaatpecificatie
- (correct) toepassen van de aangewezen instellingen en parameters voor de langetermijnarchivering van digitale informatie.

4.4 Tests

Alvorens de migratie in de praktijk wordt geïmplementeerd, is het belangrijk om de migratie uitvoerig te testen zodat de archivaris een goed zicht heeft op het eindresultaat en de accuraatheid van het migratieproces. Bij deze tests worden bij voorkeur ook de risico's onderzocht. De tests kunnen op diverse wijzen worden uitgevoerd:

- ❑ testen op een representatief staal bronbestanden uit de praktijk
- ❑ testen op speciaal samengestelde bestanden zodat de beperkingen van bron- en doelformaat en van de migratietool worden onderzocht
- ❑ validatie van de doelbestanden met een validatietool
- ❑ openen en bekijken van de doelbestanden met geselecteerde applicaties.

4.5 Migratie van de archiefdocumenten

- ❑ waak erover dat de digitale archiefdocumenten tijdens het migratieproces hun integriteit behouden:
 - overname van alle essentiële componenten
 - identieke inhoud en identificerende informatie
 - intacte band tussen document en identificerende informatie
- ❑ zorg voor een fijnmazige foutopsporing- en afhandeling: onverwachte of nieuwe fouten dienen systematisch en gedetailleerd te worden gedocumenteerd in een migratielog
- ❑ besteed aandacht aan de metadata over het digitaal archiefdocument:
 - behoud de contextuele informatie die de volledige pathaanduiding van een digitaal archiefdocument eventueel verschaft:
 - behoud de bestandsnaam van de bronbestanden, wijzig enkel de extensie of voeg de nieuwe extensie gewoon achteraan toe (bijv. document.doc.odf)
 - plaats de gemigreerde bestanden in dezelfde map als de bronbestanden
 - zorg ervoor dat de metadata ingekapseld in de bronbestanden niet verloren gaan:
 - blijven die ingebed in de doelbestanden?
 - worden die geëxtraheerd uit de bronbestanden en centraal opgeslagen?
 - worden er metadata aan de gemigreerde bestanden toegevoegd?
 - denk aan de koppeling tussen centraal opgeslagen metadata en de gemigreerde archiefdocumenten

OPMERKING: Migratie en/of inkapseling van metadata leidt onvermijdelijk tot het wijzigen van de bitstream van de digitale documenten. Betrouwbaarheidswaarborgen die gebaseerd zijn op de bitstream functioneren niet meer op de gemigreerde documenten (bijv. validatiefunctie digitale handtekening). De digitale handtekening kan wel worden gearhiveerd, maar valideert de inhoud van het gemigreerde archiefdocument niet langer na omzetting.

4.6 Validatie van de gemigreerde archiefdocumenten

Voer een kwaliteitscontrole uit op de gemigreerde documenten:

- controleer de gemigreerde documenten op hun conformiteit met de formaatspecificaties
- voer deze controle uit met een andere tool/bibliotheek dan diegene die voor de migratie werd gebruikt
- ga op voorhand na welke afwijkingen tav de formele bestandsspecificatie de validatietool al dan niet toelaat
- bewaar het logbestand met de validatieresultaten bij de gemigreerde documenten.

4.7 Documenteren

- identificeer de gebruikte bestandsformaten door volgende gegevens vast te leggen:
 - naam bestandsformaat
 - versienummer van het bestandsformaat
 - gebruikte codecs
 - ingekapselde metadata.
Let op: MIME-typing is te weinig gedetailleerd (bijv. geen onderscheid tussen de verschillende versies of identificatie van de gebruikte compressies) en laat een overlap tussen de verschillende categorieën toe.
- documenteer het migratiepad en het migratieproces:
 - bronformaat
 - doelformaat
 - gebruikte migratie- en validatietool
 - uitgevoerde kwaliteitscontrole
 - datum en tijdstip van de migratie
 - opsomming van de componenten die:
 - werden overgenomen
 - werden gewijzigd
 - verloren gingen
 - uitvoerder van de migratie
- houd de geannoteerde broncode van open source of zelf gecreëerde migratietools bij.

5. Beheer van de digitale archiefdocumenten

Om digitale documenten correct te archiveren en migraties goed te laten verlopen, zijn volgende beheershandelingen essentieel:

- houd documentatie bij over de geschikte archiveringsformaten die door de archiefdienst worden ondersteund
- zorg ervoor dat:
 - de set geselecteerde archiveringsformaten voldoende alle types digitale documenten met archiefwaarde van de organisatie dekt
 - de archiveringsformaten verschillende types digitale documenten als inhoud kunnen hebben: beperken van de archiveringsformaten die op lange tijd worden ondersteund
 - eenvoudige archiveringsformaten de voorkeur krijgen op complexe formaten
 - digitale archiefdocumenten raadpleegbaar zijn zonder dat hiervoor (individuele) paswoorden zijn vereist.
- volg de technologische evolutie op zodat je tijdig weet wanneer:

- bepaalde formaten in onbruik raken
- meer geschikte archiveringsformaten beschikbaar worden
- koppel deze digitale bewaarstrategie aan procedures voor bitpreservatie: kwaliteitscontroles van de media, verversen van de media, berekenen checksums, enz.

6. Meer informatie

- F. BOUDREZ, *Standaarden door digitale archiefdocumenten*, Antwerpen, 2004.
- G.W. LAWRENCE e.a., *Risk Management of Digital Information: A File Format Investigation*, Washington, 2000.
- TESTBED DIGITALE BEWARING, *Migratie: context en huidige stand van zaken*, Den Haag, 2001.
- C. BECKER, H. KULOVITS, M. GUTTENBRUNNER, S. STRODL, A. RAUBER en H. HOFMAN, *Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans*, in: *International Journal on Digital Libraries*, december 2009.

7. ? Vragen ? Suggesties ?



Met al uw vragen en suggesties kan u terecht bij
eDAVID: <mailto:info@edavid.be>



EINDNOTEN:

- ¹ M. HEDSTROM en S. MONTGOMERY, *Digital Preservation Needs and Requirements in RLG Member Institutions*, 1998, p. 15; M.V. CLOONAN en S. SANETT, *Preservation Strategies for Electronic Records: Where We Are Now—Obliquity and Squint?*, in: *American Archivist*, 2002, nr. 65, 70-106; D. BEARMAN, Moments of Risk: Identifying Threats to Electronic Records, *Archivaria* 62.
- ² F. BOUDREZ, *Bewaarstrategieën*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2003.
- ³ Voor een opsomming van alle voor- en nadelen van migratie en andere bewaarstrategieën: F. BOUDREZ, *Bewaarstrategieën*, in: F. BOUDREZ en H. DEKEYSER, *Digitaal archiveren in de praktijk. Handboek*, Antwerpen-Leuven, 2003.