

AUSPACKEN, IDENTIFIZIEREN, ANPACKEN: ÜBERLEGUNGEN ZUR KONVERGENZ VON LINKED DATA UND PERSISTENTER IDENTIFIKATION



*Cartoon courtesy of Jørgen Stamp,
Digitalbevaring.dk. [CC BY 2.5](https://creativecommons.org/licenses/by/2.5/).*

RENÉ SCHNEIDER

ORCID 0000-0003-4897-8561

JULIEN A. RAEMY

@JULSRAEMY - ORCID 0000-0002-4711-5759

HAUTE ECOLE DE GESTION DE GENÈVE

INFORMATIONEN – DIGITAL VERPACKT

ZENTRALBIBLIOTHEK ZÜRICH, 22. SEPTEMBER 2020

ÜBERBLICK

1. Einleitung
2. Auspacken & Identifizieren
3. Anpacken
4. Fazit

1. EINLEITUNG



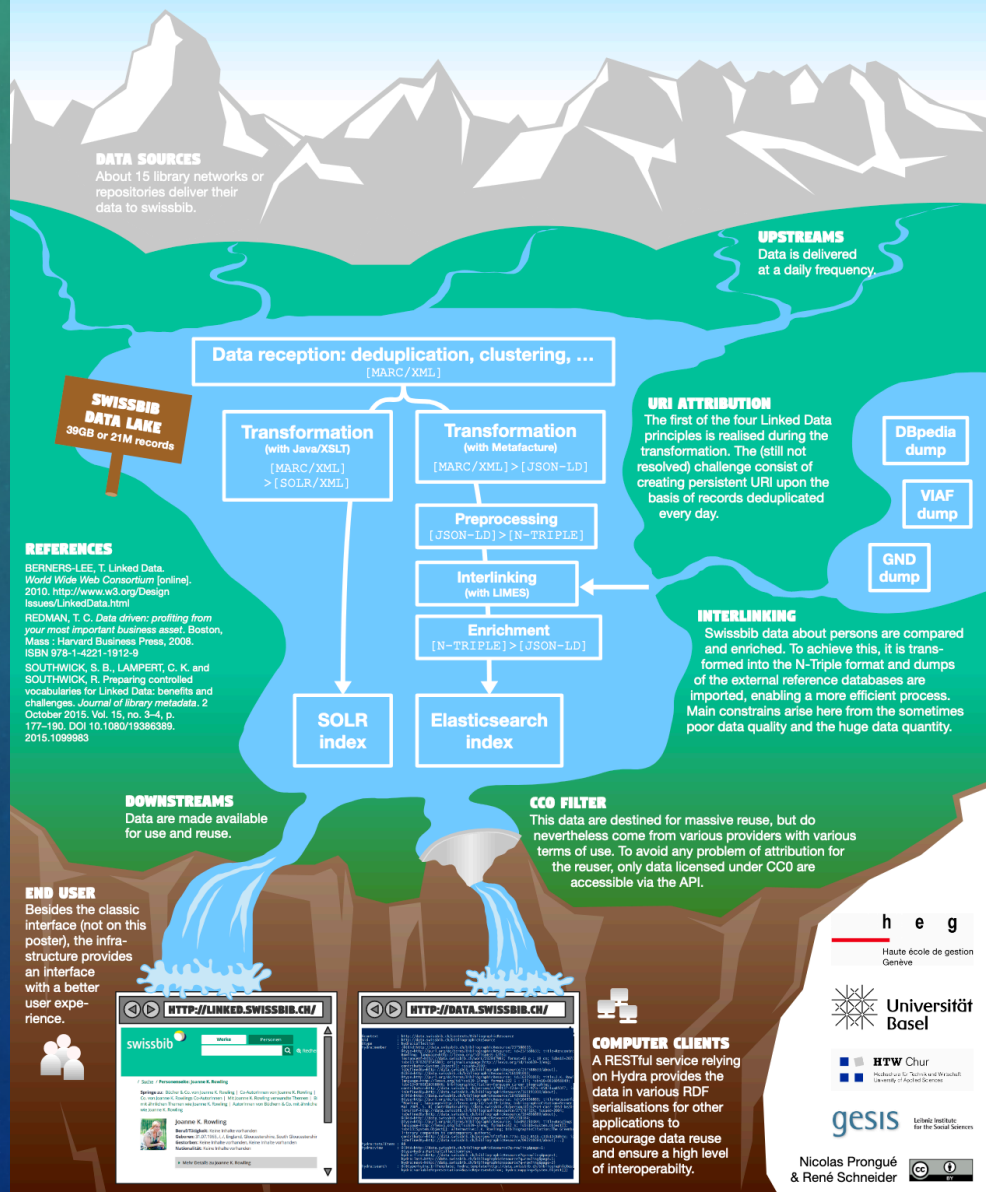
https://www.reddit.com/r/socialism/comments/2yq08w/redesigning_the_hammer_and_sicklexpost/

LINKED.SWISSBIB.CH

DATA STREAMS IN LINKED.SWISSBIB.CH

THE SWISS METACATALOG IN THE LINKED OPEN DATA CLOUD

The project linked.swissbib.ch integrates the metacatalog swissbib into the semantic web, together with the creation of a Linked Data service. This implies the transformation of the metadata from different Swiss library networks into an RDF based format, its enrichment and its publication. Main challenges arise from the requirement of fully automated processes to allow daily updates.



Hipler, Prongué, Schneider 2018

<https://arodes.hes-so.ch/record/2619/>

Data reception: deduplication, clustering, ...

[MARC/XML]

Transformation

(with Java/XSLT)

[MARC/XML]
> [SOLR/XML]

Transformation

(with Metafacture)

[MARC/XML] > [JSON-LD]

Preprocessing

[JSON-LD] > [N-TRIPLE]

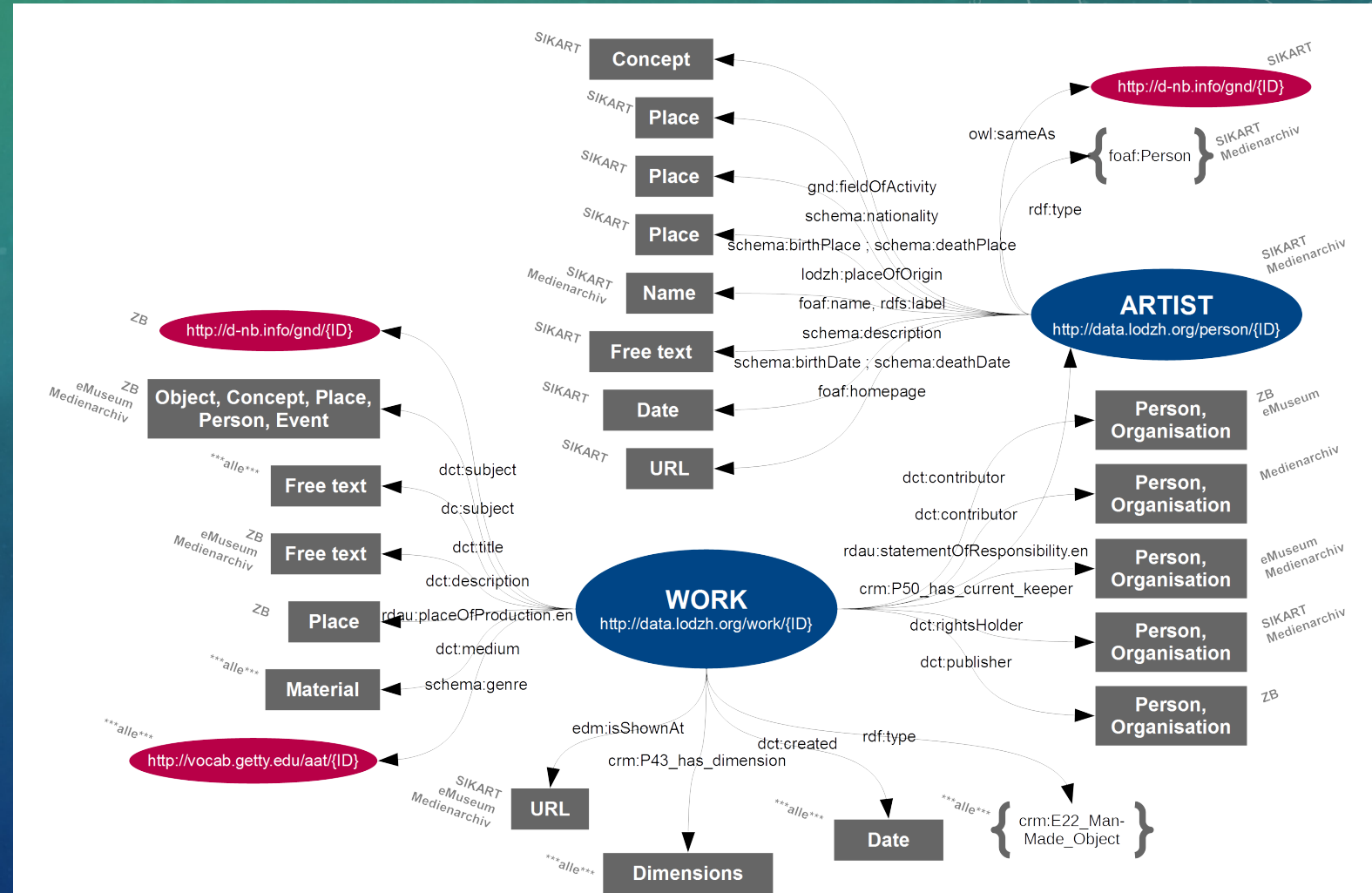
URI ATTRIBUTION

The first of the four Linked Data principles is realised during the transformation. The (still not resolved) challenge consist of **creating persistent URI** upon the basis of records deduplicated every day.

Problem: Notwendigkeit persistenter Identifikation,
im Idealfall dynamisch persistenter Identifikation

LINKED OPEN DATA ZURICH (LODZ)

Problem: Diversität und
feine Granularität ohne
absolute Notwendigkeit
einer semantischen
Spezifikation



Prongué et al. 2017

<https://arodes.hes-so.ch/record/1972>

2. AUSPACKEN UND IDENTIFIZIEREN

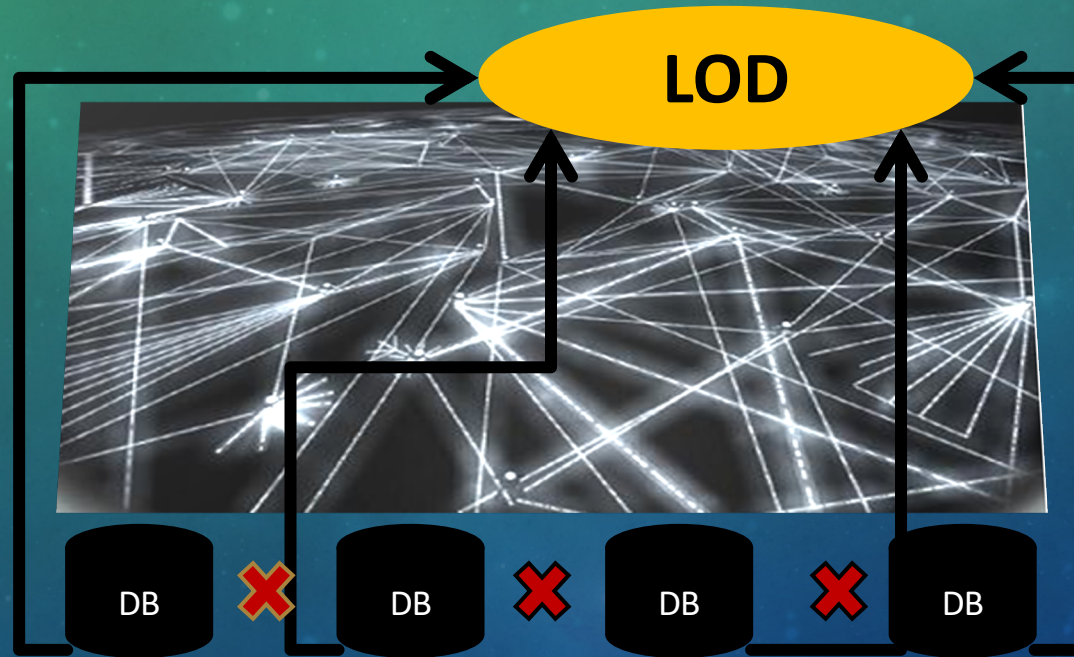


Created by Sarah Tan
from Noun Project



Created by Chris Homan
from Noun Project

AUSPACKEN - DAS PROBLEM DER DATENSILOS



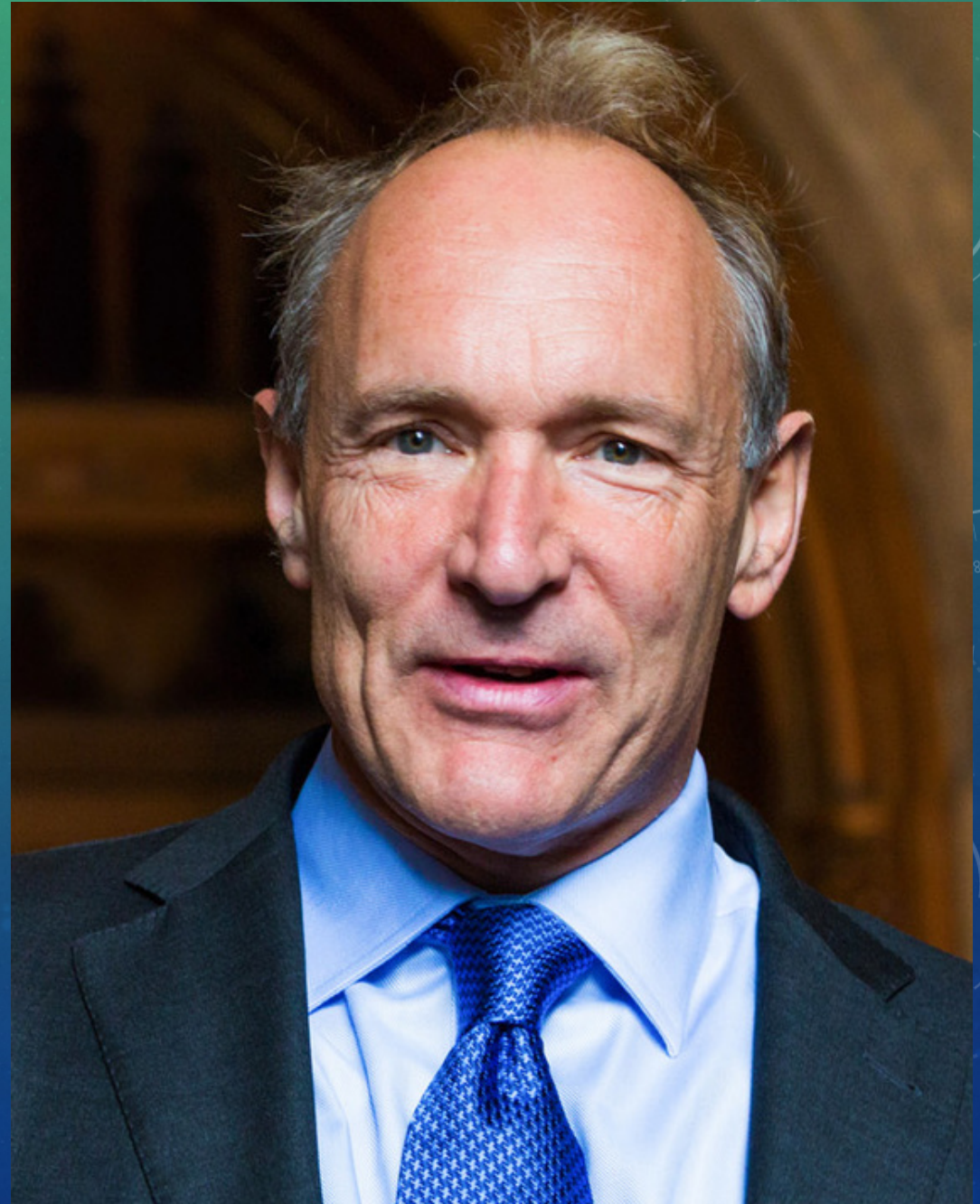
AUSPACKEN – COOL URIS

*“Cool URIs don’t change”**

Tim Berners-Lee

<https://www.w3.org/Provider/Style/URI>

**or do they?*



PERSISTENT UNIFORM RESOURCE LOCATORS (PURLS)

Metadata

URI	http://purl.org/dc/terms/
Namespace	http://purl.org/dc/terms/
homepage	http://dublincore.org/documents/dcmi-terms
Description	an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative, including properties, vocabulary encoding schemes, syntax encoding schemes, and classes. @en
Language	English en
Creator	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI
Publisher	Dublin Core Metadata Initiative http://purl.org/dc/aboutdcmi#DCMI
Comment	(2013-03-07) Bernard Vatant : Prefix restored to dcterms (2014-03-14) Bernard Vatant : This vocabulary is one of the most used in the LOD cloud, and here to stay, even if the purl redirection is sometimes down, like at the time I write this review. (2015-03-24) Bernard Vatant : Annual review OK (2016-05-10) Ghislain Atemezing : Annual review OK (2018-08-02) Ghislain Atemezing : Annual review - OK

PURL CRISIS



Created by Fauzan Adiima
from Noun Project



Mehr Informationen: https://docs.google.com/document/d/17TBUja8z8EJGx5ZEyknP3gWuPITf6lt-XkXDHU_CKaY/edit

IDENTIFIZIEREN: PERSISTENT IDENTIFIER (PID)

A persistent identifier

is a long-lasting and biunique reference to a digital resource.*

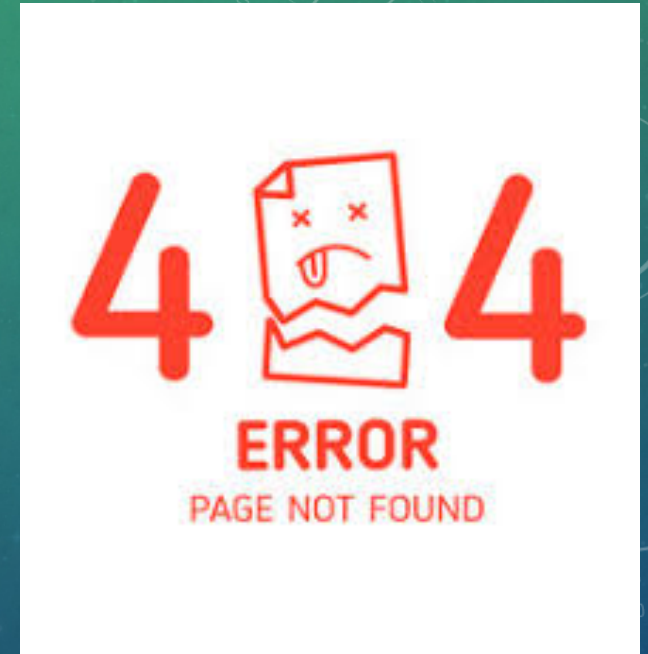
**biunique: unique in both directions (bijective) / eineindeutig*

Ein Art von ISBN für Daten und alle digitalen Ressourcen.



ZWECK

- Dauerhafter (nicht permanenter) Zugang
- Verhindern von Fehlermeldungen



<https://www.interserver.net/tips/kb/404-error-fix/>

PERSISTENT NICHT PERMANENT

10 persistent myths about PIDs by John Kunze

- PIDs guarantee access. Get real. All PID services run on evolving software/hardware that no vendor warranties.
- PIDs rarely break. Nonsense. Millions of PIDs are broken. Updating redirection tables is real work for you and your successors.
- ...

<https://n2t.net/ark:/13030/c7gb1xh09>

John Kunze @jakkbl

10 persistent myths about persistent identifiers (PIDs), whether ARK, DOI, Handle, URN, PURL, or "Cool URL" (from my experience running [EZID.cdlib.org](https://ezid.cdlib.org) and [N2T.net](https://n2t.net)).

EZID Home
EZID service creates long-term identifiers and is a key component for keeping data and digital objects safe while ...
ezid.cdlib.org

8:30 PM · Aug 24, 2018 · [Twitter Web Client](#)

51 Retweets 90 Likes 8 Quotes

John Kunze @jakkbl · Aug 24, 2018
Replying to @jakkbl
Myth 1: PIDs guarantee access. Get real. All PID services run on evolving software/hardware that no vendor warranties. How could anyone, let alone poor non-profits, change that?

John Kunze @jakkbl · Aug 24, 2018
Myth 2: PIDs rarely break. Nonsense. Millions of PIDs are broken. Updating redirection tables is real work for you and your successors. (You do have a succession plan, right?)

John Kunze @jakkbl · Aug 24, 2018
Myth 3: PIDs must not be URLs. What a crock. PIDs not carried inside clickable URLs are irrelevant or only of academic interest.

PIDs und cool URIs (Bazzanella, Bortoli, Bouquet 2013)

Feature	PIDs	Cool URIs
Resolver	YES	NO
Authority	YES	NO
Naming authorities	YES	NO
Level of trust	HIGH	LOW
Policies	YES	NO
Persistence	YES	NO
Actionability of IDs	Partially	YES
Uniqueness	YES	NO
Content change	NO	YES
Content negotiation	NO	YES
Cross linkage	NO	YES
Effort for implementation	HIGH	LOW
Costs for users	Potentially HIGH	LOW
Sustainability issues	MANY	FEW
Identified entities	Mainly digital objects	Everything
Bridge metadata	NO	YES

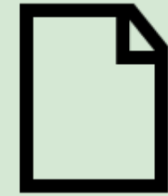
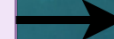
PROBLEMATIK



ID



Linking



object and/or associated metadata

Broken PIDs

Uncool URIs

Zombie PIDs

3. ANPACKEN



Created by Alfredo @ IconsAlfredo.com
from Noun Project

BIBLIOTHÈQUE NATIONALE DE FRANCE (BNF)

Linked Data

Data.bnf.fr

PID

Archival Resource Key (ARK)

DATA & SERVICE CENTER FOR THE HUMANITIES (DASCH)

PID with timestamps

Archival Resource Key (ARK)

Linked Data

Knowledge Organization, Representation, and Annotation (Knora)

OPENCITATIONS

PIDs

Linked Data

OpenCitations Data Model (OCDM)

PIDs

INTERNATIONAL STANDARD MANUSCRIPT IDENTIFIER (ISMI)

ISMI IDs

Archival Resource Key (ARK) or bespoke scheme

Ontology (to be created)

PIDs

ABSTRAKTION DES ANWENDUNGSFALLES

BNF

DASCH

OPENCITATIONS

ISMI

Linked Data
Data.bnf.fr

PID with timestamps
*Archival Resource Key
(ARK)*

PID

ISMI IDs
*Archival Resource Key
(ARK) or bespoke scheme*

PID
*Archival Resource Key
(ARK)*

Linked Data
*Knowledge Organization,
Representation, and
Annotation (Knora)*

Linked Data
*OpenCitations Data Model
(OCDM)*

Ontology (to be created)

PID

PIDs

4. FAZIT

- Hohes Potenzial in der Kombination von LOD und PIDs
- Eine Fusion der Aspekte “Trust” und Persistenz auf Basis einer klareren und logischeren Struktur
- Bessere Modularität (“thinking in blocks and layers”)



BIBLIOGRAPHIE (1)

- BAZZANELLA, Barbara, BORTOLI, Stefano and BOUQUET, Paolo, 2013. Can persistent identifiers be cool? *International journal of digital curation*. 14 June 2013. Vol. 8, no. 1, p. 14–28. DOI [10.2218/ijdc.v8i1.246](https://doi.org/10.2218/ijdc.v8i1.246).
- BERMÈS, Emmanuelle, 2006. *Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF* [online]. Paris, France: Bibliothèque nationale de France. [Accessed 17 September 2020]. Available from: https://web.archive.org/web/20181006042857/http://www.bnf.fr/documents/ark_presentation_bermes_2006.pdf
- CEVEY, Matthieu, RAEMY, Julien A. and SCHNEIDER, René (dir), 2020. *Swiss PID Hub : creation pour la gestion des identifiantes pérennes en Suisse* [online]. Genève, Suisse : Haute école de gestion de Genève. [Accessed 17 September 2020]. Available from: <https://doc.rero.ch/record/328460/>
- ESPASANDIN, Kate, JAQUET, Aurélie, LEFORT, Lise and SCHNEIDER, René (dir), 2018. TRMASID 14: *Panorama et modélisation d'identifiants pérennes pour la création d'identités de confiance* [online]. Genève, Suisse: Haute école de gestion de Genève. [Accessed 30 August 2020]. Available from: <https://doc.rero.ch/record/309479>
- EU. DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION, 2018. KI-06-18-206-EN-N: *Turning FAIR into reality. Final Report and Action Plan on FAIR Data* [online]. Brussels, Belgium. [Accessed 17 September 2020]. Available from: <https://doi.org/10.2777/1524>
- HILSE, Hans-Werner and KOTHE, Jochen, 2006. *Implementing persistent identifiers: overview of concepts, guidelines and recommendations*. London: CERL. ISBN 978-90-6984-508-1.

BIBLIOGRAPHIE (2)

- HIPLER, Günter, PRONGUÉ, Nicolas and SCHNEIDER, René, 2018. Swissbib und linked.swissbib.ch: Leistung und Potenziale einer offenen Plattform für Schweizer Bibliotheksdaten. In: *Bibliotheken der Schweiz: Innovation durch Kooperation* [online]. Berlin, Boston: De Gruyter. p. 160–172. [Accessed 18 September 2020]. ISBN 978-3-11-055379-6. Available from: <https://doi.org/10.1515/9783110553796-008>
- KUNZE, John, CALVERT, Scout, DEBARRY, Jeremy D., HANLON, Matthew, JANÉE, Greg and SWEAT, Sandra, 2017. Persistence Statements: Describing Digital Stickiness. *Data Science Journal*. 14 August 2017. Vol. 16, no. 0, p. 39. DOI [10.5334/dsj-2017-039](https://doi.org/10.5334/dsj-2017-039).
- KUNZE, John, 2018. Ten persistent myths about persistent identifiers. *UC Office of the President: California Digital Library* [online]. 24 August 2018. [Accessed 18 September 2020]. Available from: <https://n2t.net/ark:/13030/c7gb1xh09>
- MEADOWS, Alice, 2017. PIDapalooza – the open festival for persistent identifiers. *Insights*. 8 November 2017. Vol. 30, no. 3, p. 161–164. DOI [10.1629/uksg.393](https://doi.org/10.1629/uksg.393).
- NICHOLAS, Nick, WARD, Nigel and BLINCO, Kerry, 2009. A policy checklist for enabling persistence of identifiers. *D-Lib magazine* [online]. January 2009. Vol. 15, no. 1/2. [Accessed 20 May 2019]. DOI [10.1045/january2009-nicholas](https://doi.org/10.1045/january2009-nicholas). Available from: <http://www.dlib.org/dlib/january09/nicholas/01nicholas.html>
- PEYRARD, Sébastien, KUNZE, John A. and TRAMONI, Jean-Philippe, 2014. The ARK Identifier Scheme: Lessons Learnt at the BnF and Questions Yet Unanswered. *International Conference on Dublin Core and Metadata Applications*. 8 October 2014. P. 83–94.

BIBLIOGRAPHIE (3)

- PRONGUÉ, Nicolas, RICCI, Fabio, SCHNEIDER, René and SCHURTE, René, 2017. Art and design as linked data: the LODZ project (Linked Open Data Zurich). *Libellarium: journal for the research of writing, books, and cultural heritage institutions* [online]. 2 March 2017. Vol. 9, no. 2. [Accessed 10 January 2020]. Available from: <http://www.libellarium.org/index.php/libellarium/article/view/256>
- RAEMY, Julien A. and SCHNEIDER, René, 2019. Towards Trusted Identities for Swiss Researchers and their Data. *International Journal of Digital Curation*. 11 September 2019. Vol. 14, no. 1, p. 303–314. DOI [10.2218/ijdc.v14i1.596](https://doi.org/10.2218/ijdc.v14i1.596).
- VAN DE SOMPEL, Herbert and NELSON, Michael L., 2015. Reminiscing About 15 Years of Interoperability Efforts. *D-Lib Magazine* [online]. November 2015. Vol. 21, no. 11/12. DOI [10.1045/november2015-vandesompe1](https://doi.org/10.1045/november2015-vandesompe1)
- VAN DE SOMPEL, Herbert, SANDERSON, Robert, SHANKAR, Harihar and KLEIN, Martin, 2014. Persistent Identifiers for Scholarly Assets and the Web: The Need for an Unambiguous Mapping. *International Journal of Digital Curation* [online]. 23 July 2014. Vol. 9, no. 1. DOI [10.2218/ijdc.v9i1.320](https://doi.org/10.2218/ijdc.v9i1.320).
- VAN DE SOMPEL, Herbert, KLEIN, Martin and JONES, Shawn M., 2016. Persistent URIs Must Be Used To Be Persistent. *arXiv:1602.09102 [cs]* [online]. 29 February 2016. [Accessed 17 September 2020]. Available from: <http://arxiv.org/abs/1602.09102>

FRAGEN

rene.schneider@hesge.ch

julien.raemy@hesge.ch



Created by iconsphere
from Noun Project