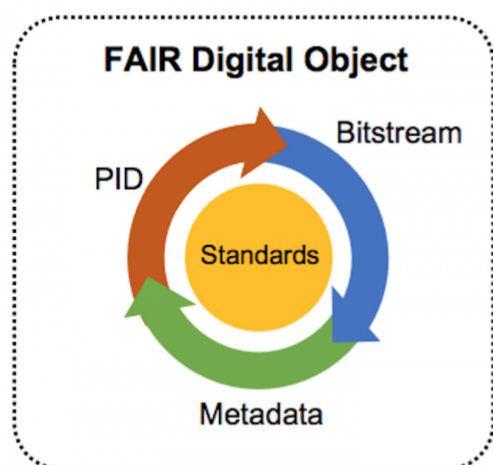FAIRsFAIR
08 Jun 2020 .

The use of persistent identifiers across the data lifecycle emerged as a topic of key interest at the recent FAIRsFAIR webinar FAIRification of Services – Two Examples which highlighted two deliverables from FAIRsFAIR, one on the subject of sustainable support for semantic interoperability and the other concerning the development of FAIR-enabling practices. This article summarises and extends the discussion.

The webinar was led by Claudia Behnke and Hylke Koers of SURFsara, and centered around the reports D2.3 Set of FAIR data repositories features, and Assessment report on FAIRness of services (Milestone 2.7), both of which are still open to comment from the community.

After the presentations the discussion turned to PIDs among other issues. PIDs are an essential part of data lifecycle management. As such their use must be planned and managed not only to facilitate the publication of research data but also to serve research workflows. In addition, to ensure the development of solutions that are unambiguous and (re)usable across domains, the use of identifiers should be envisaged in complex and dynamic contexts.

From webinar presentation by Hylke Koers, SURFsara, FAIR Asessment for Data Services

Identifiers in the Publication Process

The FAIR Data Object (FDO) by definition requires both metadata and a Persistent Identifier (PID). In

practice, these elements are often created when data is published. In fact, data publication may be defined as data being made available (with needed restrictions in place) accompanied by complete metadata and documented PIDs, effectively constituting an FDO. Conceptually publication is therefore separate from data sharing, which involves active research data during the actual research process and less rigorous documentation.

Identifiers in the Pre-publication Phase

The definition of data publication as the creation of an FDO does not eliminate the need for metadata and PIDs before publication. Both are often necessary for machine actionable solutions or provenance tracing and the associated requirements regarding technical authenticity and integrity, rights management, and reproducibility.

Challenges in Citation

The dynamic character of data creates challenges for citation in the sense that there is sometimes the need to cite elements on a level for which metadata does not exist. Typical cases which are each resolved in a different way include subsets, queries, individual data objects and data collections. Also, humans and machines have different needs as users. For PIDs to work in the long run and keep their trustworthiness, they should be used correctly and coherently across communities. In this regard, a refinement in the vocabulary applied to PIDs is required.

Identifying Identifiers

The necessity to assign PIDs to different PID systems and objects will become increasingly relevant as the creation of PID graphs proceeds. This issue will probably be addressed in the upcoming report on FAIR requirements for persistence and interoperability (D2.4) due after the summer. In the mean time, comments and suggestions to the current iteration of the report (D2.1 Report on FAIR requirements for persistence and interoperability) are most welcome and can be made directly on the transcript here.

The webinar presentations and a recording of the session are available here.

```
<?php// print render($content['links']); ?>
```