



notitie

FORUM STANDAARDISATIE Concept COLLEGE NOTITIE

Agendapunt:	05 Lijst open standaarden		
Bijlagen:	Additioneel onderzoek OAI-PMH		
Aan:	College Standaardisatie		
Van:	Forum Standaardisatie		
Datum:	November 2010	Versie	0.2
Betreft:	Advies over opname OAI-PMH op lijst met open standaarden voor 'pas toe of leg uit'		

Waarom is een keuze belangrijk?

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is een internationale standaard voor het verzamelen van metadata over objecten uit digitale bibliotheken (ook wel "repositories" genoemd). Met metadata moet gedacht worden aan zaken als titel, thema, auteur, etc. Een digitale bibliotheek is een elektronisch systeem met daarin objecten, zoals documenten, multimedia of andere vormen van (digitale) content. OAI-PMH maakt het mogelijk om op één plek aan de hand van metadata naar objecten in verschillende bibliotheken te zoeken.

Op steeds meer plaatsen komen documenten en andere objecten in digitale bibliotheken beschikbaar. De noodzaak om tegelijkertijd in verschillende bibliotheken op een slimme, maar gestructureerde manier te kunnen zoeken wordt daardoor steeds groter. Daarom is het van belang deze standaard op te nemen. De standaard wordt reeds toegepast in de erfgoed, de onderwijs- en de bibliotheeksector.

Kunt u met een gerust hart "ja" zeggen?

Het voorliggende advies is het resultaat van een uitgebreid expertonderzoek, een publieke consultatie en bespreking in het Forum Standaardisatie. Naar aanleiding van de publieke consultatie is op verzoek van het Forum een aanvullend onderzoek (zie bijlage) uitgevoerd. Op basis hiervan is het toepassingsgebied verder verduidelijkt. Eventuele technische of inhoudelijke conflicten met andere standaarden of ontwikkelingen zijn daardoor weggenomen.

Zijn er risico's verbonden aan de keuze?

Een aantal organisaties heeft in de consultatieronde aangegeven niet bekend te zijn met OAI-PMH en daarom niet in staat te zijn een goed oordeel te kunnen geven over de eventuele risico's. Organisaties krijgen echter enkel te maken met OAI-PMH indien bewust voor het gedefinieerde toepassingsgebied wordt gekozen. Bovendien interfereert OAI-PMH niet met bestaande ontwikkelingen op het gebied van digitalisering van informatie, maar is vooral aanvullend daarop.

Doel

Het College Standaardisatie wordt gevraagd in te stemmen met:

Datum

1 oktober 2010

1. de opname van de standaard OAI-PMH versie 2.0 (Open Archives Initiative Protocol for Metadata Harvesting) op de lijst met open standaarden voor 'pas toe of leg uit'.
2. het volgende functioneel toepassingsgebied: "Het vraaggestuurd aanbieden en ophalen van verzamelingen metadata uit bibliotheken met (digitale) documenten of andere objecten, met als doel het opnemen van deze metadata in een centrale bibliotheek. Uitgezonderd zijn die toepassingen waarvoor op basis van de lijst voor 'pas toe of leg uit' het gebruik van OSB (nu: Digikoppeling) verplicht is."
3. het volgende organisatorisch werkingsgebied: "Overheden en instellingen uit de (semi-) publieke sector".

Toelichting*Ad 1. Opname op de "pas toe of leg uit"-lijst*

Het Open Archives Initiative Protocol for Metadata Harvesting versie 2.0 (hierna te noemen: OAI-PMH) is een standaard voor het verzamelen van metadata uit digitale bibliotheken (zgn. 'repositories' of collecties). In een digitale bibliotheek bevinden zich digitale objecten, zoals documenten, foto's, video, etc. In metadata worden bepaalde kenmerken van deze digitale objecten weergegeven. Te denken valt aan de titel, de auteur, het onderwerp, een samenvatting, het jaartal, etc.

Door middel van OAI-PMH kan deze metadata worden opgevraagd uit meerdere digitale bibliotheken en vervolgens worden samengebracht in één geïntegreerde digitale bibliotheek (de digitale objecten uit de bibliotheek worden zelf niet overgezet). Vanuit de centrale bibliotheek kunnen gebruikers vervolgens zoeken in de inhoud van alle aangesloten digitale bibliotheken. In de educatieve sector wordt dit ondermeer gebruikt voor het bijeenbrengen van lesmateriaal van verschillende uitgevers. Musea gebruiken dit voor het integreren en doorzoekbaar maken van collecties.

Toevoeging van OAI-PMH aan de lijst met open standaarden voor 'pas toe of leg uit' betekent dat van alle (semi-) publieke organisaties wordt verwacht dat zij voor deze standaard een 'pas toe-of-leg uit' beleid gaan toepassen.

Door een expertgroep is de standaard beoordeeld op de vastgestelde criteria: openheid, potentieel, bruikbaarheid en impact. Over alle vier de criteria is positief geadviseerd.

In de consultatieronde is een reactie ontvangen van 16 partijen. 8 partijen reageerden neutraal of positief, 8 partijen reageerden negatief of hadden specifieke kritiekpunten. Deze kritiek spitste zich toe op drie inhoudelijke punten ten aanzien van OAI-PMH:

- Er zou beter gedefinieerd moeten worden voor welke vormen van uitwisseling van metadata OAI-PMH ingezet zou moeten worden.
- De relatie met Digikoppeling (een andere standaard op de lijst) zou beter onderzocht moeten worden, om interferentie daarmee te voorkomen.
- Een aantal organisaties gaf aan niet bekend te zijn met OAI-PMH en daarom niet in staat te zijn een goed oordeel te kunnen geven over de eventuele risico's. Er zou duidelijker gedefinieerd moeten worden wat het potentieel is en wat de risico's zijn buiten de archief-/educatie-/bibliotheekwereld (de oorsprong van OAI-PMH).

Datum
1 oktober 2010

Aan de hand hiervan is een additioneel onderzoek uitgevoerd. Dit onderzoek heeft geleid tot een aanpassing van het functioneel toepassingsgebied. Deze aanpassingen zijn beschreven onder punt 2.

In het onderzoek is gekeken naar het potentieel en de risico's buiten de archief-/educatie-/bibliotheekwereld. Feit is dat op tal van plaatsen binnen en buiten de overheid wordt gewerkt aan digitalisering van documenten en andere vormen van content. Voorbeelden zijn de digitalisering bij rechtbanken (vonnissen, dossiers), bij archieven (foto's, oude documenten), van brondossiers (bijvoorbeeld aktes van de burgerlijke stand), etc. Ook worden er nieuwe geheel digitale bibliotheken aangelegd door burgers, bedrijven en overheden. Denk aan online foto bibliotheken of de mogelijkheid om wetenschappelijke rapporten te downloaden.

In al deze situaties ontstaat vroeger of later de noodzaak om meerdere van deze digitale bibliotheken met elkaar te verbinden om daarin integraal te kunnen zoeken naar informatie. OAI-PMH speelt een belangrijke rol bij het leggen van deze verbinding en maakt mogelijk metadata uit al deze bibliotheken uit te wisselen. De standaard wordt hiertoe aanbevolen door o.a. Digitaal Erfgoed Nederland.

Toch speelt deze specifieke ontwikkeling (nl. het verbinden van digitale bibliotheken t.b.v. geïntegreerd zoeken) nog niet bij alle organisaties. Dat maakt het verklaarbaar dat OAI-PMH nog niet overal een gangbare en bekende standaard is, zoals bleek in de consultatie. Tegelijkertijd is het toepassingsgebied zo gekozen dat organisaties pas te maken krijgen met OAI-PMH indien bewust de keuze wordt gemaakt meerdere digitale bibliotheken met elkaar te verbinden met als doel te zoeken in metadata van deze gekoppelde bibliotheken. Hierdoor is het risico op interferentie met lopende ontwikkelingen binnen organisaties zeer gering. In het aanvullend onderzoek is dit getoetst op basis van een case bij de Kamer van Koophandel.

Op basis van deze afweging van potentieel en risico's en gegeven het feit dat de standaard voldoet aan de gestelde criteria wordt geadviseerd OAI-PMH op te nemen op de lijst.

Ad 2 en 3. Functioneel toepassingsgebied en organisatorisch werkingsgebied

Het functioneel toepassingsgebied is:

"Het vraaggestuurd aanbieden en ophalen van verzamelingen metadata uit bibliotheken met (digitale) documenten of andere objecten, met als doel het opnemen van deze metadata in een centrale bibliotheek. Uitgezonderd zijn die toepassingen waarvoor op basis van de lijst voor 'pas toe of leg uit' het gebruik van OSB (nu: Digikoppeling) verplicht is."

Het organisatorische werkingsgebied is: "Overheden en instellingen uit de (semi-) publieke sector".

Het toepassingsgebied is aangescherpt aan de hand van het aanvullende onderzoek:

- Er is gekozen voor de term 'vraaggestuurd aanbieden en ophalen van verzamelingen metadata'¹ om onderscheid te maken van andere vormen van uitwisseling van metadata, die niet door OAI-PMH worden ondersteund. Een voorbeeld hiervan zijn de projecten van e-Overheid voor Burgers, waarbij

¹ Het gaat dus om een interactief pull-model (actieve bevraging). Niet om een niet-interactief model ('spidering' van online content) of een push-model waarbij het initiatief van de aanbieder van metadata uitgaat.

vergunninginformatie van gemeenten wordt verzameld. Dit valt nu buiten het toepassingsgebied.

Datum
1 oktober 2010

- Er is gekozen voor een doelbinding: het doel van de uitwisseling van metadata moet zijn 'het opnemen ... in een centrale bibliotheek'. Hierdoor wordt voorkomen dat veel lopende ontwikkelingen op het gebied van digitalisering, waarbij metadata wordt uitgewisseld, onbedoeld binnen het toepassingsgebied komen te vallen. Hierdoor kunnen veel zorgen uit de consultatieronde worden weggenomen.
- In het expertonderzoek bleek dat er een interferentie kan ontstaan indien (op basis van de lijst) zowel OAI-PMH als Digikoppeling toegepast moet worden. Hoewel de kans hierop klein is, wordt dit door opname van een bepaling over Digikoppeling voorkomen.
- Tot slot, het gaat om "verzamelingen metadata uit bibliotheken met (digitale) documenten of objecten". Buiten het toepassingsbereik valt het uitwisselen van alleen gestructureerde data die is vastgelegd in databases, zoals bijvoorbeeld bij Basisregistraties het geval is.

Welk probleem wordt daarmee opgelost?

OAI-PMH maakt het mogelijk om geïntegreerd te zoeken in meerdere digitale bibliotheken. Dit kunnen verzamelingen documenten, afbeeldingen, video of andersoortige digitale content zijn.

Er zijn grofweg twee manieren waarop dit ingericht kan worden:

- Een zoekstelsel vraagt alle digitale content op uit de gekoppelde digitale bibliotheken en indexeert dit. Dit is de manier zoals de meeste zoeksystemen op internet werken (bijvoorbeeld: Google).
- Een digitale bibliotheek biedt een 'index' aan van de opgenomen digitale content. Deze index bevat metadata (vergelijk dit met het 'bibliotheekkaartje' in de voorflap van een bibliotheekboek) van de opgenomen digitale content. Het zoekstelsel leest deze index in en stelt gebruikers in staat om te zoeken in deze index.

OAI-PMH richt zich op de tweede mogelijkheid. Periodiek wordt een index opgevraagd uit meerdere digitale bibliotheken. Vervolgens wordt dit ingelezen in een overkoepelend zoekstelsel van waaruit de gebruiker kan zoeken.

Dit biedt een aantal belangrijke voordelen:

- De inhoud (die erg omvangrijk kan zijn; denk aan video) hoeft niet integraal overgezonden of geïndexeerd te worden.
- Er hoeft geen permanente koppeling te zijn tussen de digitale bibliotheken. Periodiek wordt immers de index verversd.
- Het kan wenselijk zijn om wel te weten waar een bepaald stuk aanwezig is, maar niet zomaar toegang te verlenen tot de uiteindelijke inhoud van dat stuk. Dat kan door alleen de metadata uit te wisselen.

Het nadeel is dat de uiteindelijke vindbaarheid van de documenten of digitale objecten volledig afhankelijk is van de kwaliteit van de metadata. OAI-PMH is daarom met name interessant als er sprake is van goed beheerde metadata.

Binnen de overheid worden steeds meer documenten digitaal opgeslagen. Ook zijn er tal van overheidsarchieven en bronregistraties die worden ingescand en als PDF/a² document worden opgenomen in een digitale bibliotheek. Vaak zijn er meerdere overheidspartijen die een gerelateerde digitale bibliotheek bijhouden.

² Archiefformaat dat is opgenomen op de lijst met open standaarden voor 'pas toe of leg uit'.

Kenmerken zijn dat:

- Naast documenten vaak metadata wordt opgeslagen.
- Voor de gebruiker relevante documenten vaak zijn opgeslagen in verschillende bibliotheken.
- Daarom via de metadata geïntegreerd gezocht moet kunnen worden over de verschillende bibliotheken.
- Niet iedereen zomaar toegang hoeft te krijgen tot alle documenten en digitale objecten uit de repository.

Datum

1 oktober 2010

Indien in die situaties gekozen wordt voor het inrichten van een geïntegreerd zoekstelsel dan is OAI-PMH de standaard die de daarvoor benodigde gegevensuitwisseling mogelijk maakt.

Een voorbeeld van deze toepassing is Edurep van Kennisnet: een zoekstelsel voor educatieve content. Meerdere uitgeverijen en andere leveranciers van educatief materiaal zijn daar op aangesloten. Leerkrachten kunnen in Edurep zoeken naar al deze content, zonder daarvoor te moeten zoeken in de individuele digitale bibliotheken van de aangesloten partijen.

Ook in het veld van cultureel erfgoed wordt OAI-PMH al ingezet. Het gaat bijvoorbeeld om het Nederlands Instituut voor Beeld en Geluid, het Rijksmuseum Amsterdam, Naturalis en de Koninklijke Bibliotheek.

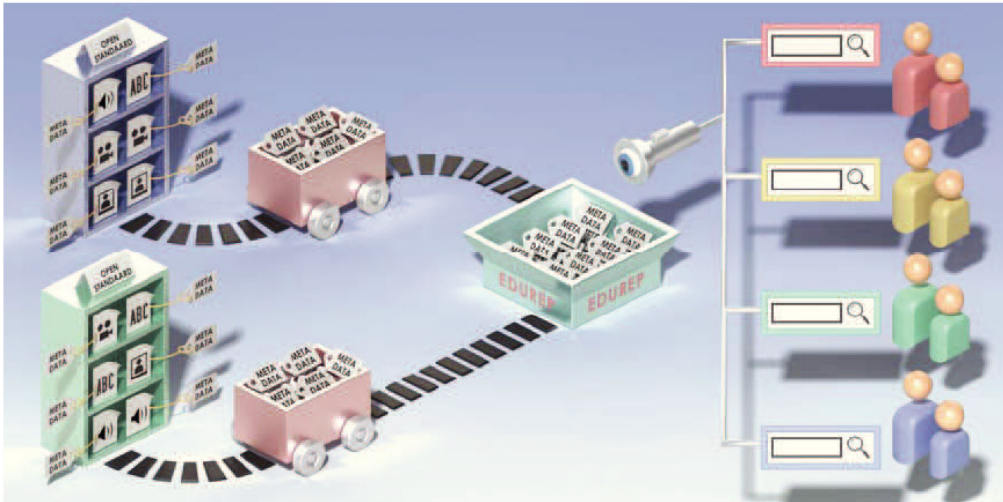
Waar gaat het inhoudelijk over?

OAI-PMH verzamelt metadata uit meerdere digitale bibliotheken (titel, auteur, thema, etc.) en maakt het mogelijk deze metadata onder te brengen in een centraal zoekstelsel.

De standaard definieert een aantal 'vragen' die aan een digitale bibliotheek gesteld kunnen worden. Dit varieert van 'welke content zit er in deze bibliotheek?' tot 'geef mij de metadata van deze verzameling documenten'. De bibliotheek levert vervolgens deze metadata aan het zoekstelsel. Het zoekstelsel kan dit vervolgens weer opnemen in een totaalindex, waarin gebruikers kunnen zoeken.

Onderstaande figuur geeft dit grafisch weer:

- Links staan twee digitale bibliotheken. Aan de digitale content in die bibliotheken hangt metadata (de labeltjes)
- OAI-PMH – de rode karretjes – zorgt voor het ophalen van deze metadata
- Het centrale zoekstelsel (de groene bak rechts) verzamelt deze metadata.
- Gebruikers (rechts) kunnen zoeken in deze verzameling metadata.



Datum
1 oktober 2010

Figuur: Schets van OAI-PMH: de standaard verzamelt metadata uit meerdere digitale bibliotheken (bron: Kennisnet)

Twee belangrijke opmerkingen hierbij:

- OAI-PMH is generiek, dat wil zeggen: het is niet toegesneden op een bepaalde sector of een bepaald domein. Het kan met alle vormen van metadata³ worden gebruikt.
- OAI-PMH specificeert *niet* hoe de digitale content uiteindelijk door de gebruiker uit de onderliggende bibliotheek wordt gehaald en welk formaat de digitale content heeft. Hiervoor zijn andere standaarden nodig (zoals bijv. FTP en PDF/A).

Zijn er alternatieven voor de voorgestelde keuze?

Gegeven de sterke focus van het toepassingsgebied zijn er geen direct voor de hand liggende alternatieven.

Wel kunnen organisaties er voor kiezen om het zoeken in digitale bibliotheken op een andere manier in te richten. Bijvoorbeeld:

- Men zou kunnen kiezen voor het inrichten van één digitale bibliotheek voor alle partijen. De noodzaak om overkoepelend te zoeken ontbreekt dan.
- Er kan gekozen worden voor een andersoortig mechanisme van gegevensuitwisseling. e-Overheid voor Burgers indexeert met haar zoekdienst bijvoorbeeld websites van gemeenten, zonder het vraag/antwoordspel zoals beschreven in de vorige paragraaf.

Deze alternatieve architecturen blijven mogelijk, ook wanneer OAI-PMH opgenomen wordt op de lijst. Hiermee wordt recht gedaan aan eventuele specifieke situaties in bepaalde sectoren of domeinen.

Schets van de expertgroep en de consultatie

De expertgroep bestond uit 10 leden, afkomstig uit de onderwijswereld, organisaties op het gebied van digitale duurzaamheid en uitgeverijen. In een bijeenkomst op 8 december 2009 heeft deze groep de standaard getoetst aan de vier gestelde criteria. Op alle criteria werd positief geoordeeld. Het expertadvies is gereviewed door Herbert van de Sompel, werkzaam bij Los Alamos National Laboratory (VS) en lid van het executive committee van het Open Archives Initiative (de beherende organisatie).

³ De minimum eis is dat de metadata voldoet aan de veelzijdige Dublin Core standaard.

Datum
1 oktober 2010

Naast de uiteindelijke leden van de expertgroep is een grotere groep overige – mogelijk betrokken - organisaties gevraagd om deel te nemen. Uit deze groep zijn vooraf enkele reacties ontvangen, die zijn betrokken bij het opstellen van het expertadvies.

In de consultatieronde is een reactie ontvangen van 16 partijen. 8 partijen reageerden neutraal of positief, 8 partijen reageerden negatief of hadden specifieke kritiekpunten. Deze kritiek spitste zich toe op drie inhoudelijke punten ten aanzien van OAI-PMH.

Het Forum Standaardisatie heeft vervolgens besloten om additioneel onderzoek te laten verrichten naar deze drie punten (zie de toelichting). Op basis van dit additioneel onderzoek is het toepassingsgebied verduidelijkt, waarmee recht wordt gedaan aan de zorgen die in de consultatieronde werden geuit.

Mogelijke consequenties van opname op de lijst met standaarden

Met de opname van OAI-PMH wordt bijgedragen aan het ontwikkelen van een infrastructuur in Nederland op het gebied van digitale bibliotheken en archieven. Geïntegreerd zoeken via metadata, door gebruik te maken van OAI-PMH, kan gezien worden als een belangrijke bouwsteen op dit vlak.

Opname op de lijst kan betekenen dat organisaties worden aangezet om verdere stappen te zetten op dit vlak.

Communicatie

Zowel het Forum Standaardisatie als het Programmabureau Nederland Open in Verbinding zullen aandacht besteden aan de opname van OAI-PMH op de lijst met standaarden.