

Final Report

Persistent Identifier Types for the Social Sciences (PITSS)

December 2017

Author: René van Horik, DANS
Email: rene.van.horik@dans.knaw.nl

Executive Summary

The RDA collaboration project PITSS (Persistent Identifiers Types for the Social Sciences) aims to introduce and assess outcomes of the RDA Working Group “PID kernel information” into the social science research data community. A workshop was organised, PID profiles (also known as PID kernel information) for social science data sets were defined and discussed and a contribution to the PID kernel information session at the 10th RDA plenary in Montreal was provided. The notes and presentations of this session are available [online](#). The PITSS project resulted in valuable information for further implementation and support for persistent identifiers in the social sciences, more specifically by the [CESSDA ERIC](#). The project made clear that a sustainable (federated) data type registry is an important component of the PID architecture as worked on in the RDA working groups. The implementation of PID kernel information in the social science research domain is not a trivial issue and requires further actions and evaluations that will be carried out by a CESSDA working group on persistent identifiers.

Objectives

The objectives of the PITSS project are to introduce and assess RDA outcomes in the field of PIDs into the social science research data community, more specifically the CESSDA consortium (<https://www.cessda.eu/>). The activities of the RDA working group “PID kernel information” (<https://www.rd-alliance.org/groups/pid-kernel-information-wg>) are of main interest, with an emphasis of the value of PID Profiles and the evaluation of data type registry.

Initial State

Within the social sciences research domain, the CESSDA ERIC (<https://www.cessda.eu/>) is important with respect to the realisation of services to support a research infrastructure. The implementation of PIDs obviously is an important part of this. The CESSDA ERIC initiated a working group on PIDs and members of this working group were involved in the PITSS project. The CESSDA working group on PIDs has formulated basic principles and policies on the usage of PIDs and is interested to learn to what extent the concept of PID Profiles (managed by a data type registry) are of value for the CESSDA service platform.

Initially, the knowledge on this topic was quite limited and the PITSS project provided the means to improve the situation.

Project Outcomes

Members of the project group that represent the social science research data community are familiar with persistent identifiers, such as the DOI to identify datasets. The work done in the RDA on PID Profiles, however, was not known in detail by the group and an important goal of the PITSS project is to get acquainted with this and to assess its value for the social science research data community. The project not only consulted RDA experts to discuss the value of its work for this specific target group, but also a couple of property identifiers were defined and discussed that are relevant for social science research data (e.g. "WasDerivedFrom", "Licence flag", "MultipleInheritance", "HasMetadata". This exercise should be seen as a proof of concept.

The main project outcomes are:

1. A report of the workshop "PID Information Types in the Social Sciences" 29/30 May 2017, The Hague.
2. Report "PID Profiles for the Social Sciences" containing "data type definitions",
3. Presentation at PID Kernel Information WG meeting at RDA plenary and discussion (September 21, 2017, Montreal)
4. Contribution to activities of the CESSDA working group on Persistent Identifiers, that created "[CESSDA PID Policies on Persistent Identifiers](#)" and "[CESSDA best practice guidelines for PID policies](#)"

Dissemination Activities / Publications

Workshop

Title "PID Information Types for the Social Sciences" May 29-30 2017, The Hague

- Workshop program:

https://docs.google.com/document/d/1bBSnV9XqqGp6Mu6SxJrhWgEVZYL6ZsS4Iwr4pSwA_EE/edit

- Workshop report:

<https://docs.google.com/document/d/1uVhVvEijdkshedAZ88OTvQM1zgHKmXhctSvRXZldd068/edit>

Report

- "PID Profiles for the Social Sciences"

<https://docs.google.com/document/d/1gVw3yuwY3FHtMT7yv301R9UsOt841PVygKu9ijw63-M/edit>

- Adjusted version of Strawman "PID Kernel Information Profile". Property Identifiers that were defined as relevant for social science research data objects were added.

<https://docs.google.com/document/d/1QTfreakLUvhrG5LDYbZrZxC5oQVz5bXp-couHVB2tqxc/edit#heading=h.zc4n5hsmyo2h>

PID Kernel WG session at 10th RDA plenary

- Presentation "PITSS Project"

<https://www.rd-alliance.org/sites/default/files/20170921%20PITSS%20presentation%20RDA.pdf>

- Agenda and notes

<https://www.rd-alliance.org/group/pid-kernel-information-wg/wiki/pid-kernel-information-wg-p10-montr%C3%A9al-session>

#	Event	Date	Publication
1	Workshop "PID information Types for the Social Science"	29-30 May, 2017. The Hague	- Workshop program - Workshop report - PID Profiles for the social sciences
2	PID Kernel WG Session at 10 th RDA Plenary	21 September, 2017. Montreal	- Presentation "PITSS Project" - Agenda and notes
3			
4			

Summary & Conclusions

The PITSS project started with an orientation on the work done by a couple of RDA Working Groups in the field of persistent identifiers. These are: RDA PID Kernel Information Working Group and related WGs: RDA Persistent Identifier Types Workgroup, RDA Persistent Identifier Types API Recommendation, RDA Data Type Registry Recommendation, RDA Data Fabric Interest Group.

Next, a workshop was organised attended by experts from the RDA and representatives from the social science research data community (total of 7 people). The RDA outcomes were presented and discussed. Specific features of the social science research data community were presented.

Based on the discussions at the workshop and RDA input, 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)

After that activities were carried out to store the property identifiers in a data type registry (and also the “strawman” document was adjusted, using a copy of the “strawman” document).

The outcomes of the workshop and related activities were presented at the RDA plenary as part of the meeting of the working group on PID kernel information. The social science use case as prepared by the PITSS project brought terminology issues to the forefront (not all components of the PID kernel are defined in an unambiguous way). Also, the issue to what extent social science data sets have specific, dedicated property identifiers, was raised. The LicenceFlag, for instance, refers to authorization / access issues and it was suggested that this is not a proper property identifier for the PID kernel. Further discussion and dialogue is required to achieve consensus on the optimal way to introduce the outcomes of the PID kernel WG in the social science research data community.

The PITSS project facilitated a thorough orientation on RDA activities in the field of PID kernel information with an emphasis on its potential value for the persistent identification of social science research data objects. The implementation of PID services in the CESSDA ERIC will take results from RDA working groups on PIDs into consideration. In the activities of the RDA working groups technological solutions are more prominent than governance and ease of practical implementation. The project enabled the discussion between experts from the RDA community and the social science research data community and this communication will certainly be extended in the future at several occasions (RDA WG, Plenaries, Projects, etc.).