



West-Brabants Archief

Overdrachtsprotocol e-depot

E-depot projectgroep WBA

Paul van den Bosch (auteur)

Chantal Damen

Albert Hoekstra

Ilse Jansen

Hans Janssen (projectleider)

Versie 1.0

December 2017

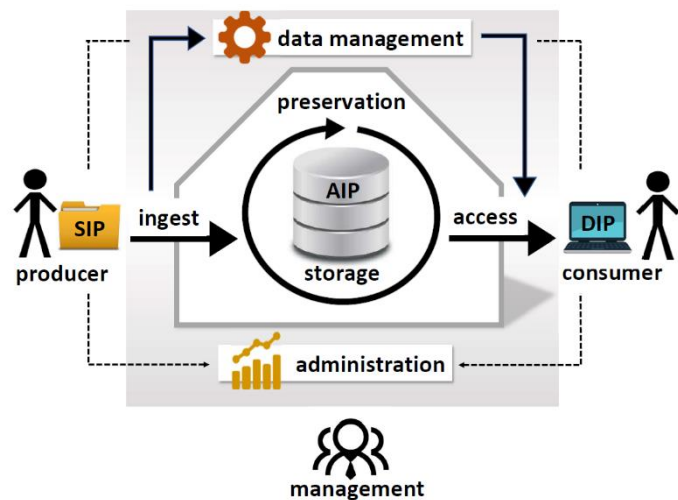
Inhoud

1.	Inleiding	3
2.	Overbrenging.....	4
2.1	Te maken keuzes en afspraken	4
2.2	Feitelijke overdracht.....	5
3.	Procedure overdracht informatieobjecten	6
3.1	Aanlevering als informatiepakket	6
3.2	Fasen in de overdracht.....	6
3.2.1	Vorbereidende fase	6
3.2.2	Formele fase	8
3.2.3	Overdrachtsfase	8
3.2.4	Afrondingsfase.....	9
3.3	Verwijderen van de informatieobjecten bij de archiefvormer	9
4	Eisen aan metadata	10
4.1	Waarom metadata?	10
4.2	Aggregatieniveau's	10
5	Eisen aan de structuur van de export.....	12
5.1	ToPX 2.3.....	12
5.2	Naamgeving mappen en bestanden	13
6	Opslagformaten.....	14
6.1	Voorkeurs- en acceptabele formaten	14

1. Inleiding

Dit overdrachtsprotocol beschrijft welke stappen doorlopen worden indien een archiefvormer digitaal archief aanbiedt ter opname in het e-depot van de archiefbeheerder. Ook worden de voorwaarden beschreven waaraan de in het e-depot op te nemen informatieobjecten moeten voldoen. Voor een visuele toelichting op het e-depot zie de volgende animatiefilm van het Nationaal Archief: https://youtu.be/vgM-CAJd_co

Als invalshoek voor de beschrijving van de processtappen voor overdracht van informatieobjecten is het OAIS-referentiemodel gebruikt. In dit model heeft de producer de rol als archiefvormer welke het e-depot voorziet van informatieobjecten in de vorm van een SIP (Submission Information Packages). De SIP is een pakketje dat bestaat uit informatieobjecten met bijbehorende inhoudelijke en technische metadata. Wanneer de SIP in het e-depot is opgenomen, wordt het getransformeerd naar een AIP (Archival Information Package). De consumer heeft de rol als burger of ambtenaar welke informatie opneemt met behulp van de zoekfunctionaliteit. Deze aangeboden informatie is een kopie van de AIP in de vorm van een DIP (Dissemination Information Package). Een DIP bestaat uit informatie in de AIP. Het management heeft de rol als e-depot beheerder welke uitgevoerd wordt door het WBA.



Informatie kan op verschillende manieren overgebracht worden naar het e-depot. In hoofdstuk 2 worden de verschillende vormen van overbrenging toegelicht. Het proces voor overdracht van informatieobjecten zal volgens de vastgestelde procedure doorlopen worden welke beschreven staat in hoofdstuk 3. Voordat een overdracht door het e-depot geaccepteerd wordt moet worden voldaan aan een aantal vastgestelde eisen waaronder aan de metadatering (hoofdstuk 4), structuur van de export (hoofdstuk 5) en de opslagformaten (hoofdstuk 6).

2. Overbrenging

Er zijn twee scenario's met betrekking tot het overdragen van archieven naar het e depot, namelijk overbrenging en uitplaatsing.

Overbrenging

Artikel 12 van de Archiefwet bepaalt dat de zorgdrager de archiefbescheiden die niet voor vernietiging in aanmerking komen en ouder zijn dan twintig jaar overbrengt naar een archiefbewaarplaats. Deze termijn behoeft aanmerkelijke verkorting gelet op de sterk toenemende behoefte aan toegankelijkheid van digitaal gevormde informatie binnen de maatschappij, zowel op internet, social media als in het kader van de Wet openbaarheid van bestuur (WOB). Het is echter toegestaan archiefbescheiden al eerder over te brengen naar een archiefbewaarplaats: vervroegde overbrenging.

Dit heeft een aantal voordelen, zowel voor de archiefvormende organisatie als voor de burger:

- De digitale informatie valt volledig onder het beheer en de zorg van de archiefbeheerder, waarmee de archiefvormer wordt ontzorgd;
- De informatie wordt (behoudens eventuele uitzonderingen) geheel openbaar en daarmee toegankelijk voor de burger, wat de transparantie van de bedrijfsvoering van de gemeente bevordert.
- De duurzame bewaring en duurzame toegankelijkheid zijn bij overbrenging beter gewaarborgd.

Uitplaatsing

Uitgeplaatste archieven zijn archiefbescheiden van een overheidsorgaan, waarvan het beheer in afwachting van overbrenging of vernietiging tijdelijk aan een andere partij wordt opgedragen. Uitgeplaatste archieven bestaan uit afgesloten dossiers waarvan de informatieobjecten zijn opgeslagen in het e-Depot. In het geval van de zorgdragers betekent dit zij de zorg voor de duurzaamheid en de beschikbaarheid van die informatie opdragen aan het WBA. Het WBA gebruikt het e-Depot en de eigen beheersystemen om deze archieven te beheren, met het oog op duurzame toegankelijkheid. Het WBA voert de ingest, preservation en datamanagement activiteiten op de archieven uit.

Op dit moment wordt deze dienst nog niet ondersteund door het WBA. Voor het afnemen van deze dienst moeten er duidelijke afspraken gemaakt worden tussen archiefvormer en archiefbeheerder. Wanneer een archiefvormer deze optie wil benutten zal er samen met de archiefbeheerder een apart traject worden opgestart.

2.1 Te maken keuzes en afspraken

Bij de overbrenging van informatieobjecten van een archiefvormer naar het e-depot van een archiefbeheerder moet een aantal afwegingen, keuzes en afspraken gemaakt worden:

- Na welke termijn worden informatieobjecten overgebracht? Met het oog op duurzaam beheer van digitaal gevormde archieven stelt het WBA voor om uiterlijk binnen vijf jaar na afsluiting van een zaak over te gaan tot overbrenging. De archiefvormer hoeft de informatieobjecten dan nog maar weinig te raadplegen, derden wellicht vaker.

- Is vastgelegd of stukken op grond van de wetgeving met betrekking tot privacy niet openbaar gemaakt mogen worden of op grond van de Auteurswet niet verveelvoudigd mogen worden?
- Een belangrijke afspraak tussen zorgdrager en archiefbeheerder betreft een exit-strategie. Kan de archiefvormer de informatieobjecten die in het e-depot zijn gezet hier ooit weer uithalen, bijvoorbeeld bij uittreding uit de gemeenschappelijke regeling WBA? Welke voorwaarden en bijvoorbeeld formaten moeten hierbij worden afgesproken?

Wanneer de overbrenging is afgerond dient de archiefvormer een 'verklaring van overbrenging' op te stellen, die door de archiefvormer en de archiefbeheerder wordt ondertekend.

2.2 Feitelijke overdracht

De term 'overbrenging' moet niet worden verward met 'overdracht'. Het laatste begrip verwijst naar de feitelijke verplaatsing, waarvan zowel bij uitplaatsing als bij overbrenging sprake kan zijn. Niet zelden worden deze termen door elkaar gehaald. Pas als een formele verklaring van overbrenging is opgemaakt is er sprake van overbrenging in de zin van art. 12 Archiefwet. Kortom eerst vindt overdracht van het digitale archief plaats. Indien deze geaccepteerd is en wordt opgenomen in het e-depot dan is deze formeel overgebracht (overbrenging).

3. Procedure overdracht informatieobjecten

De navolgende procedure voor opname van informatieobjecten in het e-depot is globaal gebaseerd op de ISO-norm 20652 (Producer-Archive Interface Methodology). Deze norm beschrijft de eerste stappen bij het opnemen van gegevens in een e-depot en is daarmee een uitwerking van het zogenaamde ingest-proces (het binnenhalen van gegevens) uit het OAIS model (Open Archival Information System Reference Model).

3.1 Aanlevering als informatiepakket

De archiefvormer levert aan de archiefdienst documenten en metadata aan. Deze worden aangeleverd als één pakket, dit wordt een Submission Information Package (SIP) genoemd. Uiteindelijk moet de metadata conform het Metadatamodel WBA worden aangeleverd en de documenten voldoen aan de eisen van aan te leveren bestandsformaten. Voor de eisen aan de metadata zie hoofdstuk 4 en voor de eisen aan de bestandsformaten van de documenten, zie hoofdstuk 6.

3.2 Fasen in de overdracht

Er worden vier fasen onderscheiden in de interactie tussen archiefvormer en de archiefbeheerder:

- de voorbereidende fase
- de formele fase
- de overdrachtsfase
- de afrondingsfase

3.2.1 Voorbereidende fase

In deze fase worden de afspraken gemaakt en verantwoordelijkheden tussen archiefvormer en archiefbeheerder bepaald en wordt duidelijk welke informatieobjecten worden overgedragen.

Deze fase bestaat uit de volgende activiteiten:

1. Het vastleggen van de contactpersonen zowel aan de kant van de archiefbeheerder als bij de archiefvormer. Voor het goed organiseren van een overbrenging naar het e-depot moeten volgende rollen ingevuld worden:

Bestuurder	Om de visie op duurzame toegankelijkheid en de bijbehorende veranderingen in de organisatie te sponsoren. Daarnaast in de rol als budgethouder om projecten rond digitaal archiveren te faciliteren.
Manager	Nieuwe doelstellingen op het gebied van duurzame digitale informatie vraagt de inzet van mensen en samenwerking door de organisatie heen. Commitment van de verantwoordelijke managers (afdelingshoofden) is een randvoorwaarde om daadwerkelijk stappen te kunnen zetten.
IT specialist	Het e-Depot moet een plek krijgen in de informatie-architectuur. Het is één van de plaatsen waar informatie in

	opgeslagen en uit ontsloten wordt. Het exporteren uit applicaties zal ook ingeregeld moeten worden.
Informatieadviseur	De informatieadviseurs van de organisatie weten als geen ander waar zich welke informatie bevindt, in welke staat het verkeerd, en of- en wanneer het overbracht zou moeten worden naar het archief. Vaak hebben zij ook een coördinerende rol in het (opnieuw) inrichten van informatiesystemen en processen. Ze vormen de spil tussen de andere interne stakeholders en kunnen als aanspreekpunt voor het WBA optreden.
Vakspecialist (Proceseigenaren)	Inhoudelijke kennis van bepaalde informatie, de beschrijving ervan en de toepassingen liggen voornamelijk bij de experts van het vakgebied zelf. Tegelijkertijd zijn zij ook de makers en bewerkers van die informatie. Dit is een essentiële doelgroep om te betrekken in het duurzaam toegankelijk houden van informatie en in het proces van digitale archivering.
Archivaris	De archivaris vervult een kernrol in het digitaal archiveren.
Projectleider	Overbrenging naar het e-depot vraagt om een projectmatige aansturing. De projectleider is hiervoor verantwoordelijk.

2. Informatie verzamelen over het over te dragen archief waaronder:

- Een beschrijving van de over te brengen informatieobjecten (titel, periode waarin het materiaal is ontstaan, etc.)
- De structuur van het archief: gestructureerd informatieobjecten met gekoppelde metadata
- De applicatie waarin de informatieobjecten zijn opgeslagen
- De wijze van export van de applicatie
- Informatie over eventuele migratie, vernietiging of compressie van informatieobjecten in het archief
- Een overzicht van de metadata (zowel de handmatig als de automatisch toegekende)
- De aanwezigheid van versleutelde bestanden
- De aanwezigheid van digitale handtekeningen
- Controle van de informatieobjecten op virussen
- De omvang van de informatieobjecten in Megabytes, Gigabytes of Terabytes
- De aanwezige bestandsformaten
- Eventuele openbaarheidsbeperkingen
- Eventueel aanwezige auteursrechten
- De manier waarop de informatieobjecten zijn verpakt (een voorbeeld-SIP)
- De manier waarop het archief wordt overgedragen via een beveiligde verbinding, zoals HTTPS

3. Planning maken voor het overdragen van het archief

4. Informatie vastleggen om te komen tot een aansluitplan

3.2.2 Formele fase

Vervolgens worden de resultaten van de voorbereidende fasen samen met de actoren en de planning van de overdracht vastgelegd in het aansluitplan. Dit aansluitplan wordt door zowel de archiefvormer als de archiefbeheerder geaccordeerd.

3.2.3 Overdrachtsfase

Deze fase omvat zowel een test van de overdracht als de daadwerkelijke overdracht van documenten naar het e-depot.

1. Uitvoeren overdrachtstest

Het uitvoeren van een overdrachtstest is conform de ISO-norm 20652. Wanneer al eerder een overdracht uit de betreffende applicatie bij dezelfde archiefvormer heeft plaatsgevonden kan deze fase mogelijk worden overgeslagen. Ten behoeve van de test dient een representatieve selectie gemaakt te worden uit het archiefmateriaal. Deze selectie moet aan de volgende criteria voldoen:

- Alle over te dragen dossiers dienen afgesloten te zijn
- Alle bestandsformaten welke aanwezig zijn dienen terug te komen in de selectie;
- De verschillende aggregatieniveaus dienen vertegenwoordigd te zijn in de selectie;
- De metadatering op verschillende niveaus van het archief moet terugkomen in de selectie.

De selectie wordt overgebracht naar de quarantaine-ruimte van het e-depot. In dit deel van het e-depot wordt vervolgens gecontroleerd of de testbestanden goed zijn overgekomen. Dit bestaat uit een controle van de goede staat en een controle van de geordende en toegankelijke staat. De controle van de goede staat vindt het eerst plaats en bestaat uit de volgende punten:

- Controle op virussen;
- Controle van de integriteit/validiteit, door te controleren of de bestanden ongewijzigd (d.m.v. hashing) en volledig in het e-depot zijn gekomen. Is het bestandsformaat conform de eisen?
- Controle op bruikbaarheid: zijn alle bestandsformaten door het e-depot geaccepteerd, zijn alle bestanden valide, zijn er versleutelde bestanden aanwezig (versleuteling verwijderen), zijn er gecomprimeerde mappen met bestanden aanwezig (bestanden uitpakken)?
- Controle op authenticiteit en betrouwbaarheid: zijn digitale handtekeningen inclusief certificaat aanwezig?

Vervolgens vindt een controle plaats van de geordende en toegankelijke staat:

- Klopt de ordening van de bestanden en de aggregatieniveaus aangegeven in de metadata?
- Zijn de minimaal vereiste metadata aanwezig?

Mocht er iets mis gaan, dan moet bepaald worden waar dat aan ligt en hoe dit kan worden opgelost (voor het gehele archief) alvorens door te gaan naar de volgende stap.

2. Uitvoeren overdracht gehele digitale archief

3.2.4 Afrondingsfase

De afrondingsfase omvat de technische en inhoudelijke kwaliteitscontrole van de overgedragen informatieobjecten. Anders dan in de overdrachtsfase gaat het nu om een controle van het geheel, niet van een selectie. De archiefvormer wordt ingelicht zodra het archief succesvol is opgenomen in het e-depot. Na een positieve validatie wordt er een verklaring van overbrenging opgesteld en ondertekend waarmee de overbrenging een feit is.

3.3 Verwijderen van de informatieobjecten bij de archiefvormer

Na overbrenging, d.w.z. volledige overname door de archiefbeheerder van de verantwoordelijkheid voor het beheer en de zorg, zorgt de archiefvormer voor het verwijderen van de relevante informatieobjecten uit het archiveringssysteem. Indien het archiveringssysteem hiertoe geen mogelijkheid biedt, is er nader overleg nodig over de mogelijke oplossingen.

4 Eisen aan metadata

Metadata zijn gegevens over gegevens, data over data. Metadata zijn gegevens die context, inhoud, structuur en vorm van informatie en het beheer ervan door de tijd heen beschrijven. Metadata ondersteunen daarbij het bedrijfsproces en de processen voor informatie- en archiefbeheer. Overheidsorganisaties zijn op grond van de Archiefregeling verplicht een metadataschema op te stellen als bedoeld in NEN-ISO 23081, waarin ze aangeven welke metadata voor de eigen organisatie minimaal nodig zijn en hoe deze worden vastgelegd. Om de informatieobjecten uit een export van het DMS juist in het e-depot op te kunnen nemen, moet naast een correcte metadatering ook de structuur hiervan volgens vastgestelde regels aangeleverd worden.

4.1 Waarom metadata?

Het vastleggen van metadata is essentieel voor het vaststellen van de betrouwbaarheid van informatie en maakt daarnaast het beheren, ontsluiten en raadplegen ervan mogelijk. Metadata zijn van groot belang voor interpretatie en uitwisseling van informatie. Zo kan in het geval van digitale informatie onder meer worden vastgelegd wat, wanneer en door wie in welk systeem gewijzigd of geraadpleegd is. Hiermee wordt de kwaliteit van de digitale informatieobjecten door de tijd heen gewaarborgd. Tot slot zijn metadata een randvoorwaarde voor de duurzame bewaring van de informatie in een e-depot.

Metadata worden verdeeld in drie hoofdcategorieën:

- Administratieve metadata (beheer), met als doel het beheren van informatieobjecten. Beheer omvat informatie over eigendom, vindplaatsgegevens, formele herkomst en verantwoording van beheersactiviteiten;
- Beschrijvende metadata (herkomst / context), met als doel het identificeren en beschrijven van informatie- objecten. Herkomst / context omvat informatie over identificatie en interpretatie;
- Technische metadata, met als doel het weergeven van technische informatie van een informatieobject. Dit omvat onder andere het bestandsformaat, de bestandsgrootte en de applicatie waarmee het bestand is gemaakt.

Gedetailleerde informatie over het gebruik van metadata is te vinden in het Metadatamodel WBA.

4.2 Aggregatieniveau's

Metadata zijn primair bedoeld om betrouwbaar hergebruik van digitale overheidsinformatie, verschillend aggregatieniveau, mogelijk te maken. We onderscheiden vier niveaus, te weten:

1. Archief
2. Dossier
3. Archiefstuk
4. Bestand

De vier niveaus kennen een hiërarchische structuur. Metadata die worden vastgelegd op het hoogste niveau worden overgeërfd tot op het laagste aggregatieniveau. Andersom werkt dat niet.

In dit overdrachtsprotocol beperken we ons tot de metadata die de archiefbeheerder nodig heeft om informatieobjecten voor de lange termijn te kunnen beheren. Het gaat om een set aan metadata voor het duurzame beheer en de toegankelijkheid van (wettelijk) overgebrachte digitale informatie en om verantwoording af te leggen over de integriteit hiervan. Deze metadata worden uiteindelijk opgeslagen en beheerd in het archiefbeheersysteem van de archiefbeheerder.

In onderstaand schema zijn de voor de archiefbeheerder relevante metadata benoemd als verplicht of verplicht indien van toepassing. De nummering en globale omschrijving zijn ontleend aan het TMLO:

Aggregatieniveau	Overerving	Verplicht	Verplicht indien aanwezig
Archief		2 Identificatiekenmerk 3 Aggregatieniveau 4 Naam	
Dossier	Archief	2 Identificatiekenmerk 3 Aggregatieniveau 4 Naam	5 Classificatie 6 Omschrijving 7 Plaats 9 Dekking 10 Externe identificatiekenmerk 12 Event geschiedenis 13 Event plan 15 Relatie 18 Openbaarheid 20 Integriteit
Archiefstuk	Archief, Dossier	2 Identificatiekenmerk 3 Aggregatieniveau 4 Naam 19 Vorm	6 Omschrijving 7 Plaats 9 Dekking 10 Externe identificatiekenmerk 11 Taal 12 Event geschiedenis 13 Event plan 15 Relatie 16 Gebruikersrechten 18 Openbaarheid
Bestand		21 Formaat	

De archiefvormer dient tenminste de zowel verplichte als verplicht indien aanwezige metadata aan te leveren bij de overbrenging van informatieobjecten. Indien blijkt dat metadata niet beschikbaar is in de beherende applicatie wordt in overleg met het WBA naar een oplossing gezocht. Dit kan een bewerking achteraf zijn of een afwijkende mapping van data.

5 Eisen aan de structuur van de export

De structuur van een export moet naast de metadatering ook voldoen aan een bepaalde indeling van mappen en bestanden. De content wordt aangeleverd in een mappenstructuur waarbij elke archief, serie, dossier en archiefstuk een map is en elk document een bestand. Elke map en bestand heeft hiernaast een eigen metadatabestand in de vorm van een ToPX-XML. Het metadatabestand van een map staat in deze map en het metadatabestand van een bestand staat naast het bestand.

De mappenstructuur wordt aangeleverd naar indeling van de vijf aggregatieniveaus:

Niveau	Aggregatietype	Aantal	Verplicht
1	Archief	Eén (1)	Verplicht
2	Dossier	Eén of meer (1...<)	Verplicht
3	Archiefstuk	Eén of meer (1...<)	Verplicht
4	Bestand	Eén of meer (1...<)	Verplicht

Het hoogste aggregatieniveau is van het type **archief**. Dan volgt er ten minste één aggregatieniveau van het type **dossier**. De betekenis van het type dossier kan in een DMS ook wel geïnterpreteerd worden als zaak. Een dossier kan deel uitmaken van een ander dossier of archief. Onder een dossier volgen één of meerdere niveaus van het type **archiefstuk**. Een archiefstuk hoeft geen deel uit te maken van een dossier en maakt dan deel uit van een archief. Ook kan een archiefstuk deel uit maken van meerdere dossiers. Een archiefstuk kan ook wel geïnterpreteerd worden als document. Zowel dossiers onderling als archiefstukken onderling kunnen aan elkaar gerelateerd worden om verbanden vast te kunnen leggen tussen deze niveaus. Het laagste niveau is van het type **bestand**. In een digitale omgeving is een archiefstuk opgebouwd uit één of meer bestanden waarin de inhoud van het archiefstuk is vastgelegd. Technische eigenschappen van een archiefstuk worden dan vastgelegd op het niveau van bestand. Een bestand kan ook wel geïnterpreteerd worden als record.

5.1 ToPX 2.3

Om het toepassingsprofiel TMLO leesbaar te maken voor het e-depot moet deze worden vertaald naar een metadatabestand in de vorm van een XML. Deze vertaling van metagegevens kan het e-depot verwerken. Het e-depot controleert automatisch de structuur en aanwezigheid van de verplichte metagegevens aan de hand van dit XML toepassingsprofiel. Voor het e-depot wordt gebruik gemaakt van het XML formaat ToPX (Toepassingsprofiel XML). De laatste versie is ToPX 2.3. Dit ToPX bestand wordt gevuld met de metadata zoals vastgelegd in de mapping waarin de vertaalslag is gemaakt van de metadata elementen in het DMS naar de ToPX metadata elementen.

De voorwaarden waaraan de ToPX metadata elementen moeten voldoen staan beschreven in een xsd schema. De controle met behulp van de xsd is één van de controles die het e-depot doet om te bepalen of de export in het e-depot mag worden ingelezen. Op de website van het Nationaal Archief zijn de ToPX bestanden te vinden om te controleren of een export voldoet aan de ToPX eisen:

<http://www.nationaalarchief.nl/digitaal-archiveren/informatie-archiveren-het-e-depot/metadata-het-e-depot>

5.2 Naamgeving mappen en bestanden

De export kan worden aangeleverd in een container van een tar-formaat. De container mag geen mappen en beperking aan de padnaamlengte stellen. Hiervoor kan de tool 7-Zip gebruikt worden.

De bestandsnamen dienen aan de volgende technische eisen te voldoen:

- Bestands- en mapnaam mag geen van deze karakters bevatten:
<>:"\|/?@#\$\$%^&*()[]{}
- Bestands- en mapnaam mag niet identiek zijn aan:
CON, PRN, AUX, NUL, COM1(t/m 9), LPT1(t/m 9)
- Let op: het e-depot is hoofdlettergevoelig, de bestands- en mapnamen moeten identiek zijn met de metadata in ToPX

6 Opslagformaten

Het e-depot kan technisch gezien een grote hoeveelheid aan diverse bestandsformaten opnemen. Echter met het oog op digitale duurzaamheid speelt de vraag in welk opslagformaat dit dient te gebeuren. De archiefregeling is hierover in artikel 26 duidelijk: digitale archiefbescheiden worden, uiterlijk op het tijdstip van overbrenging, opgeslagen in een valideerbaar en volledig gedocumenteerd bestandsformaat dat voldoet aan een open standaard, tenzij dit redelijkerwijs niet van de zorgdrager kan worden verlangd. Gebruikmaking van compressietechniek is slechts toegestaan voor zover daarbij niet zodanig verlies van informatie optreedt dat niet langer aan de bij deze regeling gestelde eisen ten aanzien van de toegankelijke en geordende staat van digitale archiefbescheiden kan worden voldaan.

In de archiefregeling zijn geen specifieke bestandsformaten benoemd maar wel kwaliteitseisen als valideerbaar en volledig gedocumenteerd bestandsformaat dat voldoet aan een open standaard. Een grote diversiteit aan formaten vraagt een grote inspanning om deze te preserven en beschikbaar te stellen. Terwijl open standaarden zich makkelijker laten preserven door een duidelijke documentatie en het beheer door een non-profit organisatie.

De archiefvormer zal bij overdracht van informatieobjecten moeten voldoen aan het volgende:

Er worden zoveel mogelijk opslagformaten gebruikt die voldoen aan open standaarden en genoemd worden op: <https://www.forumstandaardisatie.nl/open-standaarden/lijsten-met-open-standaarden>. Open bestandsformaten zijn ook onderdeel van de DUTO-kwaliteitseisen, ontwikkeld door het Nationaal Archief, waarmee overheidsorganen kunnen bepalen welke maatregelen ze moeten nemen om de digitale informatie die ze ontvangen en creëren, duurzaam toegankelijk te maken.

Belangrijk is hierbij dat documenten vindbaar, toegankelijk en leesbaar blijven. Omzetten van een Excel bestand met veel kolommen naar een PDF komt de leesbaarheid en het gedrag ervan bijvoorbeeld niet ten goede. In zulke gevallen dient het originele bestand naar het e-depot overgebracht te worden.

6.1 Voorkeurs- en acceptabele formaten

Om de digitale duurzaamheid zo veel mogelijk te kunnen waarborgen hanteert het WBA twee categorieën formaten:

- Voorkeursformaat; dit zijn open bestandsformaten en bestandstypes welke veelvuldig worden gebruikt en voldoende gedocumenteerd zijn, zogenaamde 'industry standards'.
- Acceptabel formaat; dit zijn niet volledig open bestandsformaten maar wel acceptabel voor het e-depot. Het betreft hier documenten welke niet zonder meer geconverteerd kunnen worden naar een open standaard zonder verlies van data.

Indien de informatieobjecten niet voldoen aan het voorkeursformaat kan worden gekozen om het acceptabele formaat op te nemen. Indien een bestandsformaat afwijkt van beide bovenstaande categorieën zal het WBA preservering advies uitbrengen voor een passende oplossing. Onderstaande overzichten zijn ontleend aan het rapport 'Voorkeursformaten' van het Nationaal Archief waarbij het WBA zich aansluit.

Voorkeurs en acceptabele bestandsformaten

Informatieobject	Voorkeur formaat	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

Voorkeur formaat

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. Het Broadcast Wave Format is een uitbreiding op WAV met een aantal extra metadatavelden.
	BWF	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 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)
Document	PDF/A-1 PDF/A-2	PDF/A is een wijdverbreide open standaard, een NEN/ISO-norm (ISO:19005).
	ODT	ODT behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC 26300:2007), dit type wordt gebruikt als standaard voor office-documenten.
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, apparaatuuronafhankelijk, besturingssysteemafhankelijk, 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 uitsluitend voor het "gebruik van grafische afbeeldingen (met 'lossless' compressie) binnen ODF-documenten."

Presentation	ODP	ODP behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC26300:2007) dat als standaard 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/A-1 en PDF/A-2 zijn opgenomen op de 'pas-toe-of-leg-uit' lijst met open standaarden van het Forum Standaardisatie. 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.
Spreadsheet	ODS	ODS behoort tot de OpenDocument standaard (ODF, NEN-ISO/IEC26300:2007) 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 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/A-1 en PDF/A-2 zijn opgenomen op de 'pas-toe-of-leg-uit' lijst met open standaarden van het Forum Standaardisatie. 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.
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 de lijst met 'gangbare' standaarden van het Forum Standaardisatie.
Video	MXF	MXF (Material eXchange Format) is een open standaard die wordt onderhouden door de AV-standaardenorganisatie Society of Moving Pictures Engineers (SMPTE). Het formaat is bestemd voor professioneel gebruik 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 BASIS van DEN.

Acceptabele formaten

Audio	MP3	MPEG1-Layer 3 is de populairste wijze van geluidsopslag van dit moment. Met het formaat worden lossy compressies 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.
Document	PDF 1.7	PDF 1.7 wordt gebruikt 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.
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.

	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 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 kleurencoderingen.
Presentation	PPT PPTX	Microsoft Office Powerpoint is marktleider op het gebied van spreadsheets.
Spreadsheet	XLS XLSX	Microsoft Office Excel is marktleider op het gebied van spreadsheets.
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 codecs bevatten. MKV heeft echter als nadeel dat het nog een relatief geringe ondersteuning door videosoftware kent.