

Persistent Identifiers

Fundamental building blocks for RI infrastructure

Clifford Tatum
SURF | Utrecht
28 September 2018

2018 ORCID/Netherlands Roundtable Meeting





Collaborative organisation for ICT in Dutch higher education and research



Leiden University, the Netherlands

Outline

What is infrastructure?

Researcher-centric data model

ORCID-NL, pilot to consortium

PID approach to RI infrastructure

Infrastructure?

What is being enacted?



Video clip 
<https://vimeo.com/238528828>

When ~~What~~ is Infrastructure?

1. ***Embeddedness*** - Infrastructure is "sunk" into, inside of, other structures, social arrangements and technologies;
2. ***Transparency*** - Infrastructure is transparent to use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly supports those tasks;
3. ***Reach or scope*** - This may be either spatial or temporal—infrastructure has reach beyond a single event or one-site practice;
4. ***Learned as part of membership*** - The taken-for-grantedness. Strangers and outsiders encounter infrastructure as a target object to be learned about. New participants acquire a naturalized familiarity with its objects;
5. ***Links with conventions of practice*** - Infrastructure both shapes and is shaped by the conventions of a community of practice, e.g. generations of typists have learned the QWERTY keyboard; its limitations are inherited by the computer keyboard and thence by the design of today's computer furniture (Becker 1982);
6. ***Embodiment of standards*** - Modified by scope and often by conflicting conventions, infrastructure takes on transparency by plugging into other infrastructures and tools in a standardized fashion.
7. ***Built on an installed base*** - Infrastructure does not grow de novo: it wrestles with the "inertia of the installed base" and inherits strengths and limitations from that base.
8. ***Becomes visible upon breakdown*** - The normally invisible quality of working infrastructure becomes visible when it breaks; the server is down, the bridge washes out, there is a power blackout.

When ~~What~~ is Infrastructure?

Infrastructure is **enacted in practice** and occurs *when* the tension between local and global is resolved.

That is, an infrastructure occurs when local practices are afforded by a larger-scale technology, which can then be used in a natural, ready-to-use fashion.

It becomes transparent as local variations are folded into organizational changes, and becomes an unambiguous home—for somebody. This is not a physical location nor a permanent one, but a working *relation*

Researcher-centric data model

ACUMEN research

assessment of scientometric indicators in performance evaluation

Common Data Strategy

assessment of webometric (and altmetric) indicators

comparative analysis of peer review systems in Europe

analysis of gender dimension in researcher evaluation

ethnographic study of important evaluation events



15 European countries

Bulgaria	France	Netherlands
Czech Republic	Germany	Poland
Denmark	Hungary	Slovenia
Estonia	Israel	Spain
Finland	Italy	United Kingdom

4 Academic Disciplines

- (a) astronomy and astrophysics
- (b) public and occupational health
- (c) environmental engineering
- (d) philosophy (including history and philosophy of science)

ACUMEN Portfolio

Career Narrative

Links expertise, output, and influence together in an evidence-based argument; included content is negotiated with evaluator and tailored to the particular evaluation

Expertise

- scientific/scholarly
- technological
- communication
- organizational
- knowledge transfer
- educational

Output

- publications
- public media
- teaching
- web/social media
- data sets
- software/tools
- infrastructure
- grant proposals

Influence

- on science
- on society
- on economy
- on teaching



Evaluation Guidelines

- aimed at both researchers and evaluators
- development of evidence based arguments (what counts as evidence?)
- expanded list of research output
- establishing provenance
- taxonomy of indicators: bibliometric, webometric, altmetric
- guidance on use of indicators
- contextual considerations, such as: stage of career, discipline, and country of residence

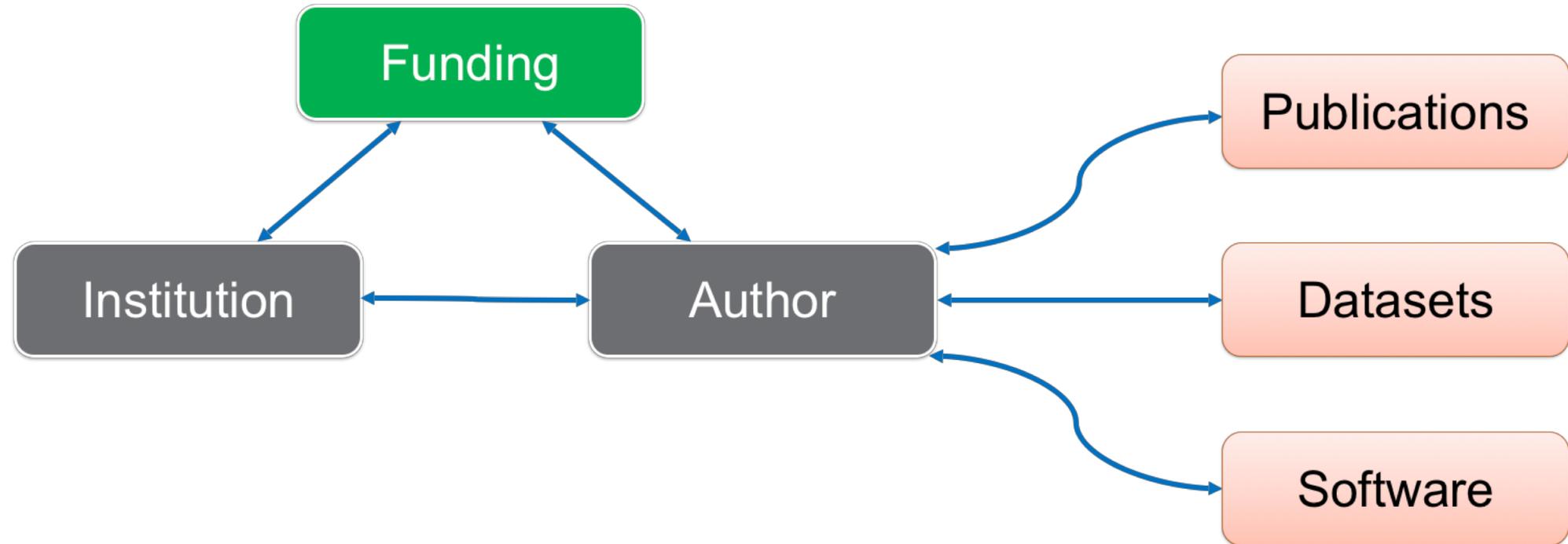
interoperability dilemma

- research *information* (RI) is in part dependent on research *infrastructure*
- local RI solutions need to first and foremost satisfy local concerns
- bottom up datasets tend to remain in institutional databases

proposal

—ignore interoperability... for now, and instead focus on portability of PIDs

Researcher-centric data model



Portable ID stack

- Portable ID Stack -

Interoperability among research metadata benefits from persistent IDs that are both *open* and *accessible* across systems

Author ID



DOI - article



- dataset



- software



- data/other



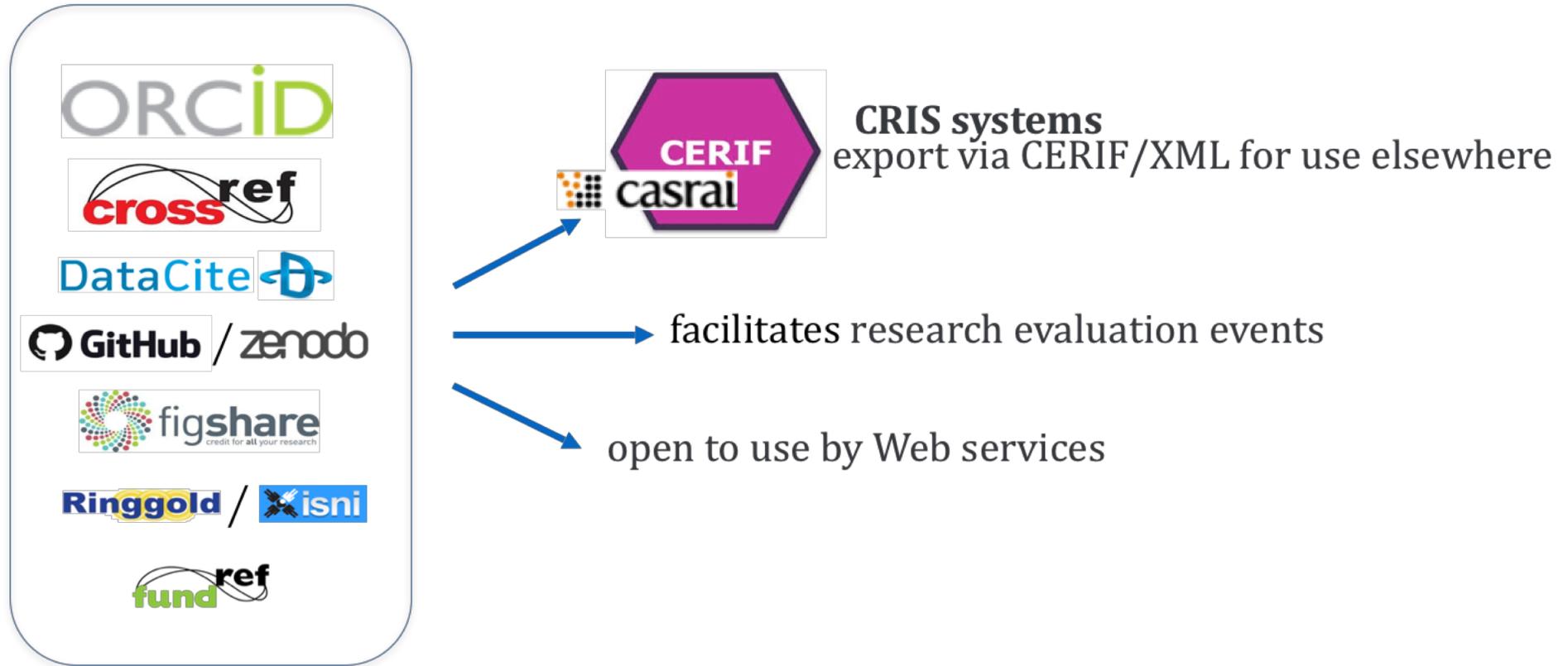
Organization



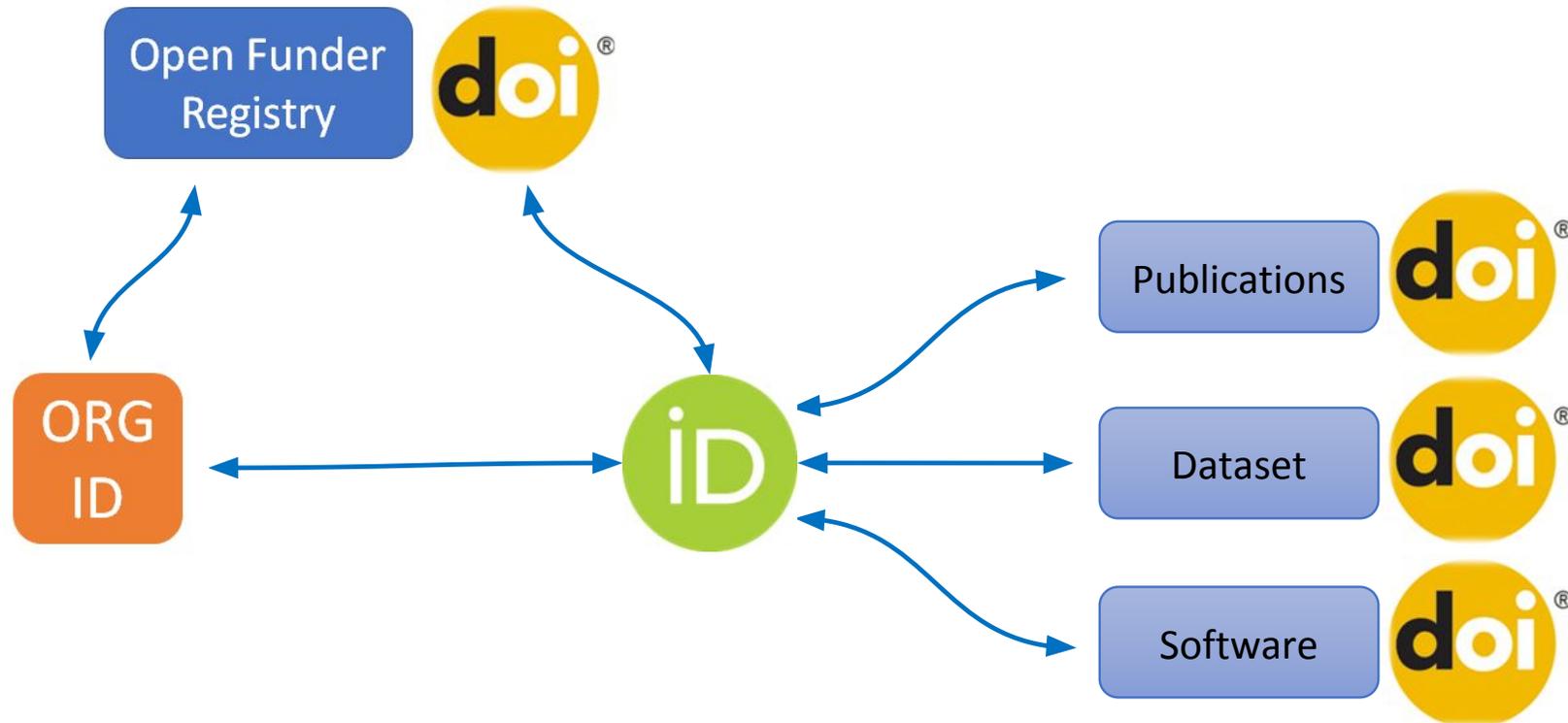
Funding source



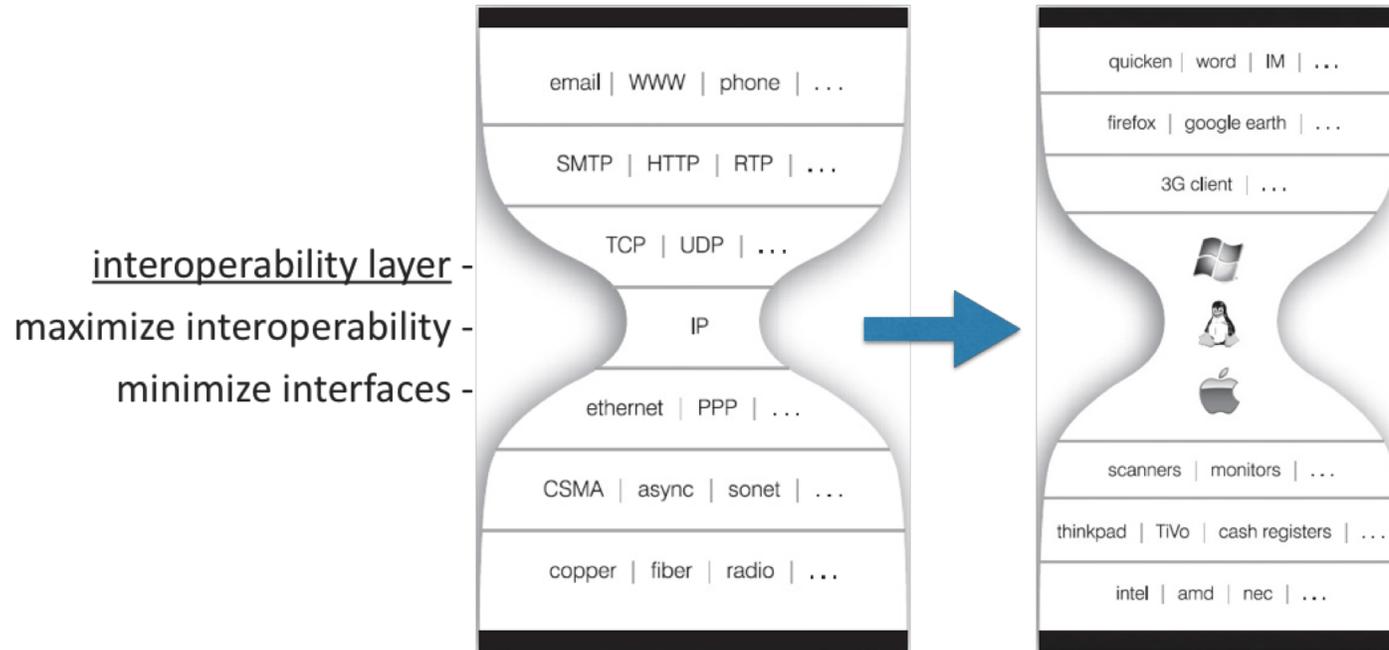
Portable ID stack



PID as interface: Researcher-centric network



Generative technologies

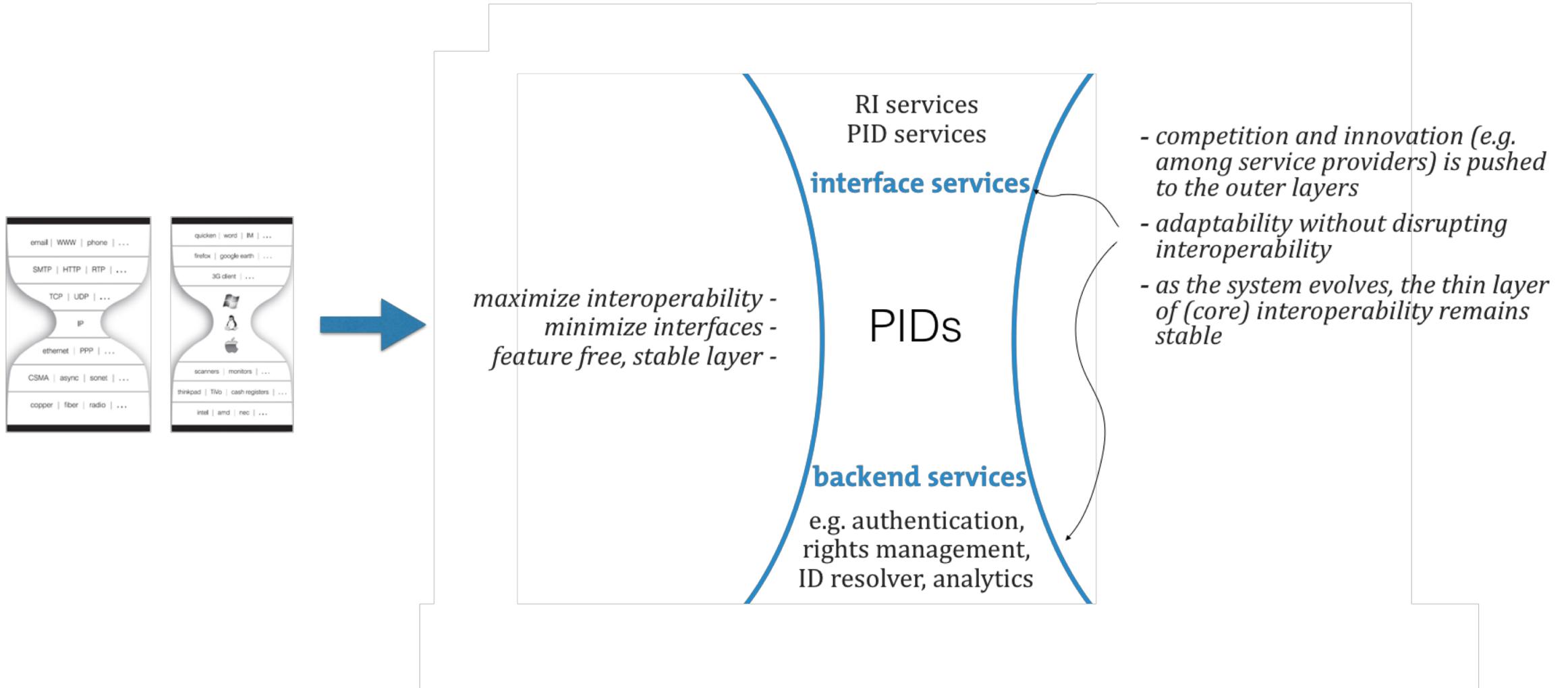


Generativity is a system's capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences (Zittrain 2009)

[1] Zittrain, Jonathan. 2009. The Future of the Internet--And How to Stop It. Yale University Press

[2] Akhshabi and Dovrolis. 2011. The Evolution of Layered Protocol Stacks Leads to an Hourglass-shaped Architecture.

PIDs and generativity



ORCID-NL, pilot to consortium

ORCID-NL, pilot (2014-2015)

- 10 out of 14 Research Universities participated in pilot consortium
 - Many/most were also implementing new CRIS systems
 - Group of 25 universities of applied science in discussion (still)
 - New consortium: 8 research universities
-
- CRIS integration sub-optimal
 - CRIS vendors not so responsive
 - Integration issues remain

ORCID-NL, pilot (2014-2015)

Lessons Learned:

- Data security and privacy issues are significant; differences between US and EU
- Diversity of relevant Stakeholders related author IDs
- Some friction with CRIS suppliers regarding ORCID integration
- ISNI/ORCID interface is promising, but requires increased interaction between organizations
- Authority for author IDs (and RI generally) is ambiguous and varies across institutions
- Business case for ORCID is based on future benefits, while stakeholders need near-term solutions
- As ORCID popularity increases (internationally), some competition from CRIS supplier
- There are opportunities to leverage existing SURF tech/services (e.g. SURFconext) with ORCID

ORCID-NL consortium (2016-present)

8 consortium members

Working group format

Thematic meetings

- 1st: Monitoring ORCID use/adoption
- 2nd: strategies for substantial expansion of registered ORCIDs
- 3rd: approaches to lobby large funders and publishers to adopt ORCID,
- 4th: Functional wishes: e.g. authentication, CRIS-content in ORCID, linking pubs in ORCID.

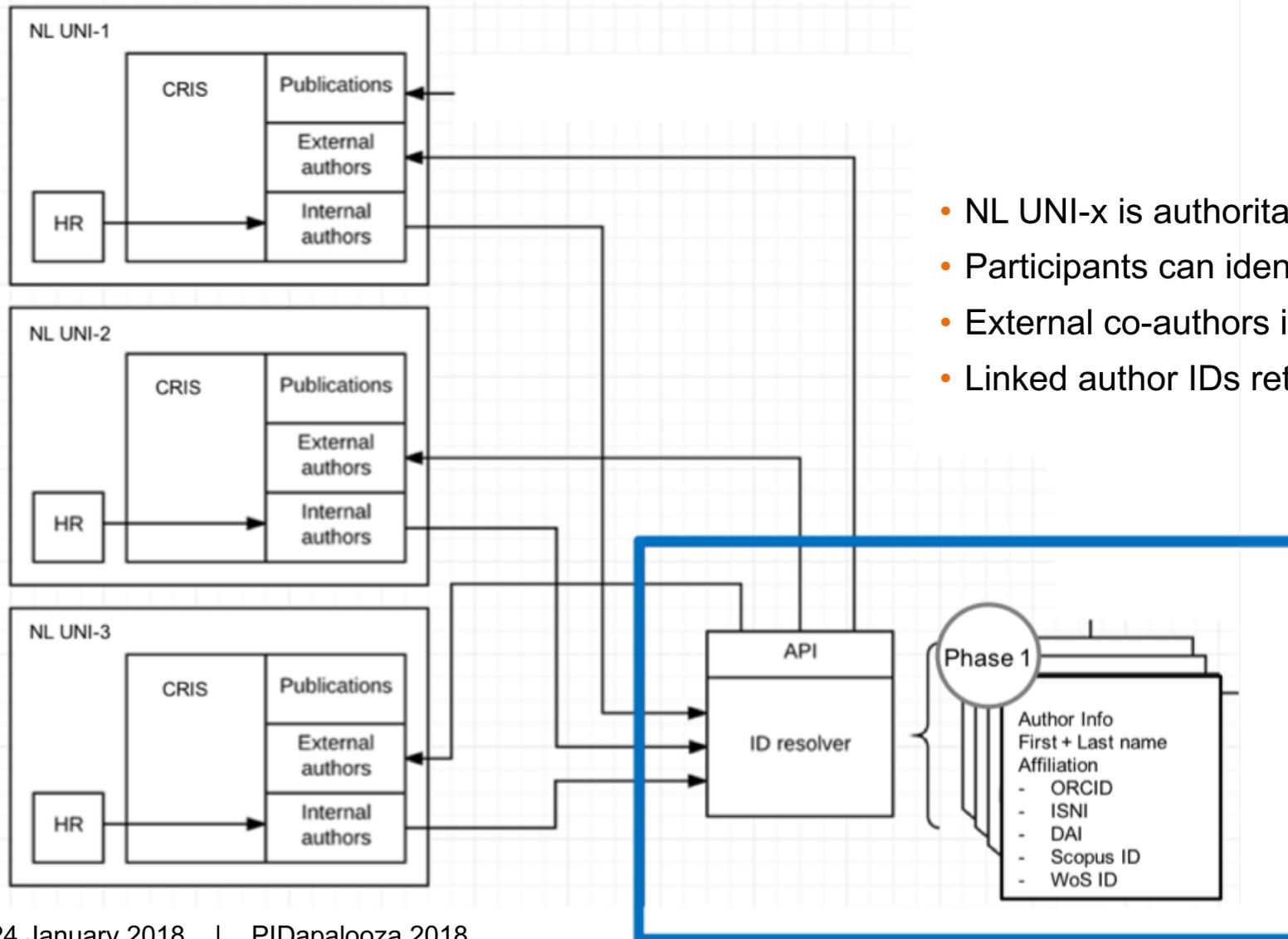
Spin-off project: ID-resolver (linking author IDs)

- Proof of concept --> develop business case
- Privacy by design --> pilot with real data

Integration issues, ORCID-NL pilot

- Identifying external (NL) co-authors
 - Institutions without CRIS systems
 - Multiple researcher IDs
 - Vendor lock-in
 - Evolving data privacy landscape
- Is there a there, there?
- What research information reality are we aiming for?
 - What are the steps to getting there?

NL Centralized ID-Resolver: phase 1, co-author disambiguation: API interface (pilot underway)



- NL UNI-x is authoritative for affiliated authors
- Participants can identify external authors only if co-authors
- External co-authors identified by authorities info
- Linked author IDs retained by ID resolver

design/concept

- Herman von Dompsele, SURFnet
- Nick Veenstra, Eindhoven University of Technology
- Clifford Tatum, SURFmarket

Towards a national research Information strategy

Centralized ID-Resolver

- phase 1, the co-author problem: API interface (pilot underway)
- phase 2, non-CRIS institutions: Web interface
- Phase 3, non-CRIS institutions: federated authentication

Future possibilities

- collect/link object and ORG IDs
- Integrate with national repository (NARCIS)
- Monitor Open Access
- Monitor output from funded research

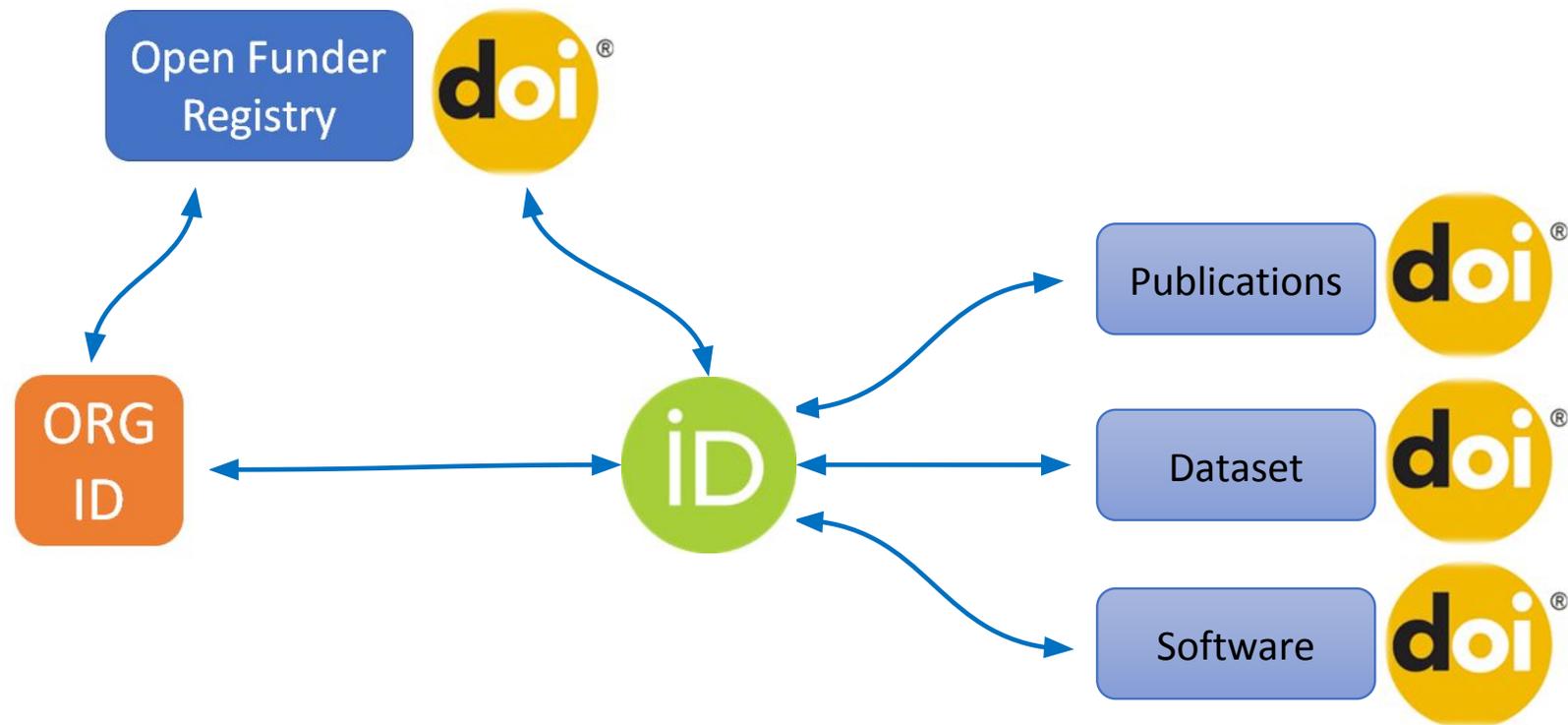
Principles and pragmatics of as open as possible: persistent identifiers as the interface between research information commons and closed systems

Clifford Tatum, SURF
<https://orcid.org/0000-0002-2212-3197>
@tatumcc

Josh Brown, ORCID
<https://orcid.org/0000-0002-8689-4935>
@joshbrown_orcid

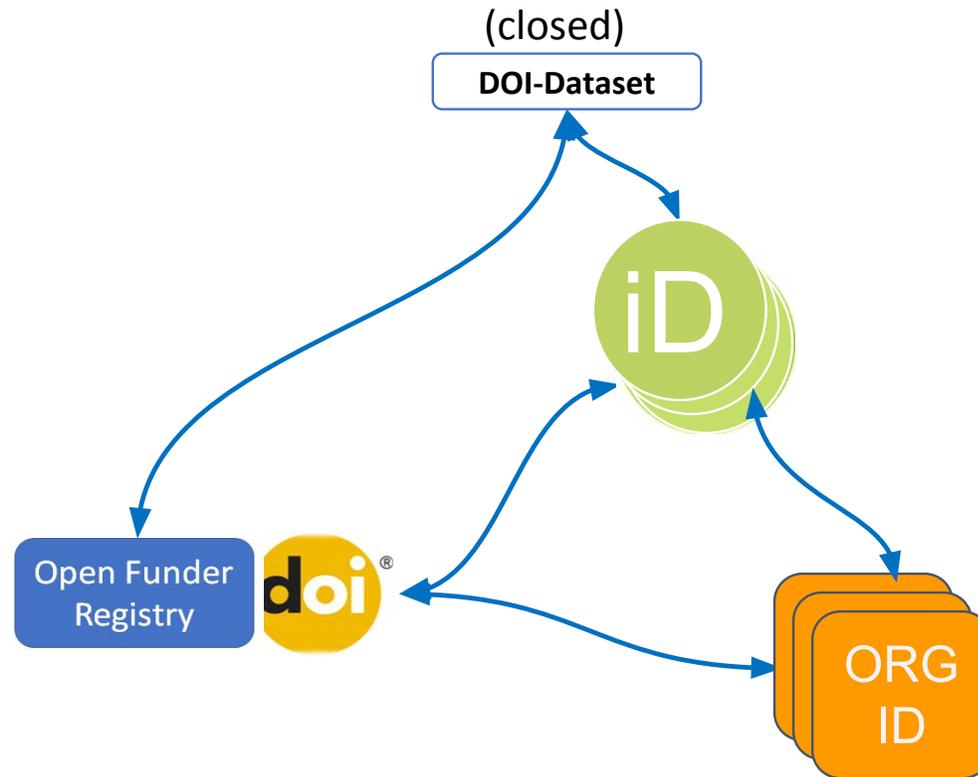
CRIS2018 - Umeå, Sweden | 14-16 June 2018
URI: <http://hdl.handle.net/11366/658>

PID as interface: Researcher-centric network

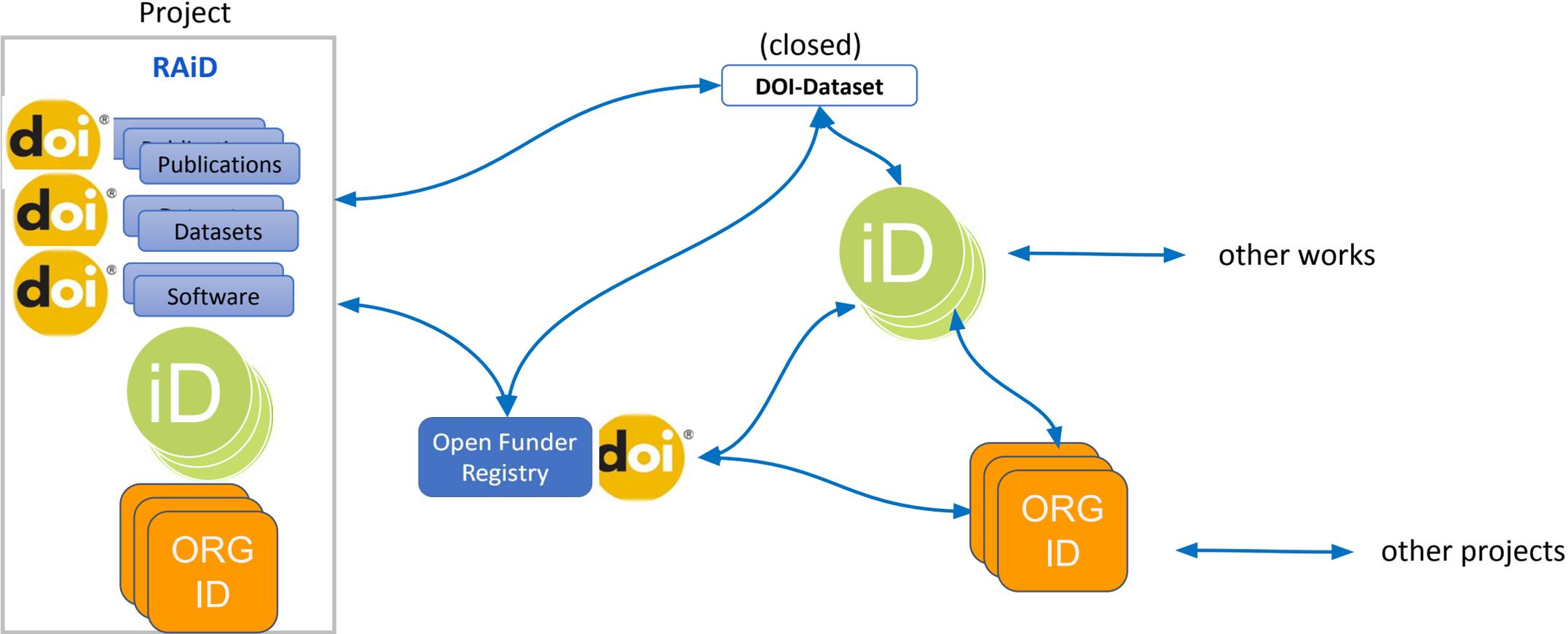


PID as interface:

Closed dataset, open PID



PID as interface: Closed dataset, open PID



Example-1: Peer review citation - closed

▼ Peer review (1) Sort

▼ review activity for [The Journal of Neuroscience\(1\)](#)

journal, The Journal of Neuroscience and eNeuro support the mission of the Society for Neuroscience. The open-access journal eNeuro embodies an emerging scientific vision that offers a new experience for authors and readers, and the highly cited Journal of Neuroscience publishes papers on a broad range of topics of general interest to those working on the nervous system.

Review date	Type	Role	Actions
2017	review	reviewer	<input type="checkbox"/> show details view

Example-1: Peer review citation - open

▼ Peer review (1) Sort

▼ review activity for [F1000Research\(1\)](#)   

journal, F1000Research

Review date	Type	Role	Actions
2017-08	review	reviewer	hide details view 

Review identifier(s): DOI: [10.5256/f1000research.13189.r24792](#)
Convening organization: F1000Research(London, United Kingdom)
Review subject: The new alchemy: Online networking, data sharing and research activity distribution tools for scientists
[version 1; referees: 1 approved] journal-article F1000Research.
DOI: [10.12688/f1000research.12185.1](#)

ID Campaign

 **nature**
International journal of science

CORRESPONDENCE · 20 JUNE 2018

DOIs and other persistent identifiers have much more to offer science

Alice Meadows 

Increasingly, ORCID, DOIs and other identifier systems that are open and community-governed are embedded in scholarly works and information systems, such as papers and citation indices. They could benefit research in many more ways than their current use in unambiguously tracking authors and published output.

Take, for example, the manuscript-submission process. Authors must create a journal account, review submission requirements and upload their manuscript, which probably contains links to other important information. Journals need to find unconflicted reviewers. Payment contacts for open access might be required. By using persistent identifiers, most of the manual processes in this workflow can be semi-automated (see go.nature.com/2lnqibu). Expertise should not be wasted on mundane administrative tasks.

Identifiers can also act as signposts and coordinates, guiding us to information sources and showing connections between research and researchers. They can increase the visibility of a study, its origins and its impact, and indicate where it is hosted and who to ask for access.

 PDF version

SUBJECTS

[Communication](#) [Publishing](#)

Search  **FOR RESEARCHERS** **FOR ORGANIZATIONS** **ABOUT** **HELP** **SIGN IN** English

Connecting Research and Researchers

Mapping the PID Landscape

 Submitted by Alice Meadows on Thu, 2018-06-07 00:00

This post was co-authored with [Christopher Brown](#) and [Neil Jacobs \(Jisc\)](#), [Josh Brown](#) and [Laure Haak \(ORCID\)](#), and [Clifford Tatum \(SURF\)](#)



The landscape of research information is largely closed to us. We rely on original research to solve many of the challenges facing humanity, to improve lives, and to advance human understanding, and we invest in it accordingly. However, when we survey the map of our research world it is filled with gaps. We pass along a few well-trodden roads (too often paying a substantial toll for the privilege) and we can only wonder about what lies just over the horizon.

We can point to many contributing factors: business models that militate against the sharing of information; aggregation of research analytics for local strategic purposes; technological barriers to linking information between sources; cultural practices that reward and privilege a small slice of research activity; and systems that emphasise hard sciences and anglophone literature. Any and all of these can, and do, hide some of the richness of research endeavour. However, these systemic challenges are not the focus of this discussion. Instead, our focus is on the gaps in our understanding of the landscape: the empty parts of the research map.

If we are to open research up, to enable and support more transparency and accountability, and to ensure that we are supporting research effectively, we must be able to survey the research landscape in its entirety. That means recognising more kinds of contributions to research, and acknowledging a broader, more diverse range of career paths. To do so, we need tools to help us to fill in the blanks. Luckily, a powerful set of these tools exists - open, community-governed identifier systems are already a well-established part of the scholarly world.

Identifiers act as coordinates on the research map. They both tell us where something is located, and also act

PID Approach to RI infrastructure



SURF PID project (proposed)

Persistent Identifiers: Fundamental building blocks for FAIR research

- Service delivery
 - Durable nationwide ORCID solution for the Netherlands (WO + HBO)
 - NL PID stakeholder strategic relationship management: UKB, NWO, VSNU, KNAW, HKI
 - NL PID strategy coordination (venue/format TBD, but probably something like a WG)
- Services development (ideas)
 - ID resolver (author ID linking database)—in progress
 - Forthcoming ORG ID (pilot with RI exchange workflows)
 - Linking funder information to projects (ORCID → exchange key)
 - Openness Profile (linked to ORCID record)—[Knowledge Exchange initiative](#)
- Knowledge exchange (international)
 - ID campaign (ORCID, SURF, Jisc, ARDC)
 - ORCID consortia forum
 - International PID/RI community

Relevant Literature

1. Aquino, Janine, John Allison, Robert Rilling, Don Stott, Kathryn Young, and Michael Daniels. 2017. "Motivation and Strategies for Implementing Digital Object Identifiers (DOIs) at NCAR's Earth Observing Laboratory – Past Progress and Future Collaborations." *Data Science Journal* 16 (0). <https://doi.org/10.5334/dsj-2017-007>.
2. Borgman, Christine L. 2012. "The Conundrum of Sharing Research Data." *Journal of the American Society for Information Science and Technology* 63 (6): 1059–78. <https://doi.org/10.1002/asi.22634>.
3. Borycz, Joshua, and Bonnie Carroll. 2018. "Managing Digital Research Objects in an Expanding Science Ecosystem: 2017 Conference Summary." *Data Science Journal* 17 (0). <https://doi.org/10.5334/dsj-2018-016>.
4. Bryant, Rebecca, Annette Dortmund, and Malpas Constance. 2017. "Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information." Dublin, Ohio: OCLC Research. <https://doi.org/10.25333/C32K7M>.
5. Dappert, Angela, Adam Farquhar, Rachael Kotarski, and Kirstie Hewlett. 2017. "Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research." *Data Science Journal* 16 (0). <https://doi.org/10.5334/dsj-2017-028>.
6. Dutch Ministry of Education, Culture and Science, (OCW). 2017. "National Plan Open Science." DOI: 10.4233/uuid:9e9fa82e-06c1-4d0d-9e20-5620259a6c65.
7. European Commission. 2016. "Programme Guidelines on FAIR Data Management in Horizon 2020."
8. European Commission. 2017. "Evaluation of Research Careers Fully Acknowledging Open Science Practices; Rewards, Incentives And/Or Recognition for Researchers Practicing Open Science." Directorate-General for Research and Innovation Open Science and ERA policy. https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf.
9. FORCE11. n.d. "The FAIR Data Principles." Accessed March 19, 2018. <https://www.force11.org/group/fairgroup/fairprinciples>.
10. Fortunato, Santo, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, et al. 2018. "Science of Science." *Science* 359 (6379): eaao0185. <https://doi.org/10.1126/science.aao0185>.
11. Golodoniuc, Pavel, Nicholas J. Car, and Jens Klump. 2017. "Distributed Persistent Identifiers System Design." *Data Science Journal* 16 (0). <https://doi.org/10.5334/dsj-2017-034>.
12. Klump, Jens, and Robert Huber. 2017a. "20 Years of Persistent Identifiers – Which Systems Are Here to Stay?" *Data Science Journal* 16 (0). <https://doi.org/10.5334/dsj-2017-009>.
13. Klump, Jens, and Robert Huber. 2017b. "What Is the Value Proposition of Persistent Identifiers?" In , 19:4567. <http://adsabs.harvard.edu/abs/2017EGUGA..19.4567K>.
14. Klump, Jens, Fiona Murphy, Tobias Weigel, and Mark A. Parsons. 2017. "Editorial: 20 Years of Persistent Identifiers – Applications and Future Directions." *Data Science Journal* 16 (0). <https://doi.org/10.5334/dsj-2017-052>.
15. Knowledge Exchange. 2011a. "Den Haag Persistent Object Identifier – Linked Open Data Manifesto." Den Haag. http://repository.jisc.ac.uk/6438/2/Den_Haag_Persistent_Object_Identifier_-_Linked_Open_Data.pdf.
16. Knowledge Exchange. 2011b. "Report on the Knowledge Exchange Seminar on Persistent Object Identifiers 14-15 June 2011, The Hague." http://repository.jisc.ac.uk/6438/1/Report_on_the_Knowledge_Exchange_Seminar_on_Persistent_Indentifiers_June_2011.pdf.
17. Kunze, John, Scout Calvert, Jeremy D. DeBarry, Matthew Hanlon, Greg Janée, and Sandra Sweat. 2017. "Persistence Statements: Describing Digital Stickiness." *Data Science Journal*. <https://doi.org/10.5334/dsj-2017-039>.
18. Leonelli, Sabina. 2017. "European Commission: MLE on Open Science: Almetrics and Rewards – Incentives and Rewards to Engage in Open Science Activities." Brussels: Directorate-General for Research and Innovation.
19. McMurry, J. A., N. Juty, N. Blomberg, T. Burdett, T. Conlin, N. Conte, M. Courtot, et al. 2017. "Identifiers for the 21st Century: How to Design, Provision, and Reuse Persistent Identifiers to Maximize Utility and Impact of Life Science Data., Identifiers for the 21st Century: How to Design, Provision, and Reuse Persistent Identifiers to Maximize Utility and Impact of Life Science Data." *PLoS Biology*, *PLoS Biology* 15, 15 (6, 6): e2001414–e2001414. <https://doi.org/10.1371/journal.pbio.2001414>, [10.1371/journal.pbio.2001414](https://doi.org/10.1371/journal.pbio.2001414).
20. Meadows, Alice, Christopher Brown, Josh Brown, Laure Haak, Neil Jacobs, and Clifford Tatum. 2018a. "Mapping the PID Landscape." Text. June 7, 2018. <https://orcid.org/blog/2018/06/21/mapping-pid-landscape>.
21. Meadows, Alice, Christopher Brown, Josh Brown, Laure Haak, Neil Jacobs, and Clifford Tatum. 2018b. "DOIs and Other Persistent Identifiers Have Much More to Offer Science." *News. Nature*. June 20, 2018. <https://doi.org/10.1038/d41586-018-05456-8>.
22. Neylon, Cameron. 2017. "Knowledge Exchange Approach towards Open Scholarship," July. <https://doi.org/10.5281/zenodo.826643>.
23. Star, Susan Leigh, and Karen Ruhleder. 1996. "Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces." *Information Systems Research* 7 (1): 111–34. <https://doi.org/10.1287/isre.7.1.111>.
24. Tatum, Clifford. 2016. "Towards Governance of PID Portability for Research Evaluation." presented at the PIDapalooza, Reykjavik, Iceland, November 10. <https://doi.org/10.6084/m9.figshare.4212732.v1>.
25. Tatum, Clifford, and Josh Brown. 2018. "Principles and Pragmatics of 'as Open as Possible': Persistent Identifiers as the Interface between Research Information Commons and Closed Systems." presented at the CRIS2018, Umeå, Sweden, June 14. <http://dSPACECRIS.eurocris.org/handle/11366/658>.
26. Tatum, Clifford, and John Doove. 2018. "Domination and Submission: The Struggle to Retain Ownership/Control of National Research Information." presented at the PIDapalooza, Girona, Spain, January 24. <https://doi.org/10.6084/m9.figshare.5817309.v1>.
27. Wimalaratne, Sarala M., Nick Juty, John Kunze, Greg Janée, Julie A. McMurry, Niall Beard, Rafael Jimenez, et al. 2018. "Uniform Resolution of Compact Identifiers for Biomedical Data." *Scientific Data* 5 (May): 180029. <https://doi.org/10.1038/sdata.2018.29>.

THANKS!

 Clifford Tatum

 E-mail: Clifford.Tatum@surfmarket.nl

 www.surf.nl

 @tatumcc

 <https://orcid.org/0000-0002-2212-3197>

Samen aanjagen van vernieuwing



SURF