Mapping NARCIS domain to OpenAIRE CERIF model

Wilko Steinhoff, technical lead NARCIS @ DANS
Lightning talk @ euroCRIS Membership Meeting
May 29, 2019, Helsinki, Finland
Outline

• Short introduction NARCIS Portal
• Mapping NARCIS domain to OpenAIRE CERIF
• Issue’s we ran into
• CERIF implementation status
NARCIS

Dutch national aggregator of scholarly output.
- Dutch universities
- Research institutes
- Dutch Research Database (NOD)

Various input models:
- NL_DIDL (EduStandaard)
- Dublin Core
- DataCite

Normalized to NARCIS internal format.
Five main entities:

- Publications: cfResultPublication
- Data sets: cfResultProduct
- Research projects: cfProject
- People: cfPerson
- Organisations: cfOrganisationUnit

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,931,349</td>
<td>Publications</td>
</tr>
<tr>
<td>227,100</td>
<td>Data sets</td>
</tr>
<tr>
<td>69,899</td>
<td>Research</td>
</tr>
<tr>
<td>60,618</td>
<td>People</td>
</tr>
<tr>
<td>2,969</td>
<td>Organisations</td>
</tr>
</tbody>
</table>
Mapping remarks

- XSLT Transformations from NARCIS internal format to CERIF
  => CERIF XML schema is subset of NARCIS internal format

- Easy mapping from ‘info:eu-repo’ namespace to COAR Resource Types

- Patent: sub-type of a NARCIS publication v.s. cfProduct type in OA-CERIF

- Person identifiers types restricted list within OA-CERIF: Extended for DAI.

- Finding right classification-scheme’s. F.i: ‘Project status’
PMH referential ID issue’s

Missing referential integrity (PMH) for embedded, 2nd level objects
   => Persons and OrgUnits related to Publications, Patents & Products

- Embedded Objects should be (PMH) retrievable by {internal ID}
- Index does not always know this {internal ID}

Consequence:
- Top-level (PMH) retrieval of this object is not possible (NO node)
- <DisplayName> tag only, not providing available supported global identifiers
Solution (WIP)

Enrich metadata at ingest time:

1. **Lookup** \{internal ID\} in Dutch Research Database for Person and OrgUnit’s by their supplied persistent-identifiers.
2. Add found {internal ID}’s to the metadata.
3. Authors will now be de-referenceable by {internal ID} in PMH-repository.

Gain:
1. Supply persistent identifiers to an Person/OrgUnit.
2. Extra node in the CERIF-graph.
NARCIS Internal normalized format

Created from descriptive metadata...
Missing PMH {internal ID}.

```
<name>
  <type>personal</type>
  <family>Borst</family>
  <given>S.C.</given>
  <mcRoleTerm>inv</mcRoleTerm>
  <nameIdentifier type="dai-nl">info:eu-repo/dai/nl/298063085</nameIdentifier>
</name>

Yields:

```
<Inventor>
  <DisplayName>Borst, S.C.</DisplayName>
  <Person/>
</Inventor>
```
Enriched Internal format

```xml
<name>
  <type>personal</type>
  <family>Borst</family>
  <given>S.C.</given>
  <mcRoleTerm>inv</mcRoleTerm>
  <nameIdentifier type="dai-nl">info:eu-repo/dai/nl/298063085</nameIdentifier>
  <nameIdentifier type="nod-prs">PRS1242227</nameIdentifier>
  <nameIdentifier type="orcid">0000-0003-3306-6447</nameIdentifier>
  <nameIdentifier type="isni">0000000397111243</nameIdentifier>
</name>
```
Referenceable Person in PMH

<Inventor>
  <DisplayName>Borst, S.C.</DisplayName>
  <Person id="PRS1242227"> => { internal ID }
      <PersonName>
        <FamilyNames>Borst</FamilyNames> => structured name
        <FirstNames>S.C.</FirstNames>
      </PersonName>
      <ORCID>http://orcid.org/0000-0003-3306-6447</ORCID> => additional identifiers
          <ISNI>http://www.isni.org/isni/000000397111243</ISNI>
          <DAI>info:eu-repo/dai/nl/298063085</DAI>
    </Person>
</Inventor>
Opportunities: Journals and Publishers

Problem:
No scientific Journals and Publishers as objects in NARCIS.
2nd level objects PublishedIn “Scientific Journals” and Publishers only available as ‘unstructured’ from descriptive metadata (text-string).

Solution (TODO/Future opportunity):
Lookup and normalize scientific Journals by identifier (i.e. ISSN, title, etc.) and add them to NARCIS as an object.
Tools: Sherpa/RoMEO, Unpaywall
## Implementation overview OA-CERIF

<table>
<thead>
<tr>
<th>OA-CERIF</th>
<th>NARCIS domain</th>
<th>Available (PMH-endpoint)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication</td>
<td>Yes. (PublishedIn: no)</td>
<td>WIP</td>
</tr>
<tr>
<td>Product</td>
<td>Yes: Datasets and Software</td>
<td>✓</td>
</tr>
<tr>
<td>Patent</td>
<td>Yes (Publication sub-genre/type)</td>
<td>✓</td>
</tr>
<tr>
<td>Person</td>
<td>Yes</td>
<td>✓</td>
</tr>
<tr>
<td>OrgUnit</td>
<td>Yes</td>
<td>✓</td>
</tr>
<tr>
<td>Project</td>
<td>Yes</td>
<td>✓</td>
</tr>
<tr>
<td>Funding</td>
<td>Little: no research output/results available for now.</td>
<td>x</td>
</tr>
</tbody>
</table>
Thank you

wilko.steinhoff@dans.knaw.nl

https://dans.knaw.nl
Links

- **NARCIS Portal**
  https://www.narcis.nl

- **EduStandaard**
  https://www.edustandaard.nl

- **OpenAIRE Guidelines for CRIS Managers**

- **NARCIS-CERIF PMH-endpoint (restricted access @ test-env)**
  http://services.nod.dans.knaw.nl/oa-cerif?verb=Identify