

Understanding URI Ecosystems

Henry S. Thompson

[ORCID: 0000-0001-5490-1347](https://orcid.org/0000-0001-5490-1347)

Presented at [PIDapalooza 2019](#)

23 January 2019

<https://doi.org/10.5281/zenodo.2551045>

Copyright © 2019 [Henry S. Thompson](#)



Table of Contents

1. [Acknowledgements](#)
2. [TL;DR](#)
3. [Changing the metaphor](#)
4. [Historical example 1: Binomial nomenclature](#)
5. [How does it work?](#)
6. [Players in the binomial nomenclature ecosystem](#)
7. [Historical example 2: Scholarly bibliographies](#)
8. [How does it work?](#)
9. [How does it work \(cont'd\)](#)
10. [Players in the scholarly bibliographies ecosystem](#)
11. [Belabouring the point](#)
12. [PIDs \(finally\)](#)
13. [Players in the scholarly publication PID ecosystem](#)
14. [The breakthrough compromise](#)
15. [A temporary \(?\) imbalance](#)
16. [Wrapping it up](#)

1. Acknowledgements

Jonathan Rees planted the first seeds of doubt in my mind regarding the 'official story' about URIs sometime in the previous century, and we've been exchanging thoughts about the nature of naming on the Web ever since.

Tim Berners-Lee disagrees with almost everything I have to say on the subject, but has always been willing to listen.

2. TL;DR

"No PID is an island, entire of itself"

- C.f. Donne, John, *Devotions upon Emergent Occasions*, Meditation XVII, 1624. Stationers' Company, London.

"A good [technology] sales pitch focusses on problems, not technology."

- Wallace, Donald (Smokey), 2001. Personal communication.
- Walk, Paul, *PIDs in Dublin Core*, 2019. PIDapalooza 19, Dublin. Abstract [available online](#).

Identifiers/URIs/PIDs function in a variety of contexts

- Wittengenstein's *Forms of Life* and their associated *Language Games*
- (Three examples from the Scholarly Reference game above)

Identifiers are used and hence understood *differently* in different games

You can't get credibility/break into such a game without paying your dues

- In particular, by identifying the different player roles and their requirements

3. Changing the metaphor

I like the **ecosystem** metaphor for thinking about identifiers, with its ideas of

- fitness
- differentiation
- interdependence
- feedback cycles

And I'm borrowing from sociology and social anthropology

- For the idea of **roles** that the participants in social systems (implicitly) adopt
- And the idea of **communities of practice** for the emergent social systems created by cooperation wrt shared tasks

What follows next then are two quick attempts to understand historically successful identifiers in these terms

4. Historical example 1: Binomial nomenclature

Homo sapiens, clostridium botulinum, quercus robur

Probably the most persistent identifier system to date

- Begins with Linnaeus in the 18th century
- Reaches its more-or-less mature form in the 19th

At least three more-or-less non-overlapping user communities/ecosystems

- Science/medicine;
- Agriculture (incl. animal husbandry);
- Landscape architecture, professional and amateur gardening

In each case we have both practioners and associated infrastructure providers/suppliers

5. How does it work?

For communication (1)

When a botanist makes a claim about *Quercus robur*, all other botanists know what she's talking about

- As do many other biologists and even many amateurs

That is, it functions as a proper name in some matrix language

For retrieval

Univocal/reliable and permanent (well, so far) retrieval

There is a well-defined, surprisingly robust, mechanism for establishing the species identified by any binomen, via

- First publication
- Type specimen

For communication (2)

Conveys *some* information to readers/hearers, even if novel to them as a whole

- if they know part of it already as a proper name
- and/or if they know the meaning of the (Latin, Greek, ...) root(s) as words

For cohesion

As with any other specialist vocabulary, signals in-group membership by its use.

6. Players in the binomial nomenclature ecosystem

This one is surprisingly simple: there really are only three roles:

minters

The (small number of) authors of the papers that introduce a new identifier;

users

The (large number of) scientists/gardeners/etc. who learn and use them;

gatekeepers

The (small number of) people who manage the system, arbitrate disputes, etc.

This tripartate structure characterises a wide range of more-or-less well-circumscribed specialist languages embeddable in natural language

- Drug names, chemical compound names, entries in the periodic table, astronomical object names, ...

7. Historical example 2: Scholarly bibliographies

Thompson, H. S. (1990a) "Parallel parsers for context free grammars---two actual implementations compared". In *Parallel Natural Language Processing*, G. Adriaens and U. Hahn, eds. Ablex, Norwood, N.J., 0-89391-869-5.

Pre-World-Wide-Web, (bibliographic) reference was "the connective tissue of scholarship" (Brian Cantwell Smith)

Only very slowly standardised

- And indeed, with many more-or-less interoperable standards in play

Similar to binomial nomenclature in being both embedded in natural language and stylised in form

- Albeit in much more restricted contexts
- And with much more variation in styling

8. How does it work?

For communication (1)

Readers may recognise the title, author and/or publication venue as proper names;

For retrieval

Readers can retrieve the referent via a univocal/reliable, permanent (albeit eroding fast now) search of

- at best, more-or-less local shelves;
- at worst, via a nearby large (university) library.

For communication (2)

Even if readers recognise nothing in the way of proper names

- the natural language words in the title and/or publication venue may give *some* clue as to whether they are worth following up

9. How does it work (cont'd)

For status (1)

Authors establish credibility by citing well-known relevant works;

For status (2)

Authors gain in reputation as their works are referenced;

For status (3)

Publication venues gain in reputation as their products are referenced.

And of course status may well translate more or less directly into wealth

10. Players in the scholarly bibliographies ecosystem

- Large numbers of **authors**
- Large numbers (one hopes :-) of **readers**
- Small numbers of **gatekeepers** (publishers)
- Modest numbers of **institutions** who employ the authors

11. Belabouring the point

In both cases, there is a web of interactions between the different functions

- made possible by the use of referential mechanisms (identifiers)

In other words they constitute a (metaphorical) ecosystem

- and require a community of practice to sustain them

12. PIDs (finally)

What are scholarly publication PIDs *for*?

- For the publication reader and search engine users: reliable persistent access to a digital object
- For the publisher, devolving responsibility for resolution persistence:

- "404? Not my problem, you should have used the DOI"
- Facilitating access improves self-esteem, by feeling you are advancing science, and status, by advancing your reputation

13. Players in the scholarly publication PID ecosystem

- (Large numbers of) **authors**;
- (Large numbers of) **readers**;
- (Modest numbers of) **gatekeepers** (publishers);
- (Small numbers of) **gatekeepers** (PID system owners);
- (Small numbers of) **gatekeepers** (funders)

The core object lesson: the breakthrough compromise (see next slide) *solved a big problem* for authors, readers and publishers, that is, **404**

- And hugely accelerated the success of DOIs
- Bringing the rest of the PID space along with

14. The breakthrough compromise

After ten years or so of "We want names, not addresses" vs. "Just use http:"

- Never *quite* becoming really nasty

Pragmatism trumped ideology and a compromise emerged

- Glossed as "PID systems should define (and support) mappings to actionable identifiers"
- Publicly 'ratified' at a meeting in London in 2010
- [Tip of my hat to [Paul Walk](#)]
- - <http://identifiers2010.jiscpress.org/4-candidate-statements-of-common-agreement/> -> **[404!]**
 - <https://web.archive.org/web/20110214151948/http://identifiers2010.jiscpress.org/4-candidate-statements-of-common-agreement/> -> [Works, phew!]

15. A temporary (?) imbalance

The PID community has focussed on the **For retrieval** functionality within its various (almost entirely scholarly) ecosystems.

More recently, by bringing 'compact identifiers' (CIDs) into play,

- We're getting *some* traction in the area of **For communication (1)**, e.g. `taxon:9606`
- but this has evident limits

16. Wrapping it up

Three different ecosystems

- And the last certainly doesn't exhaust ones where PIDs have made/are making inroads
- I've said nothing about PIDs for data, and almost nothing about CIDs
- to say nothing of the many, *much* larger web-hosted ecosystems, from travel ticketing to video streaming, where PIDs are not just unseen, but completely unheard of.

But I hope the *idea* of thinking about ecosystems and where identifiers fit may prove useful to you

- Particularly with respect to thinking about change
- You can't bully an ecosystem
- You have to motivate its community of practice to *want* to change