

Persistent identifiers

Ondřej Košarko

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University

Why PIDs?

- Referencing and interlinking
 - resources
 - semantic definitions
- Reproducibility & reuse of results
- Location on the web is not a good identifier
 - It will stop working
 - The content can be changed at will
 - Names can lead to certain expectations (final_version.pdf)

What is a good identifier

- Persistence
 - Longevity
 - Commitment
- Uniqueness
 - One id identifies one “object”
 - What about the other way?
- Opacity

PID != DOI

A question for you: Have you ever seen or used a different PID than DOI?

PID Systems

- URI/URN (IETF standard)
- Handles
 - DOIs
- PURL
- ARK
- Info-URI
- XRI
- ...

Why so many ... different concepts

- Naming schema
 - urn:oid:0.9.2342.19200300.100.1.3
 - What can you do with it?
 - Eventually <http://oid-info.com/get/0.9.2342.19200300.100.1.3>
 - urn:oasis:names:specification:docbook:dtd:xml:4.1.2
- Resolution system
 - Provides a way to give you a location, e.g. web driven proxies
 - doi:10.1000/182 (or <http://dx.doi.org/10.1000/182>)
 - hdl:loc.gdc/scd0001.00162117695 (or <http://hdl.handle.net/loc.gdc/scd0001.00162117695>)

Why so many ... different extras

- Services around DOIs
 - <http://citation.crosscite.org/>
 - Crossref's Auto-Update for ORCID records
 - ...
- Possibility to attach generic (typed) values
 - Available metadata prior to resolution
- Parts/fragments
 - Useful for continuous data

Why so many ...

- Costs & Ease of use
 - Purl vs (raw) handle vs DOI

Time for another question

DOI[®] Handbook: The DOI name 10.1000/182 identifies the currently applicable latest version of the handbook.

Time for another question

DOI® Handbook: The DOI name 10.1000/182 identifies the currently applicable latest version of the handbook.

If I have assigned a DOI name and I make a change to my material, should I assign a new DOI?

Time for another question

DOI® Handbook: The DOI name 10.1000/182 identifies the currently applicable latest version of the handbook.

If I have assigned a DOI name and I make a change to my material, should I assign a new DOI?

The IDF does not have any rules on this. Individual RAs adopt appropriate rules for their community and application. As a general rule, if the change is substantial and/or it is necessary to identify both the original and the changed material, assign a new DOI name.¹

1: From DOI faq <https://www.doi.org/faq.html>

Versions and PIDs

- The rules are up to the PID providers
- What's a substantial change
 - Who's to say?
 - There was something bothering you enough to invest time into modifying the resource
- If the point of pids is persistence and uniqueness, shouldn't new versions "automatically" get different PID?
- The versions can be linked in metadata

Granularity

- What “deserves” a PID
 - Sentences, paragraphs, articles collections
- Generally not limited by PID systems
 - Might be cumbersome if you get PIDs through a repository
 - You should be able to properly describe the resource

On persistence

- Assigning a PID to a resource doesn't make the resource itself persistent
- Digital archives are meant for that
- PIDs make it easier for you if you are committed to keeping resources available
- In certain cases it makes sense to withdraw a resource. The PID should remain as well as the descriptive metadata

Moving away from repositories to citing arbitrary views of data

“Case study”

- The dataset and the querying service are online
- <https://lindat.mff.cuni.cz/services/pmltq/#!/treebank/pdt30/query/IYWgdg9gJgp gBAEgK4AcUwE5wFwF44DaAUAM7AC2MIALhjFcCYgEYzUDu9YANEaJLE QAbCJyx5CcALrSA3H3DR4CVhy458BOAEI4wMAGMYJahCzI0mXgAo41s7 AwgUdI7CYX0GPmChwHmCAAZhBCluwe4ZgAIERmdgFOIWGiHqhePn6Jzq 4w7sKiMXDR0kREQA/result/svg>
- Let's use shortener!
- Still a URL

shortref.org

- Assign a PID (handle) to that url
- Possible to change the location
 - We keep a track
- Additional metadata
- Healthcheck on the location

Shortref.org for users

- <http://shortref.org>
- Fill the form
 - Url
 - Metadata
- Save the token for future updates
- Use the handle as reference

Shortref.org for services

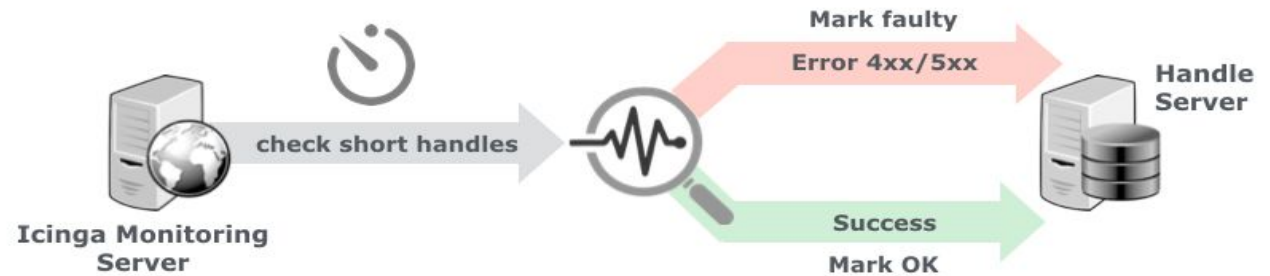
- Use REST API
 - Provide metadata
 - Provide url
 - Store the token
- Show the PID
- Just a click away

Overview

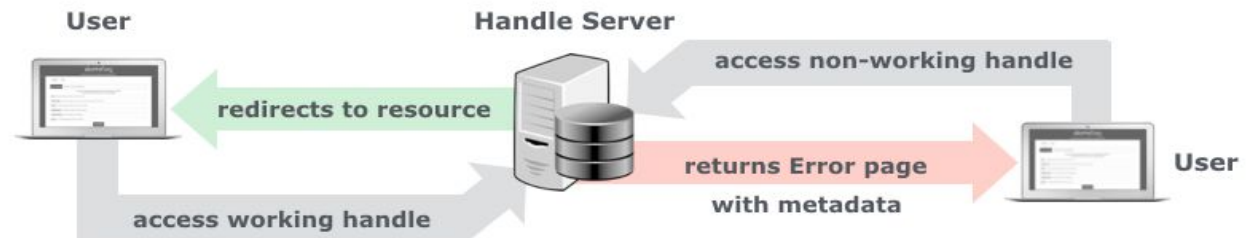
1 Creation



2 Monitoring



3 Resolution



Healthcheck

- The locations are periodically polled
- After certain number of failures an error page is shown
- <http://hdl.handle.net/11346/fff-TN7H>

Error page

shortref.org

Create Easy-to-Cite persistent dynamic URLs

The handle you've tried to access was registered with shortref.org. In order to provide the best services, we at shortref.org actively monitor the URLs handles re

The URL <http://ffff> your handle resolves to is **down since 2016-08-24T07:53:19+02:00**.
Until it comes up again, use the information provided below to gather details about what this URL was.

TITLE fff

POSITORY LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague

IBMITDATE 2016-08-24T13:16:09Z

ORTEMAIL fff

In case you have any question about why you are seeing this page please contact [our support](#).

Powered by  LINDAT
CLARIN

Copyright (c) 2015 UFAL MFF UK. All rights reserved.

Thank you

Ondřej Košarko

kosarko@ufal.mff.cuni.cz

<http://lindat.cz>

<http://shortref.org>

