Use of Remote Open Access in repositories

⚠ **WORK IN PROGRESS!!**

Omgaan met Remote Open Access publicaties in de repository

Notitie Arjan Hogenaar, 16 november 2010

Table of contents

- Table of contents
- Document information
- Probleemstelling
- Welke oplossing is haalbaar?
- ROA in de repositories
  - A. ROA tot final author versions
  - B. ROA tot Published Version

Document information

Title: Use of Remote Open Access in Repositories
Subject: DARE repositories; metadata, DIDL, vocabulary
Moderator: Hoogenaar, Arjan
Version: 0.1
Date published: 2011-02-01
Excerpt: Addition to the DIDL-MODS application profile to express the phenomenon "Remote Open Access"; to provide end-users a link to the Open Access file registered in the repository, yet remotely located at another location.

(Optional information)
Type: Internal report
Format: Text/richtext
Identifier: none yet
Language: NL
Rights: Copyright Stichting SURF. The text of this document may be used freely, without permission of Stichting SURF.
Tags:

Probleemstelling

Normaliter wordt de repository gevuld door het deponeren van objecten, gekoppeld aan bibliografische beschrijvingen. Deze bibliografische beschrijvingen komen niet zelden uit Metis, maar kunnen natuurlijk ook op andere wijze in de repository worden ingevoerd.

Na deponeren kan gekozen worden voor open, beperkte (in de regel: alleen toegang binnen de eigen instelling) of gesloten toegang tot de objecten.

Het vullen van de repository loopt vaak parallel met het streven naar Open Access tot publicaties, maar niet altijd. Dit laatste doet zich voor in die gevallen waarin er elders in de wereld Open Access repositories zijn met daarin versies van publicaties [van artikelen van onderzoekers die aan Nederlandse wetenschappelijke instellingen verbonden zijn] die niet mogen worden opgenomen in een van de institutionele repositories.

Met andere woorden: een ieder kan de publicaties wel inzien, maar de repository managers mogen deze niet kopiëren en deponeren in de eigen repository. Voorbeelden zijn: arXiv en PubMedCentral (PMC). Daarnaast komt het voor dat uitgevers gratis toegang bieden tot de content van een of meer tijdschriften, onder dezelfde restrictieve voorwaarden.

Het beste is dit te illustreren aan de hand van een voorbeeld.

Robbert Dijkgraaf heeft geschreven:

Instantons on ALE spaces and orbifold partitions, verschenen in Journal of High Energy Physics (2008). Dit artikel is zowel via arXiv (final author version) als via IOP press (officiële versie) vrij beschikbaar. Echter, beide versies mogen niet in de repository worden opgenomen. Daarom valt de beschrijving van dit artikel in NARCIS formeel onder de toegangscategorie ‘closed access’.

Zie: http://www.narcis.info/publication/RecordID/oai:uva.nl:300388/ en http://dare.uva.nl/record/300388 voor de formele beschrijvingen en

http://arxiv.org/abs/0712.1427 plus

http://iopscience.iop.org/1126-6708/2008/03/013/pdf/1126-6708_2008_03_013.pdf

voor de full-text versie via arXiv resp. IOP press.

Wat voor bibliotheekmensen juist is, werkt uitermate verwarrend en frustrerend voor een eindgebruiker. Door deze indeling zal die eindgebruiker die het artikel via NARCIS detecteert denken dat het artikel niet vrij toegankelijk is en het ibl belasten met een onnodige aanvraag.

Dit voorbeeld toont ook aan dat een oplossing op Surfshare-niveau moet worden gevonden. Dijkgraafs output is afkomstig van de repository van de UvA . De UvA geeft in de door NARCIS geharveste metadata aan dat de full-text closed access is.

Als je ditzelfde record bekijkt in de repository van de UvA zie je dat daar een link naar arXiv is opgenomen in het 'Note-field' en dat de UvA-linker een link naar de IOP-site genereert.

Ander voorbeeld:

In de UvT-repository is de beschrijving van het rapport "Evaluatie Zorgverzekeringswet en Wet op de Zorgtoeslag"

http://dbiref.uvt.nl/iPort?request=full_record&db=wo&language=eng&query=doc_id=4112728] opgenomen.

Het rapport zelf is niet gedeponeerd in de repository van de UvT. Daardoor valt de beschrijving van het rapport binnen NARCIS onder de categorie `'closed access'`.

Het rapport is echter wel degelijk openbaar en te vinden op de site van ZonMW .

De huidige repository infrastructuur en categorie-indeling in NARCIS is dus uitsluitend gerelateerd aan wat er in de repository is gedeponeerd en aan welke metadata aan het record zijn toegekend. Formeel juist, maar losstaand van de werkelijkheid van gebruikers van systemen als NARCIS.

Dit is een ongewenste situatie. Voor een gebruiker maakt het niet uit of hij de full-text van de repository, PMC, arXiv of van de uitgeverssite haalt. De categorie ‘closed access’ zet hem op het verkeerde spoor.

Welke oplossing is haalbaar?

Een mogelijke oplossing is de introductie van een nieuwe categorie in de repositories met de naam 'remote open access (ROA)'. Betere namen zijn welkom.

In de ROA-categorie vallen de beschrijvingen van die artikelen waarvan de full-text versies niet in de repository mogen worden opgenomen, maar die elders wel open access beschikbaar zijn. Door in de repository een link naar de locatie op te nemen waar het artikel opgehaald kan worden wordt de gebruiker maximaal geholpen en laten we hem/haar niet ongewild in de kou staan.

Wat op repositoryniveau wellicht gekunsteld overkomt, is een uitkomst voor NARCIS. In NARCIS kunnen de ROA en OA records beide voorzien worden van het label ‘Open Access’. Zo zal in eerder genoemd voorbeeld de gebruiker zien dat hij het artikel **Instantons on ALE spaces and orbifold partitions** vrij kan benaderen.

Verdere verfijning is evenwel nodig. In de voorbeelden van de links naar IOP resp. naar ZonMW wordt de gebruiker naar de uitgeversversie van de publicatie geleid. Versies van publicaties die via bijv. arXiv of PubMedCentral (PMC) toegankelijk zijn, zijn te beschouwen als ‘final author versions’ (hoe mooi ze in de regel ook vormgegeven zijn).

Dit is te illustreren aan de hand van een voorbeeld. De EUR had in haar repository het artikel *"Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study"* kunnen deponeren.

De full-text van de final-author version is toegankelijk via PubmedCentral (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2803340).

De uitgeversversie staat op de site van 'The Lancet' (http://www.thelancet.com/journals/lancet/article/PIIS0140673608613434/abstract).

De toegang tot de full-text op de uitgeverssite is restricted access (veelal gebaseerd op IP-adres), die op PMC is Open Access.

ROA in de reposities

In de metadata van de bibliografische records moeten bij het aangeven van ROA twee zaken worden aangegeven:

1. Er is sprake van Remote Open Access
2. Er is sprake van toegang tot de 'published version' of tot de 'final author version'.

Deze extra metadata moeten in het DIDL/MODS-document worden opgenomen.

In principe zou de informatie over de toegang tot de versies op de sites van de uitgevers (de published versions) in de MODS kunnen worden opgenomen (het gaat immers over informatie die van de uitgever afkomstig is).

Omdat de informatie over de vrij toegankelijke final author versions niet in de MODS thuishoren, zouden de repository managers gedwongen worden de toegang tot beide vormen van Remote Open Access op verschillende manieren vast te leggen.

In overleg met Thomas Place doe ik daarom het voorstel de toegang tot remote published versions en die tot remote final author versions op dezelfde wijze te beschrijven.

Wat verder nodig is, is het opnemen van het Access rights type "Remote Open Access".

In de huidige constructie zal Remote Open Access aldus beschreven kunnen worden (aanpassing van **Eprints AccessRights Vocabulary Encoding Scheme** nodig!)

```
<didl:Descriptor>
    <didl:Statement mimeType="application/xml" >
        <dcterms:accessRights>http://purl.org/eprint/accessRights/RemoteOpenAccess</dcterms:accessRights>
     </didl:Statement>
</didl:Descriptor>
```

let wel: Remote Open Access bestaat natuurlijk nog niet.

Alternatief:

```
<didl:Descriptor>
    <didl:Statement mimeType="application/xml">
        <rdf:type rdf:resource="info:eu-repo/semantics/remoteopenAccess"/>
       </didl:Statement>
</didl:Descriptor>
```

Let wel: hiervoor is aanpassing van info-eu-repo nodig.

A. ROA tot final author versions

Dit is informatie die niet van een uitgever kan komen. Daarom wordt deze informatie niet weggezet in MODS. Om deze informatie toch op te nemen in de metadata kan daarom gebruik worden gemaakt van `DIDL/Item/Item`.

```
<didl:Item>
   <didl:Item>
      <didl:Descriptor>
         <didl:Statement mimeType="application/xml" >
           <dcterms:accessRights>http://purl.org/eprint/accessRights/RemoteOpenAccess</dcterms:accessRights>
         </didl:Statement>
    </didl:Descriptor>
     <didl:Descriptor> <!-- Item type -->
        <didl:Statement mimeType="application/xml">
          <rdf:type rdf:resource="info:eu-repo/semantics/objectFile" />
        </didl:Statement>
    </didl:Descriptor>
    <didl:Descriptor>
       <didl:Statement mimeType="application/xml">
          <rdf:type rdf:resource="info:eu-repo/semantics/acceptedVersion"/>
       </didl:Statement>
</didl:Descriptor>
  <didl:Component> <!-- Actual resource of Item -->
     <didl:Resource
         mimeType="application/pdf"
        ref="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2803340/pdf/nihms108887.pdf " />
  </didl:Component>
</didl:Item>
</didl:Item>
```

Omwille van de eenvoud (en om de verwerking door NARCIS simpeler te maken) verdient het aanbeveling om de ROA tot published version ook gewoon in de DIDL op te nemen *(tip van Thomas Place)*.

B. ROA tot Published Version

Ook hiervoor wordt `DIDL/Item/Item` gebruikt. Uitgeschreven wordt dit:

```
<didl:Item>
   <didl:Item>
      <didl:Descriptor>
         <didl:Statement mimeType="application/xml" >
           <dcterms:accessRights>http://purl.org/eprint/accessRights/RemoteOpenAccess</dcterms:accessRights>
         </didl:Statement>
    </didl:Descriptor>
     <didl:Descriptor> <!-- Item type -->
        <didl:Statement mimeType="application/xml">
          <rdf:type rdf:resource="info:eu-repo/semantics/objectFile" />
        </didl:Statement>
    </didl:Descriptor>
    <didl:Descriptor>
       <didl:Statement mimeType="application/xml">
          <rdf:type rdf:resource="info:eu-repo/semantics/publishedVersion"/>
       </didl:Statement>
</didl:Descriptor>
  <didl:Component> <!-- Actual resource of Item -->
     <didl:Resource
         mimeType="application/pdf"
        ref="http://iopscience.iop.org/1126-6708/2008/03/013/pdf/1126-08_2008_03_013.pdf " />
  </didl:Component>
</didl:Item>
</didl:Item>
```