

PID Non-User Stories: ADP – Slovenian Social Science Data Archives

Irena Vipavc Brvar, ADP

CESSDA PID, Cologne, February 18th, 2016

This work is licensed under [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).



CESSDA – Questionnaire on PIDs

- ADP classify as non-user of PID.
- Difficult to fill it in.
- Obviously there is more to PID than what we would normally thought about.
- Some items not clear.

Some of the issues are more for „registration agencies“ and are taken for granted. They describe quality of their services.

Requirements

What are specific requirements concerning the use of PIDs in the framework of the data archive?

What we need and / or what users need? -> solve both issues
All DA have some sort of identification system.

CITATION - What is needed is some common PID for all DA (or similar services) so users will be able to cite data / materials that we distribute.

-> data discovery

-> future linking Publication and Data

Possible need for more technical support.

What metadata do we link to PID? Should we use one schema (DataCite)?

Challenges

What are the biggest challenges to introduce a PID system?

There is a need for a system that will generate PID.

Connecting ADP and PID registration agency – National University Library (uses URN). Need to use specification offered by registration agency – are we prepared?

Simple solution vs. Complex (whole process) solution.

Integrating PID in existing applications used in DA.

DOI – either there are none in a country or it's commercial service

Shall we use both URN and DOI in a future?

There is no national registration service available, at least not related to research data.

Plans / Wishes

What are the future plans / wishes concerning PIDs?

At this point we are developing Java application on top of FEDORA.

- > connecting to FEDORA at National and University Library (NUK)
- > at the point of deposit (materials to NUK) – URN will be given to material – granulation!!
- PIDS are not main goal – parallel to development of main application.

URN:SI:UNI-LJ-FDV:ADP:StudyID_MaterialTypeNo_Language_Version_Subversion

URN:SI:UNI-LJ-FDV:

Prefix

SJM12_

Study ID

P1_

Type of material:

Questionnaire (VP)

Data file (P)

Data Description (DD)

Related materials (RM)

Related publication (RP)

Other materials (OM)

EN_

Language
code in ID

V1_

Version

R1

Subversion

~~URN:NBN~~



DC for DATA

title

creator / publisher / date / type / format / identifier / language /
coverage / rights

```
<oai_dc:dc
```

```
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/  
http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
```

```
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

```
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
```

```
  xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```
    <dc:title id="title">Title</dc:title>
```

```
    <dc:title id="subtitle">Subtitle</dc:title>
```

```
    <dc:creator> Creator/Author 1</dc:creator>
```

```
    <dc:creator> Creator/Author 2</dc:creator>
```

```
    <dc:creator>...</dc:creator>
```

```
  <dc:publisher> Publisher 1</dc:publisher>
```

```
    <dc:publisher> Publisher 2</dc:publisher>
```

```
    <dc:publisher>...</dc:publisher>
```

DC Data, Part 2

```
<dc:date>Date</dc:date>
  <dc:type> Type of material </dc:type>
  <dc:format id="mimetype">Format</dc:format>
  <dc:format id="cq">number of cases</dc:format>
  <dc:format id="vq"> number of variable</dc:format>
  <dc:format>Format</dc:format>
  <dc:identifier identifier_type="URN">URN</dc:identifier>
  <dc:identifier identifier_type="FILENAME">File name</dc:identifier>
  <dc:identifier identifier_type="ISBN"> ISBN</dc:identifier>
  <dc:identifier identifier_type="ISSN"> ISSN</dc:identifier>
  <dc:language>Language</dc:language>
  <dc:coverage>Coverage</dc:coverage>
  <dc:rights> Rights/restrictions </dc:rights>
</oai_dc:dc>
```

We need to clean and verify all metadata that will be fed into URN system.



DC Study

title

creator / subject / description / publisher / date / type / format/
identifier/ language / coverage / rights

```
<?xml version="1.0"?>
```

```
<oai_dc:dc
```

```
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/  
http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
```

```
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

```
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
```

```
  xmlns:dc="http://purl.org/dc/elements/1.1/">
```

```
    <dc:title id="title">Title</dc:title>
```

```
    <dc:title id="subtitle">Subtitle</dc:title>
```

```
    <dc:creator>Creator/Author 1</dc:creator>
```

```
    <dc:creator> Creator/Author 2</dc:creator>
```

```
    <dc:creator>...</dc:creator>
```

```
    <dc:subject id="cerifid">Cerif ID</dc:subject>
```

```
    <dc:subject id="sifrasc"> SC Code</dc:subject>
```

```
    <dc:subject id="sifrasc"> SC Code</dc:subject>
```

```
    <dc:subject id="sifrasc">...</dc:subject>
```



DC Study, part 2

```
<dc:description>Abstract</dc:description>
  <dc:publisher>Publisher 1</dc:publisher>
  <dc:publisher>Publisher 2</dc:publisher>
  <dc:publisher>...</dc:publisher>
  <dc:date>Date</dc:date>
  <dc:type>Study description</dc:type>
  <dc:format id="mimetype">Format</dc:format>
  <dc:identifier identifier_type="URN"> URN</dc:identifier>
  <dc:identifier identifier_type="IDNO"> study ID</dc:identifier>
  <dc:identifier identifier_type="FILENAME">File name</dc:identifier>
  <dc:identifier identifier_type="COBISSID">Cobiss ID</dc:identifier>
  <dc:identifier identifier_type="FEDORAID">Fedora ID</dc:identifier>
  <dc:language>Language</dc:language>
  <dc:coverage>Coverage</dc:coverage>
  <dc:rights> Rights/restrictions </dc:rights>
</oai_dc:dc>
```

CESSDAs role

What should be CESSDAs role in the field of PIDs?

Need for more clear role.. Guidance

One agency – Can CESSDA play a role of registration agency?

Can one SP do it for whole CESSDA SP?

What about metadata schemas – use one / interoperability / CV?

Do we need to change something in our system (metadata schemas) at this point?

When organizations register a DOI for a resource, they should not use metadata that might change over time (e.g. publisher, archive, owner) (Ball & Duke, 2015)

Define granularity of PID (study, data, version?)

By Ball and Duke *the dataset to which a DOI points should remain unchanged*-> granularity to at least data file level.



CESSDAs role

PIDs can be complex, we need something simple - > searching for service that will provide whole package

Will we use DOI, URN or something else?

Eliminate duplication

Can and why would we use multiple identifiers?

(DA will only register a study if the „own“ data, single distributor needed)

Future vision needed

How will this be visible to users?

Data Portal

Citation Index ?

Contact

University of Ljubljana
Faculty of Social Sciences
Social Science Data Archives
Kardeljeva ploščad 5
1000 Ljubljana
SI - Slovenia



www.adp.fdv.uni-lj.si



arhiv.podatkov@fdv.uni-lj.si



Arhiv.Druzboslovnih.Podatkov



@ArhivPodatkov

Univerza v Ljubljani



cessda saw



cessda



DwB
Data without Boundaries