Opzet proefopstelling – basale koppeling metadata tussen preserveringsvoorziening en CBS

Door: Antoinet, Peter, Maarten  
Datum: 23 April 2020

# Introductie

Het Noord-Hollands Archief is bezig met het opzetten van een eigen preserveringsvoorziening voor digital born materiaal. In April 2020 heeft het archief gekozen voor het product Preservica als technische infrastructuur om deze voorziening op te bouwen. Dit product is naar verwachting per juli 2020 gereed voor gebruik.

In de huidige situatie van het archief wordt alle ManifestatieOnafhankelijke Metadata (MOM) van een archiefstuk beschreven in het CollectieBeheerSysteem (CBS) Mais Flexis. In een digitale preserveringsvoorziening zal in principe alle binnenkomende MOM in de context van de oorspronkelijke manifestatie opgeslagen worden. Dus voor digital born in de preserveringsvoorziening.

Om toch een compleet overzicht te behouden van de collectie heeft het MT, na een intern onderzoek (uitgevoerd door Peter Brand), gekozen voor een basale koppeling van metadata tussen preserveringsvoorziening en CBS.

Nu deze keuze voor de infrastructuur en de koppeling gemaakt is komen meer concrete vragen naar boven over deze koppeling. Dit document beschrijft de vragen hoe deze koppeling ingericht gaat worden en het proces om deze vragen te beantwoorden.

Omdat in eerste instantie nog geen gestructureerde digital born archieven binnenkomen voor alle verschillende casussen zal een proefopstelling ontwikkeld worden. Deze proefopstelling dient om verdere vragen rondom de koppeling van het CBS te beantwoorden. Hierin wordt gebruik gemaakt van de testomgeving van de preserveringsvoorziening en de testomgeving van MAIS.

# Randvoorwaarden

Voordat deze proefopstelling gerealiseerd kan worden, en daarmee ook de vragen hieronder beantwoord kunnen worden zullen er aan een aantal randvoorwaarden voldaan moeten worden.

Bovenal moet er voldoende ontwikkelcapaciteit om potentiele middleware te moeten bouwen tussen de preserveringsvoorziening en het CBS. Ook moet er voldoende personeel zijn om de vragen tijdig te kunnen beantwoorden.

## Preserveringsvoorziening / Preservica

De voorwaarden die aan de gekozen preserveringsvoorziening gesteld worden zijn:

* Er moet een metadataschema geregistreerd en indexbaar zijn in Preservica, gebaseerd hierop moeten exports gedaan worden[[1]](#footnote-8522).
* Importworkflows met koppeling naar relevante metadataschema’s zijn ingericht.
* Een werkende API toegang op de preserveringsvoorziening (bijv. OAI-PMH of CMIS).
* Er moet een aangestelde systeembeheerder / functioneel beheerder beschikbaar zijn die met bovenstaande voorwaarden overweg kan.

## CBS / MAIS-FLEXIS

De voorwaarden die aan ons CBS gesteld worden zijn:

* Een testomgeving in MF9, zonder een testomgeving in MF9 kunnen we de koppeling niet onderzoek die daadwerkelijk toekomstgericht is.
* Werkende importfunctionaliteit, zonder importfunctie of import API zal er geen koppeling gelegd kunnen worden.
* Functioneel beheerders / systeembeheerder die beschikbaar zijn om bovenstaande voorwaarden te verwezenlijken danwel te onderhouden.

## Datamodel

Voordat de vragen hieronder beantwoord kunnen worden is een stabiel en vastgesteld datamodel nodig. Momenteel wordt TMLO gebruikt als datamodel gebruikt voor overheidsarchieven. Er wordt op dit moment gewerkt aan een opvolger hiervan: MDTO. Ons voorstel is te kiezen voor een stabiele variant van MDTO[[2]](#footnote-25108), zodat we toekomstgericht werken.

# Onderzoeksvragen

De onderstaande vragen worden gedurende het proces beantwoord. Deze vragen zullen voor vier verschillende casussen beantwoord moeten worden:

* Volledig digitale archieven
  + Van overheden
  + Van particulieren
* Hybride archieven
  + Van overheden
  + Van particulieren

Per casus wordt een testset gedefinieerd. In eerste instantie zal hier een theoretische testset gebruikt worden, om de complexiteit van de metadata overzichtelijk te houden. Vervolgens wordt er met de adviseur digitaal archiveren gekeken naar de daarbij aansluitende praktijkvoorbeelden.

## Keuze metadata dat geregistreerd wordt binnen het CBS

Afhankelijk van het gekozen metadataschema voor de digital born archieven wordt een selectie van MOM-informatie gemaakt die opgeslagen worden binnen het CBS. We gaan vooralsnog uit van MDTO. Deze zal waarschijnlijk op het moment van uitvoeren van het onderzoek nog niet zijn vastgesteld.

MDTO zal waarschijnlijk voor de toekomst vastgesteld worden voor overheidsarchieven. Voor particuliere archieven zal deze waarschijnlijk niet makkelijk toepasbaar zijn. Een apart onderzoek is bezig met het opstellen van een schema voor deze archieven. Er zal hier afgestemd moeten worden hoe dat ingericht gaat worden.

Vervolgens wordt onderzocht hoe deze vertaald kunnen worden naar een metadatastandaard die in MAIS opgeslagen kan worden. Hiervoor moeten wellicht nieuwe archiefeenheidssoorten opgesteld worden. Vooralsnog zou dat gaan omde volgende twee:

* Collectie (digital born)
* Digital born record

Dit kan een aandachtspunt zijn, voornamelijk wanneer er binnen de preserveringsvoorziening meerdere metadatastandaarden gebruikt zouden kunnen worden. De term ‘record’ heeft binnen TMLO en MDTO bijvoorbeeld andere betekenissen.

Relevante subvragen voor keuzes in het bovenstaande zijn of:

* Op welke manier de (semi-)handmatige koppeling exact gelegd wordt
* Of en hoe we naderhand gaan verrijken in MAIS
* Worden er kruisverwijzingen gemaakt vanuit CBS naar preserveringsvoorziening en vice versa en op welke manier?

## Manier van exporteren en importen van metadata

Het is tot dusver complex tot onmogelijk gebleken om automatisch metadata toe te voegen aan MAIS. Deze koppelingen zijn tot dusver niet structureel tot stand gekomen. De mogelijkheden van de exportfunctionaliteiten van de preserveringsvoorziening en de importmogelijkheden van het CBS worden in kaart gebracht.

Gedurende het onderzoekproces worden ook andere manier onderzocht zoals (semi) handmatig overzetten van metadata.

## Tijdsplanning

Het onderzoek kan beginnen nadat aan de voorwaarden, zie boven, voldaan is. Op zijn vroegst kan daarmee dit proces in juli pas van start gaan. De geschatte doorlooptijd van dit onderzoek zal minimaal 6 maanden en maximaal 12 maanden in beslag nemen.

## Personele bezetting

* Peter Brand zal als specialist informatiemanagement de projectleiding op zich nemen en rapporteren over de voortgang, hij is tevens de linking pin richting het implemenatietraject van de preserveringsvoorziening
* Informatiemanagement zal een secretariële rol uitoefenen voor dit project.
* De dataspecialist zal kijken een voorstel doen voor metadata mapping en inrichting van MAIS. Functioneel beheerder van MAIS worden hierbij betrokken.
* De functioneel beheerder/systeem beheerder van de preserveringsvoorziening zal de export van metadata onderzoeken en inregelen.
* Voor digital born en hybride particuliere archieven wordt Harco Gorter betrokken.
* De adviseur digitale archieven wordt betrokken voor het ontwikkelen van testsets.

1. Het huidige gangbare metadataschema voor zorgdragers van het NHA, TMLO wordt in 2020 vervangen door MDTO. Ten tijde van de inrichting van Preservica is MDTO nog niet gereed [↑](#footnote-ref-8522)
2. HIerbij gaat het NHA telkens uit van de laatste vastgestelde versie [↑](#footnote-ref-25108)