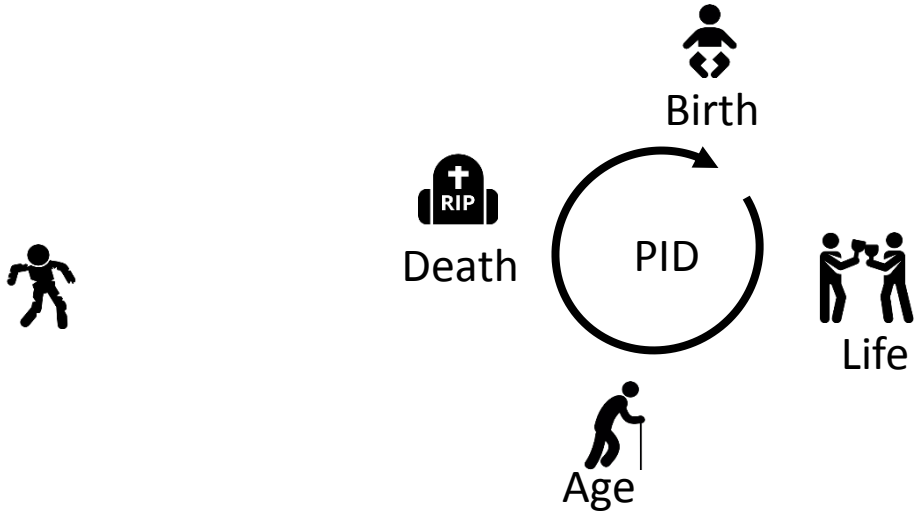
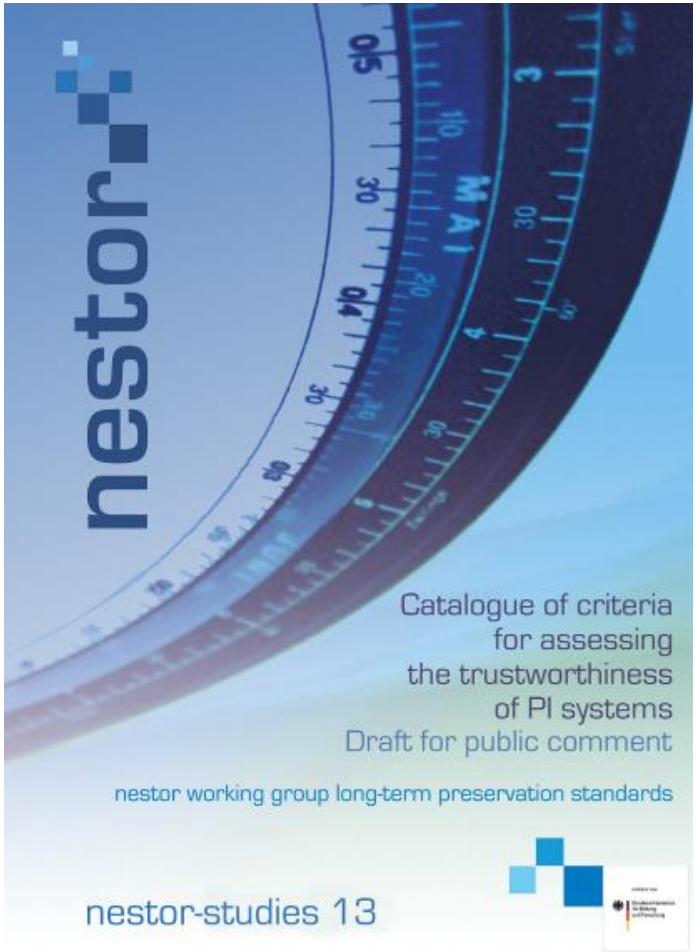
How dead is dead in the PID Zombie Zoo?

Robert Huber[1] & Jens Klump[2]
[1]Universität Bremen (marum, PANGAEA), Bremen, Germany
[2]CSIRO (Mineral Resources), Perth, Australia

Views about PID Systems
Munich, 2. September 2016



What makes a PID System trustworthy?

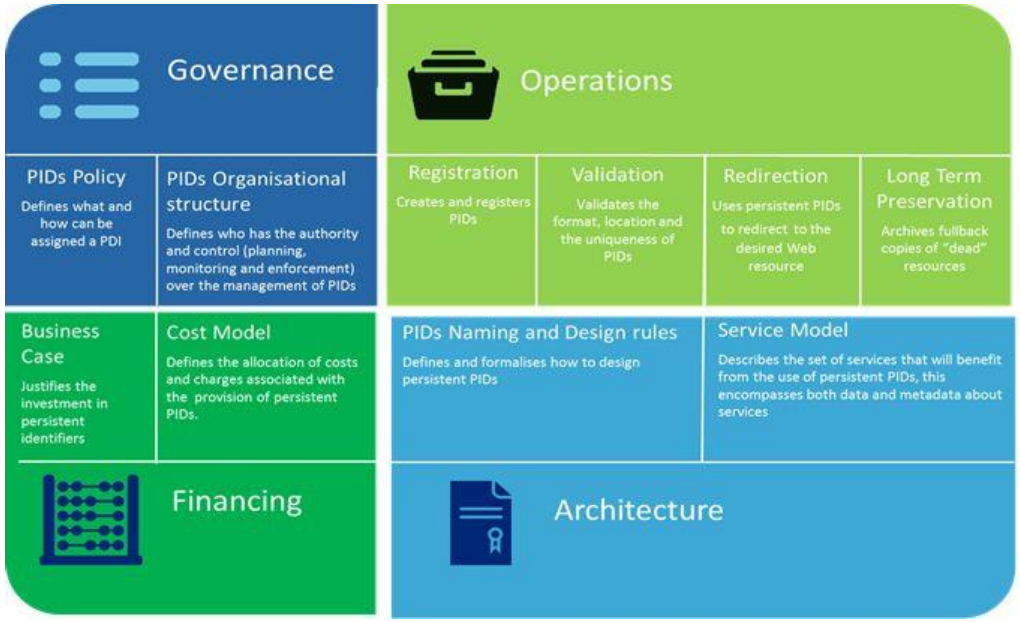


- Reliable operation (creation, update, resolution of IDs)
- Long-term stability of the PID schema
- Long-term stability of the resolving procedure
- Versioning of resources
- Organisational stability and commitment

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

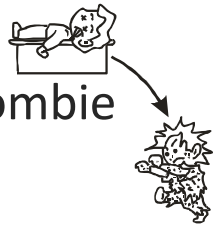
- **Status:** Paralysis

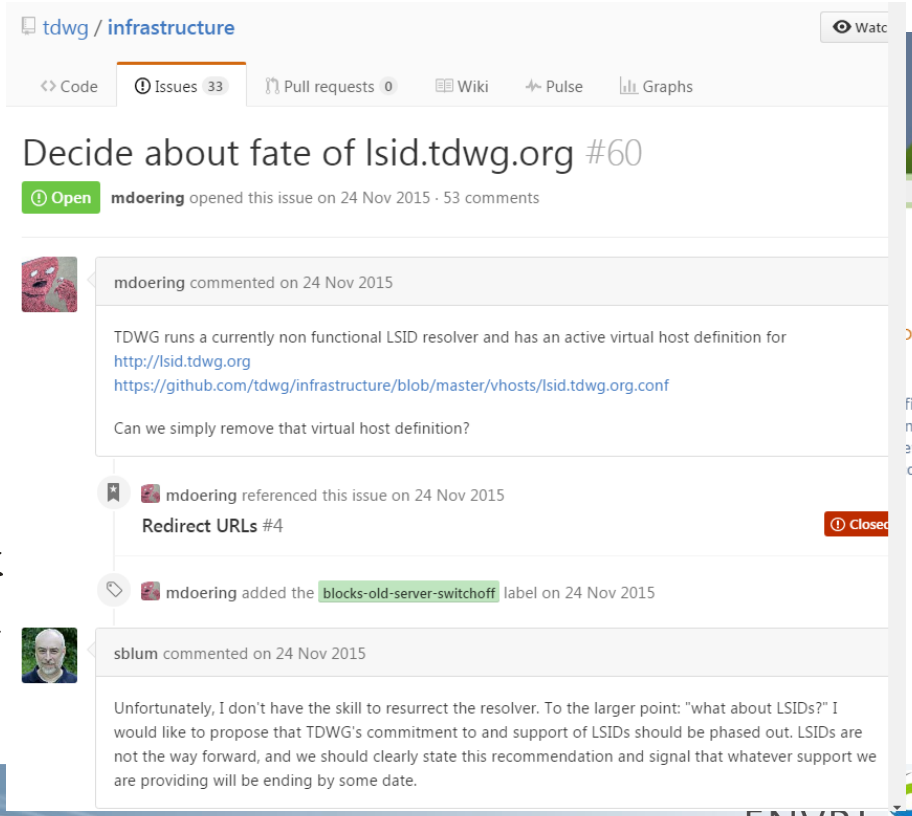
Life Science Identifier- LSID

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

- Resolver lsid.tdwg.org shut down in 2015
- Community switches to HTTP URIs

- **Status:** Paralysis - Zombie





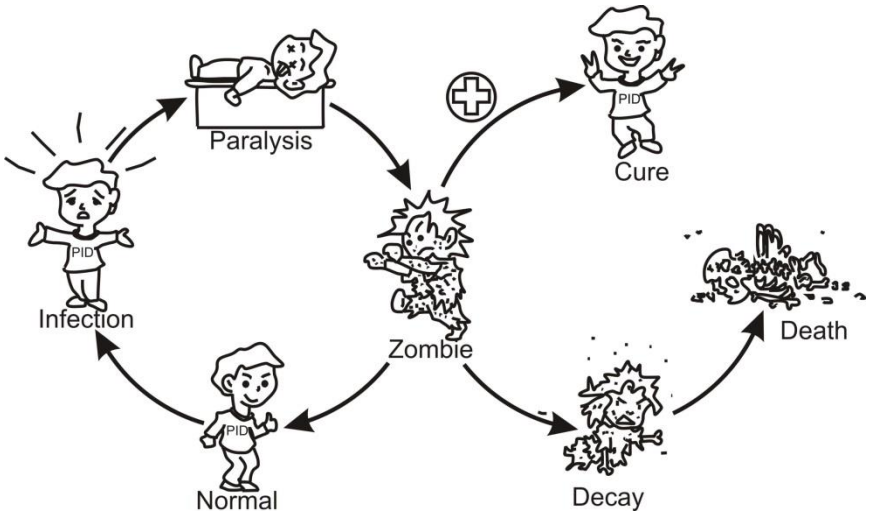
The PID Zombie (Life) Cycle

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.

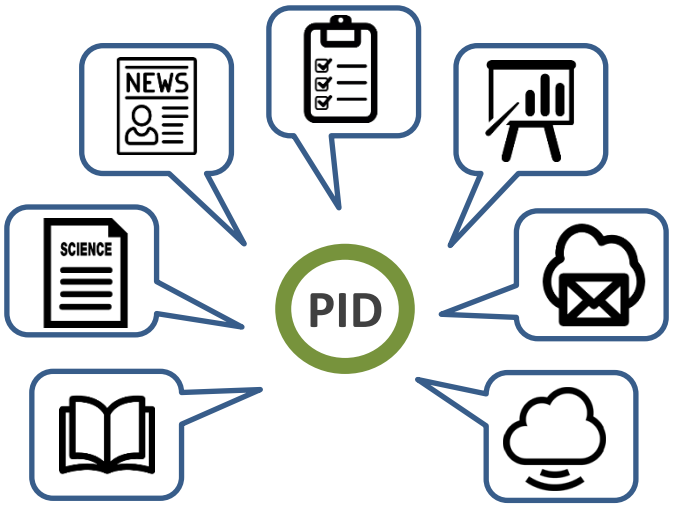
Paralysis
The PID system loses one or more of its critical components such as query tools, APIs or PID creation GUIs.

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)

Sachs & Finin (2010): What does it mean for a URI to resolve?. Proceedings of the AAAI Spring Symposium on Linked Data Meets Artificial Intelligence

The Schema.org Cure

JSON-LD & Schema.org

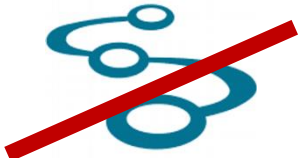
- URIs never die
- Describe: JSON-LD & Schema.org (Dataset)
- Link: e.g. Schema.org properties @id & URL
- Resolve: Google & Co Structured Data APIs

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



Sindice

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!

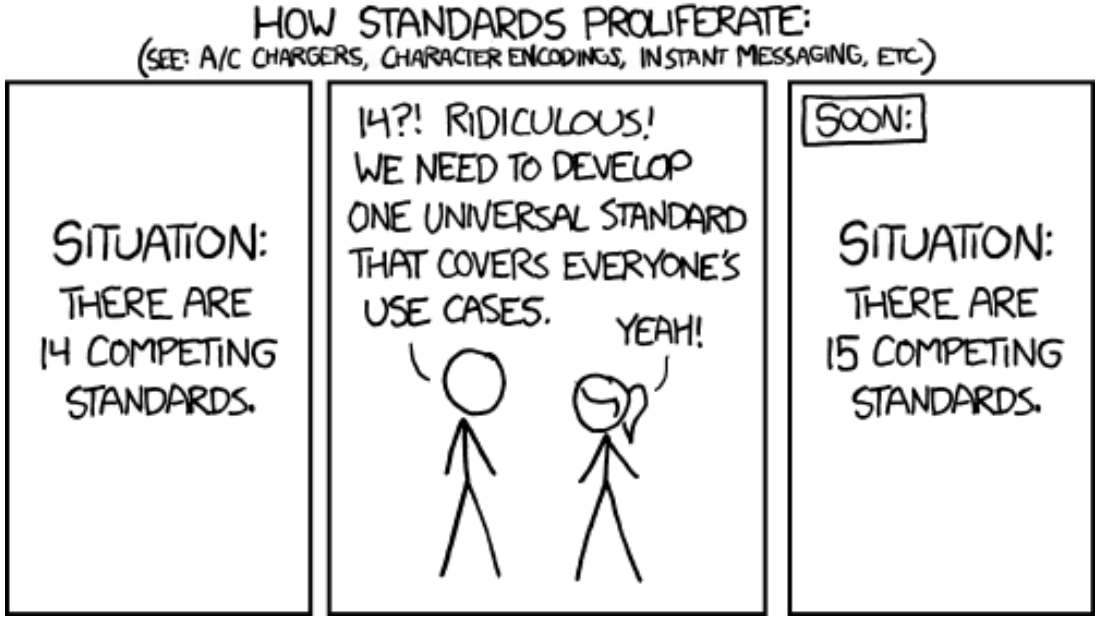
Adoption of Persistent Identifiers

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

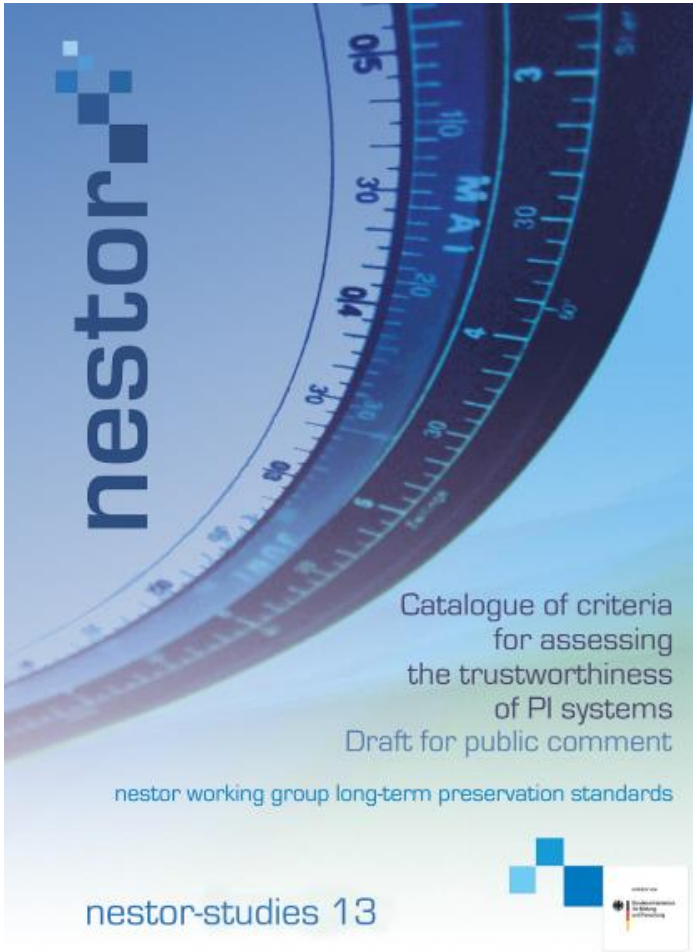
- A survey across 1421 repositories listed by re3data.org shows that DOI is by far the most widely adopted PID.
- There is a significant number of “other” IDs in use.
- PID systems advocated by the national libraries (URN) and other information providers (PURL, ARK) show no widespread adoption.
- A more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

How many systems do we need?



Archival Resource Keys (**ARK**s), Digital Object Identifiers (**DOI**s), Persistent Uniform Resource Locators (**PURL**s), Uniform Resource Names (**URN**s), and Extensible Resource Identifiers (**XRI**s), 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))