



-Voorkeursformaten e-depot
Brabants Historisch Informatie Centrum-

|| || || bhic

BRABANTS HISTORISCH INFORMATIE CENTRUM

Inhoud

1. Voorkeursformaten Brabants Historisch Informatie Centrum	3
2. Voor wie is dit document bedoeld?	3
3. Ontwikkeling en review	3
4. Welke informatiesoorten onderscheidt het BHIC?	4
5. Waarom voorkeursformaten?	4
6. Beheer en beschikbaarstelling: makkelijker bij zo min mogelijk diversiteit.....	4
7. Open standaarden en interoperabiliteit: zo weinig mogelijk hindernissen	5
8. Voorkeursformaten en acceptabele formaten	5
8.1 Twee categorieën formaten	6
8.2 Wat als er sprake is van een ander formaat?	6
Bijlage 1 - Voorkeursformaten en acceptabele formaten in het kort.....	6
Bijlage 2 - Overzicht voorkeursformaten inclusief onderbouwing	7

1. Voorkeursformaten Brabants Historisch Informatie Centrum

Het E-depot van het BHIC kan digitale informatie in allerlei vormen en in allerlei formaten ontvangen, duurzaam bewaren en beschikbaar stellen. Maar met het oog op digitale duurzaamheid heeft het BHIC wel een aantal voorkeuren voor formaten waarin archiefvormers informatie aanleveren. Dit document benoemt die voorkeursformaten, en geeft ook een onderbouwing daarvoor.

Bij de overbrenging van digitale informatie naar het e-depot wordt als uitgangspunt genomen dat de originele bestanden worden gebruikt.¹ Voor zover sprake is van een legacyomgeving², kunnen deze formaten afwijken van die welke worden genoemd in onderstaande lijst van voorkeursformaten. In dat geval kan binnen de e-depotomgeving een preserveringskopie worden gemaakt in één van de gangbare voorkeursformaten. Deze aanpak is gekozen om ervoor te zorgen dat de authenticiteit en integriteit van de bestandsinhoud zo goed mogelijk behouden blijven, denk bijvoorbeeld aan het gedrag van archiefbescheiden.³ Bovendien legt dit de meest solide basis voor het bepalen van een mogelijke migratiestrategie. Als de archiefvormer naast de originele bestanden beschikt over een representatieve versie (bijvoorbeeld een afgeleide in PDF van een DOCX bestand), zal tijdens het proces van selectie en waardering de afweging worden gemaakt of deze versies al dan niet voor opname in het e-depot in aanmerking komen, of worden voorgedragen ter vernietiging.

2. Voor wie is dit document bedoeld?

Dit document is bedoeld om archiefvormers handvatten te geven waarmee zij al bij de inrichting van (zaak)systemen waarin digitaal archief gevormd wordt, rekening kunnen houden met de voorkeursformaten van het BHIC. Zij kunnen dan in een vroeg stadium keuzes maken in de duurzaamheid van overheidsinformatie. Archiefvormers kunnen dit document echter ook gebruiken bij reeds gevormd archief. Ze kunnen ermee vaststellen welke acties nodig zijn om de duurzaamheid van het archief te bevorderen, en hoe zij het archief goed kunnen overdragen naar het BHIC. Dit document is dan ook met name bedoeld voor medewerkers zoals informatiearchitecten, (zaak)systeembeheerders en DIV-medewerkers.

3. Ontwikkeling en review

Dit document is een weergave van de huidige inzichten van het BHIC. Maar de (technologische) ontwikkelingen staan niet stil. Daarom zal regelmatige actualisering nodig zijn. Dat betekent onder meer dat de bestandsformaten geregeld worden gereviewd en waar nodig herzien. Zo kunnen formaten verschuiven van 'voorkeur' naar 'acceptabel', of andersom. Er kunnen nieuwe formaten bij komen, of formaten kunnen komen te vervallen. Leveren archiefvormers formaten aan die niet meer staan op de lijst van voorkeurs- of acceptabele formaten, dan gaat het BHIC met hen in gesprek over te nemen maatregelen.

¹ Dit zijn bestanden die de oorspronkelijke versie van een document of ander type digitale inhoud bevatten.

² In de context van informatiebeheer verwijst een legacyomgeving naar verouderde systemen, software, processen of praktijken die binnen een organisatie worden gebruikt om informatie en documenten te beheren.

³ Het geheel van dynamische en interactieve kenmerken van archiefbescheiden.

4. Welke informatiesoorten onderscheidt het BHIC?

Vanuit de aangesloten partijen / archiefvormers verwacht het BHIC de komende jaren vooral digitale – en steeds meer digital born – informatieobjecten te ontvangen, in verschillende bestandsformaten, bijvoorbeeld TIFFscans (rasterplaatjes), (Outlook) e-mails, PDF-(tekst)documenten en Microsoft Office-bestanden (tekstdocumenten, spreadsheets en presentaties).

Het is belangrijk dat het overzicht wordt bewaard in de omgang met de vele soorten informatie en bestandsformaten. Daarom hanteert het BHIC een onderverdeling in informatiesoorten die is afgestemd met (inter)nationale best practices, en die verder wordt afgestemd met best practices uit het werkveld. Dat is goed voor de (inter)nationale samenwerking en standaardisatie, en daarmee voor de eigen efficiëntie.

Het BHIC hanteert de volgende onderverdeling in informatiesoorten:

1. Audio (geluid)
2. Database
3. Document (tekst gebaseerde documenten)
4. E-mail (berichten)
5. Image (rasterplaatjes)
6. Presentatie
7. Spreadsheet
8. Vector image (schaalbare plaatjes)
9. Video (bewegend beeld)
10. Websites

5. Waarom voorkeursformaten?

Het BHIC legt in haar conserveringsplan geen beperkingen op wat betreft bestandsformaten. Ook is de e-depotvoorziening technisch gezien in staat om een grote verscheidenheid aan bestandsformaten op te nemen en te bewaren. Maar met het oog op digitale duurzaamheid heeft het BHIC wel een voorkeur voor een beperkt aantal bestandsformaten die voldoen aan open standaarden.

6. Beheer en beschikbaarstelling: makkelijker bij zo min mogelijk diversiteit

Hoe groter de diversiteit aan aangeleverde bestandsformaten, des te groter de inspanning om al die verschillende soorten formaten goed in de tijd te beheren en beschikbaar te stellen. Daarom verdient het de voorkeur om het aantal bestandsformaten zo klein mogelijk te houden. Het BHIC heeft ervaring opgedaan met diverse typen bestanden, en kan nu zeggen dat zij deze typen bestanden “voor de eeuwigheid” kan bewaren en beschikbaar stellen voor het publiek. Bij ‘exotische’ formaten

– formaten die vrij onbekend zijn en weinig in gebruik – zullen het beheer en de beschikbaarstelling een veel grotere inspanning vergen en dus meer geld kosten.

Een keuze voor niet-open standaarden heeft mogelijk gevolgen voor:

- de bijbehorende conserveringsstrategie, dat wil zeggen de actieve en passieve conservering
- de vorm van de beschikbaarstelling van de informatie, in verband met de noodzakelijke viewers of vrij te verkrijgen software.

7. Open standaarden en interoperabiliteit: zo weinig mogelijk hindernissen

Preservering – digitale informatie meenemen door de tijd heen – gaat makkelijker met open standaarden. Het BHIC volgt hierin het Forum Standaardisatie⁴, dat de interoperabiliteit en de toepassing van open standaarden binnen de Nederlandse overheid bevordert. Waar het Forum Standaardisatie (nog) geen standaard heeft gedefinieerd, volgt het BHIC de Handreiking voorkeursformaten van het Nationaal Archief en de Wegwijzer Voorkeursformaten van het Netwerk Digitaal Erfgoed (NDE)⁵.

Een ‘open standaard’ omvat volgens het Forum Standaardisatie een laagdrempelige beschikbaarheid van documentatie, geen hindernissen op basis van intellectuele eigendomsrechten (zoals patent royalties), inspraakmogelijkheden en onafhankelijkheid en duurzaamheid van de standaardisatie-organisatie. Open standaarden worden beheerd door non-profit organisaties. Iedereen kan onbeperkt gebruik maken van open standaarden. Belangrijk is ook dat de besluitvorming over open standaarden voor alle betrokken partijen toegankelijk is. Overheden en semi-overheden zijn verplicht om de open standaarden die op de lijst staan, bij aanschaf of (ver)bouw van ICT-systemen of -diensten te eisen.

Het BHIC heeft omwille van de interoperabiliteit een duidelijke voorkeur voor de ‘open’ formaten op de lijst van het Forum Standaardisatie. Interoperabiliteit betekent enerzijds het vermogen van organisaties (en hun processen en systemen) om effectief en efficiënt informatie te delen met hun omgeving. En anderzijds om deze informatie ook in de loop van de tijd duurzaam toegankelijk te houden. Open formaten zijn ook onderdeel van de DUTO-kwaliteitseisen waarmee overheidsorganen kunnen bepalen welke maatregelen ze moeten nemen om de digitale informatie die ze ontvangen en creëren, duurzaam toegankelijk te maken en te houden.

8. Voorkeursformaten en acceptabele formaten

Het BHIC krijgt digitale informatie voor het E-depot het liefst aangeleverd in ‘open’ formaten. Naast deze voorkeursformaten onderscheidt het BHIC ook acceptabele formaten. Hieronder staat een overzicht van de voorkeursformaten en de acceptabele formaten.

⁴ <https://www.forumstandaardisatie.nl/>

⁵ <https://www.wegwijzervoorkeursformaten.nl/index.php/Hoofdpagina>

8.1 Twee categorieën formaten

Het BHIC hanteert twee categorieën formaten:

- 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.
- Acceptabele formaten: dit zijn formaten die niet (volledig) ‘open’ zijn, maar die toch acceptabel zijn voor het BHIC. De reden hiervoor kan zijn dat het BHIC (of andere (inter)nationale instanties) ervaring heeft met een bepaald formaat, en er een goede preservingsstrategie voor heeft.

8.2 Wat als er sprake is van een ander formaat?

Formaten die niet in één van bovenstaande categorieën vallen, worden niet bij voorbaat volledig uitgesloten voor opname in het E-depot. Mocht het zo zijn dat er informatie ter overbrenging wordt aangeboden in een formaat dat buiten deze twee categorieën valt, dan gaat het BHIC in overleg met de aanbieder. Er wordt dan een preservings-advies voor een passende oplossing geboden.

Bijlage 1 - 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
Presentatie	ODP, CSV, PDF/A	PPT, PPTX
Spreadsheet	ODS, CSV, PDF/A	XLS, XLSX
Vector Image	SVG	In overleg
Video	Uncompressed 10-bit AVI	MXF, MPEG-4, MKV
Website	WARC	ARC

Bijlage 2 - Overzicht voorkeursformaten inclusief onderbouwing

Informatiesoort	Formaat	Onderbouwing
Audio	WAV, WAVE	WAV (Waveform Audio File Format) is een bestandsformaat voor de opslag van (veelal) ongecomprimeerde audiodata. Het is een de 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 preservingformaat voor audiobestanden van het Nederlands 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 Independent 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 databases.
	ODB	ODB behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC 26300:2007) dat als standaard 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 standaard (ODF, NEN-ISO/IEC 26300:2007) dat als standaard 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 als MIME-inhoud worden opgeslagen in die bestanden, of als apart bestand waarnaar vanuit het EML-bestand gelinkt wordt. EML is steeds meer een de facto (industry) standaard aan het worden.
Image	TIFF	Tagged Image File Format is een eenvoudig, apparatuuronafhankelijk, 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.
Presentatie	ODP	ODP behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC 26300:2007) dat als standaard voor office-documenten staat op de 'pas-toe-of-leg-uit' lijst van het Forum Standaardisatie.
	PDF/A	<p>PDF/A is een wijd verbreide open standaard, een NEN/ISO norm (ISO:19005). PDF/A-1 en PDF/A-2 zijn opgenomen op de 'pas-toe-of-leg-uit' lijst met open standaarden van het Forum Standaardisatie.</p> <p>Let wel: bepaalde (interactieve) functionaliteit 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.</p>
Spreadsheet	ODS	ODS behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC 26300:2007) dat als standaard 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 komma gescheiden (.csv) tekstbestand als alternatief voor een spreadsheet worden gekozen.
	PDF/A	<p>PDF/A is een wijd verbreide open standaard, een NEN/ISO norm (ISO:19005). PDF/A-1 en PDF/A-2 zijn opgenomen op de 'pas-toe-of-leg-uit' lijst met open standaarden van het Forum Standaardisatie.</p> <p>Let wel: bepaalde (interactieve) functionaliteit 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</p>
Vector Image	SVG	<p>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.</p> <p>SVG staat op de lijst met 'gangbare' standaarden van het Forum Standaardisatie.</p>
Video	AVI/MOV	Audio Video Interleave (ook Audio Video Interleaved), bekend onder de initialen AVI, is een multimediacontainerformaat dat door Microsoft in november 1992 is geïntroduceerd als onderdeel van de Video for Windows-software. AVI-bestanden kunnen zowel audio- als

videogegevens bevatten in een bestandscontainer die synchrone audio-met-video-weergave mogelijk maakt. AVI-bestanden ondersteunen streaming audio en video, hoewel deze functies zelden worden gebruikt. De meeste AVI-bestanden gebruiken ook de bestandsextensie-uitbreidingen die in februari 1996 door de Matrox OpenDML-groep zijn ontwikkeld. Deze bestanden worden door Microsoft ondersteund en worden officieus "AVI 2.0" genoemd. AVI is een subformat van het Resource Interchange File Format (RIFF), dat de gegevens van een bestand verdeelt in blokken, of 'chunks'. Elke "brok" wordt geïdentificeerd door een FourCC-tag. Een AVI-bestand neemt de vorm aan van één "brok" in een RIFF-geformatteerd bestand, dat vervolgens wordt onderverdeeld in twee verplichte "brokken" en één optionele "brok".

Het eerste subblok wordt geïdentificeerd door de tag "hdrl". Dit subgedeelte is de bestandskop en bevat metagegevens over de video, zoals de breedte, hoogte en framesnelheid. Het tweede subblok wordt geïdentificeerd door de "movi" -tag. Dit blok bevat de feitelijke audio / visuele gegevens waaruit de AVI-film bestaat. Het derde optionele subblok wordt geïdentificeerd door de tag "idx1" die de offsets van de gegevensblokkeringen in het bestand indexeert.

Via het RIFF-formaat kunnen de audiovisuele gegevens in de "movi"brok worden gecodeerd of gedecodeerd door software die een codec wordt genoemd, wat een afkorting is voor (en) coder / decoder. Bij het maken van het bestand vertaalt de codec onbewerkte gegevens naar het (gecomprimeerde) gegevensformaat dat in het blok wordt gebruikt. Een AVI-bestand kan in de chunks audio- / visuele gegevens bevatten in vrijwel elk compressieschema, inclusief Full Frame (niet-gecomprimeerd), Intel Real Time (Indeo), Cinepak, Motion JPEG, bewerkbare MPEG, VDOWave, ClearVideo / RealVideo, QPEG en MPEG -4.

Een MOV-bestand is een standaard bestandsindeling voor multimediabestanden die is ontwikkeld door Apple en compatibel is met zowel Macintosh- als Windows-platforms. Het kan meerdere tracks bevatten die verschillende soorten mediagegevens opslaan en wordt vaak gebruikt voor het opslaan van films en andere videobestanden. MOV-bestanden gebruiken gewoonlijk de MPEG-4-codec voor compressie.

MOV-bestanden kunnen video-, audio-, tijdcodes- en teksttracks bevatten, die kunnen worden gebruikt voor het opslaan en synchroniseren van verschillende soorten media. Elke mediaspoor kan worden gecodeerd met behulp van een van verschillende codecs. Om een MOV-bestand af te spelen, moet een toepassing de specifieke codec (s) die in het bestand worden gebruikt, kunnen herkennen.

MOV-bestanden zijn te converteren naar meer standaard mediabestandsindelingen met verschillende mediatoeepassingen.

Websites	WARC	<p>WARC (ISO 28500:2009) is een open bestandsformaat voor het opslaan van websites.</p> <p>Hoewel websites uit meerdere digitale bronbestanden bestaan kunnen ze met het WARC formaat in één archiefbestand opgeslagen worden.</p> <p>Het WARC formaat is een herziening van het Internet Archive's ARC File Format [ARC_IA]. Het formaat wordt onder andere gebruikt voor web-archivering door de Internet Archive (Wayback Machine) en de Library of Congress.</p>
----------	------	--