

Presentatie mapping MDTO-PREMIS

KIA Thematiendaagse e-depot

25 maart 2024



Kerngroep Duurzame Toegankelijkheid & Common Ground

Spreekers Léon Masselink en Eva van den Hurk – van 't Klooster

Uitgeschreven presentatie behorende bij de presentatie Kerngroep MDTO-PREMIS

Voordat ik inga op de mapping, wil ik eerst de twee standaarden toelichten. Het MDTO is bij de meesten van jullie waarschijnlijk wel bekend. MDTO is een metagegevensstandaard voor duurzame toegankelijkheid van overheidsinformatie. Het Nationaal Archief onderscheidt twee toepassingen voor MDTO: als informatiemodel voor een informatieapplicatie in de dynamische en semi-statische fase van informatiebeheer (vanuit de gedachte dat zo vorm wordt gegeven aan digitale duurzaamheid en archiving by design), en als uitwisselstandaard voor de migratie van overheidsinformatie van de archiefvormer naar de archiefdienst ten behoeve van uitplaatsing of overbrenging. MDTO maakt onderscheid tussen verschillende objecten waar metagegevens aan verbonden zijn. Informatieobject en Bestand worden genoemd en uitgewerkt in de standaard. Bedrijfsactiviteit, Actor en Locatie worden genoemd, maar zijn niet verder van metagegevens voorzien in de standaard. MDTO biedt de mogelijkheid om hiervoor andere standaarden te gebruiken.

PREMIS is misschien wat minder goed bekend. De PREMIS Data Dictionary is een metagegevensstandaard voor digitale preservatie, bedoeld om de preservatie van digitale objecten te ondersteunen en hun toekomstige raadpleegbaarheid te waarborgen. Het kan daarnaast dienen als informatiemodel in een preservatiesysteem, zoals een e-depotvoorziening of een ander soort digitale objectenbewaarplaats. PREMIS heeft daarmee een nauwe relatie met het Open Archival Information System (OAIS)-model, het conceptuele model dat aan de grondslag ligt van de huidige generatie e-depotsystemen. PREMIS onderscheidt vier entiteiten, namelijk Agents, Rights, Events en Objects. Deze vier entiteiten kennen een onderlinge relatie. Een Agent met zekere Rights mag een Event uitvoeren op een Object. Iedere entiteit is voorzien van een eigen, uitgebreide set aan metagegevens.

Met deze onderdelen van de standaarden in het achterhoofd, kunnen we doorgaan naar de mapping. De afgelopen periode hebben we gewerkt aan een mapping MDTO-PREMIS en PREMIS-MDTO. We hebben heel bewust gekozen voor een mapping beide kanten op, omdat wegens de diversiteit aan velden in beide standaarden hiermee het overzicht beter bewaard blijft. Naast deze twee mappings heeft de Kerngroep de volgende bijbehorende producten opgeleverd: een algemeen introductiedocument met uitleg over de standaarden, implementatiescenario's en een aantal praktische observaties met betrekking tot de standaarden en de mapping en een overkoepelend document met uitleg over de gemaakte mappingkeuzes, bijvoorbeeld waarom wij vinden dat twee velden één op één gemapt kunnen worden of waarom de mapping deels een match is.

Bij het werken aan de mapping had de Kerngroep een aantal uitgangspunten: we wilden dat de standaarden afzonderlijk van elkaar te herleiden bleven, we wilden dubbele registratie van gegevens voorkomen en we wilden zo min mogelijk afbreuk doen aan de interne logica van de twee standaarden. Alle drie de punten brachten hun eigen uitdagingen met zich mee. Op technisch vlak zouden we één attribuut, met waarde, aan meerdere standaarden willen koppelen en in meerdere standaarden willen kunnen exporteren. Een link tussen matchende velden in een linked data-omgeving kan via SKOS gelegd worden, maar de technische uitwerking in de componenten moet nog uitgedacht worden. SKOS kan ook een oplossing bieden voor niet exacte matches in de mapping in een linked data-omgeving.

Daar waar een aantal velden voor de Kerngroep een exacte match zijn tussen MDTO en PREMIS, bestaat het merendeel van de mapping namelijk uit redelijke matches of near matches. Hoe het kan dat de matches zo moeilijk te maken zijn, heeft te maken met het derde punt: het verschil in de interne logica van de twee standaarden. Het MDTO kent bijvoorbeeld 'Informatieobjecten' en 'Bestanden', terwijl PREMIS spreekt over 'Intellectual Entities', 'Representations', 'Files' en 'Bitstreams'. Dit verschil in gelaagdheid maakt dat de metadatavelden die op het eerste gezicht identiek lijken te zijn, binnen de twee standaarden een andere invulling krijgen en op een andere wijze gerelateerd worden aan de beschreven entiteit.

De volgende stap is de vertaling van deze mappings en bevindingen naar de technische praktijk. Een van de doelen van de mappingexercitie is om MDTO en PREMIS gezamenlijk te implementeren in een component- en linked data-gebaseerd systeem voor duurzame toegankelijk, maar we wilden ook graag weten of een dergelijke mapping te implementeren zou zijn in een relationele database. Omdat binnen Common Ground oud naast nieuw kan bestaan en werken met standaarden gegevensuitwisseling vergemakkelijkt, ligt het voor de hand om de mapping van de landelijke en internationale standaarden MDTO en PREMIS op zowel de huidige e-depotsystemen als op de in ontwikkeling zijn component-gebaseerde systemen voor duurzame toegankelijkheid toe te passen, in de aanloop naar de toekomstige uitwisseling en samenwerking tussen de twee verschillende typen systemen. De mapping implementeren in een component- en linked data-gebaseerd systeem lijkt op dit moment praktisch uitvoerbaar, dankzij de inzet van SKOS als relatieverklarende koppeling tussen de twee standaarden. De implementatie van de mapping in een relationele database wordt ingewikkelder. Relationele databases zijn vaak voor een speciaal onderwerp en een specifieke doelgroep ontwikkeld, waardoor deze gericht is op gebruikersgemak maar waardoor uitwisseling of migratie bemoeilijkt wordt.

Welke implementatiescenario's zouden er, met deze informatie in het achterhoofd, dan zijn voor de mapping? We zien drie richtingen voor ons. Op volgorde van impact van implementatie: de mapping kan gebruikt worden als uitwisselingsstandaard. Dit zou betekenen dat bij een export van informatie van het ene naar het andere systeem zowel MDTO als PREMIS worden meegenomen. In het geval van een relationele database zou de mapping tussen de twee databases gezet kunnen worden als een soort concordans. Bij een export naar een component-gebaseerd systeem of tussen twee component-gebaseerde systemen zou een dergelijke mapping in de derde laag van het Common Ground model, de uitwisselingslaag, passen. Dit scenario valt op te knippen in de implementatie van MDTO als uitwisselingsstandaard voor overheidsinformatie naast de implementatie van PREMIS voor digitale preservatie en een implementatie van MDTO als uitwisselingsstandaard en informatiemodel naast de implementatie van PREMIS voor digitale preservatie.

Een tweede implementatiescenario zou kunnen zijn de gezamenlijke implementatie van MDTO en PREMIS in relationele databases. Dit vergt voornamelijk inspanning van leveranciers van traditionele e-depotsystemen bij het updaten van bestaande systemen en de bouw van nieuwe systemen. Het vraagt ook om een naar deze standaarden herleidbare inrichting van inhoudelijke metadata en preservatiemetadaten. Een derde implementatiescenario is de inrichting van een component-gebaseerd systeem voor duurzame toegankelijkheid met in de toegangslaag de implementatie van de mapping in de ETL. Hiermee wordt de brondata in deze standaarden geregistreerd bij opname in het systeem. We zien dat de laatste twee scenario's veel ontwikkeling vragen van meerdere partijen en dat de inzet van de mapping voor de export tussen systemen een logische eerste stap zou zijn.

Dan nu terug naar het proces. Met onze uitgewerkte mapping en bijbehorende producten op zak hebben we een aantal feedbacksessies georganiseerd, om zo te zien of we inhoudelijk op de juiste weg zijn en of de implementatiescenario's die wij voor ogen hebben ook daadwerkelijk uitvoerbaar zijn. In december 2023 heeft de Kerngroep gesproken met Wout van der Reijden, coordinating recordkeeping advisor bij het Nationaal Archief en ontwikkelaar van MDTO en met Marjolein Steeman, information officer bij Beeld en Geluid en tevens lid van de PREMIS Editorial Committee. Tijdens deze feedbackronde zijn we dieper met Wout en Marjolein gedoken in de informatiemodellen achter de standaarden en waar MDTO en PREMIS elkaar wel en vooral ook niet raken. Ook hebben ze ons bijgepraat over de ontwikkelingen die op de planning staan voor MDTO en PREMIS. De sessie heeft ons een beter inzicht gegeven in de relaties tussen de verschillende entiteiten in zowel MDTO als PREMIS.

Een tweede feedbackronde is tussen januari en maart 2024 georganiseerd. Tijdens deze ronde zijn we in gesprek gegaan met drie marktpartijen die ons meer zouden kunnen vertellen over de haalbaarheid van de implementatie van de mapping in zowel traditionele e-depotsystemen als in component- en linked data-gebaseerde systemen. Hiervoor hebben we gesproken met een medewerker van ArchiXL, een bedrijf gespecialiseerd in enterprise- en informatiearchitectuur, medewerkers van DiVault, een leverancier van een e-depot gebaseerd op een relationele database en een medewerker van Triply, leverancier van knowledge graph stores en daarmee ondersteuner van linked data. Het werden drie totaal verschillende gesprekken.

In het eerste gesprek, met ArchiXL, werd het accent gelegd op het grote verschil in logica tussen de twee standaarden. Zoals eerder besproken zorgt de onderverdeling van MDTO in 'Informatieobjecten' en 'Bestanden', terwijl PREMIS spreekt over 'Intellectual Entities', 'Representations', 'Files' en 'Bitstreams', voor voornamelijk redelijke of near matches in de mapping. ArchiXL adviseerde de Kerngroep om een overkoepelend conceptueel model te ontwikkelen, waarin de Kerngroep aangeeft vanuit welke werkelijkheid hij denkt en opereert om daar vervolgens de twee standaarden aan te relateren.

Met DiVault besprak de Kerngroep de praktische uitdagingen van het implementeren van de twee standaarden in relationele databases. Zo is PREMIS op dit moment niet ingebed in gemeentelijke systemen en is sommige informatie in de systemen statisch, een beschrijving verandert na overbrenging niet meer, maar is andere informatie dynamisch, zoals gegevens omtrent preserving. Het feit dat relationele databases hard-coded zijn maakt de implementatie van een mapping met redelijke of near matches een uitdaging. Daarnaast kreeg de Kerngroep tips om concreet uit te werken waar de mapping voor is en wat de valkuilen en risico's bij implementatie zijn.

In het gesprek met Triply kregen we terug dat de mapping prima te implementeren is in een linked data-omgeving, zeker wanneer SKOS als relatieverklarende koppeling toegevoegd zou worden aan de mapping. Ook kregen we een aantal voorbeelden te zien van hoe zo'n mapping er uit zou zien.

Met de feedback van de drie sessies is het nu zaak om te bepalen hoe de Kerngroep het product verder wil gaan ontwikkelen. Een van de openstaande vragen is gerelateerd aan de feedback van ArchiXL: welk conceptueel model hangen we boven de mapping? Het uitdenken van zo'n model betekent definiëren wat onze werkelijkheid is en hoe de standaarden daarin passen. Omdat een informatieobject in principe van alles kan zijn, van een stuk papier tot een database zelf, moeten we in gesprek over hoe we de verschillende vormen van entiteit tussen de twee standaarden aan elkaar gaan relateren. De implementatie van het uitgewerkte model kan vervolgens in verschillende malen gegoten worden, want implementatie als uitwisselingsstandaard, implementatie als onderdeel van een relationele database of als onderdeel van een component-gebaseerd systeem vragen verschillende dingen. Hoewel het een mapping van standaarden betreft, is er dus duidelijk niet één perfecte oplossing.

Als Kerngroep gingen we in den beginne uit van een volledig carte blanche Common Ground-omgeving, maar er moet ook rekening gehouden worden met oud naast nieuw en migratie naar een component-gebaseerde oplossing. Op basis van het uitgedachte model en de feedback van DiVault, over het verconcretiseren van onze notitie, kunnen we onze randvoorwaarden en implementatiescenario's aanscherpen. Hierbij zullen we uitgaan van implementatie van en uitwisseling tussen relationele databases en component-gerelateerde systemen, daar een implementatiescenario dat uitgaat van een volledig Common Ground ingerichte informatiearchitectuur nog te veel luchtftietserij is. De randvoorwaarden voor een linked data-systeem zijn relatief concreet: slaan we meerdere triples op of gebruiken we SKOS als mapping? Dergelijke randvoorwaarden zijn lastiger voor een relationele database, daarvoor zullen we bredere definities moeten gaan vastleggen.

Voor ons als Kerngroep liggen er, naast deze concrete vervolgstappen, nog een aantal conceptuele vragen waar we, tussen de bedrijven door, rekening mee willen houden in de uitwerking van deze vervolgstappen. Vanuit welke werkelijkheid werken wij als archiefprofessionals? Hoe verenigen we archivistische, preserverings- en technische entiteiten? En wat betekent het om onder standaarden te werken? Betekent dit dat een database conform een standaard ingericht moet zijn of dat je door een filter van de standaard informatie moet kunnen vinden? Moet je je gegevens vastleggen in een standaard maar vanuit een andere omgeving kunnen raadplegen? Maakt het uit of we standaardisering inrichten in de gegevensuitwisselingslaag van Common Ground of dat we die inrichten in de data laag? Welke gevolgen hebben onze mappingskeuzes daadwerkelijk voor systemen? We hebben het tot nu toe alleen over theoretische implementatie, een daadwerkelijke implementatie is er nog niet. En stel nou dat onze mapping vanuit ons vakgebied op goedkeuring kan rekenen, wat kan de sector er dan mee gegeven de huidige, bestaande zaaksystemen, vakapplicaties en e-depotsystemen? En wat als er een nieuwe e-depotleverancier opstaat, die volledig aan alle standaarden voldoet?

Voor ons in ieder geval genoeg voer om over na te denken, terwijl we alle feedback verwerken. Wat we ook willen doen, is jullie om feedback vragen. We zullen de mapping ter review aanbieden, in de hoop dat jullie natuurlijk in groten getale reageren en we zo de mapping kunnen aanscherpen. Het introductiedocument en de overkoepelende uitleg zullen meegeleverd worden als achtergrondinformatie. Wanneer we de review zullen starten, kan ik nu nog niet zeggen maar ik hoop zo snel mogelijk. Voor nu wil ik u hartelijk danken voor uw aandacht.

Meer informatie is te vinden op de KIA Thematiendaagse informatiepagina: [Mapping-MDTO-Premis](#)