

Leidraad Duurzame bestandsformaten

Gemeente Eindhoven
Facilitaire en Administratieve Ondersteuning
DIV – Documentaire informatievoorziening

Versie 1.3

Oktober 2022

Colofon

Uitgave

Gemeente Eindhoven

Facilitaire en Administratieve Ondersteuning

DIV – Documentaire informatievoorziening

Datum

Oktober 2022

Inhoud

1	Inleiding	4
2	Stelling	5
3	Knelpunten bij het duurzaam opslaan van digitale informatie	5
4	Oplossing voor het duurzaam opslaan van digitale informatie	6
5	Standaard bestandsformaten	7
6	Voorkeursformaten en acceptabele formaten	7
7	Toetsingscriteria voor voorkeurs- en acceptabele duurzame archiveringsformaten	8
8	Bijlage A - Overzichten van voorkeursformaten en acceptabele formaten	9
9	Bijlage B - Bestandsformaten die voorkomen in het DMS (eDocs)	13

1 Inleiding

Als overheidsorgaan is de gemeente Eindhoven conform de Archiefwet 1995 verplicht de onder hen berustende archiefbescheiden in goede, geordende en toegankelijke staat te brengen en te bewaren, alsmede zorg te dragen voor de vernietiging van de daarvoor in aanmerking komende archiefbescheiden.

Voor het beheren en toegankelijk maken van digitale documenten binnen de gemeente Eindhoven heeft de gemeente de beschikking over een Document Management Systeem (DMS) eDocs. In dit DMS kunnen verschillende soorten bestandsformaten worden opgenomen. Bestandsformaat is de aanduiding voor het soort computerbestand, vaak vastgelegd in extensies van de bestandsnaam, waarmee wordt aangegeven welk programma een computer moet hebben om dat bestand te verwerken.

Er wordt bij bestandsformaten en archiefbescheiden ook vaak gesproken over een informatieobject en dit is een op zichzelf staand geheel van gegevens met een eigen identiteit. Een informatieobject is een verzameling aan elkaar gerelateerde gegevens die als eenheid wordt behandeld. Vanuit het perspectief van DUTO betekent dit dat een informatieobject als eenheid aan de gebruikers wordt gepresenteerd. Informatieobjecten zijn de eenheden van informatie waar de DUTO-eisen betrekking op hebben. Bijvoorbeeld zoekresultaten verwijzen naar informatieobjecten, informatieobjecten kunnen opgevraagd worden, informatieobjecten hebben metagegevens, etc. Een eigen identiteit betekent dat een informatieobject een naam of identificatiekenmerk heeft waardoor het te onderscheiden is van andere informatieobjecten. Hiermee kan een gebruiker eenduidig verwijzen naar het informatieobject.

De documenten in het DMS kunnen via verschillende kanalen worden opgenomen:

- Fysieke documenten (ontvangen, uitgaande en interne) die via scannen omgezet zijn naar een digitale vorm. De gescande documenten worden in PDF formaat opgenomen.
- Digital born¹ documenten. Deze documenten worden opgenomen in het formaat van de applicatie waarmee ze zijn gecreëerd.

In de life cyclus van een digitaal documenten zijn er momenten waarop het formaat van het document kan veranderen:

- Om te kunnen voldoen aan de archiefwet is het van belang dat de inhoud van een informatieobject na afsluiten van de zaak niet meer muteerbaar is. De registratie wordt bevroren en in sommige gevallen (zie bijlage B) wordt het bestaande formaat omgezet naar een niet muteerbaar standaard formaat.
- Een overheidsorgaan is conform de archiefregeling 2010 (artikel 25) verplicht indien gereede kans bestaat dat als gevolg van wijziging of in onbruik raken van besturingsprogrammatuur of toepassingsprogrammatuur niet langer voldaan kan worden aan de gestelde eisen ten aanzien van toegankelijke en geordende staat van digitale archiefbescheiden dat conversie of migratie van die archiefbescheiden plaatsvindt.

Over het aanleveren van digitale archiefbescheiden zegt de archiefregeling, in hoofdstuk 3 “Geordende en toegankelijk staat van archiefbescheiden”, artikel 26

“Digitale archiefbescheiden worden, uiterlijk op het tijdstip van overbrenging, opgeslagen in een valideerbaar en volledig gedocumenteerde bestandsformaat dat voldoet aan een open standaard, tenzij dit redelijkerwijs niet van de zorgdrager kan worden verlangd. Alsdan vindt met de beheerder van de voor overbrenging aangewezen archiefbewaarplaats overleg plaats over een alternatief bestandsformaat.”

¹ Digital Born geeft aan dat een document in een digitaal formaat is gecreëerd en enkel in digitale vorm bestaat.

2 Stelling

Het is niet wenselijk dat bestandsformaten in het DMS worden opgenomen die geen documenten zijn, zoals bijvoorbeeld .exe of .dll bestanden. Er worden alleen bestandsformaten opgenomen in het DMS die voldoen aan onderstaande stelling:

In het DMS worden alleen bestandsformaten opgenomen die een document (kunnen) vertegenwoordigen.

Voor het borgen van deze stelling en het garanderen van toekomstige toegankelijkheid maakt de gemeente Eindhoven gebruik van een vastgestelde lijst met bestandsformaten. Alleen de bestandsformaten opgenomen op deze lijst kunnen/mogen voorkomen in het DMS.

Het DMS is zo ingericht dat het alleen bestandsformaten accepteert die zijn opgenomen in bijlage B. Als een gebruiker een bestandsformaat wil opnemen die niet in de lijst voorkomt krijgt de gebruiker een signaal. De gebruiker kan dan via de gebruikelijke kanalen een verzoek indienen om het bestandsformaat toe te voegen.

3 Knelpunten bij het duurzaam opslaan van digitale informatie

De gemeente Eindhoven genereert, ontvangt en slaat digitale informatie op in verschillende bestandsformaten. Bijvoorbeeld tekstdocumenten, power-point presentaties, spreadsheets, e-mails, databases, (audio)visuele informatie en computer aided design (CAD) bestanden. De gegenereerde of ontvangen informatie moet duurzaam worden opgeslagen. De duurzaamheid van de informatie komt in gevaar door:

- Technologische ontwikkelingen

De technologie ontwikkelt zich constant, onder andere om aan de steeds veranderende behoefte van de gebruiker te voldoen. Bestandsformaten blijven constant evolueren met als gevolg dat bestandsformaten in ongebruik raken. Bestandsformaten raken bijvoorbeeld in onbruik doordat:

- Software veroudert of omdat er een nieuwe versie van de software op de markt komt;
- Er leesbaarheidsproblemen ontstaan;
- Er betere ondersteuning is voor nieuwe bestandsformaten of omdat een andere bestandformaat meer functionaliteiten biedt;
- Bestandsformaten niet wijd gedistribueerd zijn;
- Bepaalde platformen en/of software het bestandsformaat niet accepteren;
- Het bestandsformaat slecht of niet gedocumenteerd is;

Informatie dat vandaag wordt opgeslagen in een specifiek bestandsformaat kan, na het in onbruik raken van dat bestandsformaat, niet meer worden gereproduceerd. Het niet kunnen reproduceren van de informatie betekent dat de informatie verloren gaat. De informatie kan niet meer worden geopend, gelezen en/of bewerkt. De informatie kan dan niet meer worden gebruikt in het primair proces, niet meer worden ingezet ter verantwoording en bewijs en kan geen cultureel- maatschappelijke functie meer vervullen. Wanneer informatie verloren gaat voordat de wettelijke bewaartermijn is verstreken, wordt er niet voldaan aan de eisen die vanuit de Archiefwet worden gesteld, namelijk het in goede, geordende en toegankelijke staat hebben en houden van de informatie.

- Vendor lock-in

Leveranciers willen hun positie op de markt behouden en zien graag dat gebruikers afhankelijk zijn van de door hen geproduceerde software. Het gevolg hiervan is dat bestandsformaten specifiek gekoppeld kunnen zijn aan één type software. Dit begrip staat beter bekend als vendor lock-in. De gebruiker is gebonden aan de mogelijkheden van de software die door de leverancier zijn vastgesteld. Dit beperkt de flexibiliteit en vrijheid van gebruikers. Wanneer van een ander softwarepakket gebruik gemaakt gaat worden of wanneer de ontvanger van de informatie geen beschikking heeft over de software waarin de informatie tot stand is gekomen is de informatie niet reproduceerbaar.

4 Oplossing voor het duurzaam opslaan van digitale informatie

Zoals hierboven is beschreven is de duurzame toegankelijkheid van de informatie niet vanzelfsprekend en is deze zelfs in gevaar als er geen maatregelen worden getroffen. Het duurzaamheidsprobleem wordt op dit moment het beste opgelost door de informatie op te slaan in een standaard bestandsformaat dat eveneens geschikt is als archiveringsformaat.

- Standaard bestandsformaat

Een bestandsformaat is een standaard formaat wanneer?

- Het bestandsformaat bijvoorbeeld een open standaard is. Open standaarden zijn niet gebonden aan een specifieke leverancier, wat vendor lock-in tegengaat. Een open standaard is uitwisselbaar tussen verschillende software/systemen;
- Het formaat onder de bewezen technologie valt; het formaat heeft een grote gebruikersgroep en heeft zijn nut en noodzaak bewezen. Bewezen technologie heeft als voordeel dat er van bestaande kennis gebruikt gemaakt kan worden. De standaarden die zijn geselecteerd als duurzaam archiveringsformaat, zoals opgenomen in tabel 1 in hoofdstuk 4, hebben in het verleden hun bruikbaarheid bewezen en bewijzen deze nog steeds.

- Standaard archiveringsformaat

Niet alle standaard bestandsformaten zijn geschikt om informatie te archiveren. Een standaard bestandsformaat is ook een geschikt archiveringsformaat wanneer het formaat tenminste aan de volgende criteria voldoet:

- Het archiveringsformaat is in principe een open standaard
- Het archiveringsformaat heeft voldoende marktondersteuning en is wijdverspreid
- Het archiveringsformaat heeft een goed gestructureerde opslag van informatie met zo min mogelijk informatieverlies
- Een (non-profit) beheerorganisatie heeft het archiveringsformaat goed gedocumenteerd
- Het archiveringsformaat is uitwisselbaar en niet afhankelijk van een specifieke leverancier
- Informatie opgeslagen in het archiveringsformaat moet geen verwijzingen bevatten naar externe informatiebronnen of -objecten.
- Het archiveringsformaat bewaart de integriteit van het document en de informatie daarin.
- Het archiveringsformaat kent een robuust foutopsporings- en verbeteringsmechanisme; fouten in bitopslag worden automatisch gedetecteerd en zijn herstelbaar

Het opslaan van de informatie in een duurzaam archiveringsformaat dient indien mogelijk in het werkproces te worden geïntegreerd. Elk definitief exemplaar van een document wordt door de archiefvormer direct in het archiveringsformaat opgeslagen.

Niet alle informatie kan in een duurzaam archiveringsformaat worden opgeslagen, omdat:

- Niet voor alle formaten een geschikt duurzaam archiveringsformaat op de markt is.
- De functionaliteiten van het originele bestandsformaat waarin de informatie is gegenereerd of ontvangen behouden moeten blijven. Bijvoorbeeld wanneer een spreadsheet (rekenblad) met complexe rekenfunctionaliteit in MS Excel is gegenereerd en de rekenfunctionaliteit duurzaam behouden moeten blijven. De functionaliteiten gaan verloren als de informatie wordt opgeslagen in PDF/A. PDF/A is een duurzaam archiveringsformaat, maar biedt geen ondersteuning voor (complexe) functionaliteiten.

In bovenstaande gevallen dient de informatie opgeslagen te worden in het formaat waarin de informatie is gegenereerd of ontvangen, ter voorkoming van onherstelbaar informatieverlies.

Het operationele DMS eDocs wat door een groot deel van de organisatie wordt gebruikt voorziet in het behoud van functionaliteit door beide versies op te slaan. Op het moment dat een document wordt afgehandeld en niet voldoet aan een standaardformaat wordt deze door een tool in het systeem omgezet naar PDF/A. Het systeem ziet de PDF/A versie als laatste versie en zal deze ook altijd tonen. Het originele formaat blijft voor hergebruik als eerdere versie beschikbaar.

5 Standaard bestandsformaten

Om de lange termijn leesbaarheid van digitale documenten te verzekeren is het verstandig om zoveel mogelijk gestandaardiseerde formaten te gebruiken. Standaarden spelen een belangrijke rol in de digitale bewaarstrategieën voor digitale archiefdocumenten zoals emulatie en migratie.

Standaard bestandsformaten hebben het voordeel dat ze compatibel zijn met meerdere softwaretoepassing. 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 tijd nieuwe viewers te maken. Het bewaren van informatie in bestandsformaten met de status van standaard dringt zich ook op om uitwisseling tussen partijen mogelijk te maken.

Vanuit het archiveringsprincipe moeten bestandsformaten op een bepaald moment worden omgezet naar een standaard bestandsformaat. Deze standaard moet voldoen aan de criteria voor:

- Duurzame opslag.
- Duurzame toegankelijkheid.
- Onveranderbaarheid van de documenten.

Het omzetten naar een standaard formaat is niet in alle gevallen mogelijk, bijv. bij audio-visuele documenten.

Het tijdstip voor het omzetten naar een standaardformaat is binnen de gemeente Eindhoven gezet op het moment een document de status afgehandeld krijgt. Als een dossier de status afgehandeld krijgt zet een tool in het systeem ook alle documenten in dat dossier die nog niet de status afgehandeld hebben op de status afgehandeld. Conform de vastgestelde lijst worden alle documenten in het dossier omgezet naar in de lijst genoemde standaardformaat.

6 Voorkeursformaten en acceptabele formaten

Het Nationaal Archief legt in haar preservation policy (Nationaal Archief, 2016) geen beperkingen op wat betreft bestandsformaten. Het e-depot moet technische gezien in staat zijn om een grote verscheidenheid aan bestandsformaten op te nemen en te bewaren. Waarbij dan wel de voorkeur wordt gegeven aan bestandsformaten die voldoen aan open standaarden.

De archiefregeling noemt geen specifieke opslagformaten voor digitale archiefbescheiden. Dit vanwege de voortdurende en frequente veranderingen van deze formaten. In plaats van standaarden bevat de archiefregeling alleen de (kwaliteits)eisen die eraan gesteld worden. Dus een valideerbaar en volledig gedocumenteerd bestandsformaat dat voldoet aan een open standaard. De wet schrijft dus criteria voor waaraan archiefbescheiden moeten voldoen op moment van overbrenging.

Vanuit haar preservation policy en haar ervaring heeft zij de formaten opgedeeld in twee categorieën:

- Voorkeursformaten: die zijn “open” formaten. Daarnaast kan het zijn dat voor bepaalde bestandstypes zogeheten “industry standards” zijn die massaal worden gebruikt en voldoende gedocumenteerd zijn.
- Acceptabele formaten: dit zijn formaten die niet (volledig) “open” zijn, maar die toch acceptabel zijn voor het Nationaal Archief. De reden hiervoor kan zijn dat het Nationaal Archief (of andere (inter)nationale instanties) ervaring heeft met een bepaald formaat, en er geen goede preservationstrategie voor heeft.

In bijlage A staat per documentsoort aangegeven wat het duurzame archiveringsformaat is waarin de informatie moet worden opgeslagen. Er wordt onderscheid gemaakt tussen het duurzame archiveringsformaat dat de voorkeur heeft en welke formaten acceptabel zijn als duurzaam archiveringsformaat.

Vanuit het perspectief van beheer en van efficiënt werken is er gekozen voor een beperkte set aan archiveringsformaten. Dit betekent dat elke medewerkers kennis kan hebben van het formaat waarin de informatie moet worden opgeslagen en dat binnen het hele gemeente dezelfde soort informatie in hetzelfde formaat wordt gearchieveerd. De informatie is dan beter toegankelijk voor iedereen. Ook hard- en software moet beheerd worden: zodat deze toegankelijk zijn en blijven. Hoe meer hard- en software er gebruikt wordt hoe complexer het beheer. Sector Informatisering & Beheer (IB) stelt de benodigde hard- en software beschikbaar en is verantwoordelijk voor het beheer van de programmatuur.

7 Toetsingscriteria voor voorkeurs- en acceptabele duurzame archiveringsformaten

De archiveringsformaten worden geplaatst op de lijst voor voorkeur of acceptabele formaten. De plaatsing van formaten op de voorkeur of acceptabele lijst is afhankelijk in welke mate zij aan de criteria voldoen.

Voorkeursformaten zijn de formaten die in de ideale situatie geschikt zijn als duurzaam archiveringsformaat.

Acceptabele formaten voldoen in mindere mate aan alle criteria, maar zijn duurzaam genoeg. Op de lijst van de acceptabele formaten kunnen ook formaten staan die bijvoorbeeld geen open standaard zijn, maar wel als een archiveringsformaat worden gezien.

Om te bepalen of een formaat voldoet als archiveringsformaat zijn onderstaande toetsingscriteria opgezet. De toetsingscriteria zijn zo opgezet om onderscheid te kunnen maken tussen de voorkeurs- en acceptabele archiveringsformaten.

Archief	Interoperabiliteit	Techniek	Vendor-lock in
Het archiveringsformaat heeft voldoende marktondersteuning en is wijdverspreid	Het archiveringsformaat is in principe een open standaard	Het archiveringsformaat geeft een goed gestructureerde opslag van informatie met zo min mogelijk informatieverlies	Het archiveringsformaat is niet afhankelijk van een specifieke leverancier
Een (non-profit) beheersorganisatie heeft het archiveringsformaat goed gedocumenteerd	Het archiveringsformaat is uitwisselbaar	Informatie opgeslagen in het archiveringsformaat moet geen verwijzingen bevatten naar externe informatiebronnen of objecten	
Het archiveringsformaat order de "bewezen technologie" valt; Het formaat heeft een grote gebruikersgroep en heeft zijn nut en noodzaak bewezen		Het archiveringsformaat biedt de mogelijkheid tot insluiten van (zelf gedefinieerde) metadatavelden	

Archiveringsformaten komen op de voorkeurslijst te staan zodra deze voldoen aan de vier toetsingscriteria. Niet voldoen betekent dat de archiveringsformaten worden opgenomen onder de acceptabele formaten.

In de bijlage A zijn respectievelijk opgenomen een verkorte lijst met voorkeursformaten en acceptabele formaten en overzichten van de voorkeursformaten en acceptabele formaten

8 Bijlage A - Overzichten van voorkeursformaten en acceptabele formaten

Voorkeursformaten en acceptabele formaten in het kort

Informatiesoort	Voorkeursformaat	Acceptabel formaat
Audio	WAV, WAVE, BWF	MP3, AAC
Database	SQL, SIARD, ODB	ACDB, MDB
Document	PDF/A-1, PDF/A-2, ODT	PDF 1.7, DOC, DOCX
E-mail	EML	MSG, PST, MBOX
Image	TIFF, PNG	JPEG, JPG, JP2, JPX
Presentation	ODP, PDF/A	PPT, PPTX
Spreadsheet	ODS, CSV, PDF/A	XLS, XLSX
Vector Image	SVG	In overleg
Video	MXF	MPEG-4 MKV

Voorkeursformaten: dit zijn de “open” formaten. Daarnaast kan het zijn dat voor bepaalde bestandstypes zogeheten “industry standards” zijn die massaal worden gebruikt en voldoende gedocumenteerd zijn.

Overzicht voorkeursformaten

Informatiesoort	Formaat	Onderbouwing
Audio	WAV, WAVE	WAV (Waveform Audio File Format) is een bestandsformaat voor de opslag van (veelal) ongecomprimeerde audiodata. Het is een facto, gepubliceerde standaard ontwikkeld door Microsoft en IBM
	BWF	Het Broadcast Wave Format is een uitbreiding op WAV met een aantal extra metadatavelden. BWF is het standaard conserveringsformaat voor audiobestanden van het instituut voor Beeld en Geluid
Database	SQL	SQL is opgenomen op de lijst met gangbare standaarden van het Forum standaardisatie. SQL is een ANSI/ISO-standaardtaal voor een relationeel “database management systeem” (DBMS). Het is een gestandaardiseerde taal die gebruikt kan worden voor taken zoals het bevragen en het aanpassen van informatie in een relationele databank. SQL kan met vrijwel alle moderne relationele databankproducten worden gebruikt
	SIARD	SIARD staat voor Software Independant Archiving of Relational Databases. Het is ontwikkeld door de Swiss Federal Archives om relationele databases duurzaam te kunnen archiveren. SIARD bestaat uit een formaat en een suite. Het SIARD-formaat is een open formaat voor het archiveren van de inhoud van relationele database
	ODB	ODB behoort tot de OpenDocument standard (ODF, NEN-ISO/IEC 26300:2007) dat als standard voor office-documenten staat op de pas-toe-of-leg-uit lijst van het Forum Standaardisatie.
Document	PDF/A-1, PDF/A-2	PDF/A is een wijdverbreide open standaard, een NEN/ISO norm (ISO:19005), PDF/A-1 en PDF/A-2 staan op de pas-toe-of-leg-uit lijst met open standaarden van het Forum Standaardisatie

	ODT	ODT behoort tot de OpenDocument standard (ODF, NEN-ISO/IEC 26300:2007) data is standard voor office-documenten op de pas-toe-of-leg-uit lijst van het Forum Standaardisatie staat.
E-mail	EML	Bij EML-bestanden is doorgaans sprake van het opslaan van individuele e-mails als individuele bestanden. Bijlagen kunnen MIME-inhoud worden opgeslagen in die bestanden, of als apart bestand waarnaar vanuit het EML-bestand gelinkt wordt. EML is steeds meer een facto (industry) standaard aan het worden.
Image	TIFF	Tagged Image File Format is een eenvoudig, appartaat onafhankelijk, besturingssysteem onafhankelijk, goed uitwisselbaar, "industry standard" bestandsformaat met een achterwaarts compatibele stabiele kern (TIFF-baseline). TIFF maakt deel uit van DE BASIS van DEN.
	PNG	PNG is opgenomen op de pas-toe-of-leg-uit lijst met open standaarden van Forum Standaardisatie, maar uitsluitend voor het "gebruik van grafische afbeeldingen (met "lossless" compressie) binnen ODF-documenten
Presentation	ODP	ODP behoort tot de OpenDocument standard (ODF, NEN-ISO/IEC 26300:2007) dat als standard voor office documenten staat op de pas-toe-of-leg-uit lijst van het Forum Standaardisatie
	PDF/A	PDF/A is een wijd verbreide open standaard, een NEN/ISO norm (ISO:19005). PDF/A1 en PDF/A2 zijn opgenomen op de pas-toe-of-leg-uit lijst met open standaarden van het Forum Standaardisatie. Let wel: bepaalde (interactieve) functionaliteiten zal na omzetting naar PDF/A formaat niet meer beschikbaar zijn. Als deze functionaliteit als essentieel wordt beschouwd, is dit een reden om niet voor PDF/A te kiezen.
Spreadsheet	ODS	ODS behoort tot de OpenDocument standard (ODF, NEN-ISO/IEC 26300:2007) dat als standard voor office documenten staat op de pas-toe-of-leg-uit lijst van het Forum Standaardisatie
	CSV	Als het alleen om het bewaren van niet-interactieve informatie uit cellen gaat, kan het kommagescheiden (.csv) tekstbestand als alternatief voor een spreadsheet worden gekozen
	PDF/A	PDF/A is een wijd verbreide open standaard, een NEN/ISO norm (ISO:19005). PDF/A1 en PDF/A2 zijn opgenomen op de pas-toe-of-leg-uit lijst met open standaarden van het Forum Standaardisatie. Let wel: bepaalde (interactieve) functionaliteiten zal na omzetting naar PDF/A formaat niet meer beschikbaar zijn. Als deze functionaliteit als essentieel wordt beschouwd, is dit een reden om niet voor PDF/A te kiezen.
Vector Image	SVG	SVG staat voor "Scalable Vector Graphics". Het is een robuust op XML gebaseerd formaat voor statistische en dynamische vectorafbeeldingen. SVG is een open standaard en de ondersteuning van het formaat is over het verloop van tijd sterk toegenomen. SVG staat op lijst met "gangbare" standaarden van het Forum Standaardisatie
Video	MXF	MXF (Material eXchange Format) is een open standaard die wordt onderhouden door de AV-standaardorganisatie Society of Moving Pictures Engineers (SMPTE). Het formaat is bestemd voor professioneel gebruikt en wordt ondersteund door een groot aantal verschillende transcoders en editing softwarepakketten. MXF is het standaard conserveringsformaat voor videobestanden van het Instituut voor Beeld en Geluid. MXF maakt deel uit van de DE BASIS van DEN

Acceptabele formaten: dit zijn formaten die niet (volledig) “open” zijn, maar die toch acceptabel zijn voor het Nationaal Archief. De reden hiervoor kan zijn dat het Nationaal Archief (of andere (inter)nationale instanties) ervaring heeft met een bepaald formaat, en er goed preservationstrategie voor heeft.

Overzicht acceptabele formaten

Informatiesoort	Formaat	Onderbouwing
Audio	MP3	MPEG1-Layer 3 is de populaire wijze van geluidsopslag van dit moment. Met het formaat worden lossy compressie gemaakt. MP3 maakt deel uit van DE BASIS van DEN
	AAC	Advanced Audio Coding (AAC) is een niet exact-omkeerbare compressietechniek voor digitale audio. AAC is gestandaardiseerd door ISO en IEC (ISO/IEC 14496-3:2009), als onderdeel van de MPEG-2 en MPEG-4 specificaties. AAC maakt deel uit van DE BASIS van DEN
Database	ACDB, MDB	Microsoft Access is marktleider op het gebied van databases op windowssystemen. Sinds Access 2007 is het standaard bestandsformaat voor Access-databases.accdb. Daarvoor was het .mdb. Het Nationaal Archief heeft
Document	PDF 1.7	PDF 1.7 is opgenomen op de “pas toe of leg uit” lijst met open standaarden van Forum Standaardisatie, voor het uitwisselen en publiceren van niet- of beperkt-reviseerbare documenten, waarbij duiding van oorsprong of functierijkheid onderdeel zijn van het document en waarbij PDF/A-1 als standaard niet kan worden ingezet. PDF 1.7 is een open standaard en is door ISO en NEN genormeerd als ISO 32000-1:2008
	DOC, DOCX	Microsoft Office Word is marktleider op het gebied van tekstverwerkers. Het Nationaal Archief heeft ervaring opgedaan met dit type bestandsformaat en beschikt over een gedegen bestandsstrategie
E-mail	MSG, PST	Microsoft Office is marktleider op het gebied van e-mailsoftware (MS Exchange, Outlook). Het standaardformaat waarin e-mails worden opgeslagen, is MSG. In een Personal Storage Table (PST) kunnen e-mailberichten worden gebundeld. Het Nationaal Archief heeft ervaring opgedaan met dit type bestandsformaten en beschikt over een gedegen bestandsstrategie
	MBOX	MBOX is een generieke term voor een groep bestandsformaten die gebruikt worden voor een collectie van e-mailberichten. Alle berichten in een MBOX-mailbox worden samengevoegd en opgeslagen als platte tekst in een enkel bestand. In het algemeen bestaat MBOX uit één bestand met de extensie .mbox of .mbx, met daarin de volledige inhoud van een e-mailfolder. Net als EML is MBOX steeds meer een de facto (industry) standaard aan het worden.
Image	JPEG, JPG	JPEG (Joint Photographic Experts Group) is een veelgebruikte methode voor zogenaamde “lossy” compressie van digitale afbeeldingen. JPEG is opgenomen op de “pas toe of leg uit” lijst met open standaarden van Forum Standaardisatie, maar uitsluitend voor het “gebruik van grafische afbeeldingen (met “lossy compressie”) binnen ODF-documenten
	JP2, JPX	JPEG2000 (JP2) is een standaard die met één compressie-algoritme zowel een lossless (reversible) als een lossy (irreversible) compressie kan uitvoeren. Hoewel JPEG 2000 als ISO-standaard is gepubliceerd, wordt ze (bijvoorbeeld in webbrowsers) nog niet volop (native) ondersteund. JPX is een uitgebreide versie van de JP2-standaard, met ondersteuning voor

		meerdere afbeeldingen en meer mogelijkheden voor kleurencodering.
Presentation	PPT, PPTX	Microsoft Office Powerpoint is marktleider op het gebied van spreadsheets. Het Nationaal Archief heeft al ervaring opgedaan met dit type bestandsformaat en beschikt over een gedegen bestandsstrategie.
Spreadsheet	XLS, SLSX	Microsoft Office Excel is marktleider op het gebied van spreadsheets. Het Nationaal Archief heeft al ervaring opgedaan met dit type bestandsformaat en beschikt over een gedegen bestandsstrategie
Vector Image	In overleg	Naast SVG zijn er geen duidelijke standaarden voor Vector Image onderkend
Video	MPEG-4	MPEG-4 (Moving Pictures Experts Group-4) is momenteel waarschijnlijk het meest gebruikte videoformaat en daarmee een industry standard. Echter, MPEG4 is geen open formaat vanwege vereiste royalties. Het is door ISO en NEN genormeerd als ISO/IEC 14496
	MKV	Matroska (.mkv) is een open source en zeer flexibel alternatief voor bestaande containerformaten zoals AVI, ASF, MOV, RM, MP4, MPG enzovoort en kan vrijwel alle codes bevatten . MKV heeft echter als nadeel dat het nog een relatief geringe ondersteuning door videosoftware kent.

9 Bijlage B - Bestandsformaten die voorkomen in het DMS (eDocs)

Een toelichting op de kolommen is onder de tabel opgenomen

Extensie	Applicatie	Toelichting (conform https://nl.wikipedia.org/wiki/Lijst_van_bestandsextensies)	Informatiesoort	Voorkeur formaat	Acceptabel Formaat	Omzetten naar PDF
BMP	Web Publishing BMP	Niet-gecomprimeerd grafisch bestand	Web	☹️	☹️	😊
DOC	Microsoft Word	Tekstdocument Word versie 97-2003	Document	☹️	😊	😊
DOCX	Microsoft Word	Document Word versie 2007 of hoger	Document	☹️	😊	😊
DWG	AutoCAD DM Extension	AutoCAD-tekeningbestand (DraWinG)	Image	☹️	☹️	☹️
EML	Microsoft e-mail	Opgeslagen mailbericht	E-mail	😊		
GIF	Web Publishing GIF	Gecomprimeerd grafisch bestand	Image	☹️	☹️	😊
HTM	Internet Explorer Browser	Bestand geschreven in Hyper Text Markup Language	Web	☹️	☹️	😊
HTML	Internet Explorer Browser	Bestand geschreven in Hyper Text Markup Language	Web	☹️	☹️	😊
JPEG	Web Publishing JPG	Gecomprimeerd grafisch bestand	Image	☹️	😊	😊
JPG	Web Publishing JPG	Gecomprimeerd grafisch bestand	Image	☹️	😊	😊
MHT	Internet Explorer Browser	Webarchive, een bestand waar een hele webpagina inclusief plaatjes in opgeslagen kan worden	Web	☹️	☹️	😊
MSG	Microsoft Outlook	E-mail bestand van Microsoft Office Outlook	E-mail	☹️	😊	😊
OUT	Microsoft Word	Documenten vanuit Centric applicatie	Document	☹️	☹️	😊
PDF	Adobe Acrobat	Portable Document Format van Adobe	Document	😊		
PNG	Web Publishing PNG	Gecomprimeerd grafisch bestand	Image	☹️	☹️	😊
PPT	Microsoft PowerPoint	Presentatiedocument voor Microsoft Powerpoint versie 97-2003	Presentation	☹️	😊	😊
PPTX	Microsoft PowerPoint	Presentatiedocument voor Microsoft Powerpoint versie 2007 en hoger	Presentation	☹️	😊	😊
RTF	Microsoft Word	Rich Text Format, documentopmaak in ASCII	Document	☹️	☹️	😊
TIF	Hummingbird Imaging	Tagged Image File, formaat voor bitmapafbeeldingen	Image	😊		😊

TXT	Microsoft Word	Text, platte tekst	Document	☹	☹	😊
VSD	Microsoft Visio	Diagram / flowcharts voor Microsoft Visio	Image	☹	☹	😊
XLS	Microsoft Excel	Spreadsheetdocument voor Microsoft Excel, versie 97-2003	Spreadsheet	☹	😊	😊
XLSX	Microsoft Excel	Spreadsheetdocument voor Microsoft Excel, versie 2007 en hoger	Spreadsheet	☹	😊	😊
XLSM	Microsoft Excel	Spreadsheetdocument voor Microsoft Excel met macro's	Spreadsheet	☹	☹	😊
XML	Web publishing XML	Extensible Markup Language	Document	☹	☹	😊
ZIP	Winzip	Gecomprimeerde bestanden	Container	☹	☹	☹

Toelichting kolommen

Extensie	Is de bestandsextensie die door de applicatie waarmee het bestand is gecreëerd wordt toekend aan het bestand
Applicatie	Is de software waarmee het bestand is gecreëerd
Toelichting	Betekenis en korte omschrijving van de extensie
Informatiesoort	Geeft weer waarvoor het bestand (hoofdzakelijk) wordt gebruikt
Voorkeursformaat	Geeft aan dat de extensie voldoet als voorkeursformaat (☹ = nee, 😊 = ja)
Acceptabel formaat	Geeft aan dat de extensie voldoet als acceptabel formaat (☹ = nee, 😊 = ja)
Omzetten naar PDF	Geeft aan dat de extensie door het DMS wordt omgezet naar PDF (☹ = nee, 😊 = ja)