

Metadatauitwisseling met OAI-PMH

Protocol en implementatie

Inleiding

Inmiddels is het al meer dan tien jaar geleden dat de eerste versie van het Open Archives Initiative Protocol for Metadata Harvesting gelanceerd werd. Het protocol voor ‘metadata harvesting’ heeft in die tijd een grote gebruikersgroep gevonden, met name in de wereld van culturele instellingen en archieven die hun materiaal beschikbaar stellen via internet. Het ‘harvesten’ is inmiddels bij vele instellingen de meest voor de hand liggende manier om metadata binnen te halen en binnen een eigen internetapplicatie beschikbaar te stellen. In Nederland gebruiken vooraanstaande culturele instellingen als het Nationaal Archief en het Rijksmuseum dit protocol om hun metadata toegankelijk te maken voor geïnteresseerden, die deze data vervolgens kunnen hergebruiken binnen eigen websites en platforms.

Wat begon als een project om onderzoeksresultaten centraal doorzoekbaar te maken binnen een relatief kleine gemeenschap is uitgegroeid tot onmisbare functionaliteit voor websites over de hele wereld die voor hun data uit vele bronnen putten en gezamenlijk complexe contentketens vormen met OAI-PMH als basis voor de uitwisseling van metadata.

Ter inleiding van dit artikel zal een korte schets gegeven worden van wat het OAI-PMH-protocol is en hoe het ontstaan is. Vervolgens wordt ingegaan op de functionaliteit binnen het protocol en op de keuzes voor de implementatie ervan. Hierbij worden ook zaken genoemd die belangrijk zijn om in gedachten te houden bij het werken met dit protocol, mede aan de hand van de ‘best practices’ die de afgelopen jaren zijn geformuleerd binnen diverse projecten. We zullen zien dat het vermijden van reeds erkende valkuilen in grote mate bijdraagt aan de bruikbaarheid van de aangeboden metadata voor potentiële consumenten. Er kan bijvoorbeeld veel tijd gaan zitten in het normaliseren van verschillen die voortkomen uit verschillende opvattingen rondom ‘metadatamapping’ of interpretatie van de standaard. Het is daarom goed om bewust om te gaan met de metadata, of het nu gaat om het aanbieden van data of het binnenhalen ervan. Ook zal aan de hand van een usecase beschreven worden hoe geavanceerde ketens van metadatauitwisseling kunnen ontstaan met behulp van dit op zichzelf eenvoudige protocol. Tenslotte wordt kort vooruitgeblikt naar mogelijke nieuwe ontwikkelingen en hun betekenis voor de toekomst van OAI-PMH.

Wat is OAI-PMH?

Het Open Archives Initiative Protocol for Metadata Harvesting is een protocol ontwikkeld door de organisatie ‘Open Archives Initiative’, die zich bezighoudt met het ontwikkelen en promoten van standaarden om beschikbaarstelling van digitale content te vergemakkelijken. Het OAI-PMH protocol is een manier voor instellingen om metadata over content beschikbaar te stellen, dan wel binnen te halen via het web. Binnen het kader van OAI-PMH bestaan er twee soorten partijen: ‘service providers’ en ‘data providers’. Data providers zijn instellingen die digitaal materiaal en de daarmee geassocieerde metadata beschikbaar stellen op het internet binnen een zogenaamde ‘repository’ ofwel een verzameling metadata over digitale bronnen. Zij bieden de functionaliteit van de OAI-provider aan service providers aan. Service providers zijn de partijen die deze metadata binnenhalen met een OAI-harvester om deze binnen een eigen omgeving te hergebruiken. Zoals we later zullen zien komt het in de praktijk natuurlijk ook voor dat partijen beide rollen tegelijk vervullen, bijvoorbeeld doordat zij metadata binnenhalen, bewerken en de aangepaste metadata vervolgens ook zelf weer beschikbaar stellen via een eigen OAI-provider. Voor het gemak zal in de rest van het artikel gesproken worden over ‘providers’ en ‘harvesters’ voor respectievelijk data providers en service providers.

De beschrijving van de functionaliteit om zowel OAI-providers als OAI-harvesters mogelijk te maken ligt vastbesloten in het OAI-PMH-protocol.

Het is in dit kader belangrijk om een vaak gehoord misverstand meteen weg te nemen. Ondanks de relatie met metadata en beschikbaarstelling van bronnen biedt OAI-PMH geen zoekfunctionaliteit. Het is de bedoeling dat een harvester de metadata van een provider zelf opslaat in een eigen database. Het is belangrijk om te beseffen dat het geen protocol is om gedistribueerde zoekopdrachten op

verschillende externe databases mee uit te voeren, zoals bijvoorbeeld SRU (Search/Retrieve via URL) dit wel is. Dit is echter geen beperking maar was in de ontstaansgeschiedenis van OAI-PMH juist een bewuste keuze.

De ontwikkeling van OAI-PMH

Het OAI-PMH protocol is ontstaan binnen de gemeenschap die zich bezighield met de zogenaamde E-prints-repositories. Dit waren archieven van resultaten van wetenschappelijk onderzoek en wetenschappelijke artikelen. Deze repositories kenden zowel verschillende manieren om nieuw materiaal toe te voegen als verschillende manieren om de content te vinden. Hierdoor moesten gebruikers vele zoekinterfaces leren gebruiken om materiaal te kunnen vinden. Eind jaren '90 werd bij een bijeenkomst in Santa Fe door een groep van technische experts uit deze gemeenschap onderkend dat deze verscheidenheid aan interfaces een probleem was.. Een ander probleem met de toenmalige architectuur van de repositories was dat er geen manier was om metadata te delen zonder dat menselijk handwerk nodig was.

Er werden twee oplossingsrichtingen besproken om dit probleem aan te pakken. De ene was het distribueren van een centrale zoekopdracht over meerdere databases, de andere oplossing was: alle metadata binnen één database opslaan en die database gebruiken om een zoekinterface op aan te sluiten. Bij overweging van de eerste oplossing werd een aantal nadelen onderkend gebaseerd op reeds bestaande ervaring met dit type zoekopdrachten. Een van de nadelen binnen dit scenario is namelijk dat de snelheid van de zoekopdracht afhankelijk is van de langzaamste database in de keten. Ook de verschillende attributen waarop gezocht moest kunnen worden en de verschillen in formulering van de zoekopdracht zouden problemen gaan opleveren. Tenslotte zou de presentatie van de verschillende soorten zoekresultaten moeilijk worden wanneer daarbij rekening gehouden zou moeten worden met zaken als 'ranking'. Deze obstakels zorgden ervoor dat men koos voor de tweede optie, waarmee de basis was gelegd voor de uitgangspunten binnen het OAI-PMH-protocol.

Binnen dit nieuwe protocol moest een aantal zaken geüniformeerd worden om de overdracht van metadata mogelijk te maken op een manier die door machines verwerkt zou kunnen worden. Dit gold onder andere het transport en het formaat van de metadata. Voor dit laatste werd in de eerste officiële publicatie van het OAI-PMH protocol gekozen voor Dublin Core als metadatastandaard, verpakt in XML. Als transportmechanisme werd gekozen voor het HTTP-protocol.¹

In 2002 werd de 2.0-versie van het protocol gepubliceerd. Deze versie is ook nu nog de meest recente versie en is in de praktijk de enige implementatie van het protocol die op grote schaal wordt gebruikt.

De bouwstenen: HTTP, XML en metadata

De genoemde designkeuzes met betrekking tot transport en formaat van de data bieden een vrij toegankelijk en laagdrempelig protocol voor mensen die al enige ervaring hebben met het uitwisselen van metadata. De data wordt binnen het OAI-PMH-protocol getransporteerd over HTTP net als bij het opvragen van een webpagina. Dit betekent dat het harvesten van een OAI-provider in zijn meest eenvoudige vorm al mogelijk is door middel van het intypen van een url in een webbrowser. Het responsformaat lijkt daarentegen niet op die van een webpagina, de data wordt namelijk verpakt in XML in plaats van in HTML. Het protocol is dan ook in de eerste plaats gericht op geautomatiseerd gebruik tussen systemen onderling zonder tussenkomst van een gebruiker.

Binnen het OAI-PMH-protocol is de gehele interface voor zowel client als server beschreven. Belangrijk binnen het principe van deze interface is het idee dat bepaalde bronnen binnen de repository afzonderlijk benoemd en geïdentificeerd zijn en dat de representatie van deze bronnen, (namelijk de metadata van het object), uitgewisseld kan worden binnen de beschreven interface. Het is

¹ Carpenter, OAI for Beginners, the Open Archives Forum online tutorial, Hfst. 2: History and development of OAI-PMH, <http://www.oaforum.org/tutorial/english/page2.htm>

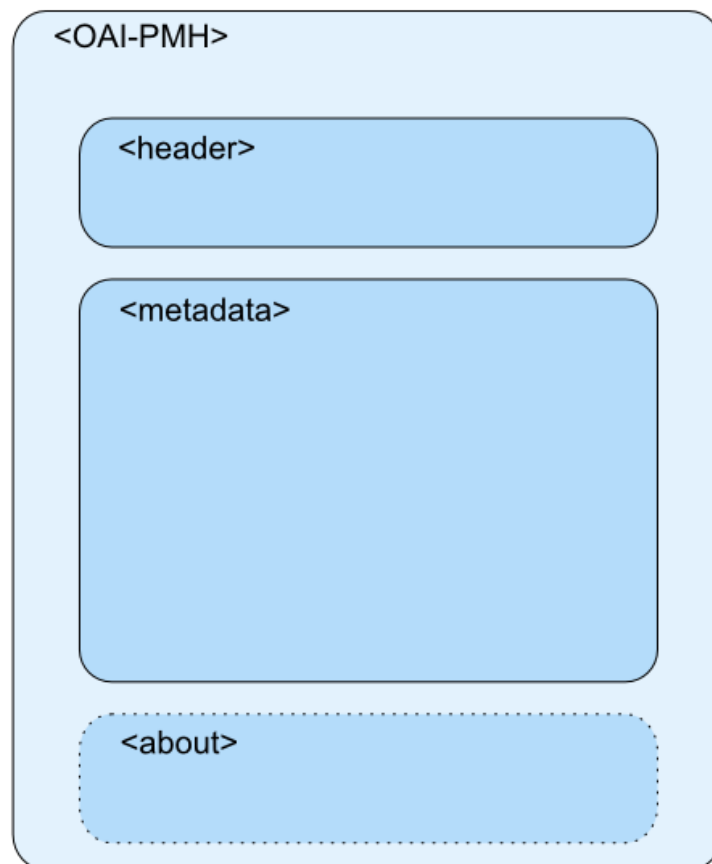
hierbij dus belangrijk te bedenken dat niet de bronnen zelf uitgewisseld worden maar een of meerdere representaties hiervan, weergegeven in een of meerdere metadataformaten.

Aangaande metadatastandaarden is binnen het protocol als enige specifieke eis gesteld dat elke OAI-provider tenminste in staat moet zijn om metadata uit te leveren in 'Dublin Core'-formaat. Naast dit formaat is er geen beperking gesteld aan de hoeveelheid en soorten metadatastandaarden die als alternatief voor Dublin Core aangeboden mogen worden door een OAI-provider. Wel is het verplicht dat elke metadatastandaard verwijst naar een xml-schema waar de betreffende metadatastandaard aan voldoet.²

Wat voor alle metadataformaten geldt is dat elke bron die als losse bron binnen een repository onderscheiden wordt als een aparte record uitgeleverd moet kunnen worden. Een record bestaat binnen OAI-PMH altijd uit een header-deel en een metadata-deel. In de header zijn enkele administratieve gegevens over de record opgenomen zoals een unieke identifier en een datum die geassocieerd is met het record. Dit kan een datum zijn waarop het record is aangemaakt, gewijzigd of verwijderd.

Het belang van het associëren van een record met een datum hangt ook samen met het principe van het selectief harvesten binnen OAI-PMH. Dit betekent dat een OAI-harvester alleen records opvraagt die gemaakt of gewijzigd zijn vanaf een bepaalde datum, tot een bepaalde datum of binnen een bepaalde tijdsspanne. Dit zorgt voor een efficiënt overdrachtsmechanisme, aangezien de harvester maar één keer alle informatie hoeft op te halen en voor de rest alleen binnen een bepaalde termijn updates op deze data hoeft te verwerken. Dit betekent dat er doorgaans per updatesessie minder data uitgewisseld hoeft te worden dan bij een reguliere, volledige harvest.

Het metadata-deel is de container voor de eigenlijke metadata van het object zelf, in het gespecificeerde metadataformaat. Een record kan verder nog een optioneel about-element bevatten, waarover later meer.



Figuur 1: Schematische weergave van een OAI-PMH-record.

²The Open Archives Initiative Protocol for Metadata Harvesting, 3 Protocol Features, <http://www.openarchives.org/OAI/openarchivesprotocol.html#ProtocolFeatures>

Het protocol

Wat is er nu eigenlijk precies mogelijk met OAI-PMH? Het antwoord op deze vraag is binnen het protocol vormgegeven aan de hand van de zogenaamde ‘verbs’ die de daadwerkelijke opdrachten van het protocol uitmaken. Deze ‘verbs’ kunnen beschouwd worden als de commando’s die een harvester kan meegeven in het verzoek aan een provider met het doel gericht de gewenste informatie op te kunnen halen.

OAI-PMH kent de volgende zes verbs waarvan de namen waarschijnlijk al grotendeels voor zich spreken: Identify, ListMetadataFormats, ListSets, ListIdentifiers, ListRecords en GetRecord.

De Identify-verb vervult de basale rol van het beschrijven van de repository zelf. Deze zal - in een geautomatiseerde workflow- dan ook niet vaak een centrale rol spelen maar is meer bedoeld om informatie te geven die voor harvesters belangrijk is om rekening mee te houden in de communicatie met de OAI-provider. Een belangrijk element dat in de Identify-respons opgenomen moet worden is bijvoorbeeld hoe er binnen de repository omgegaan wordt met verwijderde records. Dit is voor de harvester belangrijk om te weten omdat er binnen OAI-PMH drie opties zijn gedefinieerd om hier mee om te gaan, elk met technische consequenties voor het proces van harvesting. Deze opties zullen hieronder uitgebreider beschreven worden.

Een andere verb die gericht is op het beschrijven van de mogelijkheden van de repository zelf is ‘ListMetadataFormats’. Zoals uit de naam al blijkt wordt deze verb gebruikt om aan te geven in welke metadataformaten de objecten binnen een repository opvraagbaar zijn.

De verbs die gericht zijn op het opvragen van de eigenlijke metadata van de bronnen binnen een repository zijn de verbs ‘ListRecords’ en ‘GetRecord’. GetRecord kan gebruikt worden voor het opvragen van de metadata van een enkel object door de identifier van het object mee te geven in het GetRecord-verzoek. ListRecords maakt eigenlijk de daadwerkelijke harvestactie uit van het OAI-PMH-protocol daar deze verb het mogelijk maakt de metadata van alle bronnen van een repository op te vragen. Met ListIdentifiers is het tenslotte mogelijk om alleen de headers van een metadata-record op te vragen.

Behalve deze kernfunctionaliteit bevat het protocol nog enkele optionele elementen die zeer nuttig kunnen zijn voor instellingen die een grote en gevarieerde collectie aan data beschikbaar willen stellen. Gezien het feit dat een ListRecords-verzoek er voor zorgt dat een export van alle objecten in een repository wordt gegenereerd door een OAI-provider is er binnen het protocol gedefinieerd dat een provider zogenaamde ‘resumption tokens’ mag verstrekken bij een verzoek. Dit betekent dat niet alle records in één keer worden uitgeleverd maar dat de provider zelf kan kiezen hoeveel records per keer uitgeleverd worden. De gebruiker krijgt een resumption token en moet deze weer meesturen in zijn volgende verzoek om de volgende lading records op te halen, net zo lang tot alle records binnen de repository uitgeleverd zijn. Dit procedé zorgt ervoor dat de respons een kleinere omvang en een snellere verwerkingstijd heeft.

Een ander meer inhoudelijk georiënteerd principe binnen OAI-PMH met betrekking tot het omgaan met grote hoeveelheden data, is de functionaliteit van ‘sets’. Een set binnen OAI-PMH is een deelverzameling objecten binnen een repository. De manier waarop deze indeling tot stand komt is ter vrije invulling van de OAI-provider en kan zowel uit een hiërarchie van collecties en subcollecties bestaan als uit een platte lijst.

Een archief kan bijvoorbeeld de OAI-provider zo inrichten dat deze sets op basis van trefwoord maakt, zodat een harvester alleen die onderwerpen kan binnenhalen die voor zijn interessegebied van belang zijn. Sets worden in de praktijk ook vaak gebruikt om deelcollecties binnen een archief afzonderlijk van elkaar te ontsluiten. Een andere vorm waar aan te denken valt is het aanmaken van sets op basis van het type object binnen de repository. Om het voor de harvester zo inzichtelijk mogelijk te maken

welke criteria zijn gekozen voor de indeling in sets is het mogelijk om aan een set een apart beschrijvingsselement mee te geven.³

Waarom OAI-PMH?

Nu dat we de ontstaansgeschiedenis en de basisfunctionaliteit hebben beschreven zou de vraag kunnen rijzen, wat is het praktische nut van OAI-PMH? Met andere woorden, wat zou een typisch scenario zijn dat we ons zouden kunnen voorstellen bij inzet van deze techniek? Waarom zouden we eigen data willen aanbieden aan anderen en waarom zouden we andermans data willen binnenhalen?

Voor providers geldt vaak dat ze hun materiaal weliswaar hebben gedigitaliseerd en beschikbaar hebben gesteld op een eigen website, maar dat het in de praktijk niet makkelijk genoeg door gebruikers gevonden wordt of dat het daadwerkelijke gebruik ervan laag is. In zo'n geval kunnen deze providers profiteren van de verspreiding van hun materiaal buiten de context van hun eigen website. Niet alleen vergroot dit de vindbaarheid van het materiaal doordat het vanaf meerdere websites te benaderen is. Door de extra zichtbaarheid worden ook nieuwe doelgroepen bereikt. Vernieuwend gebruik binnen een andere website of door middel van een app kan daardoor toenemen. Dit laatste speelt bijvoorbeeld een rol binnen het initiatief Open Cultuur Data dat organisaties stimuleert om digitale content beschikbaar te maken en nieuwe toepassingen met deze content te realiseren.⁴ Het kan ook zijn dat een instelling het delen van erfgoed ziet als een kernactiviteit. Op de site van het Rijksmuseum, waar ook een OAI-provider aangeboden wordt, valt bijvoorbeeld te lezen "Het Rijksmuseum verbindt mensen, kunst en geschiedenis. Ons doel is om de collectie beschikbaar te stellen en dichterbij het publiek te brengen."⁵

Vanuit de optiek van de contentprovider biedt OAI-PMH kortom, een mogelijkheid om een bredere doelgroep te bereiken en nieuwe vormen van gebruik te stimuleren.

Hoe zit het met de harvesters? Wat voor nut heeft het om een externe collectie beschikbaar te stellen via een eigen website?

Een belangrijke en voor de hand liggende mogelijkheid die dit harvesters biedt is het centraal doorzoekbaar maken van materiaal met een gemeenschappelijk thema of doel. Een voorbeeld van een dergelijke repository is de dienst Edurep van Kennisnet; deze dienst verzamelt door middel van OAI-PMH bronnen van instellingen en uitgeverijen met als gemeenschappelijk thema materiaal dat relevant is voor het onderwijs. Deze dienst heeft geen eigen webinterface maar stelt door middel van webservices andere partijen in staat om eigen zoekportals te ontwikkelen voor het zoeken naar onderwijsmateriaal. Ook een grote partij als OCLC biedt met Worldcat en OAIster dergelijke diensten maar dan voor een internationaal publiek en meer gericht op hoger onderwijs. Door middel van een dergelijke dienst is het mogelijk om verschillende typen materiaal binnen één catalogussysteem doorzoekbaar te maken, bijvoorbeeld binnen de mediatheek van een universiteit.

Een andere reden om als harvester OAI-PMH te gebruiken is om eigen diensten aantrekkelijker te maken door het aanbieden van externe content. Sommige partijen bieden namelijk online tools en portals aan maar hebben zelf geen collectie om aan gebruikers aan te kunnen bieden als verwerkingsmateriaal. Elektronische leeromgevingen zoals die tegenwoordig in het onderwijsveld worden gebruikt zijn hiervan een goed voorbeeld. Deze omgevingen bevatten vaak de mogelijkheid om eigen lesmateriaal te ontwikkelen maar hebben niet altijd de beschikking over grote hoeveelheden eigen beeldmateriaal of multimedia om bij het maken van lesmateriaal te gebruiken. OAI-PMH kan voor dit soort partijen een gestandaardiseerde manier bieden om extra bronmateriaal te kunnen leveren aan gebruikers. Uiteraard zijn andere scenario's om materiaal te importeren ook denkbaar. De instellingen waarvan materiaal in een dergelijke omgeving beschikbaar wordt gesteld kunnen hiervan op hun beurt weer profijt hebben doordat zij wellicht zelf niet de middelen hebben om tools speciaal gericht op het onderwijs te ontwikkelen.

³ Het protocol in zijn geheel is beschreven in: The Open Archives Initiative Protocol for Metadata Harvesting: Document Version 2008-12-07T20:42:00Z, <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁴ <http://www.opencultuurdata.nl/waarom-open-cultuur-data/>

⁵ <https://www.rijksmuseum.nl/api/uitleg>

Best practices rondom implementatie van OAI-PMH

Doordat OAI-PMH voor ICT-begrippen alweer een tijd meegaat is er in de loop der jaren veel kennis opgebouwd rondom het gebruik en de implementatie van het protocol. Er is onderzoek gedaan naar verschillende protocollen om data beschikbaar te stellen en naar de manier waarop OAI-PMH geïmplementeerd is bij verschillende instellingen. Hierbij zijn er verscheidene beperkingen en nadelen onderkend en op basis daarvan zijn er best practices en adviezen geformuleerd voor mensen die dit protocol willen gaan gebruiken in hun diensten.

Het uitgebreid behandelen van de resultaten en conclusies van al deze onderzoeken voert wellicht te ver voor dit artikel; enkele conclusies uit deze onderzoeken bieden echter belangrijke lessen die tijd en geld kunnen besparen bij implementatie van een provider of harvester. De organisatie 'Open Archives Initiative' zelf heeft - nog voor de publicatie van versie 2.0 van OAI-PMH- een document met implementatierichtlijnen doen uitgaan. Dit document bestaat uit richtlijnen voor providers, voor harvesters en voor aggregaten. Tenslotte zijn er verschillende artikelen en blogposts op het net te vinden waarin mensen hun ervaringen met OAI-PMH hebben gedeeld. Enkele nuttige adviezen op basis van al deze gegevens zullen hieronder beschreven worden.

De keuze voor OAI-PMH

Het beste is om te beginnen bij de keuze voor OAI-PMH. Het is belangrijk om te begrijpen welke rol een OAI-provider kan spelen binnen een bepaalde systeemarchitectuur en waar de techniek zich minder goed voor leent. Omdat selectief harvesten uitermate geschikt is voor het ontvangen van updates wordt de techniek soms vergeleken met de wellicht bekendere techniek van de RSS-feed. Een belangrijk verschil is echter dat OAI-PMH ontworpen is voor het delen van metadata-records van losstaande digitale objecten, terwijl RSS ook gebruikt kan worden voor het ontvangen van updates over bijvoorbeeld nieuwsberichten of websitewijzigingen. Dit zijn dus stukken content die zelf geen losstaand digitaal object met metadata vormen. OAI-PMH is in dit opzicht dus met name interessant voor instellingen die een collectie van digitale objecten inclusief metadata beheren en minder voor instellingen met websites waarbij de site zelf de content is. De nadruk op gestandaardiseerde metadata bij OAI-PMH maakt dat het protocol uiteindelijk niet precies te vergelijken is met RSS.

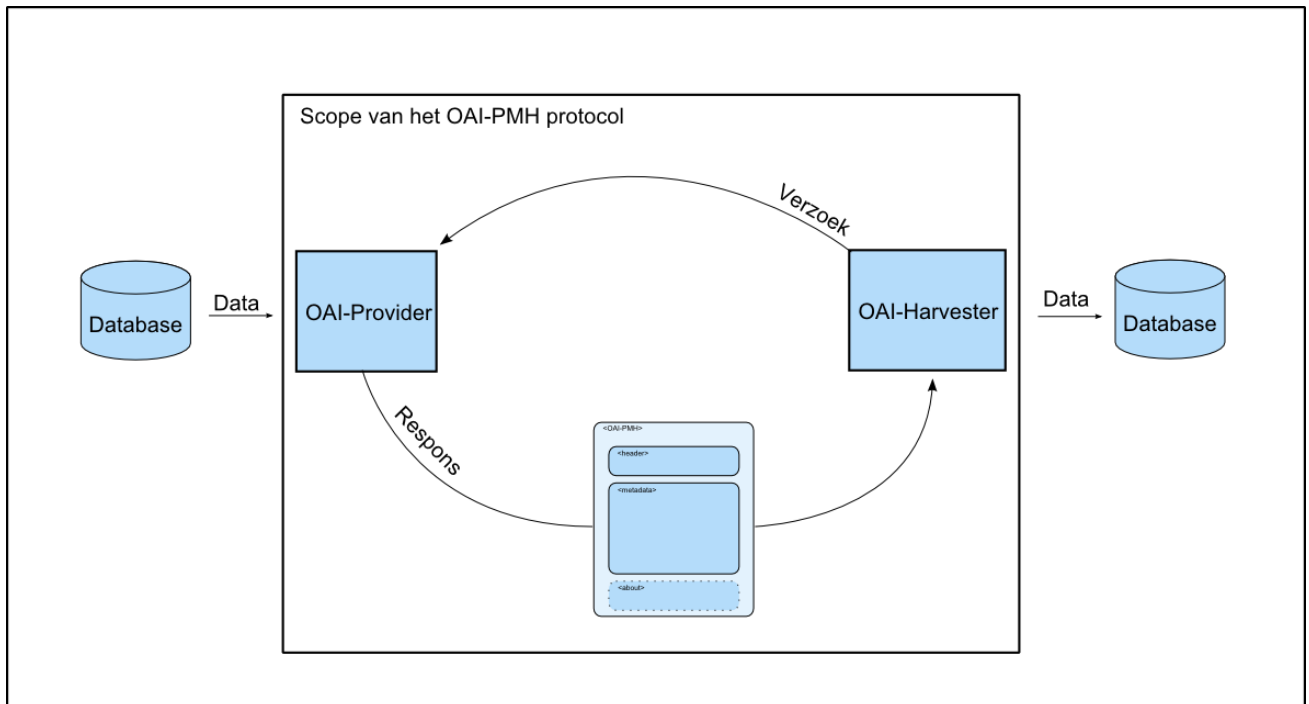
Ook is het goed om te beseffen dat een OAI-provider geen uitkomst biedt wanneer men op zoek is naar een webservice die externe portals de mogelijkheid geeft live-zoekopdrachten te doen. Het OAI-PMH-protocol biedt immers geen mechanismes om zoekopdrachten te formuleren. Instellingen die hiernaar op zoek zijn kunnen beter gebruik maken van het eerder genoemde SRU. Harvesters zullen altijd de metadata moeten binnenhalen en zelf doorzoekbaar maken binnen een eigen database. Vanuit de filosofie van OAI-PMH is dit geen beperking maar juist een voordeel doordat de snelheid van de zoekopdracht op die manier niet gekoppeld is aan de snelheid van de site waar de metadata vandaan komt en ook het oorspronkelijke systeem daarmee minder belast wordt.

Voor een instelling die haar collecties beschikbaar wil stellen is een goed uitgangspunt om beleid op te stellen rondom de implementatie en het onderhoud van een provider. Het moet duidelijk zijn wat de rol wordt van deze techniek en hoe een en ander past binnen de digitale infrastructuur en de missie van de instelling in het algemeen. Beleid geeft daarmee ook richting aan de inhoud en de omvang van de collectie zoals deze beschikbaar gesteld wordt.⁶ Voor de kwaliteit van de service en de metadata is het van belang dat het beleid van de instelling waarde hecht aan het delen van data. Wanneer dit niet het geval is bestaat de kans dat er niet goed gereageerd kan worden op technische problemen of dat er onvoldoende documentatie beschikbaar is rondom de dienst.⁷ Wanneer een OAI-PMH provider voor een instelling slechts een beperkte waarde heeft zal dit een negatief effect hebben op potentiële

⁶ Weenink e.a., A DRIVER's Guide to European Repositories 2007, <http://dare.uva.nl/aup/en/record/260224>, blz. 60

⁷ Muriel Foulonneau and Francis André, Investigative Study of Standards for Digital Repositories and Related Services 2007, <http://dare.uva.nl/aup/nl/record/260224>, blz 25

harvesters. Met een provider die geen prioriteit geeft aan beschikbaarstelling kunnen harvesters niet op een betrouwbare manier nieuwe diensten ontwikkelen. Het is belangrijk dat instellingen die willen fungeren als provider hun implementatie zoveel mogelijk conform standaarden inrichten. Doordat providers vaak niet weten wie hun data gebruikt is het belangrijk dat de verschillende implementatiekeuzes gedocumenteerd zijn en op een transparante manier inzichtelijk gemaakt kunnen worden aan harvesters.



Figuur 2: Een typisch harvest-scenario. OAI-PMH vormt de communicatielaag voor de overdracht van metadata van de ene database naar de andere.

Deletes

Veel best practices rondom OAI-PMH zijn geformuleerd om nadere invulling te geven aan de vele optionele zaken die in het protocol staan gedefinieerd. Een van die zaken betreft de delete-functionaliteit. Binnen het protocol bestaan er drie verschillende niveau's die gebruikt kunnen worden om aan harvesters duidelijk te maken dat een bepaalde record verwijderd is. De meest simpele vorm is om de record niet meer aan te bieden. Dit heeft als nadeel dat harvesters eigenlijk geen mogelijkheid hebben om vast te stellen dat een bepaalde record verwijderd is anders dan de hele collectie binnen te halen en de nieuwe set met de oude te vergelijken. Dit is een moeizaam en intensief proces dat in de praktijk waarschijnlijk niet snel geïmplementeerd zal worden. De tweede optie is om gedurende een bepaalde tijd door te geven dat een record verwijderd is, zodat harvesters de kans hebben dit te ontdekken. Na verloop van tijd wordt het record dan stilzwijgend weggelaten uit de collectie. Dit scenario gaat ervan uit dat harvesters na een bepaalde tijd de update wel zullen hebben verwerkt en is met name geschikt voor providers die zeker weten dat al hun gebruikers regelmatig updates ophalen. De derde mogelijkheid bestaat uit het voor onbepaalde tijd doorgeven dat een record verwijderd is. Dit is eigenlijk in alle gevallen de meest klantvriendelijke optie, aangezien hier de minste eisen worden gesteld aan de harvester. Voor harvesters is het uiteraard van belang om iets te doen met de melding dat een record verwijderd is, bijvoorbeeld door de verwijdering ook in de eigen database door te voeren. Het verwijderen van records is meestal niet het eerste waaraan gedacht wordt bij het ontwikkelen van een mechanisme dat tot doel heeft om records juist beschikbaar te maken. Het is echter doorgaans makkelijker om met deze zaken tijdens de implementatie al rekening te houden dan om ze na de implementatie te moeten wijzigen.

Sets

Een ander optioneel element binnen het protocol is het definiëren van sets. Een set wordt binnen het protocol beschreven als een optionele constructie voor het groeperen van objecten met als doel selectief te kunnen harvesten. Anders gezegd: het is een manier om een grote collectie op te kunnen delen in verschillende deelcollecties zodat harvesters een deelcollectie naar keuze kunnen binnenhalen.

In het protocol staat verder gespecificeerd hoe deze sets hiërarchisch gestructureerd kunnen worden zodat er verschillende deelniveau's aangeboden kunnen worden. Het protocol bevat echter geen richtlijnen over de inhoudelijke vraag wat een set uitmaakt, anders dan de suggestie dat hierover binnen individuele communities wellicht afspraken gemaakt kunnen worden.⁸ De vaagheid van deze definitie is inmiddels een erkend probleem, evenals de gevolgen die deze vaagheid kan opleveren: instellingen stellen soms sets samen die alleen voor intern gebruik van betekenis zijn.⁹ Bij implementatie is het dus belangrijk om goed na te denken over consistente manieren van objecten opdelen in sets op een zodanige manier dat ze ook voor externe partijen van nut kunnen zijn. Ook is het belangrijk om het aantal sets niet onoverzichtelijk groot te maken omdat dit de werkbaarheid van het listSets-commando vermindert.¹⁰ In de officiële richtlijnen van OAI wordt aangeraden om sets te voorzien van een beschrijving. Uit een onderzoek van Muriel Foulonneau e.a. in het kader van het Europese DRIVER-project naar de OAI-inplementaties in vijf Europese landen bleek bijvoorbeeld dat indeling op basis van instelling/afdeling, onderwerp en bestandstype de meest voorkomende zijn.¹¹ Indeling op basis van onderwerp kan voor externe harvesters interessant zijn, maar indeling op basis van bestandstype kan al snel tot onduidelijkheid leiden, bijvoorbeeld wanneer het bestandstype als html wordt weergegeven omdat het een foto op een webpagina betreft.

Voor harvesters is het belangrijk om rekening te houden met het feit dat binnen het protocol geen methode is gedefinieerd om mee te delen dat een record uit een bepaalde set is gehaald. Het is mogelijk, door middel van 'deletes', om door te geven dat een object uit de repository is verwijderd, maar als een object slechts uit een deelcollectie is gehaald maar nog wel beschikbaar is binnen de gehele repository dan biedt het protocol geen mogelijkheid dit aan een harvester door te geven.

Metadata

Wat betreft metadata laat het protocol veel ruimte voor eigen invulling. Het enige dat hieromtrent verplicht gesteld is betreft het feit dat alle data tenminste in Dublin Core beschikbaar moet zijn en dat andere metadataformaten aan een vast schema moeten voldoen. Eerstgenoemde eis betekent in de praktijk vaak dat Dublin Core ook meteen het enige metadataformaat is dat repositories aanbieden.¹² Dublin Core is vanwege de bekendheid en overzichtelijkheid van het formaat populair voor uitwisseling, maar het is vaak niet het formaat dat instellingen zelf gebruiken als metadatastandaard voor hun interne catalogus. Dit maakt dat in de praktijk vaak rijkere data wordt 'platgeslagen' tot Dublin Core, hetgeen ten koste kan gaan van de begrijpelijkheid van de data voor een harvester. Een voorbeeld is het DC-veld 'coverage'. Dit element is binnen de standaard bedoeld voor zowel spatiële als temporele aanduidingen. Dit betekent dat binnen hetzelfde formaat zowel de term 'middeleeuwen' als de term 'Amsterdam' voor kan komen. De meeste archieven zullen in hun eigen catalogus plaatsnamen en tijdsaanduidingen niet onder dezelfde noemer opslaan. De simpele opzet van Dublin Core dwingt de aanbieder om deze inhoudelijk zeer verschillende velden binnen hetzelfde element te plaatsen. Dit probleem kan echter eenvoudig opgelost worden door naast Dublin Core ook een

⁸ <http://www.openarchives.org/OAI/openarchivesprotocol.html#Set>

⁹ Tennant, Roy. Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers, http://web.archive.org/web/20090607100835/http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

¹⁰ Ibid.

¹¹ Foulonneau, Muriel, Francis André, and Anne-Marie Badoloato, 'Digital repositories infrastructure vision for Europe research – Review of technical standards.' Report, DRIVER project, 2007, http://www.driver-support.eu/documents/DRIVER_Review_of_Technical_Standards.pdf, blz 13

¹² Tennant, Roy. Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers, http://web.archive.org/web/20090607100835/http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

uitgebreider metadataformaat aan te bieden. DCterms, de uitbreiding van Dublin Core, ook wel bekend als ‘Qualified Dublin Core’, biedt bijvoorbeeld de losse velden ‘coverage_spatial’ en ‘coverage_temporal’ aan om dit probleem te voorkomen.

Het is dan ook aan te raden om, naast Dublin Core als een verplicht minimum, andere, rijkere indelingen van metadata aan te bieden als alternatief.

Verder is het voor interoperabiliteit belangrijk om data op een consistente manier in te vullen. Velden voor bijvoorbeeld ‘datum’ kunnen het beste ingevuld worden aan de hand van de betreffende standaard ISO-8601 zoals in de Dublin Core-standaard zelf wordt aangeraden.¹³ Het is daarnaast ook belangrijk om de semantische betekenis consistent te houden en deze te communiceren naar externe partijen. Bij het datum-veld is het bijvoorbeeld belangrijk om te weten *welke* datum het betreft. De oorspronkelijke datum van publicatie van het object is vaak het zinnigst om hier in te vullen maar het gebeurt ook dat hier de datum van toevoeging aan de repository of een andere datum wordt genoemd.¹⁴ Een ander aandachtspunt rondom metadata heeft niet zozeer betrekking op de inhoudelijke mapping van velden maar heeft te maken met de technische representatie van metadata.

Veelvoorkomende obstakels zijn namelijk dat er tekens in de data zitten die niet in XML mogen voorkomen of dat er HTML-opmaak, zoals
 in plaats van enters, in de data worden meegeleverd.¹⁵ Wanneer er onvoldoende rekening gehouden wordt met de tekstcodering van de data kan het ook zijn dat diakritische tekens niet goed getoond worden waardoor er vreemde karakters in de data voorkomen.¹⁶ Deze problemen kunnen ervoor zorgen dat de tekst in de data minder begrijpelijk wordt of erger nog, dat deze importfouten veroorzaken aan de kant van de harvester.

About-element

Doordat repositories de functie van zowel harvester als provider kunnen vervullen, is het binnen complexe ketens van metadatastromen soms niet meer duidelijk waar metadata vandaan komt en wie de oorspronkelijke eigenaar van een metadata-record is. Het about-element, een optioneel onderdeel van een metadata-record, kan hier uitkomst bieden. Het enige dat in het protocol vermeld wordt als doel van het about-element is dat het informatie dient te bevatten over de metadata zelf. Ook dient het te voldoen aan een bepaald xml-schema. Een van de aanbevolen doelen voor dit optionele element is om deze te vullen met data over de herkomst van de metadata, in het Engels ‘provenance’ genoemd.¹⁷ Dit idee wordt concreter uitgewerkt in de implementatierichtlijnen. Het provenance-element zoals deze is uitgewerkt in de richtlijnen bevat een overkoepelend element ‘originDescription’ dat een aantal verplichte elementen bevat waaronder:

- url van de repository die geharvest is
- de oorspronkelijke identifier van het item in de oorspronkelijke repository
- de oorspronkelijke datering van het record.¹⁸

Aangezien het element ‘originDescription’ zelf ook meerdere ‘originDescription’-elementen kan bevatten is het mogelijk om een hele geschiedenis van een record bij te houden zelfs als deze meerdere keren geharvest en weer uitgeleverd is. Binnen het provenance-element is het bovendien mogelijk om aan te geven of een metadata-record al dan niet aangepast is binnen een repository zodat ook precies bij te houden is welke schakels in de keten content hebben toegevoegd of gewijzigd aan de oorspronkelijke metadata.

¹³ <http://dublincore.org/documents/dces/>

¹⁴ Muriel Foulonneau and Francis André, Investigative Study of Standards for Digital Repositories and Related Services 2007, <http://dare.uva.nl/aup/nl/record/260224>, blz 25

¹⁵ Ibid. Blz. 25 en Tennant, Roy. Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers, http://web.archive.org/web/20090607100835/http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

¹⁶ http://web.archive.org/web/20090608174441/http://www.kathagedorn.com/SP_issues.html

¹⁷ <http://www.openarchives.org/OAI/openarchivesprotocol.html#Record>

¹⁸ <http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm>

Wanneer geen rekening gehouden wordt met de herkomst van metadata kunnen er in complexe ketens van harvesters en providers al snel dubbelingen ontstaan waarbij dan ook niet duidelijk is wie de oorspronkelijke eigenaar van de bron is. Data uit dezelfde bron kan via verschillende providers uiteindelijk weer in een overkoepelende repository samenkomen waardoor dezelfde metadata meerdere keren opgeslagen is. Om dit te voorkomen is het belangrijk dat er voor elke record een ‘persistente identifier’ wordt aangemaakt.¹⁹ Een persistente identifier is een wereldwijd uniek label dat aan een record gegeven kan worden en dat het object uniek identificeert ongeacht specifieke digitale locatie of formaat. Wanneer deze identifier wordt doorgegeven door providers, bijvoorbeeld in het provenance-element, kan op basis hiervan bepaald worden of een metadata-record van dit object reeds aanwezig is in de repository vanuit een andere bron. Zo kunnen dubbelingen voorkomen worden.

Rechten

Bij het bepalen van een strategie rondom beschikbaarstellen of hergebruiken van data via OAI-PMH is het altijd noodzakelijk om rekening te houden met de rechten. Het gaat hierbij zowel om de rechten op het materiaal zelf als om de rechten op de metadata. Voor de vraag welke rechten hierop gelden zijn met name het auteursrecht en het databaserecht van belang. Hoewel deze rechten per land verschillen komen ze op de belangrijkste punten in grote lijnen overeen. Auteursrecht is in de basis het unieke recht van de auteur om zijn eigen werk te reproduceren en te verspreiden. Het auteursrecht geldt vanaf het moment van creatie en is geldig tot 70 jaar na de dood van de auteur.²⁰ In een aantal Europese landen, waaronder Nederland, is het tevens van belang dat een werk origineel is en dat het een persoonlijk stempel van de auteur bevat.²¹

Het databaserecht, dat eind jaren '90 is ingevoerd binnen de Europese unie, beschermt databases tegen ongeoorloofd hergebruik. Dit gebeurt met name om de investeringen van het verzamelen van data te beschermen.²²

Metadata zelf bevat lang niet altijd een persoonlijk, creatief kenmerk van de auteur en komt daardoor niet in alle gevallen in aanmerking voor bescherming onder het auteursrecht. De database waarin deze metadata zich bevindt wordt echter beschermd door het databaserecht. Hoewel het aanbieden van een OAI-PMH-interface impliciet aangeeft dat de data gedeeld mag worden, is het - om misverstanden te voorkomen - toch aan te raden om expliciet te vermelden wat een harvester wel en niet met de metadata mag doen.²³ Harvesters kunnen deze informatie behouden in hun eigen metadata om hun gebruiksrecht desgevraagd te kunnen aantonen.

Een manier om dit op een gestandaardiseerde manier te doen is door gebruik te maken van een rechtenbeschrijving als uitbreiding in het about-element van een record zoals dat hierboven beschreven is voor herkomst-informatie. In de richtlijnen wordt een aantal voorbeelden gegeven van de wijze waarop deze rechteninformatie over de metadata zelf, vormgegeven kan worden.²⁴

Afscherming

De ‘o’ van open in ‘OAI-PMH’ betekent niet dat het protocol altijd voor iedereen gratis toegankelijk gemaakt moet worden.²⁵ Wanneer het niet wenselijk of mogelijk is om als provider de metadata beschikbaar te stellen aan de hele wereld kan het toch de moeite lonen een OAI-provider aan te bieden aan een selecte groep van partners. Wanneer de metadata niet door iedereen hergebruikt mag worden

¹⁹ Dit advies wordt bijvoorbeeld gegeven in Foulonneau, Open Archives Initiative – Protocol For Metadata Harvesting Practices of cultural heritage actors, http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf, blz. 16.

²⁰ Weenink e.a., A DRIVER's Guide to European Repositories, blz. 104

²¹ Ibid. 108

²² Mark Bide, Open Archives and Intellectual Property: incompatible world views?, http://www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf, blz. 12

²³ Ibid. 21-22

²⁴ <http://www.openarchives.org/OAI/2.0/guidelines-rights.htm>

²⁵ Dit is ook het standpunt van de Open Archives Initiative zelf getuige de informatie in de FAQ op de officiële site:

<http://www.openarchives.org/documents/FAQ.html#What%20do%20you%20mean%20by%20%22Open%22?>

kan dit in een rechtenveld opgenomen worden zoals hierboven beschreven. Voor veel instellingen zal dit echter niet voldoende zekerheid bieden. Men zal eerder gebruik willen maken van een soort afscherming zodat de harvester ook echt alleen toegankelijk is voor door de instelling geselecteerde gebruikers.

In het protocol van OAI-PMH zelf is geen voorziening opgenomen voor afscherming door middel van 'verbs' of attributen. Echter de onderliggende laag van HTTP maakt wel enkele scenario's mogelijk die een vorm van afscherming kunnen bieden.

Dit kan op verschillende manieren, waaronder afscherming door middel van http-authenticatie of op basis van IP-adres. In het geval van http-authenticatie moet de harvester gebruikersnaam en wachtwoord meesturen bij elk verzoek, hetgeen een kleine aanpassing aan de kant van de harvester vereist.²⁶

Afscherming op basis van IP-adres betekent dat de provider door middel van configuratie van de webserver bepaalt welke IP-adressen toegang kunnen krijgen tot de provider. Dit heeft als voordeel dat de harvester geen extra functionaliteit hoeft in te bouwen om de afgeschermd data te kunnen binnenhalen. Het IP-adres wordt altijd verstrekt bij een verzoek voor data aan de provider en kan in dit geval dan door de provider gebruikt worden om toegang toe te staan of te weigeren.

Beide scenario's zijn uiteraard geen maatregelen van het hoogste veiligheidsniveau. Ze moeten meer worden gezien als een basale afscherming waarmee gebruik extra gereguleerd kan worden.

Linken naar objecten

Hoewel OAI-PMH in eerste instantie bedoeld is als een protocol om metadata uit te wisselen, wordt het ook vaak gebruikt om de locatie van bijbehorende digitale objecten te verstrekken. Metadata beschrijft uiteraard altijd een object. Wanneer dit object ook in digitale vorm, open beschikbaar is, kan het voor de instelling wenselijk zijn om het object zelf ook te delen. De manier waarop dit het beste kan, is - in verband met de nadruk binnen het protocol op metadata - echter niet opgenomen in de beschrijving van OAI-PMH. In het rapport 'Investigative Study of Standards for Digital Repositories and Related Services' opgeleverd in het kader van eerdergenoemde DRIVER-project is dit punt ook onderkend als probleem. Een metadata-record hoeft niet per se een *digitaal* object te beschrijven. Het kan ook gerelateerd zijn aan *meerdere* digitale objecten, mogelijk in een bepaalde hiërarchische verhouding.²⁷ Hierdoor is het op voorhand lastig te weten of een metadata-record verwijst naar een object dat wel of niet digitaal beschikbaar is. Uiteindelijk is de eenduidigheid van deze informatie sterk afhankelijk van het metadata-formaat dat gekozen wordt. Voor het verplichte Dublin Core geldt in ieder geval dat een url naar een digitaal object opgenomen moet worden in het identifier-element. Dit zorgt echter voor problemen wanneer er bijvoorbeeld meerdere digitale formaten beschikbaar zijn. Binnen Dublin Core is niet aan te geven naar welk type objecten de verschillende identifiers verwijzen.²⁸ Ook is het hierbinnen niet mogelijk om een hiërarchie van digitale objecten weer te geven zoals bijvoorbeeld in het geval van een metadata-record van een boek met als digitale objecten losse afbeeldingen die per bladzijde gescand zijn.

Dit probleem is aangepakt door de Open Archives Initiative zelf en heeft geleid tot het opstellen van een nieuw protocol dat speciaal gericht is op het delen van objecten, optioneel te gebruiken in combinatie met OAI-PMH. Dit zal hieronder uitgebreider beschreven worden.

Harvesters

Veel van de hierboven beschreven best practices betreffen met name ontwerpkeuzes bij het implementeren van een OAI-provider. Er zijn echter ook best practices geformuleerd speciaal gericht op OAI-harvesters, met name op het gebied van de workflow van het harvesten.

²⁶ Muriel Foulonneau, Open Archives Initiative – Protocol For Metadata Harvesting Practices of cultural heritage actors, http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf, blz. 19.

²⁷ Muriel Foulonneau and Francis André, Investigative Study of Standards for Digital Repositories and Related Services 2007, <http://dare.uva.nl/aup/nl/record/260224>, blz 33

²⁸ Ibid., blz 34

Allereerst zijn er ook voor harvesters de officiële richtlijnen zoals uitgegeven door de Open Archives Initiative. Deze bevatten vooral maatregelen die ervoor moeten zorgen dat een geautomatiseerde harvest-sessie niet onbedoeld leidt tot onbereikbaarheid van de webserver doordat er teveel verzoeken op worden afgevuurd. Om dit te voorkomen is het van belang dat eventuele foutmeldingen van de webserver niet leiden tot nog meer pogingen om de data alsnog te verkrijgen. Als laatste punt van de richtlijnen wordt ook een mogelijke strategie besproken om alle beschikbare data van een repository geautomatiseerd binnen te kunnen halen door middel van het achtereenvolgens doen van de hierboven beschreven Identify-, ListMetadataFormats en ListRecords-verzoeken.²⁹

In de praktijk is dit echter wellicht een enigszins optimistisch scenario. In de meeste gevallen zal de data bekeken en de workflow getest moeten worden voordat deze op een productie-omgeving binnengehaald kan worden. Roy Tennant beschrijft in zijn artikel 'Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers' een workflow voor harvesters die met veelvoorkomende problemen rekening houdt. Hij gaat er vanuit dat data die binnen een eigen database instroomt altijd enige vorm van normalisering moet ondergaan. Binnen de voorgestelde workflow wordt dan ook uitgegaan van drie fases. In de eerste fase wordt de data geanalyseerd om te kijken welke transformaties toegepast moeten worden en worden de noodzakelijke transformaties vastgesteld. In de tweede fase wordt de data onderverdeeld in subsets. De data wordt tenslotte in de derde fase echt binnengehaald en achtereenvolgens door de processen van normalisering en onderverdeling bewerkt alvorens in te stromen in de eigen database.³⁰ Deze analyse vooraf zorgt ervoor dat er geen onverwachte data de database instroomt. Ondanks het feit dat in het OAI-PMH-protocol een uniforme interface is gedefinieerd voor harvesters zorgt de vrijheid van de metadataformaten (en verschillende interpretatie-mogelijkheden van die formaten) ervoor dat er bij het aansluiten van een nieuwe databron altijd mensenwerk te pas komt.

Usecase: Digitaal videomateriaal voor het onderwijs

Zoals eerder genoemd kunnen er complexe contentketens gecreëerd worden door middel van OAI-PMH. Doordat harvesters hun binnengehaalde materiaal zelf ook weer beschikbaar kunnen stellen ontstaan er verschillende knooppunten waarbij data wordt binnengehaald, eventueel aangevuld en weer beschikbaar gesteld.

Als usecase wordt hieronder een korte beschrijving gegeven van de manier waarop OAI-PMH gebruikt wordt door het Nederlands Instituut voor Beeld en Geluid om videomateriaal beschikbaar te stellen voor onderwijs en onderzoek.

Aan het begin van de workflow wordt - vanuit de catalogus van Beeld en Geluid - een videobestand met metadata geëxporteerd naar een 'Media-Asset-Management-systeem' op internet. Dit systeem biedt een OAI-PMH-interface aan die door verschillende onderwijsplatforms gebruikt wordt om de metadata van de video's binnen te halen. Binnen het onderwijsplatform Academia, een videoplatform voor het hoger onderwijs, wordt deze data vervolgens verrijkt en opnieuw beschikbaar gesteld via een eigen OAI-provider. Deze data wordt geïmporteerd door verschillende zoeksystemen, waaronder die van grote automatiseringsorganisaties gericht op het bibliotheekwezen zoals OCLC en Ex Libris, die deze data doorzoekbaar maken binnen hun eigen mediacatalogi voor instellingen binnen het hoger onderwijs.

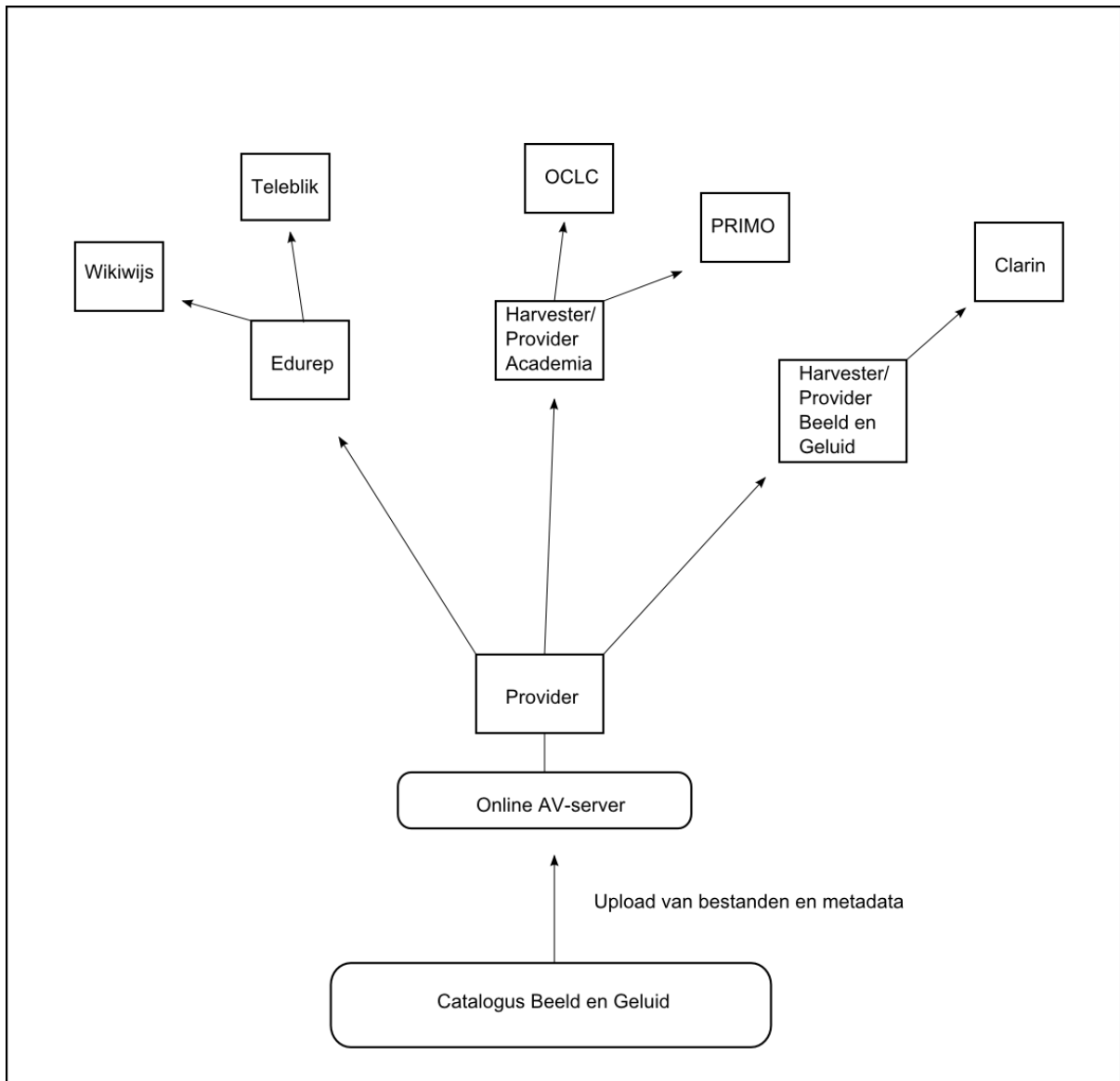
Voor het lager en middelbaar onderwijs wordt de data geharvest door de educatieve repository Edurep die de data indexeert en zelf ook beschikbaar stelt door middel van webservices. Via deze dienst wordt de data doorzoekbaar gemaakt voor partijen die digitaal materiaal gericht op onderwijs beschikbaar willen stellen via eigen websites zoals Wikiwijs, een zoekportaal voor digitaal lesmateriaal, en Teleblik, een videodienst voor lager en middelbaar onderwijs. Tenslotte wordt de data ook geharvest door een service van Beeld en Geluid zelf. Deze service koppelt de locatie van de video weer terug naar de oorspronkelijke catalogusrecords en maakt hier een persistente identifier voor aan. Deze data wordt tenslotte ook weer in een OAI-provider aangeboden en beschikbaar gesteld aan het Clarin-

29 Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting, Document Version 2005/01/19T19:27:00Z, <http://www.openarchives.org/OAI/2.0/guidelines-harvester.htm>

³⁰ Tennant, Roy. Bitter Harvest: Problems & Suggested Solutions for OAI-PMH Data & Service Providers, http://web.archive.org/web/20090607100835/http://www.cdlib.org/inside/projects/harvesting/bitter_harvest.html

netwerk voor onderzoek. Clarin is een netwerk dat zich richt op beschikbaarstelling van data en tools rondom taal.

Aan de hand van deze usecase is te zien hoe, met behulp van OAI-PMH, data op een centrale plek beheerd kan worden en zich tegelijkertijd kan verspreiden over een veelheid aan websites. Het geeft harvesters de vrijheid, voor zover de rechten dit toestaan, om het materiaal te bewerken en aan te bieden zoals ze dit zelf willen zonder dat het voor de instelling die de oorspronkelijke records beheert extra lasten met zich meebrengt of een onbeheersbare situatie oplevert.



Figuur 3: Een overzicht van de datastromen voor de onderwijsplatforms van Beeld en Geluid. OAI-PMH maakt een belangrijk deel hiervan uit bij de overdracht van metadata.

De toekomst: OAI-ORE en linked data

Zoals in het begin aangegeven is de laatste versie van het OAI-PMH-protocol nu al meer dan tien jaar oud en dateert de laatste wijziging aan het document zoals deze op internet te vinden is ook al van enkele jaren geleden. Dit betekent dat het protocol zoals we dit kennen als zeer stabiel beschouwd kan worden, wat op zich natuurlijk vertrouwenwekkend is. Aan de andere kant rijst de vraag: wat is de toekomst van het protocol? Zijn er al nieuwe technieken die OAI-PMH overbodig maken? Kan het protocol nog mee in de snel veranderende ontwikkelingen van het internet?

Hoewel er aan het document van het OAI-PMH protocol zelf niet veel is veranderd, zijn de ontwikkelaars zoals gezegd wel verder gegaan met het ontwikkelen van een nieuw protocol. Dit protocol heet OAI-ORE, dat staat voor 'Objects Reuse and Exchange'. OAI-ORE zorgt voor een duidelijke oplossing van het probleem rondom het delen van digitale objecten zoals dat is onderkend in het gebruik van OAI-PMH. Het protocol schept bijvoorbeeld mogelijkheden voor het op een eenduidige manier aanbieden van hiërarchisch samengestelde digitale objecten en voor het aanbieden van meerdere formaattypes binnen een record. Hiermee vervult het duidelijk een functie waarvoor geen plaats was binnen het OAI-PMH protocol.

In lijn met ontwikkelingen op het internet maakt OAI-ORE gebruik van linked data om relaties tussen digitale objecten op internet te kunnen leggen. Binnen OAI-ORE worden de verschillende digitale representaties die bij een bepaald object horen, samengebracht in een aggregatie. Dit is een overkoepelende constructie waarbinnen de verschillende objecten en hun relaties tot elkaar en tot het overkoepelende object beschreven kunnen worden. Zo kan er bijvoorbeeld een aggregatie zijn die verwijst naar het object van een wetenschappelijk artikel en kan er binnen deze aggregatie verwezen worden naar de urls van verschillende bestandstypen van hetzelfde artikel. De onderlinge relatie tussen de objecten wordt dan duidelijk gemaakt binnen de beschrijving van de aggregatie. Door middel van metadata kan precies duidelijk gemaakt worden welke url naar welk bestandsformaat verwijst, ook al is dit aan de url zelf niet te zien. De beschrijving van de aggregatie, inclusief de verschillende objecten die tot deze aggregatie behoren, wordt een resource map genoemd.³¹ Deze resource maps moeten op een bepaalde manier beschikbaar gesteld kunnen worden en hiervoor kan weer OAI-PMH gebruikt worden.³² Op deze manier vormt de techniek eerder een aanvulling op het OAI-PMH-protocol dan een vervanging ervan.

Zo kan OAI-PMH ook in de toekomst een belangrijke rol blijven spelen bij het uitwisselen van data op internet.

³¹ Een algemene inleiding is te vinden op: <http://www.openarchives.org/ore/1.0/primer>

³² http://www.openarchives.org/ore/1.0/discovery#OAI_PMH