

Dataviittaus ja tunnisteet

Jessica Parland-von Essen

13.5.2020

The use of Persistent Identifiers for Research Datasets

Recommendation by the Finnish Scientific Community for Open Research

We recommend that all research organisations write a PID policy to promote Open Science and the FAIR principles. This recommendation may be used as a supporting document.

1. The use of identifiers is documented and supports the needs of the research community.
2. All research datasets that are opened, or of which the metadata is published when research outputs are published, are allocated a unique persistent identifier, preferably a DOI or a URN.
3. The persistent identifier directs the user to sufficient metadata.
4. If a dataset becomes unavailable, the persistent identifier directs the user to the metadata on a tombstone page.
5. A research dataset can have persistent identifiers from several systems.
6. DataCite relation types are used to express relations.
7. Semantic meaning should be used mindfully, for instance a persistent identifying element can be used.
8. The structure of the identifiers is defined.
9. Persistent identifiers for human users are user friendly.
10. Creating superfluous persistent identifiers should be avoided.

RESEARCH DATASET

A resource that is produced or used by a researcher during the research process and that underpins the results of the research. Information resources usually include both data that is produced by the research and data that are available for use. The data needs to be complemented with descriptive and technical information about its content. Research data is associated with a lot of information about how it is structured and coded, how it has been produced and how it has been processed. This information should be stored in the metadata, in code books and/or in other documentation. Together with the data file(s) this forms a research dataset.

IDENTIFIERS

The term identifier refers to a unique string that unambiguously identifies an object within its context.

Persistent identifier (PID) refers to a machine readable unique and persistent identifier that resolves in the web and that is openly findable.

RESOLVING

A centrally managed redirect to a human readable web page that represents the content of the identifier and that offers a way to access to the content, if it is digital.

RELATION

Relations here refer to the relations between identifiers and the relations between research datasets and publications.

USER FRIENDLY

This means that the identifier is recognizable to the user and it is easy to use in a citation. For instance, it shouldn't be unreasonably long and it should be possible to create a citation with a suitable coarseness. It should for instance be possible to type a PID from a printed article to the browser manually.

Työtila

<https://wiki.eduuni.fi/x/GYv6B>

<https://doi.org/10.5281/zenodo.3560738>

Ohjeita pysyvien tunnisteiden käyttöönottoon

- Ryhmä haluaa tuottaa käytännönläheistä materiaalia, josta on konkreettista apua organisaatioille tunnisteiden käyttöönotossa
- Eri tyyppiset pysyvät tunnisteet ja niiden ominaisuudet
- Eri tyyppiset käyttötapaukset



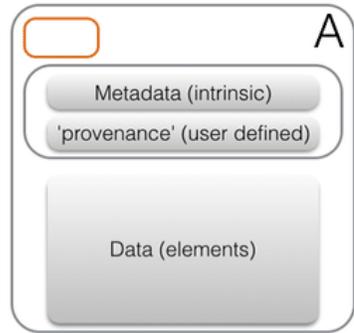
Mahdollisia aiheita

- Vaatimusten kartoittaminen
- Ratkaisujen vertaileminen
- Ylläpidon vastuut
- Eri tunnistejärjestelmät
- Relaatioiden kuvaaminen
- Esimerkit
 - Organisaation datanhallintapalvelu
 - Tutkimushankkeen tietokanta

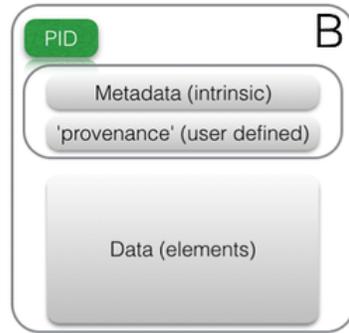


Data as increasingly FAIR Digital Objects

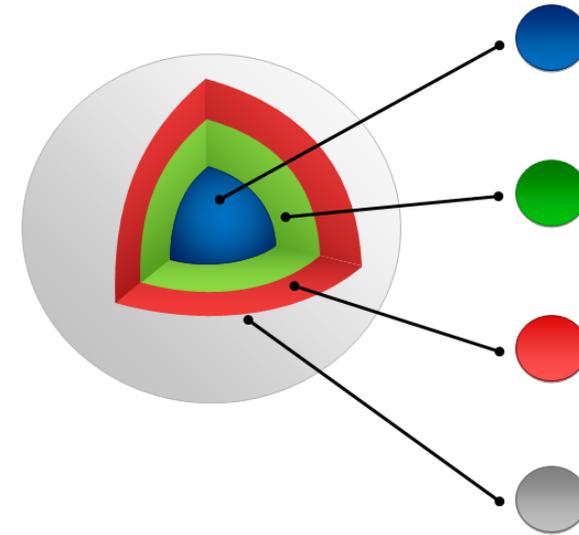
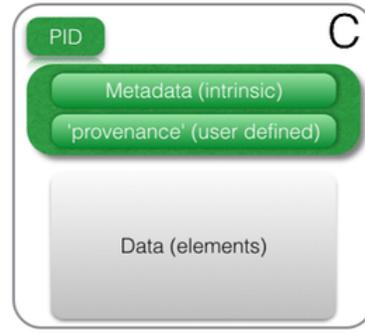
Re-useless data (80%)



Findable



FAIR metadata



DATA

The core bits
At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)
Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCID), projects (RAIDs), funders and associated research resources (RRIDs).

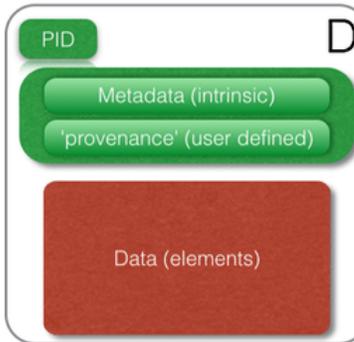
STANDARDS & CODE

Open, documented formats
Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to process and analyse the data.

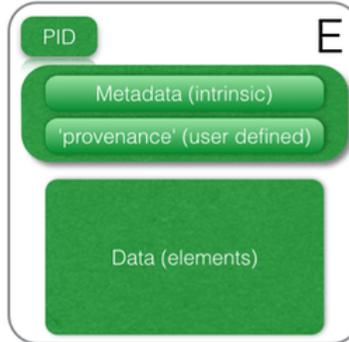
METADATA

Contextual documentation
In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a plurality of relevant attributes and a clear and accessible data usage license.

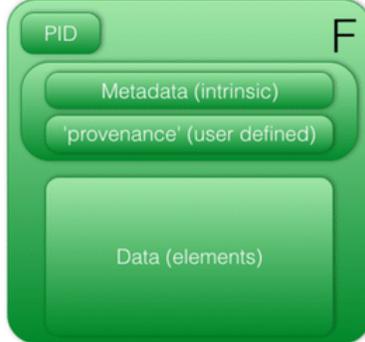
FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked

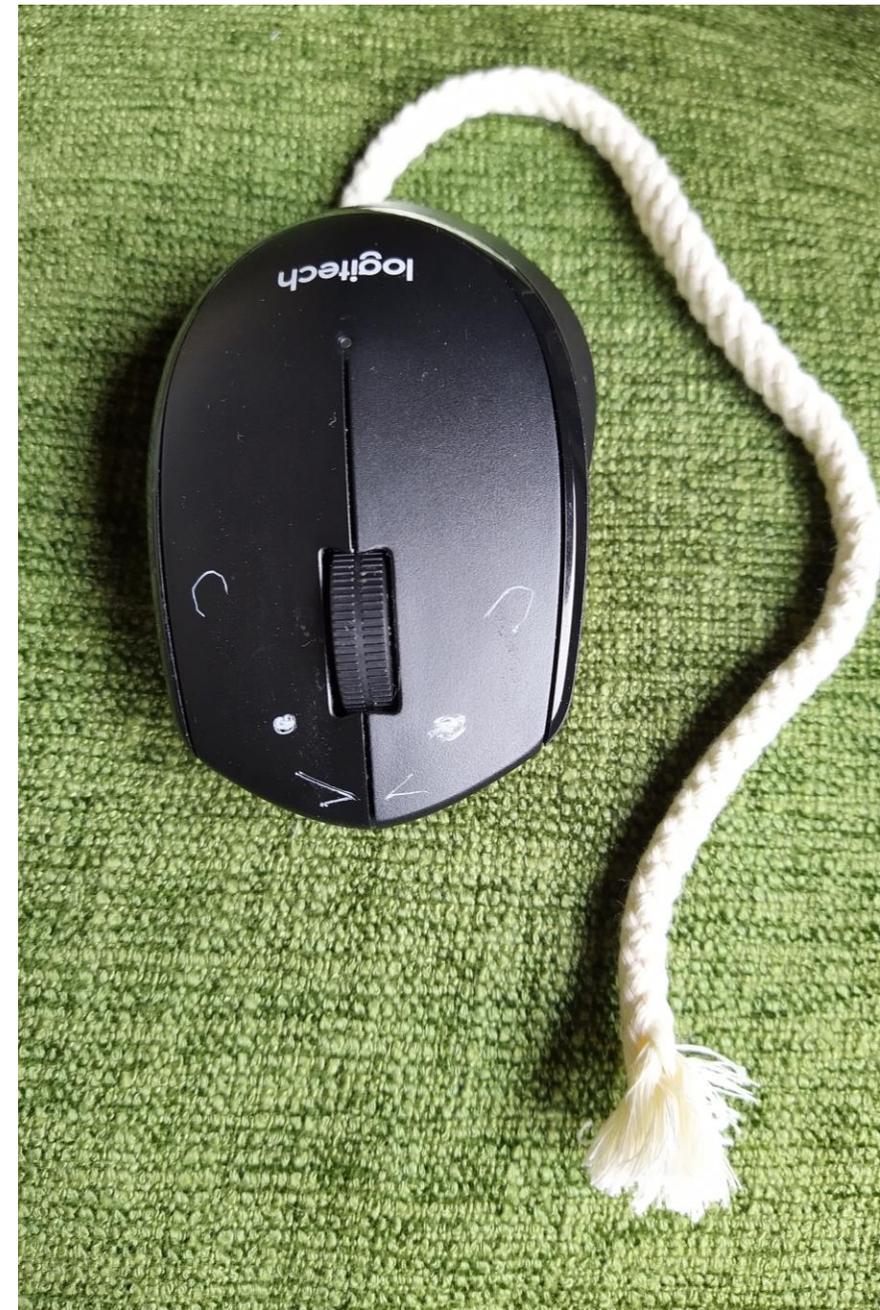


TFIR <https://doi.org/doi:10.2777/1524>

Mons, Barend & Neylon, Cameron & Velterop, Jan & Dumontier, Michel & Bonino da Silva Santos, Luiz Olavo & Wilkinson, Mark. (2017). Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 37. 1-8. <https://doi.org/doi:10.3233/ISU-170824>

Tunnistejärjestelmät (esimerkkejä)

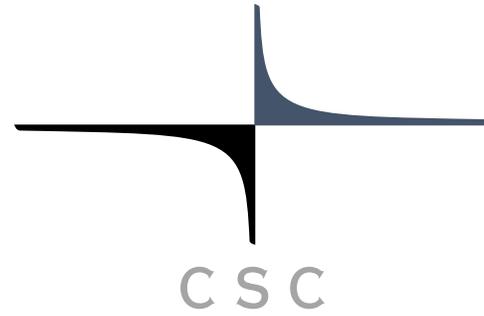
Järjestelmä	Objekti	Ylläpitäjä Suomi
Cool URI	käsitteet	Kansalliskirjasto
Crossref DOI	artikkeli	TSV
DataCite DOI	tutkimusaineist o	Konsortio (CSC)
ePIC handle	tiedosto	(CSC)
URN	infrastruktuuri	Kansalliskirjasto, TTV
ISSN	julkaisusarja	Kansalliskirjasto
ISNI	henkilö	Kansalliskirjasto
ORCID	tutkija	Konsortio (CSC)
RAID	projekti	?



Aikataulu



- Kokous joka toinen viikko
- Yhteiskirjoitusta wikissä
 - <https://wiki.eduuni.fi/x/0iypC>
- Syksyksi helposti lähestyttäviä tuotoksia
- Seurataan nopeaa kehitystä



Jessica Parland-von Essen

Twitter handle @jpve

parland@csc.fi



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc---it-center-for-science



github.com/CSCfi