TOWARDS TRUSTED IDENTITIES FOR SWISS RESEARCHERS AND THEIR DATA

DOI: 10.5281/zenodo.2415996

REÑÉ SCHNEIDER
ORCID: 0000-0003-4897-8561

JULIEN A. RAEMY
ORCID: 0000-0002-4711-5759

HES//SO – HAUTE ECOLE DE GESTION, GENEVA
GRID ID: GRID.483301.D
What?

Why?

How!
A persistent identifier

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

*biunique: unique in both directions (bijective)
DEFINITION II

A kind of ISBN for data and more,
for everything that can represent a resource.
COMPOSITION

Usually it has two parts:

1. A unique identifier (ensures the provenance of a digital resource)
2. A Location for the resource over time (ensures that the identifier resolves to the correct location)

PID ≠ PID

• Publications
• Data
• Persons
• Organisations
• Citations

and more: (antibodies, fictious characters, places, plants, e-books, ...)

---

**Image Description:** The image contains icons representing various identifiers and registries such as DOI (Digital Object Identifier), ORCID, ARK (Authority Record Key), Handle.Net Registry, HINARI (Health Information Network), PURL (Persistent URL Registry), IGSN (International Geoscientific Name Registry), GRID (Global Research Identifier), and NLA party ID.
RISKS

PID Crisis: PURL -> long-lasting PIDs need long-lasting institutions!

Zombie PIDs -> PIDs need Data Curation!
What?

Why?

How!
Why?

• Why PIDs?
• Why this project?
Why?

• Why PIDs?
• Why this project?
IN ORDER TO

- Create long lasting (not permanent) access
- Avoid error messages

https://www.interserver.net/tips/kb/404-error-fix/
Dataset Papers in Science

Volume 2014 (2014), Article ID 172182, 7 pages
http://dx.doi.org/10.1155/2014/172182

Dataset Paper

Mapping the Slums of Dhaka from 2006 to 2010

Oliver Gruebner,1 Jonathan Sachs,2 Anika Nockert,3 Michael Frings,3 Md. Mobarak Hossain Khan,4 Tobia Lakes,3 and Patrick Hoster3

1Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 517, New York, NY 10032, USA
2Department of Tropical Medicine, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA 70112, USA
3Geography Department, Humboldt-Universitat zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
4Department of Public Health Medicine, School of Public Health, University of Bielefeld, P.O. Box 100131, 33501 Bielefeld, Germany

Received 30 September 2013; Accepted 4 March 2014; Published 25 June 2014
### Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at [http://dx.doi.org/10.1155/2014/172182/dataset](http://dx.doi.org/10.1155/2014/172182/dataset). In addition, it can be downloaded from the spatial data infrastructure at Humboldt University of Berlin ([http://gdi.geo.hu-berlin.de/results.php?searchterm=dhaka](http://gdi.geo.hu-berlin.de/results.php?searchterm=dhaka)).
FAIRNESS

http://www.dit.ie/dsrh/data/fairdata/

PIDs are essential and indispensable to create fair data.

F1 Principle:

(meta)data are assigned a globally unique and eternally persistent identifier
Rec. 3: A model for FAIR Data Objects
Implementing FAIR requires a model for FAIR Data Objects which by definition have a PID linked to different types of essential metadata, including provenance and licencing. The use of community standards and sharing of code is also fundamental for interoperability and reuse.
CONCLUSION I

The quality of data repositories stands and falls with PIDs.

PIDs
- are to be considered as a *conditio sine qua non* of (research) data management
- function as both
  - a hinge to create trusted identities
  - a fulcrum to create fair data
Why?

• Why PIDs?
• Why this project?
SWISS PID LANDSCAPE
USUALLY PEOPLE TEND TO SAY

«We have DOIs, that’s far enough!»

1. Costing Model. (In Switzerland!)

2. Mainly for articles and some data...

3. with low granularity.
WHAT IS REALLY WANTED (WITHOUT REALLY HAVING IN MIND)
CONCLUSION II

The situation in Switzerland

(almost) exclusively DOIs and (almost) exclusively free for one federal institution

is unsufficient and unsatisfactory in any case!

It takes more!
What?

Why?

How!
ICOPAD PROJECT  Identités de confiance pour les données de l’art et du design

- Haute Ecole de Gestion Geneva – Instigator and Project Manager
- Zentralbibliothek Zürich / Zurich Central Library
- Zürcher Hochschule der Künste / Zurich University of the Arts
- Schweizerisches Institut für Kunstwissenschaft / Swiss Inst. for Art Research
# PROJECT PARTNER AND THEIR DATA SETS

<table>
<thead>
<tr>
<th>Institution</th>
<th>Data set types/entities</th>
<th>Needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIK-ISEA</td>
<td>Artists&lt;br&gt;Artworks&lt;br&gt;Dictionary entries</td>
<td>Diverse PIDs and links to normed data.</td>
</tr>
<tr>
<td>ZB</td>
<td>Digital surrogates</td>
<td>Fine level of granularity.</td>
</tr>
<tr>
<td>ZHdK</td>
<td>Artists&lt;br&gt;Artworks&lt;br&gt;Events&lt;br&gt;Films&lt;br&gt;Glossary entries&lt;br&gt;Projects&lt;br&gt;Research Data</td>
<td>Further development of applications such as eMuseum and Medienarchiv.</td>
</tr>
</tbody>
</table>
Espasandin, Kate / Jacquet, Kate / Lefort, Lise: Panorama et modélisation d’identifiants pérennes pour la création d’identités de confiance.
http://doc.rero.ch/record/309479
Espasandin, Kate / Jacquet, Kate / Lefort, Lise
PID DECISION TREE (FLOWCHART)

• Inspired by prior work done @ ANDS
• Adopted
• Adapted
• Developed
CONCLUSIONS III

DOI +  
\[ C(\text{doi}) = x \]

DOI + 1  
\[ C(\text{doi}) = a \]

DOI + n  
\[ C(\text{doi}) = (x_1, x_2, \ldots, x_n) \]

DOI + 1 + LD  
\[ C(\text{doi}) = a \rightarrow \text{owl: sameAs} (x_1, x_2, \ldots, x_n) \]

\[ a = \text{ark} \]

*(Archival Ressource Key provided by California Digital Library)*
• ARK identifiers are free
• ARKs are built using a completely different theoretical model, consisting of a decentral and domain (i.e. DNS) agnostic approach
• ARKs allow to use with ease LOD on top of them
• ARKs can effortlessly be combined with other specifications such as the International Image Interoperability Framework (IIIF) canonical URI syntax
SOLUTION APPROACHES

If NOT DOI @ ETH | FORS AND if NOT data archived @ DaSCH

Your own PID Attribution Service

PID service request

PID service request

existing ark service

You

-existing ark service

Swiss PID Hub

ark service to create

DaSCH Uni Bas

multitude of PIDs
# Hostname/PID Authority Matrix

<table>
<thead>
<tr>
<th>PID Authority</th>
<th>One PID Authority for all organisations</th>
<th>Each institution is its own PID Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>One hostname per domain/activity</td>
<td>1) ARK via its own means</td>
<td>N/A</td>
</tr>
<tr>
<td>One hostname per institution</td>
<td>2) ARK via DaSch</td>
<td>1) ARK via its own means</td>
</tr>
<tr>
<td>One hostname for all types of data</td>
<td>3) ARK via a national Hub</td>
<td>4) RRID-like (ADID) via its own means</td>
</tr>
<tr>
<td>Hostname - Identifier Service Provider (NAMH)</td>
<td>5) RRID-like (ADID) via a national Hub</td>
<td>5) RRID-like (ADID) via a national Hub</td>
</tr>
</tbody>
</table>

- **1) ARK via its own means**
  - [ark:/52za.ark.eu/11111/0000336](http://ark:/52za.ark.eu/11111/0000336)
  - [ark:/52za.ark.eu/22222/0000336](http://ark:/52za.ark.eu/22222/0000336)
  - [ark:/52za.ark.eu/33333/0000336](http://ark:/52za.ark.eu/33333/0000336)

- **2) ARK via DaSch**
  - [ark:1234/56789](http://ark:1234/56789)
  - [ark:1234/56789](http://ark:1234/56789)
  - [ark:1234/56789](http://ark:1234/56789)

- **3) ARK via a national Hub**
  - [ark:1234/56789](http://ark:1234/56789)
  - [ark:1234/56789](http://ark:1234/56789)
  - [ark:1234/56789](http://ark:1234/56789)

- **4) RRID-like (ADID) via its own means**
  - [example:1234/56789](http://example:1234/56789)
  - [example:1234/56789](http://example:1234/56789)

- **5) RRID-like (ADID) via a national Hub**
  - [example:1234/56789](http://example:1234/56789)
  - [example:1234/56789](http://example:1234/56789)
<table>
<thead>
<tr>
<th>Hostname - Identifier Service Provider (NMAH)</th>
<th>One PID Authority for all organisations</th>
<th>Each institution is its own PID Authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostname for all types of data</td>
<td>2) ARK via DaSCH</td>
<td>1) ARK via its own means</td>
</tr>
<tr>
<td></td>
<td><a href="http://ark.dasch.swiss/ark:72163/lex4000336z">http://ark.dasch.swiss/ark:72163/lex4000336z</a></td>
<td><a href="http://n2t.sikart.ch/ark:/11111/lex4000336z">http://n2t.sikart.ch/ark:/11111/lex4000336z</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://ark.dasch.swiss/ark:72163/ema45784v">http://ark.dasch.swiss/ark:72163/ema45784v</a></td>
<td><a href="http://n2t.zb.ch/ark:/22222/ema45784v">http://n2t.zb.ch/ark:/22222/ema45784v</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://ark.dasch.swiss/ark:/72163/bfic02001t">http://ark.dasch.swiss/ark:/72163/bfic02001t</a></td>
<td><a href="http://n2t.zhdk.ch/ark:/33333/bfic02001t">http://n2t.zhdk.ch/ark:/33333/bfic02001t</a></td>
</tr>
<tr>
<td></td>
<td>3) ARK via a national Hub</td>
<td>4) RRID-like (ADID) via its own means</td>
</tr>
<tr>
<td></td>
<td><a href="http://sikart.icopad.ch/ark:/99999/lex4000336z">http://sikart.icopad.ch/ark:/99999/lex4000336z</a></td>
<td><a href="http://n2t.sikart.ch/aidid:lex4000336z">http://n2t.sikart.ch/aidid:lex4000336z</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://zb.icopad.ch/ark:/99999/ema45784v">http://zb.icopad.ch/ark:/99999/ema45784v</a></td>
<td><a href="http://n2t.zb.ch/aidid:ema45784v">http://n2t.zb.ch/aidid:ema45784v</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://zhdk.icopad.ch/ark:/99999/bfic02001t">http://zhdk.icopad.ch/ark:/99999/bfic02001t</a></td>
<td><a href="http://n2t.zhdk.ch/aidid:bfic02001t">http://n2t.zhdk.ch/aidid:bfic02001t</a></td>
</tr>
<tr>
<td></td>
<td>5) RRID-like (ADID) via a national Hub</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td><a href="http://sikart.icopad.ch/aidid:lex4000336z">http://sikart.icopad.ch/aidid:lex4000336z</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://zb.icopad.ch/aidid:ema45784v">http://zb.icopad.ch/aidid:ema45784v</a></td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://zhdk.icopad.ch/aidid:bfic02001t">http://zhdk.icopad.ch/aidid:bfic02001t</a></td>
<td></td>
</tr>
<tr>
<td>One hostname per institution</td>
<td>3) ARK via a national Hub</td>
<td>1) ARK via its own means</td>
</tr>
<tr>
<td></td>
<td><a href="http://artist.icopad.ch/ark:/99999/lex4000336z">http://artist.icopad.ch/ark:/99999/lex4000336z</a></td>
<td><a href="http://artist.sikart.ch/ark:/11111/lex4000336z">http://artist.sikart.ch/ark:/11111/lex4000336z</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://manuscripts.icopad.ch/ark:/99999/ema45784v">http://manuscripts.icopad.ch/ark:/99999/ema45784v</a></td>
<td><a href="http://manuscripts.zb.ch/ark:/22222/ema45784v">http://manuscripts.zb.ch/ark:/22222/ema45784v</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://medienarchiv.icopad.ch/ark:/99999/bfic02001t">http://medienarchiv.icopad.ch/ark:/99999/bfic02001t</a></td>
<td><a href="http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t">http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t</a></td>
</tr>
<tr>
<td></td>
<td>5) RRID-like (ADID) via a national Hub</td>
<td>4) RRID-like (ADID) via its own means</td>
</tr>
<tr>
<td></td>
<td><a href="http://artist.icopad.ch/aidid:4000336z">http://artist.icopad.ch/aidid:4000336z</a></td>
<td><a href="http://artist.sikart.ch/aidid:lex4000336z">http://artist.sikart.ch/aidid:lex4000336z</a></td>
</tr>
<tr>
<td></td>
<td><a href="http://medienarchiv.icopad.ch/aidid:bfic02001t">http://medienarchiv.icopad.ch/aidid:bfic02001t</a></td>
<td><a href="http://medienarchiv.zhdk.ch/aidid:bfic02001t">http://medienarchiv.zhdk.ch/aidid:bfic02001t</a></td>
</tr>
</tbody>
</table>
CONCLUSION IV

Research Data Management without PIDs is possible but senseless!

In Switzerland:
Need for at least one more free PIDs and ideally a hub.
CONCLUSION IV

• PID situation still similar to the wild west
• Switzerland is definitely not the wild west
  • very specific constellation
  • lots of work to be done
  • need for coordination
NEVER FORGET!:
PIDS ARE BASED ON A SOCIAL CONTRACT

“Persistence is not dependent on the identifier itself, but on legal, organisational and technical infrastructure”.

HAKALA, Juha, 2005. *Persistent identifiers: the 7 levels of identification.*