PID kernel information and the Social Sciences

RDA Collaboration Project – PITSS (PID Information Types for the Social Sciences)

René van Horik / Vyacheslav Tykhonov

@DANSKNAW

September 22, 2017. WG PID Kernel Information @ RDA Plenary, Montreal.
Outline of presentation

1. Context
2. PITSS Project
3. PID Profiles for social science data sets
4. Next steps
Context

DANS -> Trusted Repository for social science data sets

Social Science data sets:
  - Surveys
  - Qualitative data
  - Longitudinal aspect
  - Standard: DDI (data documentation initiative)
  - Curated by social science data archives (related to CESSDA)

CESSDA ERIC -> European Social Science Data Infrastructure (16 members)

CESSDA Working group on PIDs
  - Formulation of PID policy
  - Best practices
PITSS Project (PID Information Types for the Social Sciences)

- RDA Collaboration project

- May – December 2017

- Aim: to introduce and assess RDA outcomes in the field of PIDs into the social science research community, more specifically the European CESSDA research infrastructure
  - Create 1 or 2 PID Profiles (A PID profile is a minimal set of information associated with each PID. This “kernel metadata” can be used to manage the data object, to discover clients and facilitate data reuse)

- Activities
  - Workshop (29/30 May, The Hague)
  - Presentations
  - PID profiles
  - Report
PID Profiles for the social sciences

Some issues:

Rationale for PID Profiles and applicability in social sciences
  - Fast filtering in huge number of objects
    - Is this applicable in the social science domain?

Type Registry (to register data types)
  - Using “typeregistry.org” for experiments
    - How to move to a sustainable registry for the social science domain?
Based on the discussion in the workshop the following Property Identifiers were formulated:

- **WasDerivedFrom** (Facet that describes where the dataset was derived from)
- **LicenceFlag** (Facet that describes if dataset has open or restricted access)
- **IsChild** (Facet that describes if dataset is original itself or has a parent. If it was nested it will point to the original dataset)
- **MultipleInheritance** (Facet that describes if dataset has a inheritance with multiple children)
- **HasMetadata** (Facet that indicates that digital object has extended metadata)
Property Identifier in JSON format (example)

```json
{
    "identifier": "",
    "name": "WasDerivedFrom",
    "description": "Facet that describes where the dataset was derived from.",
    "standards": [
        {
            "issuer": "",
            "name": "",
            "natureOfApplicability": "extends"
        }
    ],
    "provenance": {
        "contributors": [
            {
                "identifiedUsing": "Text",
                "name": "Vyacheslav Tykhonov"
            },
            {
                "identifiedUsing": "Text",
                "name": "Rene van Horik"
            }
        ],
        "creationDate": "2017-09-14 13:55",
        "lastModificationDate": "2017-09-14 13:55"
    },
    "representationsAndSemantics": [
        {
            "expression": "Format",
            "value": "PROPERTY_DEFINITION"
        }
    ],
    "expectedUses": ["In case of dataset has multiple children."],
    "relationships": ["has-relation",
        {
            "name": "has-relation",
            "relativeNames": ["Parent", "Child"
```
Next steps

• Discuss / evaluate outcomes in CESSDA WG on PIDs
  • Concept of PID Kernel is very abstract for social science community
  • Not clear how PID Kernel information / Type Registry fits in social science infrastructure

• Monitor / contribute to RDA initiatives on PIDs
Thanks for your attention

rene.van.horik@dans.knaw.nl
Twitter: @DANSKNAW
dans.knaw.nl/en