Bijlage A. Generiek metadatamodel

P. Bakker, R. Plugge

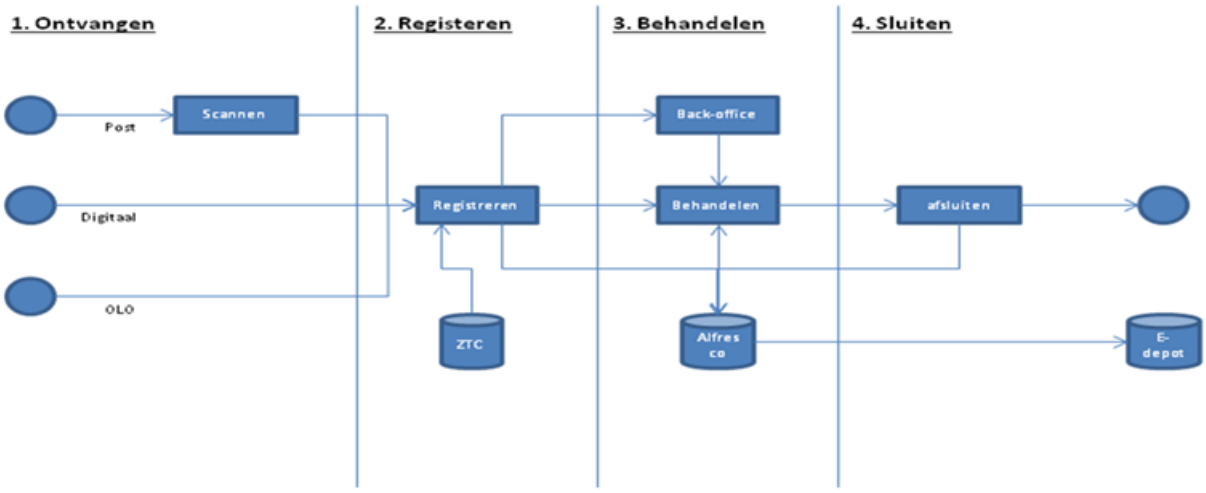
Tijdens de pilot Archief2020 heeft het RHCL in samenwerking met de Gemeente Maastricht een generiek metadatamodel op basis van TMLO tot stand gebracht. Dit schema kan gehanteerd worden als invuloefening per zaaktype of om het XML-metadataschema op te bouwen. Op deze wijze kan het Submission Information Package (SIP) op succesvolle wijze worden ge-ingest in het e-Depot als Archive Information Package (AIP).

Hoe dit product tot stand is gekomen (proces) staat beschreven in de eindrapportage van de pilot. Hetzelfde geldt voor de geleerde lessen tijdens het proces. In de oplevering van het product wordt alleen in grote lijnen beschreven hoe het proces verlopen is. Vervolgens zal er een uitleg bij het schema zijn.

I. Doel : Gewenste document- en processtroom vanaf 2017

De overheid zal vanaf 2017 zoveel mogelijk digitaal werken. De Gemeente Maastricht heeft zijn processen daar reeds op ingericht of zal de komende tijd zijn processen daar op inrichten.

In onderstaande figuur is de gewenste situatie opgenomen. Er zijn drie mogelijkheden voor het binnenkomen van de documentatie bij de gemeente via externen: per post (analoog), digitaal of via het omgevingsloket (OLO). Vervolgens vindt er een registratie plaats in het zaaksysteem (i.c. Exxellence). Daarna wordt de zaak behandeld in een back-office applicatie en verwerkt in het DMS (Alfresco). De zaak wordt gesloten in het zaaksysteem en daarmee via het DMS uitgeplaatst in het e-Depot.



II. Achtergrondinformatie

Processen aansprakelijkheidsstelling en WABO niet monumentale bouwvergunning

Aan de hand van twee zaaktypen is het generieke model voor zaaktypen tot stand gekomen. Het betreft de zaaktypen Aansprakelijkheidstelling van de Gemeente Maastricht en WABO niet monumentale bouwvergunning. Het doel van het generiek model is een handvat te bieden aan overheden, zodat zij hun *digital born archives* op een goede wijze kunnen aanleveren aan archiefdiensten. Op deze manier kunnen overheden meer *lean* gaan werken en zal de archivering steeds meer geautomatiseerd kunnen verlopen.

In de pilot is uitgegaan van uitgeplaatst *digital born archive*, waarin het *digital born* leidend is en er in principe geen analoge stukken worden gecreëerd. Er is daarom gekeken hoe een succesvolle SIP gecreëerd kon worden vanuit het DMS van de zorgdrager. Het was noodzakelijk om metadata volgens het TMLO-model duidelijk te krijgen, zodat de juiste minimale set aan metadata uit het DMS geëxporteerd kon worden. Er is in dit geval eerder sprake van het inrichten van de metadata bij de zorgdrager dan het achteraf mappen. Het moment van uitplaatsing is tevens het 'bevroren' moment van metadata. Dit betekent dat de eventgeschiedenis gevuld wordt tot aan het moment van uitplaatsing.

De opbouw van het TMLO schema

Het generieke metadataschema is ontstaan door eerst een relatief eenvoudig zaaktype als Aansprakelijkheidstelling te fileren en TMLO toe te passen. De processen, informatiestromen en gevormde documenten binnen dit zaaktype zijn in kaart gebracht. TMLO bestaat uit vier aggregatieniveaus, te weten: Archief, Serie, Dossier en Archiefstuk. Deze indeling vertoont overeenkomsten met de opbouw van een klassieke inventaris. Voor een gemeente die zaakgericht werkt, komt het niveau Serie overeen met zaaktype; het niveau Dossier is vergelijkbaar met Zaak en Archiefstuk met de stukken die zich in de zaak bevinden. Alleen het niveau Archief is vooralsnog niet te verbinden aan het zaakgericht werken.

Voor het aggregatieniveau Archief moet er een archivistische keuze gemaakt worden. In het geval van de Gemeente Maastricht is ervoor gekozen om de Gemeente Maastricht als één geheel te zien en als archief vast te leggen. Er had ook voor gekozen kunnen worden om elke onderdeel binnen de gemeente als zelfstandige archiefvormer te zien en op dit niveau Archief vast te leggen. Hier is dus niet voor gekozen.

Binnen het TMLO kunnen er verschillende entiteiten bestaan. Dit houdt kort gezegd in dat je verschillende metadataschema's hebt die je aan elkaar relateert. Zo kunnen de actoren en ook de activiteiten in een apart metadataschema vastgelegd worden en gerelateerd worden aan Archiefstuk, Dossier, Serie en Archief. Hier is niet voor gekozen. Er is besloten om een één-entiteitmodel-record te creëren. Dit betekent dat in het metadataschema de informatie moet worden vastgelegd, die normaal in een ander register wordt bijgehouden, zoals de eerder genoemde actoren en activiteiten.

Belangrijke conclusie uit de analyse van de zaaktypen is dat veel informatie en velden uit de zogenaamde TMLO Recordsmanagementactiviteitenvelden en Recordsvelden uit het DMS systeem (i.c. Alfresco) gehaald kunnen worden.[1] Uit een zogenoemde masterdatabase (i.c. I-navigator) kan de informatie voor de velden uit

[1] King Gemeenten e.a., Toepassingsprofiel metadatering Lokale overheden, versie 1.1, 12. https://archief2020.nl/downloads/toepassingsprofiel-metadatering-lokale-overheden-doc (23-5-2016).

TMLO Recordsmanagementactiviteiten, Bedrijfsactiviteiten, Actor en Records tot stand komen. Dit geldt in het bijzonder voor de aggregatieniveaus Serie, Dossier en Archiefstuk. De masterdatabase is ingericht op zaaktypen. Voor het aggregatieniveau Archief betekent dit dat er vooralsnog geen informatie gehaald kan worden uit de masterdatabase en de informatie elders moet worden bijgehouden. Een wens richting de leverancier kan zijn dat in de masterdatabase dergelijke informatie op metadataniveau eveneens kan worden vastgelegd. Vele velden uit Records voor de aggregatieniveaus Dossier en Archiefstuk kunnen gevuld worden vanuit het zaaksysteem (i.c. Exxellence) of Content Management Systeem (CMS).

De uitleg van TMLO is niet altijd eenduidig, omdat in de beschrijving van KING er naar gestreefd is om een generieke omschrijving of uitleg te geven. Hierbij was het niet altijd concreet hoe een veld geïnterpreteerd moest worden per aggregatieniveau. Het Generiek model hoopt daarmee tevens een bijdrage te leveren in de verduidelijking van de uitleg van TMLO door KING.

III. Opname van de metadata in het e-Depot

Ingest generiek model: TOPX en TMLO

Aan de hand van het TOPX XML schema van het Nationaal Archief zal de op TMLO gestoelde metadata ge-ingest worden in het Digitale Depot. Tot nu toe is er alleen een testbestand opgenomen. Dit testbestand is handmatig geüpload. Daarmee is de doelstelling voor de pilot gerealiseerd. In een vervolgtraject zal er naar gestreefd worden om de ingest zoveel mogelijk geautomatiseerd te laten verlopen. Dit houdt in dat de zorgdrager XML en bestanden op een dergelijke wijze zal moeten aanleveren, zodat ze ge-ingest kunnen worden in het Digitaal Depot.

Hierbij is van belang dat de zogenaamde *sidecar* wordt meegestuurd. De zorgdrager levert een mappenstructuur volgens de aggregatieniveaus op en levert daarbij de metadata op per map (sidecar). Voor het e-Depot houdt dit in dat de metadata per archiefstuk in de map van de Dossier wordt ondergebracht in de volgende structuur: [bestandsnaam].[extensie].metadata, terwijl voor de Voor de sidecars geldt: [mapnaam].metadata. Zie ook de bijlage ‘Structuur voor ingesten’.

Op dit moment is de functionaliteit binnen het e-Depot nog niet aanwezig om ook de sidecar op het hoogste niveau (Archief) mee te nemen. De wens is bij de leverancier (Preservica) neergelegd en zal naar verwachting gerealiseerd worden, zodat de ingest geautomatiseerd verlopen kan. Een andere mogelijkheid is dat er een generieke workflow wordt gerealiseerd om bulk bestanden te ingesten. Dit wordt op dit moment nog onderzocht.

Vanuit het DMS zal de zorgdrager de bestanden versleuteld aanleveren via een SFTP-verbinding. De zorgdrager levert daarmee het archief aan in een beveiligde omgeving bij het RHCL en vanuit deze omgeving zal het archiefmateriaal ge-SIPt worden en in de Quarantaine omgeving van het Digitaal Depot worden geplaatst.

TOPX biedt voor drie aggregatieniveaus (Archief, Dossier en Archiefstuk) XML-schema’s aan. Het Digitaal Depot neemt, indien er een sidecar Serie is, deze wel op. Zoals te zien is in het Generieke model komen de benodigde metadatavelden van Dossier en Serie nagenoeg overeen. Door het sjabloon metadatabestand van Dossier voor Serie te gebruiken en aan te passen aan Serie kan er op eenvoudige wijze een sjabloon *sidecar* gerealiseerd worden voor het aggregatieniveau Serie.

IV. Legenda generiek metadatamodel (bijlage als spreadsheet)

In de bijlage is het Generieke model opgenomen. In de kolommen A t/m G is de omschrijving volgens TMLO per veld opgenomen. Vervolgens komen de aggregatieniveaus aanbod:

- Aggregatieniveau Archief: kolom H t/m J;
- Aggregatieniveau Serie: kolom L t/m N;
- Aggregatieniveau Dossier: kolom P t/m R en
- Aggregatieniveau Archiefstuk: kolom T t/m V.

Per aggregatieniveau zijn drie kolommen opgenomen: Waarde Gemeente Maastricht, toelichting/motivatie en systeem. De kolom Waarde Gemeente Maastricht is gevuld met de verwachte waarde bij het veld. Bij toelichting/motivatie wordt een beknopte uitleg gegeven hoe de waarde tot stand komt of hoe tot de keuze is gekomen. De kolom Systeem geeft aan uit welk systeem de informatie van de waarden komt.

De gele velden geven aan welke velden per aggregatieniveau moeten worden meegenomen en gevuld worden. De witte (blanco) velden per aggregatieniveau zijn daarom niet van toepassing.