

# TOWARDS TRUSTED IDENTITIES FOR SWISS RESEARCHERS AND THEIR DATA

DOI: 10.5281/zenodo.2415996



*Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.*

RENÉ 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!

DOI: [10.5281/zenodo.2415996](https://doi.org/10.5281/zenodo.2415996)

# What?

## Why?

## How!



Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

DOI: 10.5281/zenodo.2415996

# DEFINITION PERSISTENT IDENTIFIER

A persistent identifier

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

\*biunique: unique in both directions (bijective)

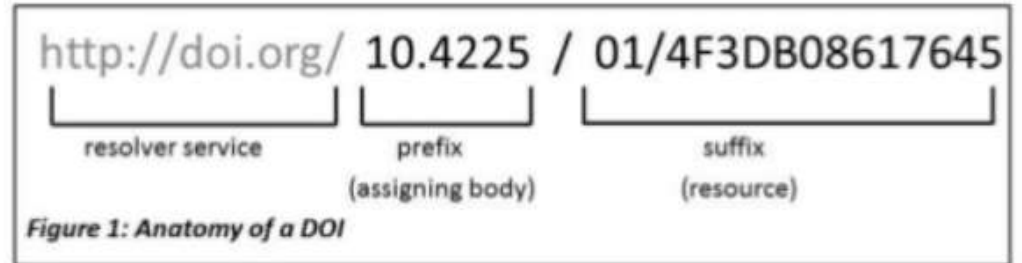
## DEFINITION II

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



# COMPOSITION

## How do DOIs work?



<https://www.slideshare.net/AustralianNationalDataService/fsci-persistent-identifiers>

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)

Eugene Barsky (<https://www.slideshare.net/ORCIDSlides/does-and-other-persistent-identifiers-in-research-data-eugene-barsky>)

## PID ≠ PID

- Publications
- Data
- Persons
- Organisations
- Citations



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

# RISKS

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

Zombie PIDs -> PIDs need Data Curation!

# What?

# Why?

# How!



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

# Why?

- Why PIDs?
- Why this project?



[www.digitalbevaring.dk](http://www.digitalbevaring.dk)

Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

# Why?

- Why PIDs?
- Why this project?

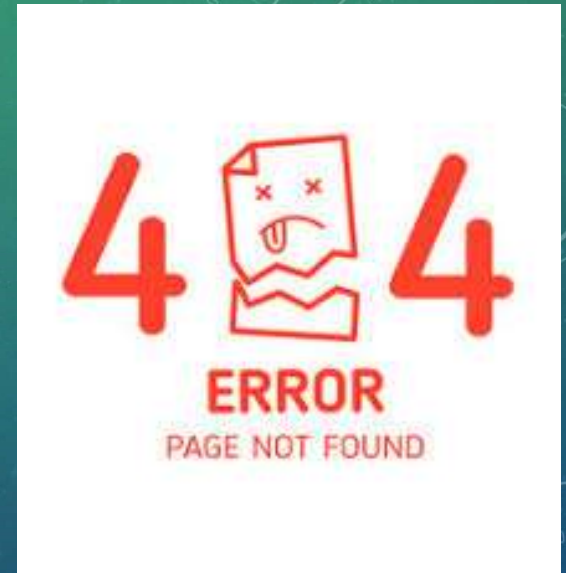


www.digitalbevaring.dk

Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

IN ORDER TO

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



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

# EXAMPLE

The image shows a screenshot of a journal article page from Hindawi Publishing Corporation. The page features a header with the Hindawi logo and the journal title 'Dataset Papers in Science'. Below the header is a navigation bar with links for 'About this Journal', 'Submit a Manuscript', and 'Table of Contents'. The main content area is divided into two columns. The left column contains a 'Journal Menu' with a list of links. The right column displays the article title 'Mapping the Slums of Dhaka from 2006 to 2010', the authors' names, their affiliations, and the publication date.

**Hindawi Publishing Corporation**

Hindawi

## Dataset Papers in Science

[About this Journal](#) | [Submit a Manuscript](#) | [Table of Contents](#)

### Journal Menu

- [About this Journal](#)
- [Abstracting and Indexing](#)
- [Advance Access](#)
- [Aims and Scope](#)
- [Article Processing Charges](#)
- [Articles in Press](#)
- [Author Guidelines](#)
- [Bibliographic Information](#)
- [Citations to this Journal](#)
- [Contact Information](#)
- [Editorial Board](#)
- [Editorial Workflow](#)
- [Free eTOC Alerts](#)
- [Publication Ethics](#)
- [Submit a Manuscript](#)
- [Subscription Information](#)

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,<sup>1</sup> Jonathan Sachs,<sup>2</sup> Anika Nockert,<sup>3</sup> Michael Frings,<sup>3</sup> Md. Mobarak Hossain Khan,<sup>4</sup> Tobia Lakes,<sup>3</sup> and Patrick Hostert<sup>3</sup>

<sup>1</sup>Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th Street, Room 517, New York, NY 10032, USA  
<sup>2</sup>Department of Tropical Medicine, School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA 70112, USA  
<sup>3</sup>Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany  
<sup>4</sup>Department 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

# ACCESS

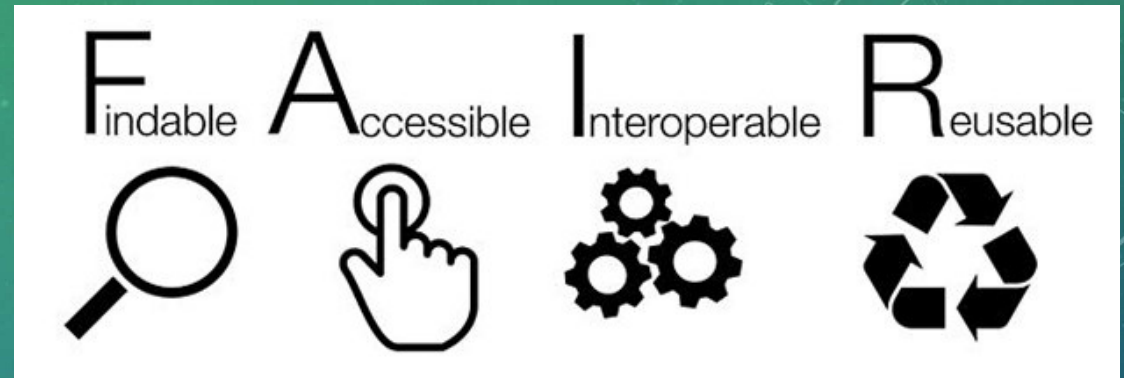
## Dataset Availability

---

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>. 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>).

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> . 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>).

# FAIRNESS

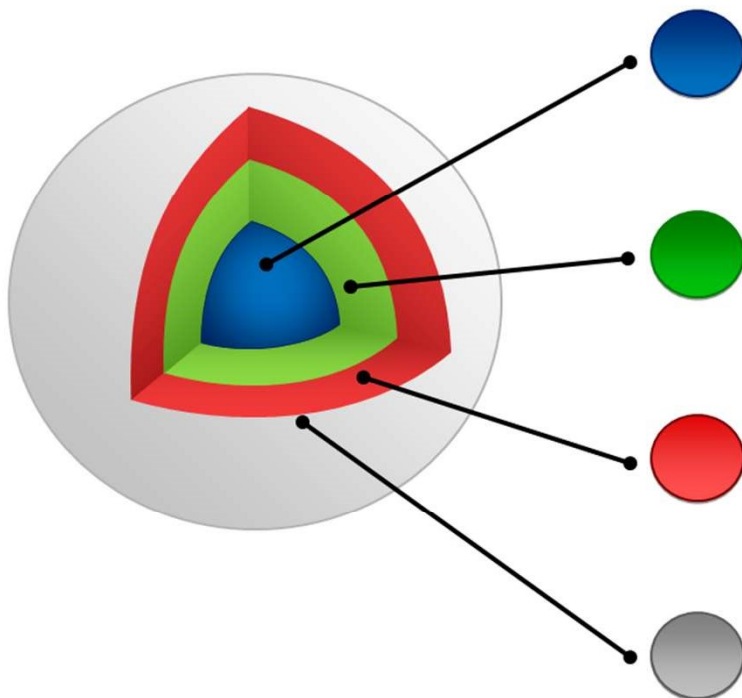


<http://www.dit.ie/dsrh/data/fairdata/>

PIDs are essential and indispensable to create fair data.

F1 Principle:

(meta)data are assigned a globally unique and eternally persistent identifier



### 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 (ORCIDs), 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 use to process and analyse the data.*

### METADATA

#### Contextual documentation

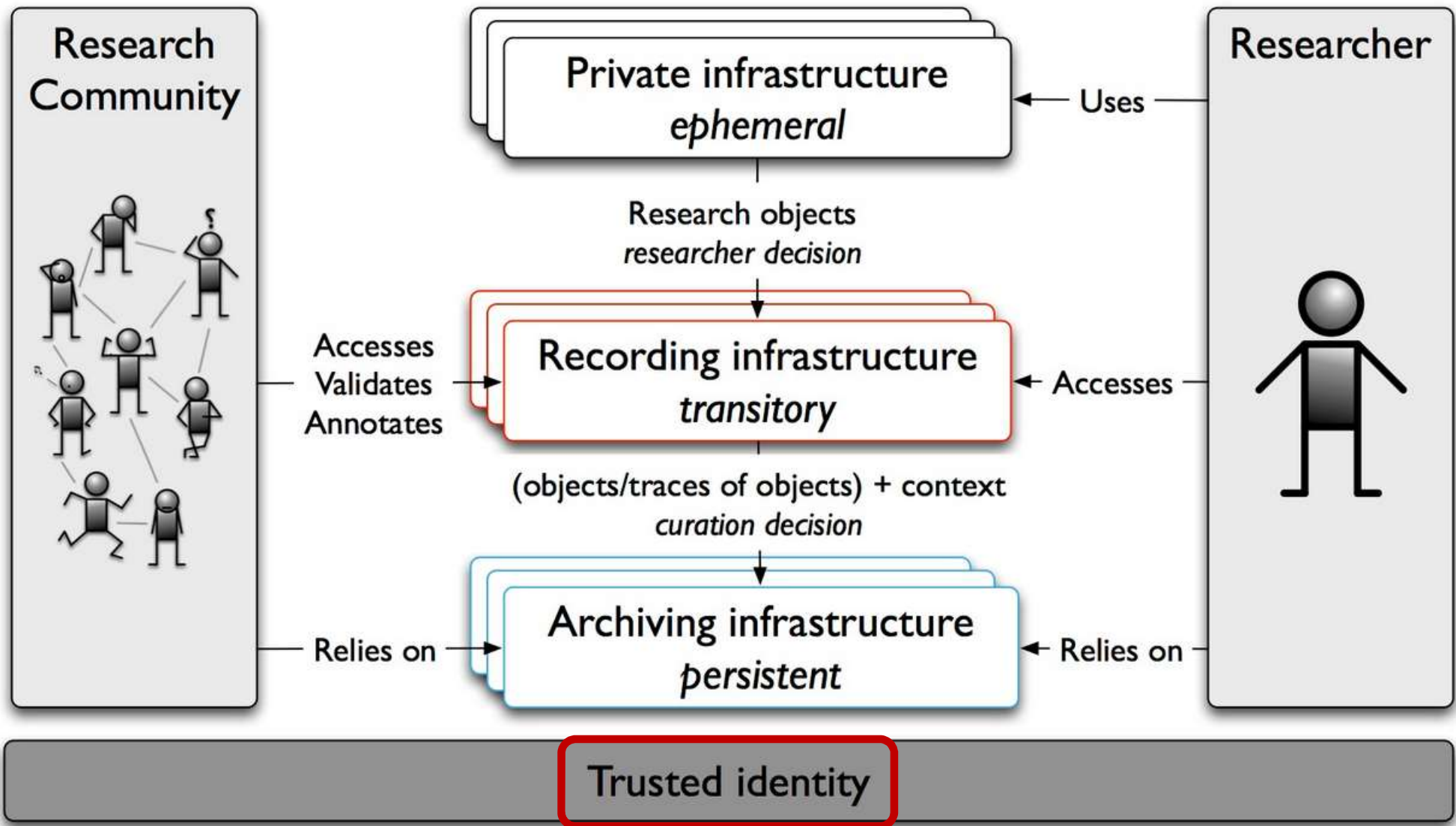
*In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest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

### 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.

## Turning FAIR Data into Reality - Report and Action Plan

<https://zenodo.org/record/1285272#.W6NSLKL4aEg>

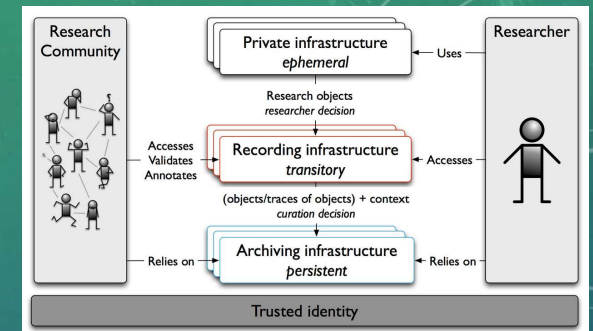


# CONCLUSION I

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?

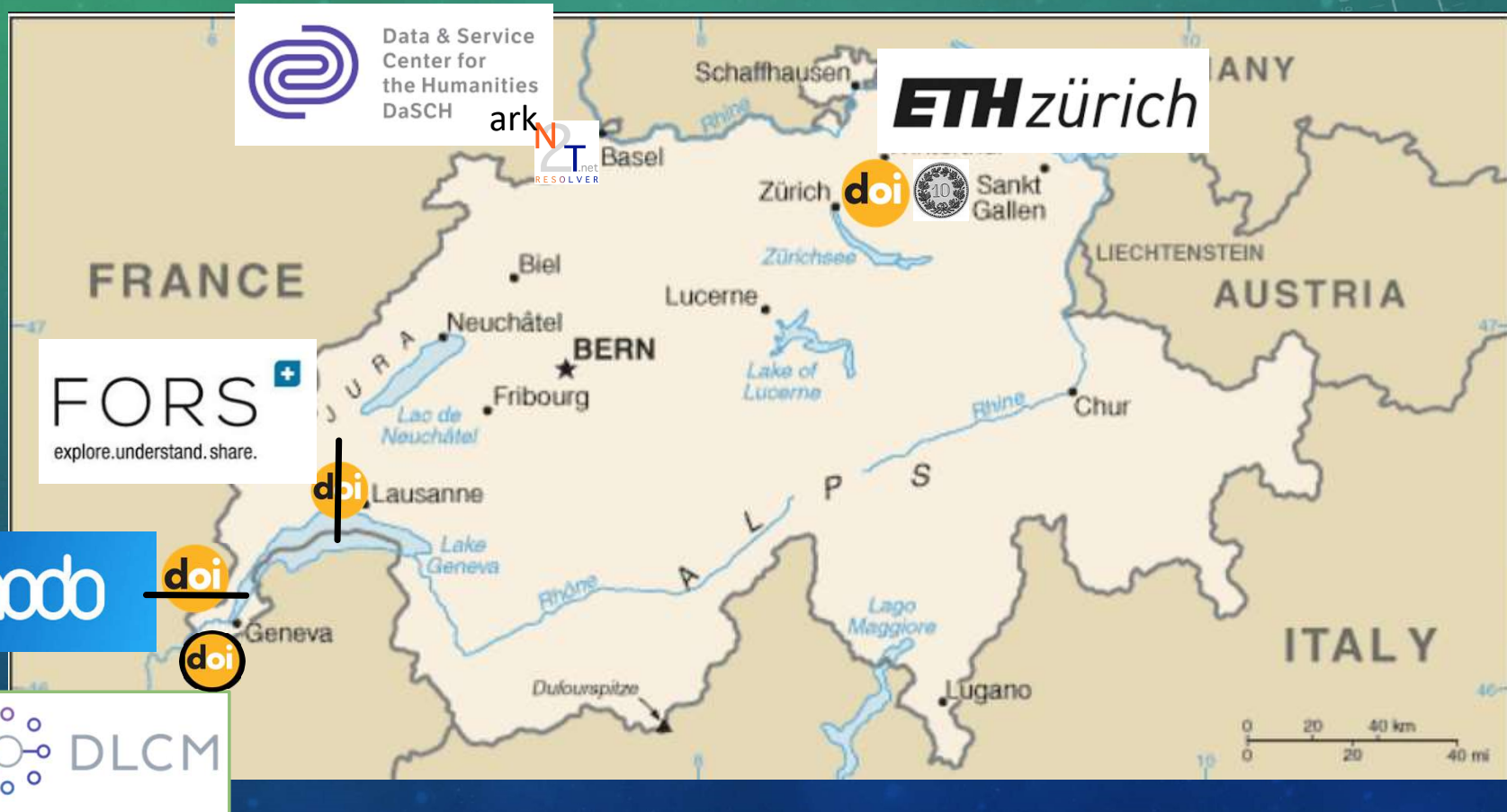
- Why PIDs?
- Why this project?



www.digitalbevaring.dk

Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

# SWISS PID LANDSCAPE



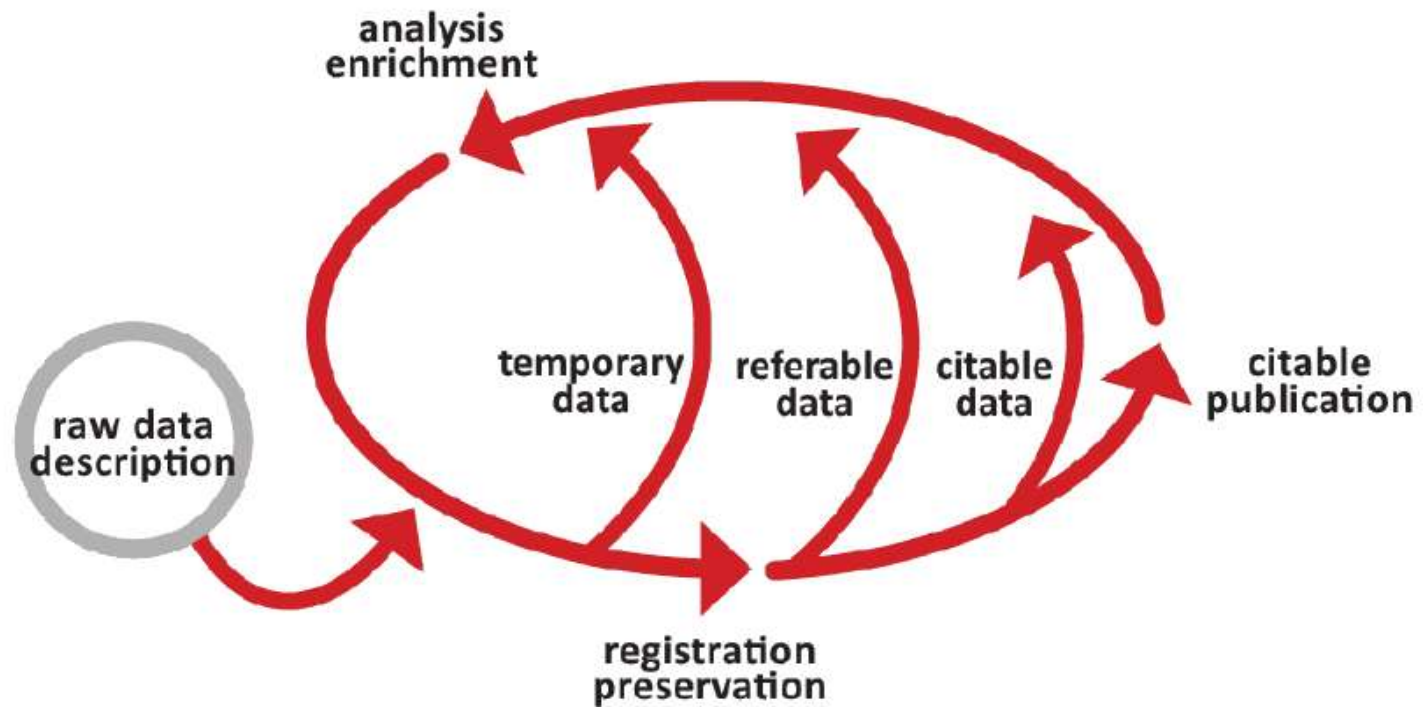
USUALLY PEOPLE TEND TO SAY

«We have DOIs, that's far enough!»

1. Costing Model. (In Switzerland!)
2. Mainly for articles and some data...
3. with low granularity.

A large, stylized logo for Digital Object Identifiers (DOIs). The letters 'd', 'o', and 'i' are rendered in a bold, sans-serif font. The 'd' is black, the 'o' is white, and the 'i' is white. They are set against a large, solid yellow circular background. The entire logo is positioned on the right side of the slide, overlapping the green and blue background.

# 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 insufficient and unsatisfactory in any case!

It takes more!



What?

Why?

How!



Cartoon courtesy of Jørgen Stamp,  
[Digitalbevaring.dk](http://Digitalbevaring.dk). CC BY 2.5.

# ICOPAD PROJECT

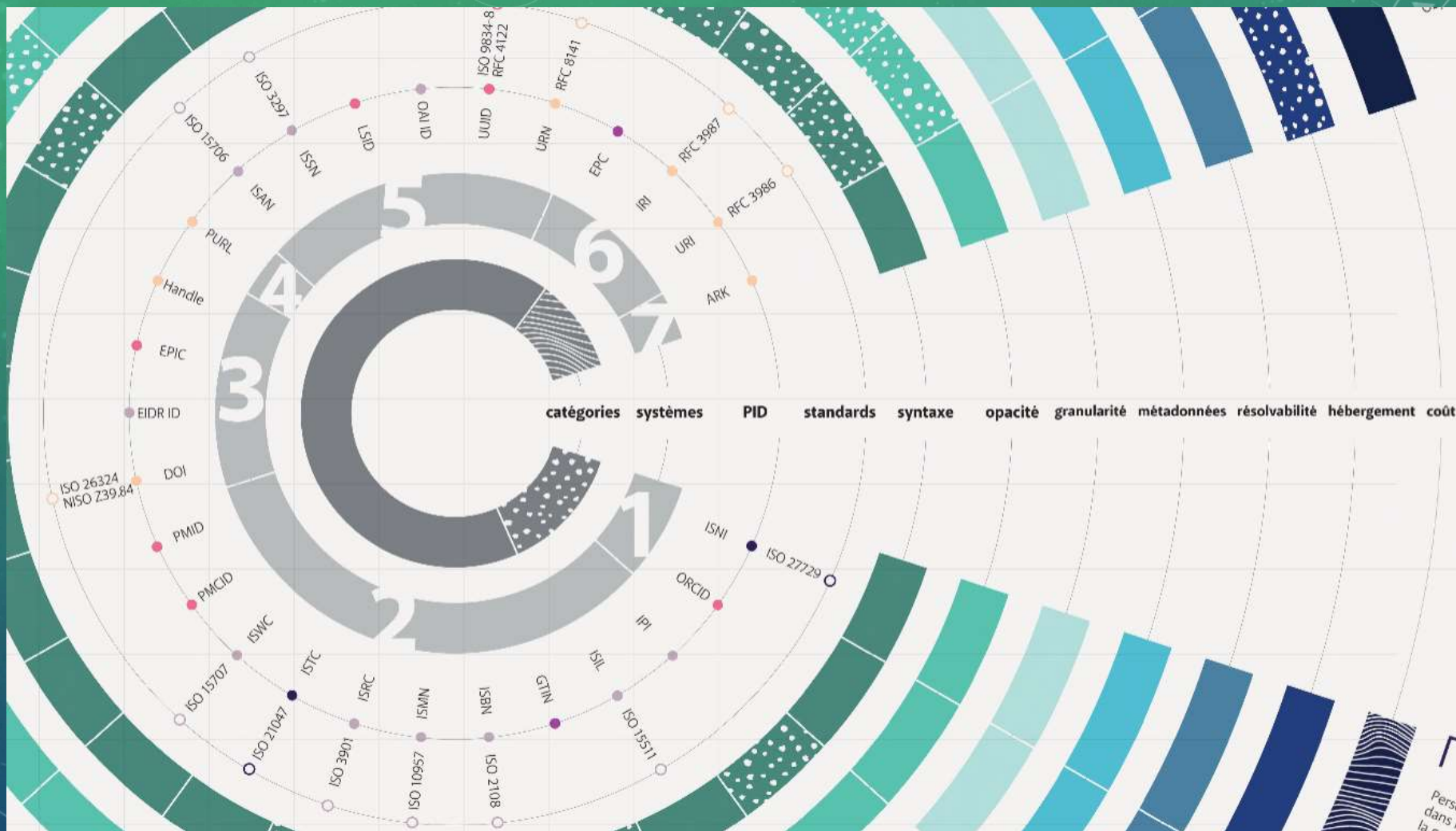
Identités de confiance pour les données de l'art et du design

- 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

Institution	Data set types/entities	Needs
SIK-ISEA	Artists Artworks Dictionary entries	Diverse PIDs and links to normed data.
ZB	Digital surrogates	Fine level of granularity.
ZHdK	Artists Artworks Events Films Glossary entries Projects Research Data	Further development of applications such as eMuseum and Medienarchiv.



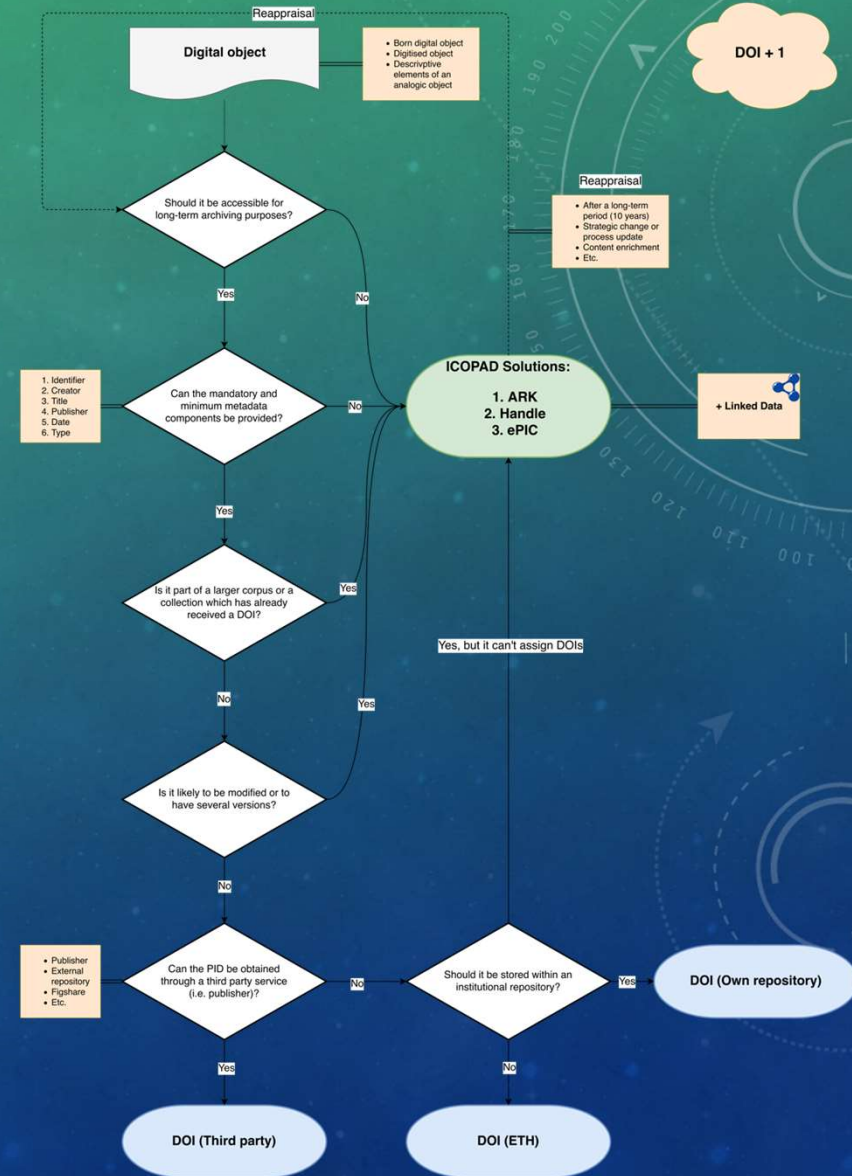


[Espasandin, Kate / Jacquet, Kate /Lefort, Lise](#)

# PID DECISION TREE (FLOWCHART)

- Inspired by prior work done @ ANDS

- Adopted
- Adapted
- Developed



## CONCLUSIONS III

DOI +  $C(doi) = x$

DOI + 1  $C(doi) = a$

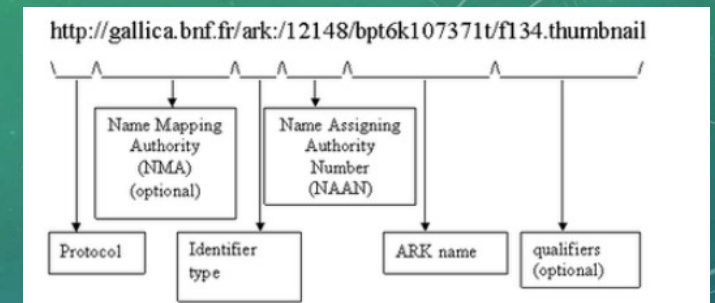
DOI + n  $C(doi) = (x_1, x_2, \dots, x_n)$

DOI + 1 + LD  $C(doi) = a \rightarrow owl:sameAs (x_1, x_2, \dots, x_n)$

$a = ark$

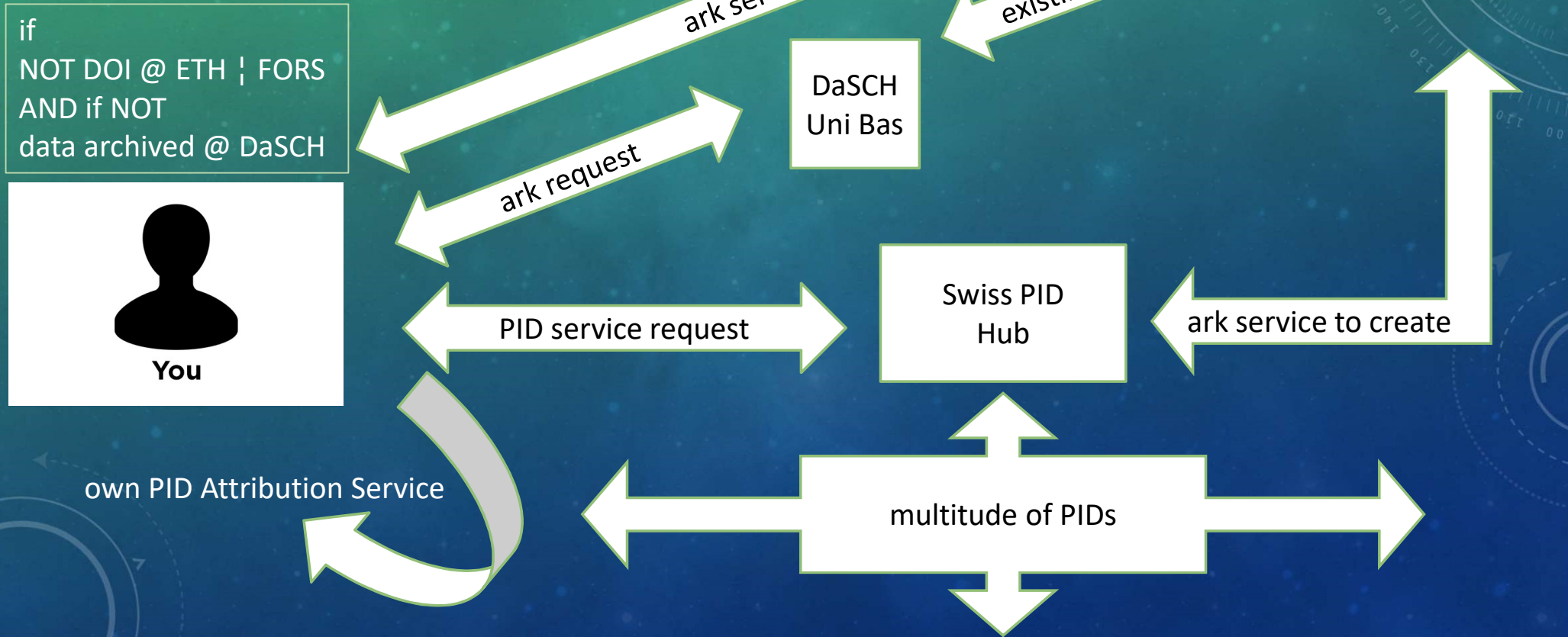
*(Archival Ressource Key provided by California Digital Library)*

# ARK, N2T



- 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



# HOSTNAME/PID AUTHORITY MATRIX

		PID Authority =Name Assigning Authority Number (NAAN)	
		One PID Authority for all organisations	Each institution is its own PID Authority
Hostname - Identifier Service Provider =Name Mapping Authority Hostport (NMAH)	One hostname for all types of data	<b>2) ARK via DaSCH</b> <a href="http://ark.dasch.swiss/ark:/72163/lex4000336z">http://ark.dasch.swiss/ark:/72163/lex4000336z</a> <a href="http://ark.dasch.swiss/ark:/72163/ema45784v">http://ark.dasch.swiss/ark:/72163/ema45784v</a> <a href="http://ark.dasch.swiss/ark:/72163/bfic02001t">http://ark.dasch.swiss/ark:/72163/bfic02001t</a>	<b>1) ARK via its own means</b> <a href="http://n2t.sikart.ch/ark:/11111/lex4000336z">http://n2t.sikart.ch/ark:/11111/lex4000336z</a> <a href="http://n2t.zb.ch/ark:/22222/ema45784v">http://n2t.zb.ch/ark:/22222/ema45784v</a> <a href="http://n2t.zhdk.ch/ark:/33333/bfic02001t">http://n2t.zhdk.ch/ark:/33333/bfic02001t</a>  <b>4) RRID-like (ADID) via its own means</b> <a href="http://n2t.sikart.ch/adid:lex4000336z">http://n2t.sikart.ch/adid:lex4000336z</a> <a href="http://n2t.zb.ch/adid:ema45784v">http://n2t.zb.ch/adid:ema45784v</a> <a href="http://n2t.zhdk.ch/adid:bfic02001t">http://n2t.zhdk.ch/adid:bfic02001t</a>
	One hostname per institution	<b>3) ARK via a national Hub</b> <a href="http://sikart.icopad.ch/ark:/99999/lex4000336z">http://sikart.icopad.ch/ark:/99999/lex4000336z</a> <a href="http://zb.icopad.ch/ark:/99999/ema45784v">http://zb.icopad.ch/ark:/99999/ema45784v</a> <a href="http://zhdk.icopad.ch/ark:/99999/bfic02001t">http://zhdk.icopad.ch/ark:/99999/bfic02001t</a>  <b>5) RRID-like (ADID) via a national Hub</b> <a href="http://sikart.icopad.ch/adid:lex4000336z">http://sikart.icopad.ch/adid:lex4000336z</a> <a href="http://zb.icopad.ch/adid:ema45784v">http://zb.icopad.ch/adid:ema45784v</a> <a href="http://zhdk.icopad.ch/adid:bfic02001t">http://zhdk.icopad.ch/adid:bfic02001t</a>	N/A
	One hostname per domain/activity	<b>3) ARK via a national Hub</b> <a href="http://artist.icopad.ch/ark:/99999/lex4000336z">http://artist.icopad.ch/ark:/99999/lex4000336z</a> <a href="http://manuscripts.icopad.ch/ark:/99999/ema45784v">http://manuscripts.icopad.ch/ark:/99999/ema45784v</a> <a href="http://medienarchiv.icopad.ch/ark:/99999/bfic02001t">http://medienarchiv.icopad.ch/ark:/99999/bfic02001t</a>  <b>5) RRID-like (ADID) via a national Hub</b> <a href="http://artist.icopad.ch/adid:4000336z">http://artist.icopad.ch/adid:4000336z</a> <a href="http://manuscripts.icopad.ch/adid:ema45784v">http://manuscripts.icopad.ch/adid:ema45784v</a> <a href="http://medienarchiv.icopad.ch/adid:bfic02001t">http://medienarchiv.icopad.ch/adid:bfic02001t</a>	<b>1) ARK via its own means</b> <a href="http://artist.sikart.ch/ark:/11111/lex4000336z">http://artist.sikart.ch/ark:/11111/lex4000336z</a> <a href="http://manuscripts.zb.ch/ark:/22222/ema45784v">http://manuscripts.zb.ch/ark:/22222/ema45784v</a> <a href="http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t">http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t</a>  <b>4) RRID-like (ADID) via its own means</b> <a href="http://artist.sikart.ch/adid:lex4000336z">http://artist.sikart.ch/adid:lex4000336z</a> <a href="http://manuscripts.zb.ch/adid:ema45784v">http://manuscripts.zb.ch/adid:ema45784v</a> <a href="http://medienarchiv.zhdk.ch/adid:bfic02001t">http://medienarchiv.zhdk.ch/adid:bfic02001t</a>

		PID Authority =Name Assigning Authority Number (NAAN)	
		One PID Authority for all organisations	Each institution is its own PID Authority
Hostname - Identifier Service Provider =Name Mapping Authority Hostport (NMAH)	<b>One hostname for all types of data</b>	<b>2) ARK via DaSCH</b> <a href="http://ark.dasch.swiss/ark:/72163/lex4000336z">http://ark.dasch.swiss/ark:/72163/lex4000336z</a> <a href="http://ark.dasch.swiss/ark:/72163/ema45784v">http://ark.dasch.swiss/ark:/72163/ema45784v</a> <a href="http://ark.dasch.swiss/ark:/72163/bfic02001t">http://ark.dasch.swiss/ark:/72163/bfic02001t</a>	<b>1) ARK via its own means</b> <a href="http://n2t.sikart.ch/ark:/11111/lex4000336z">http://n2t.sikart.ch/ark:/11111/lex4000336z</a> <a href="http://n2t.zb.ch/ark:/22222/ema45784v">http://n2t.zb.ch/ark:/22222/ema45784v</a> <a href="http://n2t.zhdk.ch/ark:/33333/bfic02001t">http://n2t.zhdk.ch/ark:/33333/bfic02001t</a>  <b>4) RRID-like (ADID) via its own means</b> <a href="http://n2t.sikart.ch/adid:lex4000336z">http://n2t.sikart.ch/adid:lex4000336z</a> <a href="http://n2t.zb.ch/adid:ema45784v">http://n2t.zb.ch/adid:ema45784v</a> <a href="http://n2t.zhdk.ch/adid:bfic02001t">http://n2t.zhdk.ch/adid:bfic02001t</a>
	<b>One hostname per institution</b>	<b>3) ARK via a national Hub</b> <a href="http://sikart.icopad.ch/ark:/99999/lex4000336z">http://sikart.icopad.ch/ark:/99999/lex4000336z</a> <a href="http://zb.icopad.ch/ark:/99999/ema45784v">http://zb.icopad.ch/ark:/99999/ema45784v</a> <a href="http://zhdk.icopad.ch/ark:/99999/bfic02001t">http://zhdk.icopad.ch/ark:/99999/bfic02001t</a>  <b>5) RRID-like (ADID) via a national Hub</b> <a href="http://sikart.icopad.ch/adid:lex4000336z">http://sikart.icopad.ch/adid:lex4000336z</a> <a href="http://zb.icopad.ch/adid:ema45784v">http://zb.icopad.ch/adid:ema45784v</a> <a href="http://zhdk.icopad.ch/adid:bfic02001t">http://zhdk.icopad.ch/adid:bfic02001t</a>	N/A
	<b>One hostname per domain/activity</b>	<b>3) ARK via a national Hub</b> <a href="http://artist.icopad.ch/ark:/99999/lex4000336z">http://artist.icopad.ch/ark:/99999/lex4000336z</a> <a href="http://manuscripts.icopad.ch/ark:/99999/ema45784v">http://manuscripts.icopad.ch/ark:/99999/ema45784v</a> <a href="http://medienarchiv.icopad.ch/ark:/99999/bfic02001t">http://medienarchiv.icopad.ch/ark:/99999/bfic02001t</a>  <b>5) RRID-like (ADID) via a national Hub</b> <a href="http://artist.icopad.ch/adid:4000336z">http://artist.icopad.ch/adid:4000336z</a> <a href="http://manuscripts.icopad.ch/adid:ema45784v">http://manuscripts.icopad.ch/adid:ema45784v</a> <a href="http://medienarchiv.icopad.ch/adid:bfic02001t">http://medienarchiv.icopad.ch/adid:bfic02001t</a>	<b>1) ARK via its own means</b> <a href="http://artist.sikart.ch/ark:/11111/lex4000336z">http://artist.sikart.ch/ark:/11111/lex4000336z</a> <a href="http://manuscripts.zb.ch/ark:/22222/ema45784v">http://manuscripts.zb.ch/ark:/22222/ema45784v</a> <a href="http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t">http://medienarchiv.zhdk.ch/ark:/33333/bfic02001t</a>  <b>4) RRID-like (ADID) via its own means</b> <a href="http://artist.sikart.ch/adid:lex4000336z">http://artist.sikart.ch/adid:lex4000336z</a> <a href="http://manuscripts.zb.ch/adid:ema45784v">http://manuscripts.zb.ch/adid:ema45784v</a> <a href="http://medienarchiv.zhdk.ch/adid:bfic02001t">http://medienarchiv.zhdk.ch/adid:bfic02001t</a>

## 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.



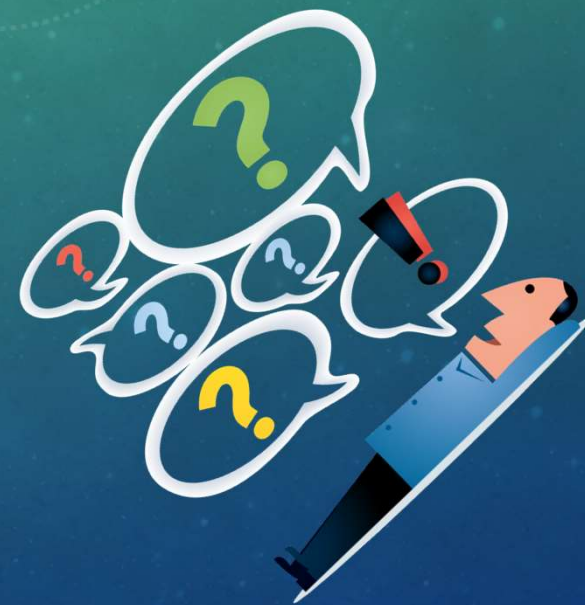
## CONCLUSION IV

- 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.\*](#)



# QUESTIONS?

DOI: [10.5281/zenodo.2415996](https://doi.org/10.5281/zenodo.2415996)

[RENE.SCHNEIDER@HESGE.CH](mailto:RENE.SCHNEIDER@HESGE.CH)