

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.

NARCIS Portal

Five main entities:

1,931,349	227,100	69,899	60,618	2,969
PUBLICATIONS	DATA SETS	RESEARCH	PEOPLE	ORGANISATIONS

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

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

```
<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/0000000397111243</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

OA-CERIF	NARCIS domain	Available (PMH-endpoint)
Publication	Yes. (PublishedIn: no)	WIP
Product	Yes: Datasets and Software	✓
Patent	Yes (Publication sub-genre/type)	✓
Person	Yes	✓
OrgUnit	Yes	✓
Project	Yes	✓
Funding	Little: no research output/results available for now.	x
Service	Yes	✓
Equipment	No	x
Event	No	x



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
<https://openaire-guidelines-for-cris-managers.readthedocs.io/en/v1.1.1/index.html>
- NARCIS-CERIF PMH-endpoint (restricted access @ test-env)
<http://services.nod.dans.knaw.nl/oa-cerif?verb=Identify>