

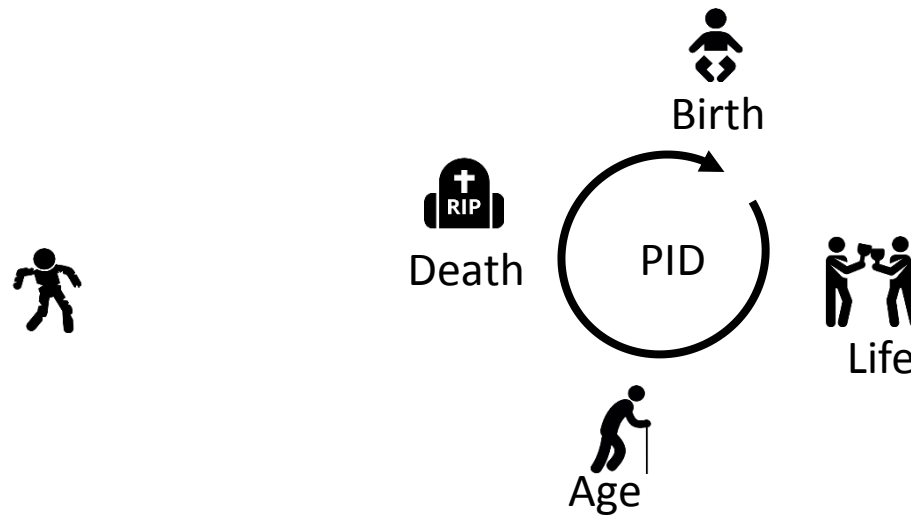
# How dead is dead in the PID Zombie Zoo?

Robert Huber<sup>1</sup> & Jens Klump<sup>2</sup>

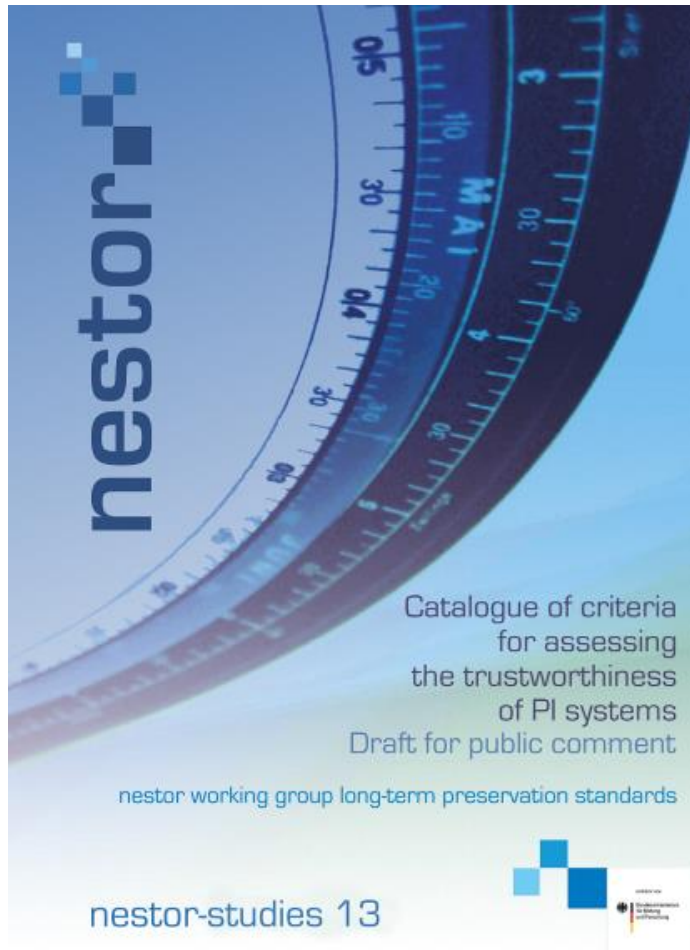
<sup>1</sup>Universität Bremen (marum, PANGAEA), Bremen, Germany

<sup>2</sup>CSIRO (Mineral Resources), Perth, Australia

Views about PID Systems  
Munich, 2. September 2016



# What makes a PID System trustworthy?

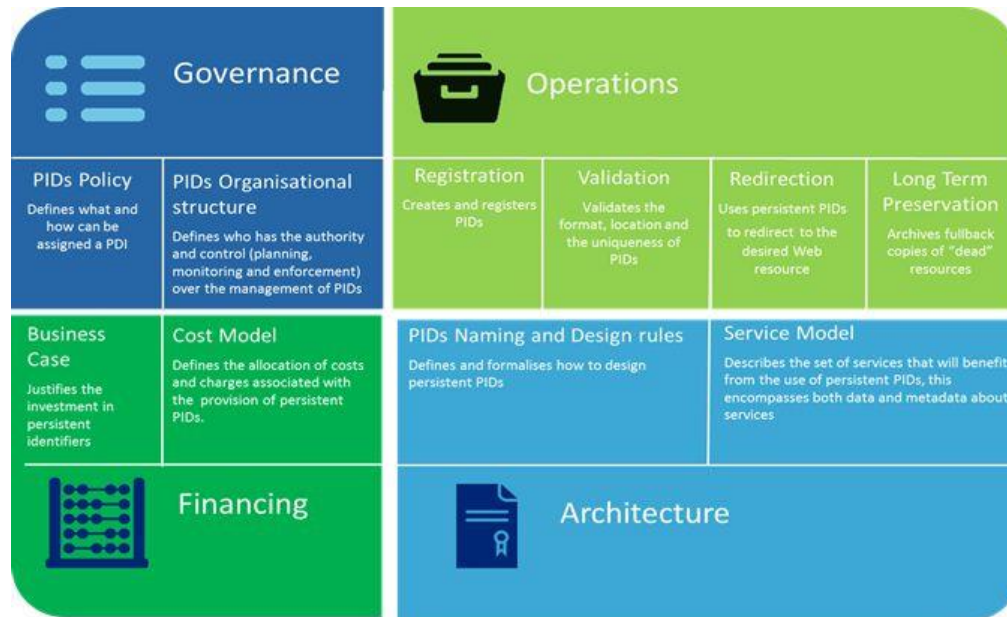


- Reliable operation (creation, update, resolution of IDs)
- Long-term stability of the PID schema
- Long-term stability of the resolving procedure
- Versioning of resources
- Organisational stability and commitment

Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, [...], and of a changing Designated Community.

**Persistent**

# Complexity of PID Systems

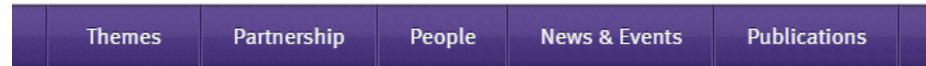


**Source:** A Reusable INSPIRE Reference Platform (ARE3NA) - Governance of Persistent Identifiers

- Complex systems are hazardous systems
- Failure consequence: orphaned, 'Zombie PIDs'
- Examples for endangered PIDs: PURLs and LSIDs

# Persistent URL - PURL

- **Initiator:** Online Computer Library Center(OCLC)
- **Available since:** 1995
- **Domain:** Multidisciplinary
- Registry in „Read-only“ mode since Nov. 2015
- **Status:** Paralysis



[Research](#) › [Themes](#) › [Data Science](#) › PURL

## PURL

**i** The purl.org service is experiencing technical difficulties and is currently operating in read-only mode. See this analysis published on the DC-ARCHITECTURE LISTSERV for details:

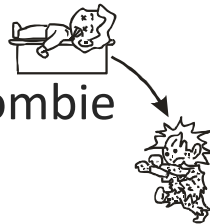
<https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=ind1511&L=DC-ARCHITECTURE&F=&S=&P=3711>

# Life Science Identifier- LSID

- **Initiator:** Object Management Group
- **Available since:** 2004
- **Promoter:** Taxonomic Database Working Group (TDWG)
- **Domain:** Life Sciences

- Resolver [lsid.tdwg.org](http://lsid.tdwg.org) shut down in 2015
- Community switches to HTTP URIs

- **Status:** Paralysis - Zombie

A screenshot of a GitHub issue titled "Decide about fate of lsid.tdwg.org #60". The issue is open and has 53 comments. The first comment by mdoering asks if the virtual host definition for http://lsid.tdwg.org can be removed. A second comment by mdoering references the issue and mentions "Redirect URLs #4". A third comment by mdoering adds the "blocks-old-server-switchoff" label. A fourth comment by sblum discusses the future of LSIDs and proposes phasing out support.

tdwg / infrastructure

<> Code Issues 33 Pull requests 0 Wiki Pulse Graphs

## Decide about fate of lsid.tdwg.org #60

Open mdoering opened this issue on 24 Nov 2015 · 53 comments

mdoering commented on 24 Nov 2015

TDWG runs a currently non functional LSID resolver and has an active virtual host definition for <http://lsid.tdwg.org>  
<https://github.com/tdwg/infrastructure/blob/master/vhosts/lsid.tdwg.org.conf>

Can we simply remove that virtual host definition?

mdoering referenced this issue on 24 Nov 2015

Redirect URLs #4

mdoering added the `blocks-old-server-switchoff` label on 24 Nov 2015

sblum commented on 24 Nov 2015

Unfortunately, I don't have the skill to resurrect the resolver. To the larger point: "what about LSIDs?" I would like to propose that TDWG's commitment to and support of LSIDs should be phased out. LSIDs are not the way forward, and we should clearly state this recommendation and signal that whatever support we are providing will be ending by some date.

# The PID Zombie (Life) Cycle

## Normal Life

The PID system is in its operational phase and strongly used within its target community

## Infection

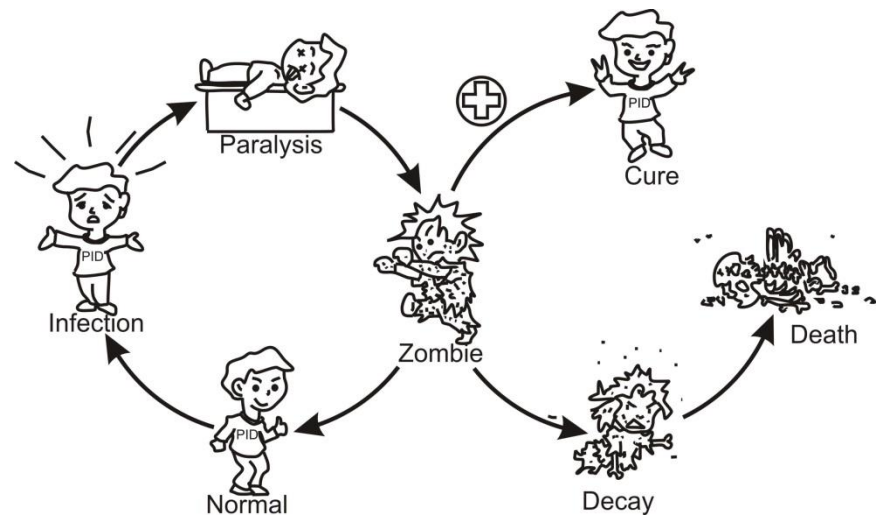
The PID system loses the interest of its target community, operating or developing agents; tools and services are temporarily unavailable.

## Paralysis

The PID system loses one or more of its critical components such as query tools, APIs or PID creation GUIs.

## Zombie stage

The PID system loses its resolution service



# Dimension of the problem

## 1) Cross media distribution



## 2) Scholarly use of PIDs

### Number of issued LSIDs & PURLs unclear:

- LSIDs: >> 5.000.000
- PURLs: >100.000

***Several millions of PID Zombie candidates!***

### Scholarly LSIDs & PURLs usage:

- Science & Nature:
  - LSID: ~40 citations
  - PURL ~25 citations
- PLOS:
  - LSID: ~1900 citations
  - PURL: ~780 citations
- The web:
  - countless

***Enough to damage overall PIDs reputation!***

# Salvation ahead?

How to deal with (PID) Zombies?

- ~~Option 1: Total extinction~~
- Option 2: Cure & vaccination





# The Semantic Web Cure

## Utilize the Semantic Web:

- URIs never die
- Describe: Resource Description Framework (RDF)
- Link: URI + Resource URL
- Resolve: Semantic Web (Browser, Search Engines)

# The Schema.org Cure

## JSON-LD & Schema.org

- URIs never die
- Describe: JSON-LD & Schema.org (Dataset)
- Link: e.g. Schema.org properties @id & URL
- Resolve: Google & Co Structured Data APIs

```
1 <script type="application/ld+json">
2 {
3   "@context": "http://schema.org/",
4   "@type": "Dataset",
5   "@id": "yaid:snet:lithologs:logs:11",
6   "name": "Outcrop: Miocene lacustrine limestone-marl alternations at the Mogente tomato field
7   outcrop, Sierra Grossa ,Valencia ,Valencia ,Ontinyent ,Spain",
8   "description": "A short section showing alternations of miocene (tortonian) gastropod-rich
9   lacustrine limestones and marls found in a small tomato field near Mogente.",
10  "datePublished": "2013-02-26 10:22:50",
11  "author": [
12    "Huber, Robert"
13  ],
14  "publisher": "Stratigraphy.net (Lithologs)",
15  "spatial": {"geo": {"@type": "GeoCoordinates", "latitude": "38.86165", "longitude": "-0.71521"}}
16  "keywords": ["litholog"]
17 }
18 </script>
```

# Risks

## Drawbacks:

- We do not yet live in the ‚Semantic Web World‘
- Outsourcing is a risk
- Services may be discontinued
- Restricted access
  - See: Google etc. structured data store
  - Querying schema.org properties not possible



# The Community Cure

- Maintain persistence of your resources & metadata
- More Transparency: Open your PID data
  - Use e.g. JSON-LD, RDF etc. to expose PIDs, their associated web links & metadata
  - As SEO & community service
  - Use and expose interfaces (RSS, OAI, sitemap)
- Register & publish your resources (e.g. GEOSS)

Help to enable cross-domain PID services

# Conclusions

- PID Zombies already exist
- Consequence: loss of confidence & reputation

But:

- URIs dont die
- Revitalisation/cure of PID Zombies is possible
- Strong community efforts required

# Thank you!

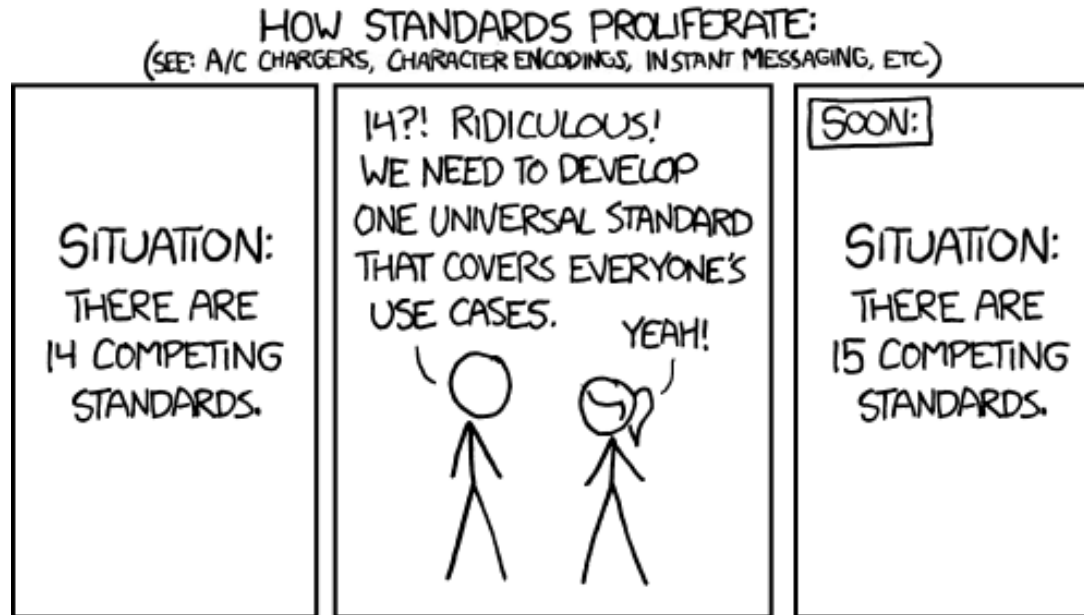
# Adoption of Persistent Identifiers

PID	count	% (all)	% (PID)
None	924	65.0	
DOI	275	19.4	55.3
Handle	102	7.2	20.5
Other	77	5.4	15.5
PURL	16	1.1	3.2
URN	16	1.1	3.2
ARK	11	0.8	2.2
LSID	9	0.6	1.8
Total	1421	100.0	100.0

Source: re3data.org, 06 December 2015

- A survey across 1421 repositories listed by re3data.org shows that DOI is by far the most widely adopted PID.
- There is a significant number of “other” IDs in use.
- PID systems advocated by the national libraries (URN) and other information providers (PURL, ARK) show no widespread adoption.
- A more detailed study is in preparation.

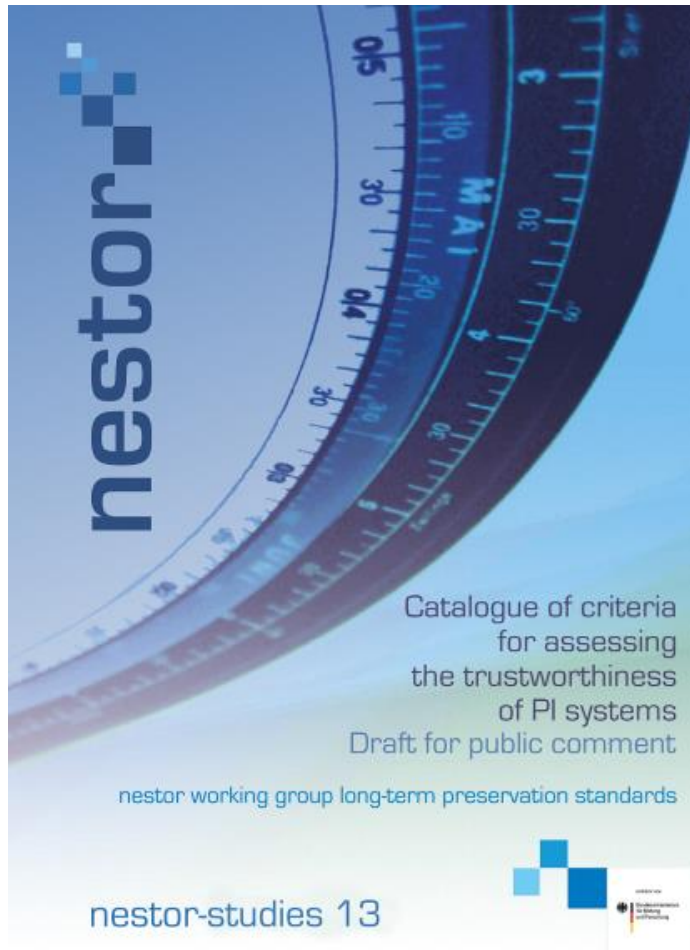
# How many systems do we need?



Archival Resource Keys (**ARKs**), Digital Object Identifiers (**DOIs**), Persistent Uniform Resource Locators (**PURLs**), Uniform Resource Names (**URNs**), and Extensible Resource Identifiers (**XRIs**), European Persistent Identifier Consortium (**EPIC**), International Geo Sample Number (**IGSN**), Life Science Identifier (**LSID**), Originator and Contributor ID (**ORCID**), Research Resource ID (**RRID**), Unique Material Identifier (**UMID**), ...



# Persistent Identifier (PID) Systems



## Definition:

A PI system is a mutually referenced combination of

- **Definitions**
- **Policies**
- **Services**
- **Data sources**

which are used for the administration and use of persistent identifiers.

## Core services:

- **Regulating the issuing of identifiers**
- **Registering**
- **Update**
- **Resolving**



(nestor studies (13))