

Aanlevervoorwaarden voor e-depot



Versie: 2.0
Datum: 26-10-2021

Versiebeheer

Versie	Opmerkingen	Datum
0.1-0.3	Conceptversie afgestemd intern Erfgoed Leiden en Omstreken	06-03-2019
0.9	Conceptversie te gebruiken bij testopnames	20-05-2019
1.0	Vastgesteld door directeur Erfgoed Leiden en Omstreken	12-12-2019
2.0	Vastgesteld door directeur Erfgoed Leiden en Omstreken	26-10-2021

Afkortingen en begrippen

Afkorting of begrip	
Archiefvormer	Persoon, groep personen of organisatie die zelfstandige archiefvorming als een van zijn of haar activiteiten heeft.
Digitaal archief	Een samenhangende verzameling van digitale bestanden en metadata, bijvoorbeeld bouwdossiers of ruimtelijke plannen.
E-depot	Het geheel van organisatie, beleid, processen en procedures, financieel beheer, personeel, databeheer, databeveiliging en aanwezige hard- en software, dat duurzaam beheren en raadplegen van te bewaren digitale archiefbescheiden mogelijk maakt.
Erfgoed Leiden	Erfgoed Leiden en Omstreken (Erfgoed Leiden) voert erfgoeddiensten uit en/of beheert historische collecties voor de aangesloten gemeenten, de Omgevingsdienst West-Holland, de RDOG Hollands Midden en de Veiligheidsregio Hollands Midden.
Ingest	Het importeren van het digitale archief in het e-depot.
Metadata	Metadata zijn gegevens die de context, inhoud, structuur en vorm van informatie en het beheer ervan door de tijd heen beschrijven.
Preservica	Leverancier van de e-depot (SAAS) oplossing van Erfgoed Leiden
OPEX	Open Exchange Format (OPEX) is een uitwisselingsformaat in gebruik door Preservica. Digitale archieven die worden opgenomen in het e-depot dienen metadata aan te bieden in het OPEX formaat om een soepele ingest mogelijk te maken.
MDTO	Metadatagegevens voor duurzaam toegankelijke overheidsinformatie (MDTO, opvolger van het TMLO) is een landelijke standaard, waarin is vastgelegd welke set van kenmerken (metadata) lokale en rijksoverheden tenminste aan informatieobjecten moeten meegeven, zodat informatie in het e-depot straks makkelijk te vinden en eenduidig te interpreteren is. De MDTO-standaard voorziet informatie op eenzelfde manier van metadata, met als doel om informatieprocessen en -systemen op elkaar aan te laten sluiten. Het draagt bij aan goede vindbaarheid en duurzame toegankelijkheid.

Inhoud

Versiebeheer	2
Afkorting en begrippen	2
Inleiding	4
1. Overbrengen of uitplaatsen	5
2. Voorwaarden	6
2.1 Goede, geordende en toegankelijke staat	6
2.2 MDTO en OPEX mapping	6
2.3 MDTO-XML schema en OPEX metadata versie 1.0	7
2.4 Sectorspecifieke metadata	7
2.5 Sidecar structuur	7
2.6 Bestandsformaten	8
2.7 Lossless compressie	10
2.8 Fixity en aantallen controle	10
2.9 Wijze van transport	10
3. Overzicht van bijlagen	11
Bijlage 1 - Guide to OPEX ingest en Monitoring	11
Bijlage 2 - OPEX xsd	11

Inleiding

Dit document is bedoeld om een overzicht te geven van de voorwaarden waaraan over te brengen of uit te plaatsen digitale archief naar het e-depot van Erfgoed Leiden moet voldoen. Het is vooral bedoeld voor de leverancier van de bronapplicatie of technisch specialist van de archiefvormende organisatie.

Erfgoed Leiden gebruikt de SAAS oplossing van Preservica als e-depot.

1. Overbrengen of uitplaatsen

Digitaal **overheidsarchief** kan op verschillende manieren aan Erfgoed Leiden worden aangeleverd:

Overbrenging

Te bewaren afgehandeld/afgesloten digitaal archief (geen wijzigingen na aanlevering) wordt overgebracht na 10 jaar (nu nog 20 jaar, naar verwachting in 2022-2023 wordt dit 10 jaar). Erfgoed Leiden wordt hierna formeel de beheerder van het archief.

Vervroegde overbrenging

Te bewaren afgehandeld/afgesloten digitaal archief wordt overbracht vóór de termijn van 10 jaar (nu nog 20 jaar, naar verwachting in 2022-2023 wordt dit 10 jaar). Erfgoed Leiden wordt bij vervroegde overbrenging formeel de beheerder van het archief.

Uitplaatsing

Te bewaren en te vernietigen digitaal archief wordt overgedragen. Een groot gedeelte van dit archief is nog niet openbaar. Na verloop van tijd zal het digitaal archief óf (gedeeltelijk) overgebracht óf (gedeeltelijk) vernietigd moeten worden. De archiefvormer blijft de beheerder van het archief, totdat het te bewaren gedeelte formeel naar Erfgoed Leiden wordt overgebracht. Dat gedeelte komt dan in het beheer van Erfgoed Leiden.

Voor alle opties hierboven, waarbij er na de initiële overbrenging vervolgoverbrengingen volgen, geldt dat in overleg wordt besloten op welke termijn deze vervolg overbrengingen zullen plaatsvinden (1 keer per jaar, 1 keer per paar jaar etc.).

Particulier archief kan ten alle tijden (in overleg) overgebracht worden. Hiervoor geldt ook dat het over te brengen digitaal archief geschoond en afgesloten moet zijn.

2. Voorwaarden

2.1 Goede, geordende en toegankelijke staat

Het over te brengen digitaal archief dient aan de normen van de goede, geordende en toegankelijke staat (Archiefregeling 2010) te voldoen.

De goede staat betekent dat het digitaal archief o.a. virus-vrij, zonder corrupte bestanden en zonder encryptie (of met de encryptiesleutel) aangeleverd wordt.

De geordende en toegankelijke staat betreft de ordening en toegankelijkheid van het digitaal archief.

Mappen- en bestandsnamen

Algemeen geldt:

- Beschrijf mapnamen en bestandsnamen betekenisvol en goed
- Kies een vaste volgorde voor de opbouw van de bestandsnamen
- Gebruik de originele extensie en gebruik die termen nooit in de naam
- Schrijf de datum volgens het stramien JJJJMMDD
- De tekens < > : " / \ ? * # & ~ % { } mogen niet in de bestandsnaam voorkomen.
- De letter- en tekencombinaties CON, PRN, AUX, NUL, COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8, COM9, LTP1, LTP2, LTP3, LTP4, LTP5, LTP6, LTP7, LTP8, LTP9, .lock, _vti_ en desktop.ini mogen niet in bestandsnamen voorkomen.
- Bestands- en padnamen mogen niet langer zijn dan 255 tekens.

2.2 MDTO en OPEX mapping

Erfgoed Leiden gebruikt het MDTO (Metagegevens voor Duurzaam Toegankelijke Overheidsinformatie, opvolger van het TMLO) en OPEX (Open Exchange Format) om metadata te importeren in het e-depot.

Voordat digitaal archief kan worden overgebracht naar het e-depot, vindt er een mapping plaats tussen het MDTO en de metadatavelden van het archiefbescheiden. Een mapping verschilt per digitaal archief. Naast de MDTO velden, kan er ook metadata worden meegenomen die niet in het MDTO staan, maar specifiek zijn voor het digitaal archief. In de OPEX metadata worden gegevens als de security tag, fixity en ID opgenomen die benodigd zijn voor een juiste ingest in het e-depot.

De MDTO en OPEX mapping inclusief de specifieke metadata voor een digitaal archiefvormen samen het metadata-schema voor de digitaal archief.

2.3 MDTO-XML schema en OPEX metadata versie 1.0

Als het metadataschema in 2.2 is samengesteld, zet de archiefvormer (en als nodig de leverancier) deze om naar het MDTO-XML schema én plaatst deze MDTO metadata in het XML document tussen de OPEX metadata zodat de velden ingelezen kunnen worden in het e-depot.

De links hieronder van het Nationaal Archief geven meer informatie:

MDTO XML schema specificatie, versie informatie en xsd:

<https://www.nationaalarchief.nl/archiveren/mdto>

XML voorbeelden per aggregatieniveau:

<https://www.nationaalarchief.nl/archiveren/mdto/xml-schema>

Meer informatie over OPEX is te vinden in bijlage 1 *Guide to OPEX ingest en Monitoring*. In bijlage 2 is de OPEX xsd te vinden.

De aangeleverde XML bestanden dienen gevalideerd te zijn aan zowel de OPEX als MDTO xsd's en met UTF-8 (8-bit Unicode Transformation Format) gecodeerd te zijn.

2.4 Sectorspecifieke metadata

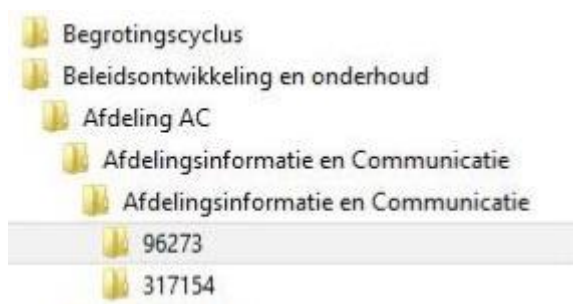
Het kan voorkomen dat niet alle benodigde metadata in het MDTO in te passen zijn. Dit gaat om sectorspecifieke metadata die het generieke karakter van het MDTO overstijgen. In overleg met Erfgoed Leiden dienen deze metadata in een separaat metadataschema geplaatst te worden. Ook deze metadata moeten in hetzelfde XML bestand als de OPEX en MDTO metadata geplaatst worden (binnen het element OPEX.descriptiveMetadata).

2.5 Sidecar structuur

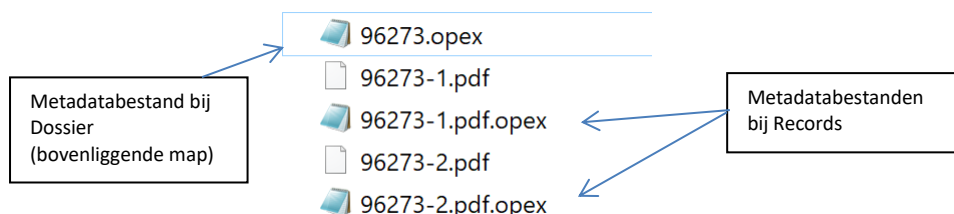
Digitaal archief dient door de archiefvormer aangeleverd te worden in een sidecar structuur.

In een sidecar structuur zit de content van het digitaal archief in een directory-structuur: dat wil zeggen elk archief, serie, dossier (zaak), subdossier en document is een map en elke versie van een document is een bestand. Elke map (archief, serie, dossier, subdossier, document) en elk bestand (versie van een document) heeft zijn eigen metadatabestand, een zogenoemde sidecar. De naam van een metadatabestand is identiek aan die van de map of bestand waar het bij hoort, aangevuld met de extensie '.opex'. Let erop dat ook de extensie van het bestand opgenomen wordt in de bestandsnaam van het metadatabestand. Bijvoorbeeld 'bestand.docx.opex'. De plaats van een metadatabestand die bij een map hoort, is in de map (dus op een lager niveau). De plaats van een metadatabestand die bij een bestand hoort, is naast dit bestand (dus op hetzelfde niveau in een map).

In de Nederlandse archiefwereld worden veelal vijf types aggregatieniveau gebruikt. Het hoogste aggregatieniveau is *archief*. Daarna kunnen er één of meerdere aggregatieniveaus van het type *serie volgen*. Dan volgt er tenminste één aggregatieniveau van het type *dossier* (ook wel zaak). Dan volgen er één of meerdere aggregatieniveaus van het type *record*, hoewel dit aggregatieniveau niet in alle gevallen verplicht is. Hieronder een voorbeeld:



Het laagste aggregatieniveau is van het type *bestand*. Bij het type bestand gaat het om de technische onderdelen van een record (de bits and bites). Voor dit aggregatieniveau hoeven geen metadatabestanden aangeleverd te worden. Het e-depot genereert deze gegevens automatisch bij ingest. Hieronder als voorbeeld de inhoud van *dossier* ofwel map '96273' in het vorige plaatje met daarin bestanden en metadata:



Het record (mapje 96273) bevat dus:

- het metadatabestand bij het *dossier* '96273.opex'
- de *records* '96273-1.pdf' en '96273-2.pdf'
- de metadatabestanden bij deze twee *records*, te weten '96273-1.pdf.opex' respectievelijk '96273-2.pdf.opex'

Het systeem is hoofdlettergevoelig. Let er dus op dat de naam van zowel het record/de map als het bijbehorende metadatabestand op exact de wijze zijn gespeld inclusief upper en lower case.

2.6 Bestandsformaten

Erfgoed Leiden kan alle bestandsformaten die in het PRONOM register staan opslaan in het e-depot. Het preservingsbeleid leidt echter tot een kleinere set van voorkeursformaten en geaccepteerde formaten. Hierbij is van de volgende uitgangspunten gebruik gemaakt:

- Open bestandsformaten hebben de voorkeur. Dit zijn bestandsformaten die niet van een bepaalde software afhankelijk zijn.
- Duurzame bestandsformaten hebben de voorkeur. Dit zijn bestandsformaten waarvan we verwachten dat deze in de toekomst ook nog te openen zijn.
- Indien een bestandsformaat niet open is, maar wel de de facto standaard wordt deze in veel gevallen geaccepteerd. Hierbij is ook in acht genomen of het in Preservica mogelijk is het bestandsformaat te normaliseren of migreren naar een open bestandsformaat.
- Erfgoed Leiden accepteert geen bestanden die in een container (bijvoorbeeld zip) worden aangeleverd. De enige uitzondering hierop is het WARC bestandsformaat.
- Bestandsformaten die niet in het PRONOM register staan worden niet geaccepteerd.

- Indien een bestandsformaat als 'endangered' is bestempeld in de BitList van de Digital Preservation Coalition wordt deze niet opgenomen in het e-depot, tenzij hierop een normalisatie of migratie kan worden uitgevoerd.

Toepassingsgebied	Voorkeursformaat	Geaccepteerd formaat	Toegestane Pronom ID's
Afbeelding	TIFF, JPEG2000	PNG, JPEG, GIF, DNG	fmt/353 , fmt/153 , fmt/154 , x-fmt/392 , fmt/11 , fmt/12 , fmt/13 , fmt/41 , fmt/42 , fmt/43 , fmt/44 , fmt/3 , fmt/4 , fmt/436 , fmt/152 , fmt/730
Vector afbeelding	SVG		fmt/91 , fmt/92
Tekst	ODF (.odt en .ott), PDF/A-1, PDF/A-2	DOC, DOCX, PDF 1.0-1.7, PDF/X, TXT, EPUB	fmt/136 , fmt/290 , fmt/291 , fmt/39 , x-fmt/275 , x-fmt/274 , fmt/412 , fmt/37 , fmt/38 , fmt/95 , fmt/354 , fmt/476 , fmt/477 , fmt/14 , fmt/15 , fmt/16 , fmt/17 , fmt/18 , fmt/19 , fmt/20 , fmt/276 , fmt/478 , fmt/157 , fmt/158 , fmt/488 , x-fmt/111 , fmt/483
Audio	WAV, WAVE, RF64	MP3, MP1	fmt/1 , fmt/703 , fmt/2 , fmt/707 , fmt/704 , fmt/527 , fmt/712 , fmt/713 , fmt/951 , fmt/6 , fmt/141 , fmt/142 , fmt/143 , fmt/134 , fmt/347
Video	MXF	MKV, mp4 (MPEG-4 Part 10/H.264)	fmt/200 , fmt/786 , fmt/569 , fmt/199 met codec H.264 LIC ID fdd000081
Email	EML	MSG	fmt/278 , x-fmt/430
Website	WARC		fmt/1355 , fmt/1281
Spreadsheet	ODF (.ods, .ots), CSV	XLS, XLSX	fmt/137 , fmt/294 , fmt/295 , x-fmt/18 , fmt/55 , fmt/56 , fmt/57 , fmt/58 , fmt/59 , fmt/61 , fmt/214
Presentatie	ODF (.odp, otp)	PPT, PPTX, PUB, VSD	fmt/138 , fmt/292 , fmt/293 , fmt/126 , fmt/215 , x-fmt/257 , x-fmt/258 , fmt/443
Database	ODF (.odb), SQL, SIARD	ACCDB, MDB	fmt/140 , fmt/444 , fmt/424 , fmt/206 , fmt/161 , fmt/995 , fmt/1196 , fmt/275 , x-fmt/66 , x-fmt/240 , x-fmt/241 , x-fmt/238 , x-fmt/239
Software en code	XML, CSS, JSON, DTD, XSL, XSD, XML, HTML		x-fmt/224 , x-fmt/315 , fmt/817 , x-fmt/280 , x-fmt/281 , fmt/101 , fmt/96 , fmt/97 , fmt/98 , fmt/99 , fmt/100 , fmt/102

3D	GLB, GLTF	X3D, OBJ, E57, 3DM, PLY	fmt/1316 , fmt/1314 , fmt/1315 , fmt/579 , fmt/580 , fmt/581 , fmt/582 , fmt/1210 , fmt/643 , x-fmt/432 , x-fmt/433 , x-fmt/434 , x-fmt/435 , fmt/831
CAD en Geo	GML, KML, GeoTIFF	DWG, APR, PRJ, ADF, AUX	x-fmt/227 , fmt/244 , fmt/155 , fmt/21 , fmt/22 , fmt/23 , fmt/24 , fmt/25 , fmt/26 , fmt/27 , fmt/28 , fmt/29 , fmt/30 , fmt/31 , fmt/32 , fmt/33 , fmt/34 , fmt/35 , fmt/36 , x-fmt/455 , fmt/434 , fmt/531 , fmt/1395 , x-fmt/317 , fmt/332 , fmt/320 , x-fmt/218 , fmt/1215

*Bestandsformaten die rood zijn kunnen in Preservica worden opgenomen, maar niet worden getoond. Dikgedrukte bestandsformaten kunnen opgeslagen worden, maar enkel na migratie in een ander bestandsformaat worden getoond.

Van de meeste bestandsformaten bestaan verschillende versies. Niet elke versie wordt geaccepteerd. Controleer daarom in de tabel welke Pronom ID's aangeboden kunnen worden.

Indien een bestandsformaat niet in de bovenstaande tabel is opgenomen, maar zich wel in het over te dragen archief bevindt dient de archiefvormer contact op te nemen met Erfgoed Leiden om de eventuele mogelijkheden te bespreken.

Je vindt het PRONOM register en meer informatie hier:
www.nationalarchives.gov.uk/PRONOM/Default.aspx.

2.7 Lossless compressie

Het is niet toegestaan gebruik te maken van een compressietechniek op het over te brengen digitaal archief.

2.8 Fixity en aantallen controle

Bij de aanlevering van archiefbescheiden aan Erfgoed Leiden dient van elk archiefstuk een hashcode (fixity) meegegeven te worden in het SHA-256 algoritme. Deze code moet zowel in het XML schema als in een aparte Excel lijst aangeleverd worden. Met deze code controleert Erfgoed Leiden of het bestand tijdens de aanlevering is beschadigd. Indien dit het geval blijkt zal het bestand opnieuw opgevraagd worden bij de archiefvormer. Tevens moet uit dit Excel bestand duidelijk worden hoeveel en welke bestanden er aan Erfgoed Leiden geleverd zijn ten behoeve van de controle op volledigheid van de aanlevering.

2.9 Wijze van transport

Het over te brengen archief dient op een server van Erfgoed Leiden geplaatst te worden. Het transport vindt plaats via een SFTP verbinding. De leverancier van de bronapplicatie of de archiefvormende instelling dient zelf zorg te dragen voor de benodigde software om deze verbinding

te kunnen leggen. De afzender partij geeft aan Erfgoed Leiden door via welke IP-adres het archief wordt verzonden. Erfgoed Leiden verschaft een gebruikersnaam en wachtwoord om de verbinding tot stand te brengen. Ook is er een handleiding beschikbaar met gedetailleerde uitleg over de werking van het transport.

Bij het transport mogen geen verborgen thumbs.db of directory.ini bestanden meegeleverd worden.

3. Overzicht van bijlagen

Bijlage 1 - Guide to OPEX ingest en Monitoring

Zie bijgevoegd document

Bijlage 2 - OPEX xsd

Zie bijgevoegd document