

# STANDAARDEN VOOR BESTANDSFORMATEN



## 1. Woord vooraf

### Belang

Om de lange termijn leesbaarheid van digitale documenten te verzekeren is het aangewezen om zoveel mogelijk standaard bestandsformaten te gebruiken. Standaarden spelen een belangrijke rol in de digitale bewaarstrategieën voor digitale archiefdocumenten (emulatie en migratie). Standaard bestandsformaten hebben hun status te danken aan (officiële) standaardiseringsinitiatieven of aan de grootschalige verspreiding.

### Voordelen

Standaard bestandsformaten hebben het voordeel dat ze compatibel zijn met meerdere softwaretoepassingen. Wanneer één applicatie niet meer beschikbaar is, moeten de computerbestanden dus niet onmiddellijk omgezet worden. Een aantal conversiestappen worden bijgevolg vermeden. De interne bestandsstructuur van veel standaard bestandsformaten is publiekelijk verspreid zodat het in principe mogelijk is om ten allen tijde nieuwe viewers te maken. Het bewaren van informatie in bestandsformaten met de status van standaard dringt zich ook op om uitwisseling mogelijk te maken. Officiële standaarden of specificaties zijn niet afhankelijk van één producent, maar worden beheerd door standaardiseringsinitiatieven. Hierdoor wordt niet alleen een stukje stabiliteit gewaarborgd, de standaarden kunnen pas na het doorlopen van een hele procedure worden gewijzigd.

### Parameters en opties

Bij de meeste bestandsformaten kunnen verschillende parameters worden ingesteld. Vanuit archiveringsperspectief zijn twee parameters belangrijk: de *ASCII-opslagmethode* en *compressieloze bewaring*.

ASCII-opslag: bestanden waarvan alle systeemdata dmv ASCII-karakters zijn opgeslagen, zijn gemakkelijker te interpreteren of te editeren door andere applicaties.

Geen compressie: compressie stemt niet overeen met de doelstelling om zo platformonafhankelijk mogelijk te archiveren. Decompressie is een essentiële schakel in het reconstructieproces. Voor het openen van het bestand heeft men immers de overeenstemmende decompressiesleutel nodig. Foutjes in gecomprimeerde bestanden resulteren bijna altijd in verlies van het digitaal archiefdocument.

### Nadelen

Het gevolg van de keuze voor ASCII-opslagmethode en compressieloze bewaring, is een grotere bestandsomvang. Omzetten naar een standaard bestandsformaat gaat soms ook met “bevrozing” of functionaliteitsverlies samen. Hierdoor kunnen de bestanden niet meer gebruikt of bewerkt worden binnen hun oorspronkelijke applicatie.

## Wanneer omzetten?

De digitale documenten kunnen op verschillende tijdstippen naar een standaard bestandsformaat worden omgezet.

1. Bewaar digitale documenten zoveel mogelijk van bij de creatie in het passende standaard bestandsformaat. Dit is vooral aangewezen bij audio-visuele documenten die je nog vele jaren nodig hebt. Bij tekstuele documenten is dit echter niet altijd mogelijk. Bijv. foto's → TIFF; audio → WAV
2. Zet de bestanden om wanneer hun onmiddellijk nut voorbij is. Bijv. WORD → PDF, XML of TIFF
3. Zet de bestanden om terwijl de oorspronkelijke software nog operationeel is. Zorg eerst voor de omzetting, alvorens de software te vervangen. Bijv. AutoCAD → DXF
4. De omzetting dient ten laatste voor overdracht naar de archiefdienst te gebeuren.

## 2. Tekstdocumenten

FORMAAT	VERSIE	PARAMETERS en OPTIES	GEBRUIKEN VOOR
XML <i>eXtensible Markup Language</i>	1.0/1.1	<ul style="list-style-type: none"><li>– bij archiefdocumenten waarvan de lay-out belangrijk is: XSL-stylesheet opstellen</li><li>– bij archiefdocumenten waarvan de structuur belangrijk is: XML Schema (of DTD) opstellen</li><li>– baseer de granulariteit van het XML-document op de interne structuur van het archiefdocument</li><li>– kies semantische XML-tags</li><li>– zorg ervoor dat de XML-documenten leesbaar zijn voor meerdere applicaties</li><li>– kies voor een XML-formaat dat vrij is van patentrechten die implementatie door derden in de weg staan of bemoeilijken</li></ul>	<ul style="list-style-type: none"><li>– e-mails</li><li>– tekstdocumenten</li><li>– inhoud websites</li><li>– databanken</li><li>– metadata</li><li>– spreadsheets</li></ul>
OD* <i>Open-Document</i>	1.0	<ul style="list-style-type: none"><li>– indien gewenst, vul metadata in de voorziene velden in of maak eigen metadatavelden aan. De metadata worden opgeslagen in het bestand ('meta.xml')</li></ul>	<ul style="list-style-type: none"><li>– tekstdocumenten met grafieken en afbeeldingen, spreadsheets, presentaties</li></ul>
Unicode	4.1.0	<ul style="list-style-type: none"><li>– duidelijke veldscheidingstekens gebruiken</li><li>– documenteer de velden in het Unicodebestand</li></ul>	<ul style="list-style-type: none"><li>– databanken</li><li>– logbestanden</li></ul>
PDF(/A) <i>Portable Document Format</i>	1.4/1.5/ 1.6	<ul style="list-style-type: none"><li>– maak PDF-documenten met Adobe PDF, Acrobat Distiller of PDFMakerA (Office VBA); gebruik niet de PDF-writer</li><li>– ASCII-opslagmethode toepassen</li><li>– pas geen compressie toe: schakel de compressiemogelijkheid uit (monochrome, grijschaal en volle kleurafbeeldingen)</li><li>– pas zoveel mogelijk instellingen van PDF/A toe:<ul style="list-style-type: none"><li>– stel PDF 1.4 als basis in</li><li>– tags gebruiken: optie "Embed tags in PDF" of "Enable accessibility and reflow"</li></ul></li></ul>	<ul style="list-style-type: none"><li>– tekstdocumenten met grafieken en afbeeldingen</li><li>– powerpointpresentaties</li><li>– drukwerk</li></ul>

- with Tagged PDF" aanvinken
- sluit alle gebruikte fonts in: gebruik geen fonts waarvoor licenties vereist zijn
- sluit de color space in
- sluit geen audio, video, javascript, Xobjects of multimedia-objecten in.
- stel geen enkele beveiliging (bijv. paswoorden, digitale handtekening, encryptie, enz.) of restricties in (bijv. openen, afdrukken, enz.)
- voor afbeeldingen: sRGB kleurenmodus

TIFF <i>Tagged Image File Format</i>	6.0	<ul style="list-style-type: none"> <li>– scans: controleer of alles goed leesbaar is veel gebruikte resolutie: 300 dpi</li> <li>– bitdiepte: <ul style="list-style-type: none"> <li>▪ zwart-wit: 1 bit</li> <li>▪ grijswaarden: 8 of 16 bits</li> <li>▪ kleur: 24 of 48 bits</li> </ul> </li> <li>– compressieloos bewaren</li> <li>– enkel baseline en vastgelegde TIFF-uitbreidingen toepassen: pas het TIFF-formaat niet intern aan</li> <li>– RGB: afbeelding op scherm (sRGB: IEC 61966-2-1)/ CYMK: drukwerk</li> <li>– byte volgorde (IBM of MacIntosh): vrij te kiezen</li> <li>– koppel metadata in de TIFF-header in: <ul style="list-style-type: none"> <li>– 269: DocumentName</li> <li>– 270: ImageDescription</li> <li>– 271: Make</li> <li>– 272: Model</li> <li>– 305: Software</li> <li>– 306: DateTime</li> <li>– 315: Artist</li> <li>– 316: HostComputer</li> <li>– 33432: CopyRight</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>– tekstdocumenten met grafieken en afbeeldingen</li> <li>– powerpointpresentaties</li> <li>– ingescande documenten</li> </ul>
(X)HTML <i>(eXtensible) HyperText Markup Language</i>	HTML 4.01 (ISO-15445) of XHTML 1.0	<ul style="list-style-type: none"> <li>– (X)HTML-syntaxregels respecteren</li> <li>– geen afgekeurde of niet-gestandaardiseerde tags en attributen gebruiken</li> <li>– leg bestandsgegevens (metadata) expliciet in de XHTML-header vast <i>(zie: Digitaal Archiveren: Richtlijn 5. Websitesbeheer voor archivering.)</i></li> </ul>	– websites

## 3. Afbeeldingen

### 3.1 Rasterafbeeldingen

FORMAAT	VERSIE	PARAMETERS en OPTIES	GEBRUIKEN VOOR
TIFF <i>Tagged Image File Format</i>	6.0	– idem als bij tekstuele documenten	– full color afbeeldingen (foto's) – ingescande documenten, – moederkopieën
EPS <i>Encapsulated PostScript</i>	3.0	– ASCII-opslagmethode toepassen – duid aan of er al dan niet een thumbnail wordt opgeslagen – variant DCS-1: CYMK-kleurscheiding – variant DCS-2: steunkleuren	– drukwerk – 2 of 3 dimensionele afbeeldingen – combinatie van tekst, grafieken en afbeeldingen
JPEG <i>Joint Pictures Experts Group</i>	JPEG JPEG2000	– bepaal de parameters in functie van de bestandsgrootte en de beeldkwaliteit – <i>lossless</i> compressie voor moederkopieën toepassen – <i>lossy</i> compressie voor werkkopieën toepassen	– foto's in websites – moederkopieën – werkkopieën
PNG <i>Portable Network Graphics</i>	1.0	– keuze tussen 8-bit en 24-bit	– afbeeldingen (in websites) – afbeeldingen met paletkleuren (8 bit) of grijswaarden (24 bit) <i>opmerking:</i> PNG past altijd compressie toe
GIF <i>Graphics Interchange Format</i>	89	– beperking op aantal kleuren = informatieverlies – LET OP: Voor de creatie van GIF-bestanden dient men in principe te beschikken over een Unisyslicentie.	– cartoons (met animatie) – logo's – zwart-wit afbeeldingen – afbeeldingen met paletkleuren en grote vlakken – afbeeldingen in websites

### 3.2 Vectorafbeeldingen

FORMAAT	VERSIE	PARAMETERS en OPTIES	GEBRUIKEN VOOR
CGM <i>Computer Graphics Metafile</i>		– pas indien mogelijk een profiel toe – WebCGM: uitwisseling van dynamische CGM-afbeeldingen met hyperlinks	2 dimensionele vector afbeeldingen
EPS <i>Encapsulated PostScript</i>	3.0	– idem als bij rasterafbeeldingen	2 of 3 dimensionele afbeeldingen
DXF <i>Drawing eXchange Format</i>	16.1.01	– gebruik de ASCII-opslagmethode – in hoogst mogelijk versie bewaren	2 en 3 dimensionele CAD/CAM-tekeningen

## 4. Geluidsdocumenten

FORMAAT	VERSIE	PARAMETERS en OPTIES	GEBRUIKEN VOOR
PCM <i>Pulse Code Modulation</i>		<ul style="list-style-type: none"> <li>– sample-rate en bitdiepte kiezen in functie van de kwaliteit van de analoge bron</li> <li>– DAT-bestand bijleveren waarin de sample-rate en sample-resolutie expliciet zijn vastgelegd</li> </ul>	digitale moederkopieën
WAV		<ul style="list-style-type: none"> <li>– sample-rate en bitdiepte kiezen in functie van de kwaliteit van de analoge bron</li> <li>– moederkopie: PCM-encoding (compressieloos) toepassen</li> <li>– uitwisseling: MP3-codec (compressie) toepassen</li> </ul>	digitale moederkopieën of uitwisseling
MP3	1.0	<ul style="list-style-type: none"> <li>– bepaal de parameters frequentie, bitrate en aantal kanalen in functie van de gewenste kwaliteit en bestands grootte</li> </ul>	uitwisseling van geluidsbestanden via netwerken. Gebruik van streaming MP3 is mogelijk.

## 5. Videodocumenten

FORMAAT	VERSIE	PARAMETERS en OPTIES	GEBRUIKEN VOOR
AAF <i>Advanced Authoring Format</i>	1.1	<ul style="list-style-type: none"> <li>– compressie: geen of <i>lossless</i> compressie met een gestandaardiseerde, open en gedocumenteerde codec (bijv. Motion-JPEG2000)</li> <li>– metadata: XML</li> <li>– zorg ervoor dat het AAF volledig zelfvoorzienig is: <ul style="list-style-type: none"> <li>▪ sla alle 'essence' en metadata in het AAF-bestand op</li> <li>▪ neem geen pointers naar externe informatie op</li> </ul> </li> <li>– neem geen producent- of softwaregebonden informatie op (bijv. edit lists, metadata, enz.)</li> </ul>	hoge resolutie digitale moederkopieën (essence + metadata)
MXF <i>Material Exchange Format</i>	1.0	<ul style="list-style-type: none"> <li>– compressie: geen of <i>lossless</i> compressie met een gestandaardiseerde, open en gedocumenteerde codec (bijv. Motion-JPEG2000)</li> <li>– metadata: XML</li> </ul>	hoge resolutie digitale moederkopieën (essence + metadata)
D5		<ul style="list-style-type: none"> <li>– sampling ratio: 4:2:2</li> </ul>	hoge resolutie digitale

		<ul style="list-style-type: none"> <li>- max. data rate: 170 Mbps</li> <li>- bits per sample: 10</li> <li>- compressie: geen</li> </ul>	moederkopieën opmerking: proprietary!
Digitale Betacam		<ul style="list-style-type: none"> <li>- sampling ratio: 4:2:2</li> <li>- max. data rate: 90 Mbps</li> <li>- bits per sample: 10</li> <li>- compression: 2:1</li> </ul>	hoge resolutie digitale moederkopieën wanneer ongecomprimeerde opslag niet mogelijk is opmerking: proprietary!
MPEG (* .MPG) <i>Moving Pictures Experts Group</i>	1.0	bepaal het aantal frames (bijv. 30), de kleurdiepte (8, 16, 24) en de resolutie (origineel, 160 x 120, 320 x 240, 640 x 480) in functie van de gewenste kwaliteit en bestandsgrootte	distributie van bewegende computerbeelden met/zonder geluid
MPEG (* .m2v)	2.0	idem als bij MPEG 1.0	digitale televisie, multimedia

## 6. ? Vragen ? Suggesties ?



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



Meer info: F. Boudrez, *Standaarden voor digitale archiefdocumenten*, (<http://www.edavid.be/davidproject>).