



Voorwaarden export

Versie	1.3 Definitief
Datum	6 februari 2017
Contact	contact@nationaalarchief.nl
Auteur	Luut Stadman, Informatieanalist, Infrastructuur en Services, Nationaal Archief
Afgestemd met (versie 1.0 of hoger)	Gelders Archief, één persoon Nationaal Archief, Collectie, meerdere personen Nationaal Archief, Digitale Archiefdiensten, meerdere personen Nationaal Archief, Infrastructuur en Services, meerdere personen Nationaal Archief, Kennis en Advies, meerdere personen Ordina, meerdere personen
Eigenaar	Diana Teunissen, Directeur, Digitale Infrastructuur en Advies, Nationaal Archief

Inhoud

1. Inleiding	3
2. Sidecar-structuur	3
3. Vijf types aggregatieniveau	3
4. Versies en bijlagen	5
5. Afgesloten dossiers	5
6. ToPX 2.3	5
7. Identificatiekenmerk	7
8. Naamgeving mappen, bestanden en paden	8
9. Hoofdlettergevoeligheid	9
10. Controle op verzending	9
11. Supplementen	9

Versie-informatie

Versie	Auteur(s)	Informatie
1.0	Luit Stadman, Wai Wong	Eerste versie
1.1	Luit Stadman, Wai Wong	<ul style="list-style-type: none">- Paragraaf 6 uitgebreid met informatie m.b.t. xsd.- Paragraaf 7 gereviseerd. Verpakken in container is verplicht geworden.- Nieuwe paragraaf 9 m.b.t. controle op verzending.- <i>Versie 1.0 is nog steeds bruikbaar.</i>
1.2	Luit Stadman, Wai Wong	<ul style="list-style-type: none">- Op diverse plaatsen in het document het verschil tussen de begrippen document en bestand verduidelijkt.- In paragraaf 3 en 4 de mogelijkheden verruimd. Er mogen nu meerdere aggregatieniveaus van het type dossier in een export voorkomen. De verplichting van een record tussen een dossier en een bestand is opgeheven in geval een document maar één versie kent en geen bijlagen heeft.- <i>De versies 1.0 en 1.1 zijn nog steeds bruikbaar.</i>
1.3	Luit Stadman	<ul style="list-style-type: none">- In paragraaf 6 een opmerking over karaktercodering UTF-8 zonder BOM toegevoegd.- Nieuwe paragraaf 7 met de technische eisen m.b.t. ToPX 2.3-metadata-element identificatiekenmerk conform de papieren mapping.- In paragraaf 11 een tweetal zinnen m.b.t. ToPX 2.3-metadata-element identificatiekenmerk toegevoegd.- <i>De versies 1.0, 1.1 en 1.2 zijn nog steeds bruikbaar.</i>

1. Inleiding

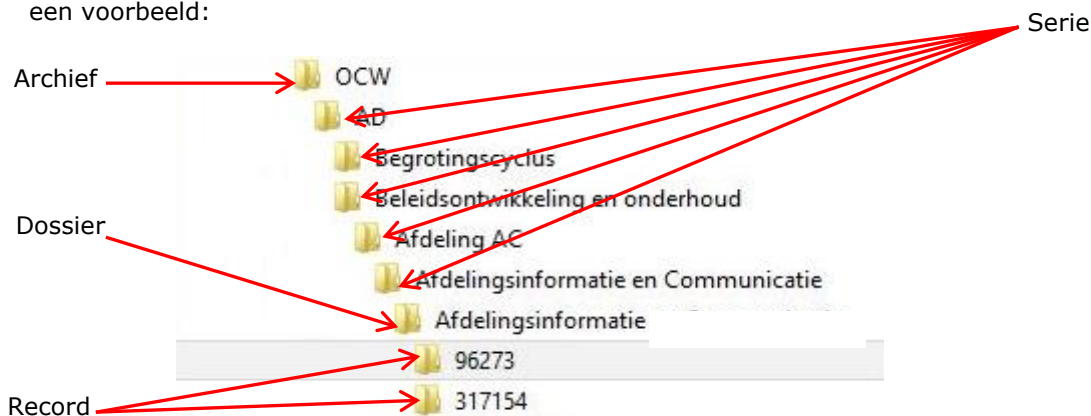
In dit document vindt u de beschrijving voor het maken van de structuur van een export uit een DMS (Document Management System) voor de overheveling van de content (in DMS-termen: rubrieken, series, dossiers, subdossiers, documenten en versies van documenten) en de metadata van het DMS naar het e-Depot. De beschrijving richt zich eerst op een initiële export uit het DMS, dus ten behoeve van een nieuw archief in het e-Depot. In de laatste paragraaf wordt beschreven hoe vervollexporten voor toevoegingen aan een bestaand archief in het e-Depot, zogenoemde supplementen, eruit moeten zien. Om bij de initiële export goed voorbereid te zijn op supplementen komen in paragraaf 7 supplementen ook al kort aan bod.

2. Sidecar-structuur

De export dient een zogenoemde sidecar-structuur te hebben. In deze structuur zit de content in een directory-structuur: elke rubriek, serie, dossier, subdossier en document is een map en elke versie van een document is een bestand. Elke map (rubriek, 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 `.metadata`. De plaats van een metadatabestand dat bij een map hoort, is in deze map. De plaats van een metadatabestand dat bij een bestand hoort, is naast dit bestand.

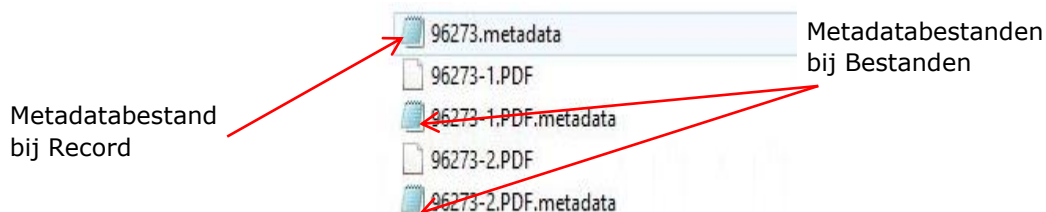
3. Vijf types aggregatieniveau

In de Nederlandse archiefwereld kent men vijf types aggregatieniveau. Het hoogste aggregatieniveau is van het type *archief*. Daarna volgen er meestal één of meerdere aggregatieniveaus van het type *serie*. Dit type aggregatieniveau is evenwel niet verplicht. Dan volgt er ten minste één aggregatieniveau van het type *dossier*. In de archiefwereld is een dossier niet per se hetzelfde als in de DMS-wereld. Dan volgen er in de regel één of meerdere aggregatieniveaus van het type *record*, hoewel dit aggregatieniveau niet in alle gevallen verplicht is. De vier tot nu toe genoemde types aggregatieniveau komen overeen met de DMS-mappen (rubrieken, series, dossiers, subdossiers en documenten) in de export. Hieronder een voorbeeld:



Het laagste aggregatieniveau is van het type *bestand*. Wat in de archiefwereld een *bestand* heet, is feitelijk niets anders dan wat in de DMS-wereld een versie van een document is.

Hieronder als voorbeeld de inhoud van *record* ofwel map '96273' in het vorige plaatje:



Het record bevat dus:

- het metadatabestand bij het *record* '96273.metadata'
- de *bestanden* '96273-1.PDF' en '96273-2.PDF'
- de metadatabestanden bij deze twee *bestanden*, te weten '96273-1.PDF.metadata' respectievelijk '96273-2.PDF.metadata'

Hieronder de vijf types aggregatieniveau in een tabelletje:

Type	Niveau	Aantal aggregatieniveaus
Archief	Hoogste	1
Serie	Tweede tot en met (n-1)-de	0..*
Dossier	n-de tot en met m-de	1..*
Record	(m+1)-de tot en met één-na-laagste	Indien er geen sprake is van meerdere versies van en/of bijlagen bij een document: 0..* Indien er wel sprake is van meerdere versies en/of bijlagen bij een document: 1..*
Bestand	Laagste	0 of 1

Een *bestand* hangt in de regel niet rechtstreeks onder een *dossier*. Er zit gewoonlijk ten minste één *record* tussen een *dossier* en een *bestand* om verschillende versies van en/of bijlagen bij een document te kunnen aggregeren. In de volgende paragraaf vindt u hierover een uitgebreidere beschrijving. Alleen in het geval dat een document maar één versie kent en geen bijlagen heeft, is het *bestand* feitelijk ook het document en mag het dus rechtstreeks onder het *dossier* hangen.

Spil in de bepaling van de types van de aggregatieniveaus in de export is de bepaling van het hoogste aggregatieniveau van type *dossier*. Op dit niveau worden alle documenten met betrekking tot één zaak geaggregeerd. In DMS-terminen zitten we dan vaak op het niveau van de dossiers, maar soms ook op het/een niveau van de subdossiers.

Zodra duidelijk is welk niveau in het DMS overeenkomt met het hoogste aggregatieniveau van het type *dossier*, kunnen de types van de andere

aggregatieniveaus eenvoudig worden afgeleid. Alle mappen boven het hoogste aggregatieniveau van type *dossier* zijn van het type *serie*, behoudens de hoogste, die van het type *archief* is. De mappen onder het hoogste aggregatieniveau van type *dossier* mogen allen van het type *record* zijn, maar ook eerst nog van het type *dossier* zolang er geen sprake is van de aggregatie van versies van en/of bijlagen bij een document, want vanaf daar dient het type *record* gebruikt te worden. Wat nu nog resteert, zijn de bestanden, die uiteraard van het laagste aggregatieniveau van type *bestand* zijn.

4. Versies en bijlagen

Aggregatieniveaus van het type *record* kunnen dienen voor het aanbrengen van een ordening in subdossiers binnen een dossier (of binnen een hoger subdossier), maar moeten in elk geval gebruikt worden voor het bijhouden van verschillende versies van en/of bijlagen bij een document. Zijn er verschillende versies van een document of zijn er bijlagen, dan zijn deze alle terug te vinden in één *record*. Zo bevat het *record* '96273' in het eerder gebruikte voorbeeld twee versies van het document '96273', te weten '96273-1.PDF' en '96273-2.PDF'. Van bijlagen is in het voorbeeld geen sprake.

In sommige DMS-en zit(ten verschillende versies van) een/de bijlage(n) bij een document in een map naast de versies van dit document. Dan is het toegestaan om in de export die map als *record* op te nemen. Zo ontstaat er een onderliggend aggregatieniveau van het type *record*. Omdat bijlagen ook bijlagen kunnen hebben, kunnen daaronder ook weer aggregatieniveaus van het type *record* ontstaan.

5. Afgesloten dossiers

De export mag alleen afgesloten *dossiers* bevatten. *Dossier* wordt hier bedoeld als type aggregatieniveau in de Nederlandse archiefwereld. Nog lopende *dossiers* in het DMS worden dus niet meegenomen in de export.

6. ToPX 2.3


De metadatabestanden worden gevuld met ToPX 2.3-metadata conform de papieren mapping waarin de vertaalslag van de metadata-elementen in het DMS naar de ToPX 2.3-metadata-elementen is vastgelegd, en die in samenwerking tussen uw organisatie en het Nationaal Archief of een RHC (Regionaal Historisch Centrum) tot stand is gekomen.

Voor het vullen kan gebruik worden gemaakt van onderstaand vijftal sjablonen (dubbelklik op de paperclip om het sjabloon te openen):


- 1) Voor metadatabestand bij het *archief*:

 ToPX-2.3_Archief_max.xml

- 2) Voor metadatabestanden bij de *series*:

 ToPX-2.3_Serie_max.xml


- 3) Voor metadatabestanden bij de *dossiers*:

 ToPX-2.3_Dossier_max.xml

- 4) Voor metadatabestanden bij de *records*:

 ToPX-2.3_Record_max.xml

- 5) Voor metadatabestanden bij de *bestanden*:

 ToPX-2.3_Bestand_max.xml

In deze sjablonen verwijdert u vervolgens alle ToPX 2.3-metadatabestanden waarvoor in de papieren mapping is afgesproken dat deze niet gevuld zullen worden, of zet u deze ToPX 2.3-metadatabestanden tussen commentaar-tags. Deze actie is nodig om te voorkomen dat bij controle met behulp van de xsd, waarin de voorwaarden staan waaraan de ToPX 2.3-metadatabestanden in de metadatabestanden moeten voldoen, fouten worden geconstateerd met betrekking tot ToPX 2.3-metadatabestanden die niet leeg mogen zijn *indien aanwezig*.

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. Hieronder vindt u een link naar deze xsd, zodat u zelf kunt controleren of uw export voldoet aan de ToPX 2.3-eisen alvorens deze naar het e-Depot te versturen:

<http://hdl.handle.net/10648/topx-2.3.xsd>

De metadatabestanden dienen te worden opgeslagen met karaktercodering UTF-8 (8-bit Unicode Transformation Format) zonder BOM (Byte Order Mark).

7. Identificatiekenmerk

Hoewel de vulling van de ToPX2.3-metadata-elementen wordt afgestemd in de papieren mapping, willen we in dit document toch ingaan op ToPX 2.3-metadata-element 2, identificatiekenmerk. Dit element dient namelijk aan een aantal technische eisen te voldoen die essentieel zijn om toevoegingen aan bestaand archief (supplementen) mogelijk te maken.

De technische eisen met betrekking tot ToPX 2.3-metadata-element 2, identificatiekenmerk, in een export zijn:

- Het identificatiekenmerk van het *archief* bevat het nummer toegang dat u van uw contactpersoon bij het Nationaal Archief of een RHC aangereikt heeft gekregen.
- Het identificatiekenmerk van een *serie* mag niet gelijk zijn aan het identificatiekenmerk van het *archief* of het identificatiekenmerk van een andere *serie* in de export.
- Het identificatiekenmerk van een *serie* moet uniek afleidbaar zijn uit gegevens in uw DMS. Indien het identificatiekenmerk niet uniek afleidbaar zou zijn, dan zou een vervollexport (supplement) niet op de juiste wijze aan bestaand archief in het e-Depot kunnen worden toegevoegd.
- Het identificatiekenmerk van een *dossier* mag niet gelijk zijn aan het identificatiekenmerk van een ander *dossier* in de export of het identificatiekenmerk van een *record* in de export .
Voor supplementen geldt daarnaast dat het identificatiekenmerk van een *dossier* niet identiek mag zijn aan het identificatiekenmerk van een *dossier* of *record* in een voorgaande export (ofwel in reeds bestaand archief in het e-Depot).
- Het identificatiekenmerk van een *dossier* moet uniek afleidbaar zijn uit gegevens in uw DMS. Indien het identificatiekenmerk niet uniek afleidbaar zou zijn, dan zou namelijk één en hetzelfde afgesloten *dossier* meerdere malen in het e-Depot kunnen worden opgenomen. Het e-Depot voorkomt dat een *dossier* met een specifiek identificatiekenmerk nooit voor een tweede keer kan worden opgenomen.
- Het identificatiekenmerk van een *record* mag niet gelijk zijn aan het identificatiekenmerk van een *dossier* in de export of het identificatiekenmerk van een ander *record* in de export.
Voor supplementen geldt daarnaast dat het identificatiekenmerk van een *record* niet identiek mag zijn aan het identificatiekenmerk van een *record* of

dossier in een voorgaande export (ofwel in reeds bestaand archief in het e-Depot).

- Een identificatiekenmerk mag uit niet meer dan 255 karakters bestaan.

Kort samengevat kan bovenstaande, over alle aanleveringen heen, ook als volgt worden verwoord:

- Het identificatiekenmerk van het *archief* is het nummer toegang dat u van uw contactpersoon heeft ontvangen.
- Het *archief* en elke *serie* binnen het archief heeft een identificatiekenmerk dat uniek is binnen deze groep.
- Elk *dossier* en elk *record* binnen het archief heeft een identificatiekenmerk dat uniek is binnen deze groep.
- Het identificatiekenmerk van een *serie* of een *dossier* moet uniek afleidbaar zijn (of in geval de *serie* of het *dossier* wegens vernietiging niet meer bestaat in het DMS: zijn geweest) uit de gegevens in het DMS.
- Elk *dossier* en alles wat eronder zit mag maar één keer worden aangeleverd.
- Een identificatiekenmerk mag uit niet meer dan 255 karakters bestaan.

8. Naamgeving mappen, bestanden en paden

Voor het creëren van de sidecar-structuur kunt u de mappen-en-bestanden-structuur van uw besturingssysteem gebruiken. Hierbij gelden de volgende technische beperkingen:

- Map- en bestandsnamen mogen niet de volgende karakters bevatten:
< > : " / \ | ? * # &
U zult de map- en bestandsnamen moeten schonen van deze karakters.
- Map- en bestandsnamen mogen niet identiek zijn aan:
**CON, PRN, AUX, NUL, COM1, COM2, COM3, COM4, COM5, COM6,
COM7, COM8, COM9, LPT1, LPT2, LPT3, LPT4, LPT5, LPT6, LPT7, LPT8
of LPT9**
Uiteraard mag een map- of bestandsnaam deze strings wel bevatten.
Map- en bestandsnamen die identiek zijn aan één van de genoemde strings, zult u moeten veranderen.

Om te voorkomen dat er problemen ontstaan bij het verzenden en het verwerken van de export door padnamen van meer dan 254 karakters, dient u de gecreëerde mappenstructuur voor verzending richting het Nationaal Archief of een RHC in een container met tar-formaat te plaatsen met behulp van een tool die geen beperking aan de padnaamlengte stelt, zoals de tool 7-Zip en de tool GNU tar. De container mag alleen de mappenstructuur bevatten, met het archiefniveau als hoogste map, en mag niet gecompriemd worden aangemaakt.

De naam van de container dient te beginnen met de ISIL-code van uw organisatie (zie: <http://hdl.handle.net/10648/isil>), gevolgd door een underscore, gevolgd door het identificatiekenmerk van het archief, gevolgd door een underscore, gevolgd door datum en tijdstip van creatie (in formaat jjjjmddThmm). Daarachter kan naar eigen invulling informatie worden toegevoegd.

9. Hoofdlettergevoeligheid

Let u op: in de infrastructuur van het e-Depot wordt Linux gebruikt. Linux is hoofdlettergevoelig. Dit betekent dat een pad-, map- of bestandsnaam die is afgeleid van een (andere) map- of bestandsnaam, qua keuze in hoofdletter/kleine letter niet mag afwijken van de map- of bestandsnaam waarvan hij is afgeleid.

10. Controle op verzending

Om te controleren of de verzending van de export naar het e-Depot goed is gegaan, heeft uw contactpersoon bij het Nationaal Archief of een RHC een tweetal gegevens nodig:

- het aantal bestanden in de export met extensie ongelijk aan .metadata en
- de omvang van de content in de export in bytes (ofwel de optelsom van de grootte van de bestanden met extensie ongelijk aan .metadata).

Hoe u deze gegevens aanlevert, spreekt u af met uw contactpersoon.

11. Supplementen

Vervolgexporten ten behoeve van toevoegingen aan een bestaand archief in het e-Depot, een zogenoemd supplement, dienen te voldoen aan alles wat hiervoor beschreven is voor een initiële export, behoudens dat een dergelijke vervolgexport in archieftermen alleen het volgende bevat:

- afgesloten *dossiers* die niet reeds in een voorgaande export naar het e-Depot zijn verzonden,
- alle *records* en *bestanden* die onder die afgesloten *dossiers* hangen en
- het *archief* en alle *series* (bestaande en nieuwe) die nodig zijn om vanaf het *archief* bij die afgesloten *dossiers* te komen.

Ook voor het *archief* en per bestaande *serie* dient een metadatabestand te worden aangeleverd. Het identificatiekenmerk in zo een metadatabestand is het identificatiekenmerk waaronder het *archief* of de *serie* reeds bekend is in het e-Depot. Let u op: eventuele wijzigingen in de metadata in zo een metadatabestand ten opzichte van de eerste aanlevering van de metadata worden door het e-Depot genegeerd.

Gewijzigde *dossiers* zijn niet toegestaan. Nadat een *dossier* afgesloten is, mag er niets meer aan worden toegevoegd of gewijzigd, of uit worden verwijderd. Indien een vervollexport toch één of meerdere *dossiers* bevat met hetzelfde identificatiekenmerk als een *dossier* dat reeds in het e-Depot aanwezig is, dan zal de vervollexport worden afgekeurd.

Tot slot de volgende opmerking: indien u sinds de vorige export wijzigingen in de structuur van uw DMS of in de naamgeving van mappen in uw DMS heeft doorgevoerd, en deze wijzigingen (deels) betrekking hebben op mappen die reeds in het e-Depot aanwezig zijn, neemt u dan alvorens u een nieuwe export stuurt eerst contact op met uw contactpersoon bij het Nationaal Archief of een RHC. Een toevoeging aan het bestaand archief is dan niet zonder meer mogelijk. Uw contactpersoon zal dan samen met u naar een oplossing zoeken.