

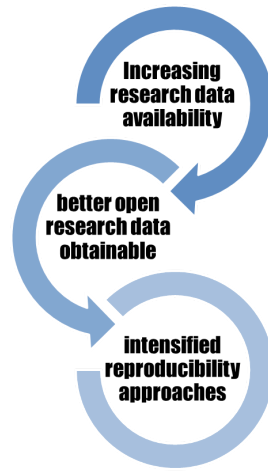
Enhancing PID services towards a more fine-grained granularity level as a base for a FAIR data infrastructure

Janete Saldanha Bach Claus-Peter Klas Peter Mutschke

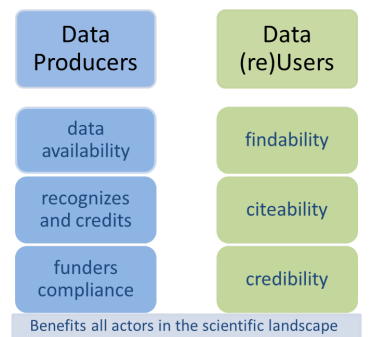
Why PIDs

- Persistent Identifiers (PIDs) are a core element of FAIR data infrastructures;
- Increasing research data availability intensifies re-use and reproducibility approaches;
- Assigning a PID to a whole dataset is not enough to unambiguously data citation;
- PID benefits data producers and data re(users) making research data easier to find and cite;
- Individual elements of the data files can be referenced and retrieved with the required metadata for machine-actionable and human access.

Social Sciences' research data availability



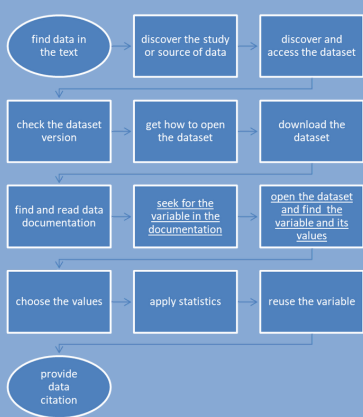
Data producers' and data users' best interests



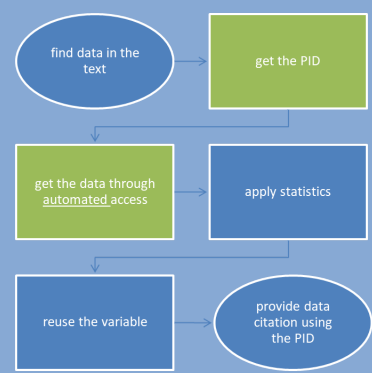
The hurdles of current data citation practice in the Social Sciences

- When re-used, variables are currently cited "in the text" without a unique identifier; usually, only the study, the dataset or questions are cited;
- When using PIDs, researchers can take advantage of those machine-actionable features using scripts (do-files, R scripts, etc.);
- Data can be get through automated access; for instance, an R code would run over the variable's PID and lead the researcher to obtain the exact variable data from that dataset;
- Consequently, this is a faster way to reach data reused, with fewer steps if a PID is unavailable.

Process of accessing a variable without PID



Process of accessing a variable with a PID



Architecture mockup of the PID registration service

- The is a da|ra service widening and assigns a PID with Handle standard (ePIC);
- The service will be upgraded to handle PIDs on variable level;
- PID is registered with the relevant metadata standards;
- An automated way to register PIDs as a bulk is available.

The infrastructure delivers:

- PID suggestion;
- landing page;
- original survey DOI, and metadata schema.

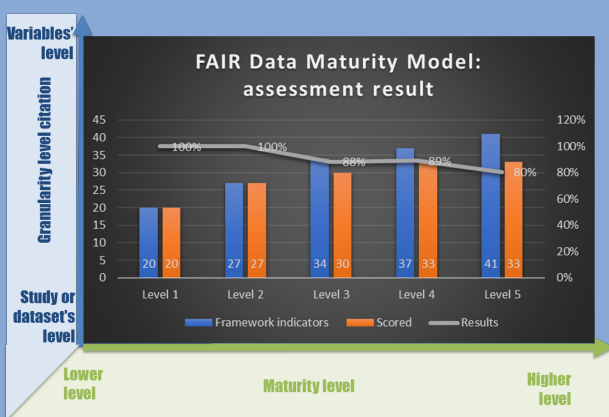
The variable registry process:

- validates the metadata;
- confirms the registered PID;
- stores the metadata.



Data citation practices can perform to a higher maturity level

FAIR Data Maturity Model



We assessed the service under the FAIR Data Maturity Model (RDA Working Group on FAIR Data Maturity Model, 2020):

- The stricter evaluation method applied on each indicator;
- The service meets all indicators classified as essential;
- The results demonstrate the achievements:
 - 100% at level 1
 - 100% at level 2
 - 88% at level 3
 - 89% at level 4
 - 80% at level 5

Given the high relevance of the service for implementing FAIR, we aim to provide reusable and generalized components as a blueprint for other projects.

The service is part of KonsortSWD project deliverable, NFDI funding number 442494171