



Merits of an Interconnected and Interoperable Repository Ecosystem

Martin Klein

martin.klein@pnnl.gov

<https://orcid.org/0000-0003-0130-2097>

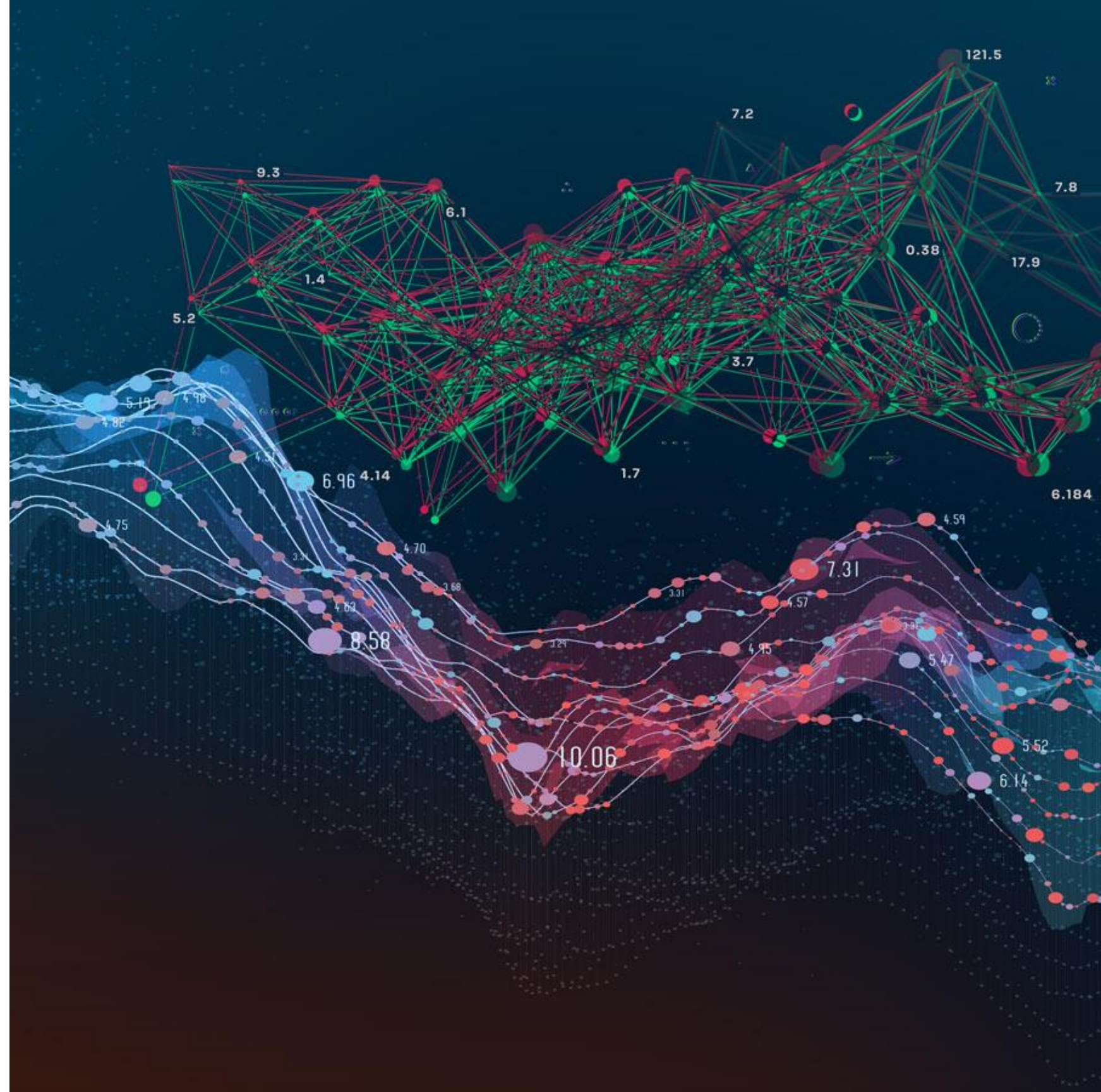
Inspired by conversations with

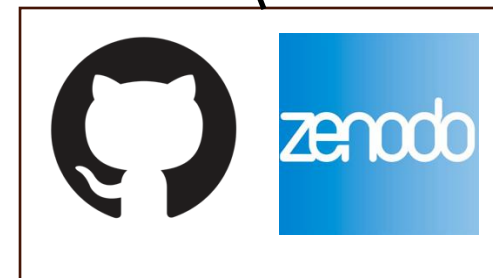
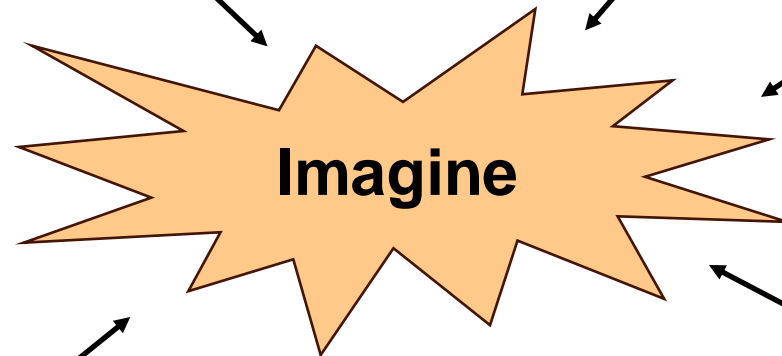
[Herbert Van de Sompel](#)



PNNL is operated by Battelle for the U.S. Department of Energy

PNNL-SA-204806





Repositories

- Active nodes in the scholarly communication ecosystem
 - Where the meaningful life of resources starts
 - Not where things go to die
- Decoupling of core value-adding services for research artifacts
 - Registration, certification, awareness, archiving
- System interoperability as the as key enabler for the creation of value-added services
 - Limit the reliance on 3rd party aggregators

Interoperability

- Siloed repository landscape
 - Indexes are incomplete and out of date
- OAI-PMH the lingua franca (!!!)
 - Technical support is lacking
- Ad-hoc APIs are not the answer
 - Though often hailed as the holy grail of interoperability

Interoperability Approach #1

- How do we support machines in navigating landing pages?
 - Find PIDs, authors, license, metadata, etc.
 - Find related resources e.g., datasets
 - Describe the resource
- **FAIR Signposting**
 - Typed links via HTTP link headers
 - Lightweight, works through paywalls, HEAD requests

<https://signposting.org/FAIR/>

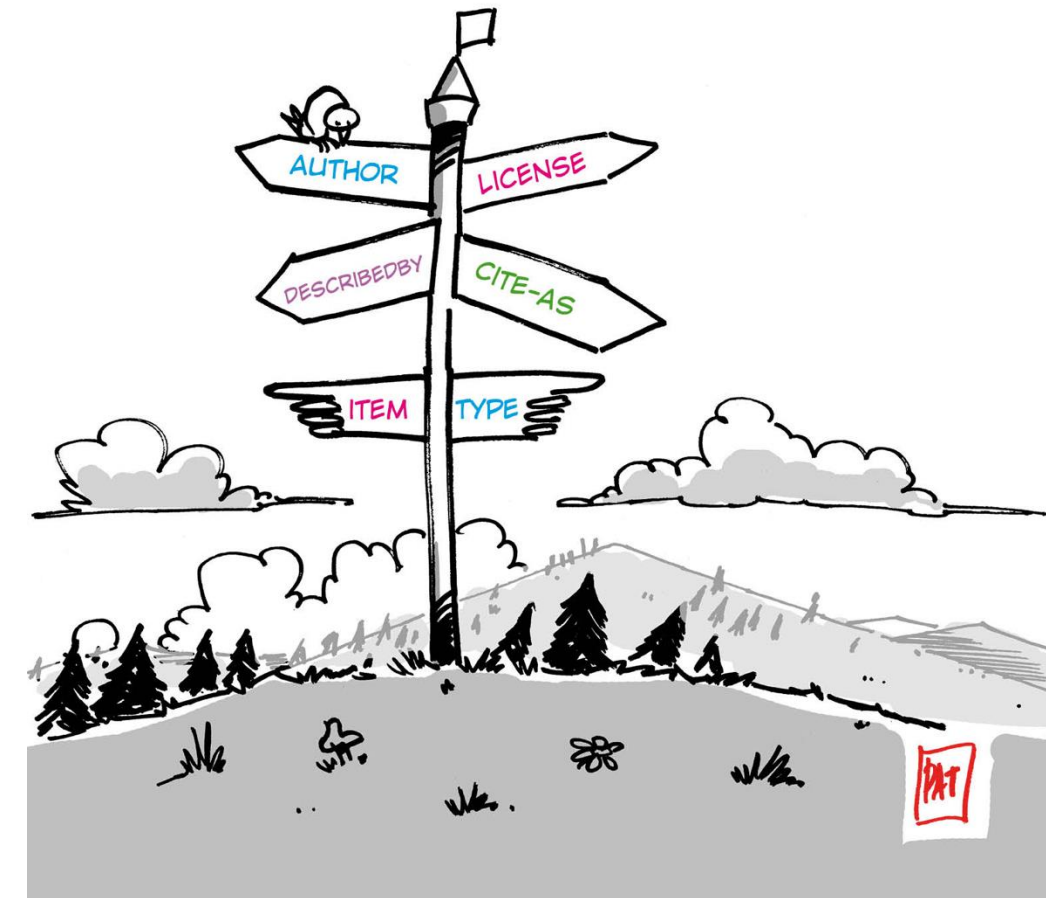


Image courtesy of Patrick Hochstenbach

Interoperability Approach #2

- How do we support machines in discovering in-scope resources?
 - PDFs
 - Datasets
 - Bibtex metadata
- **Signmap**
 - Built on Sitemaps
 - Signposting typed links added

<https://signposting.org/Signmap/>

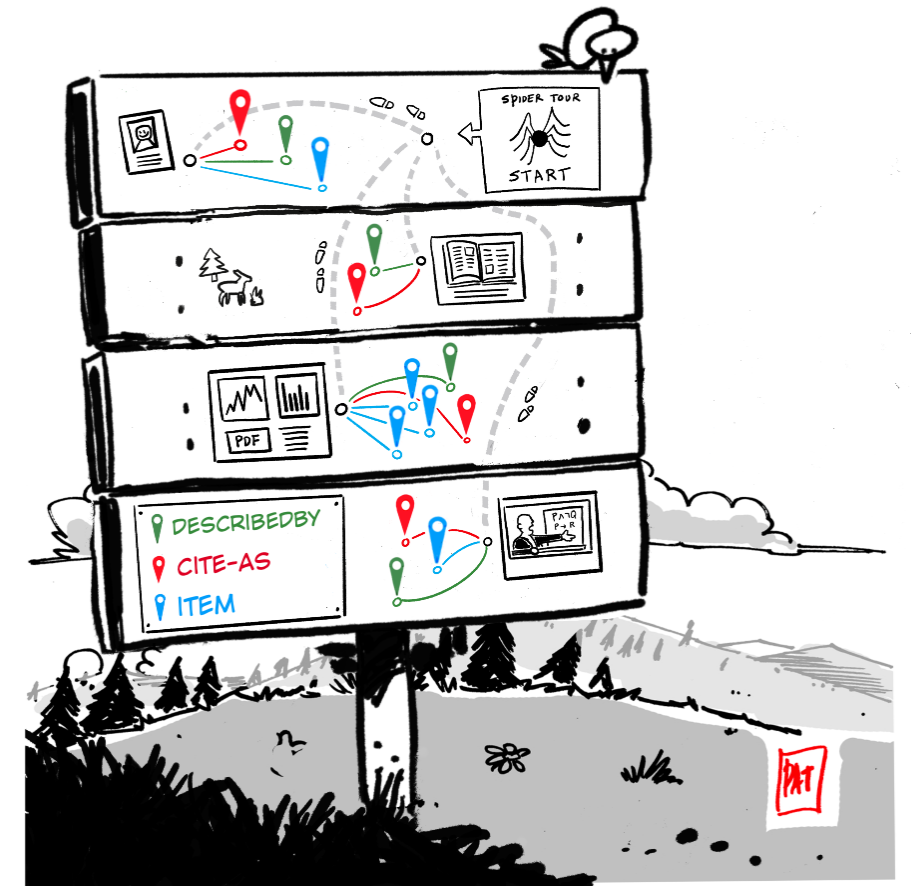


Image courtesy of Patrick Hochstenbach

Interoperability Approach #3

- How do we know what FAIR affordances a repository supports?
 - SPARQL
 - OAI-PMH
 - FAIR Signposting
 - APIs
- **FAIRiCat**
 - FAIR interoperability catalog
 - Human- and machine-interpretable

<https://signposting.org/FAIRiCat/>

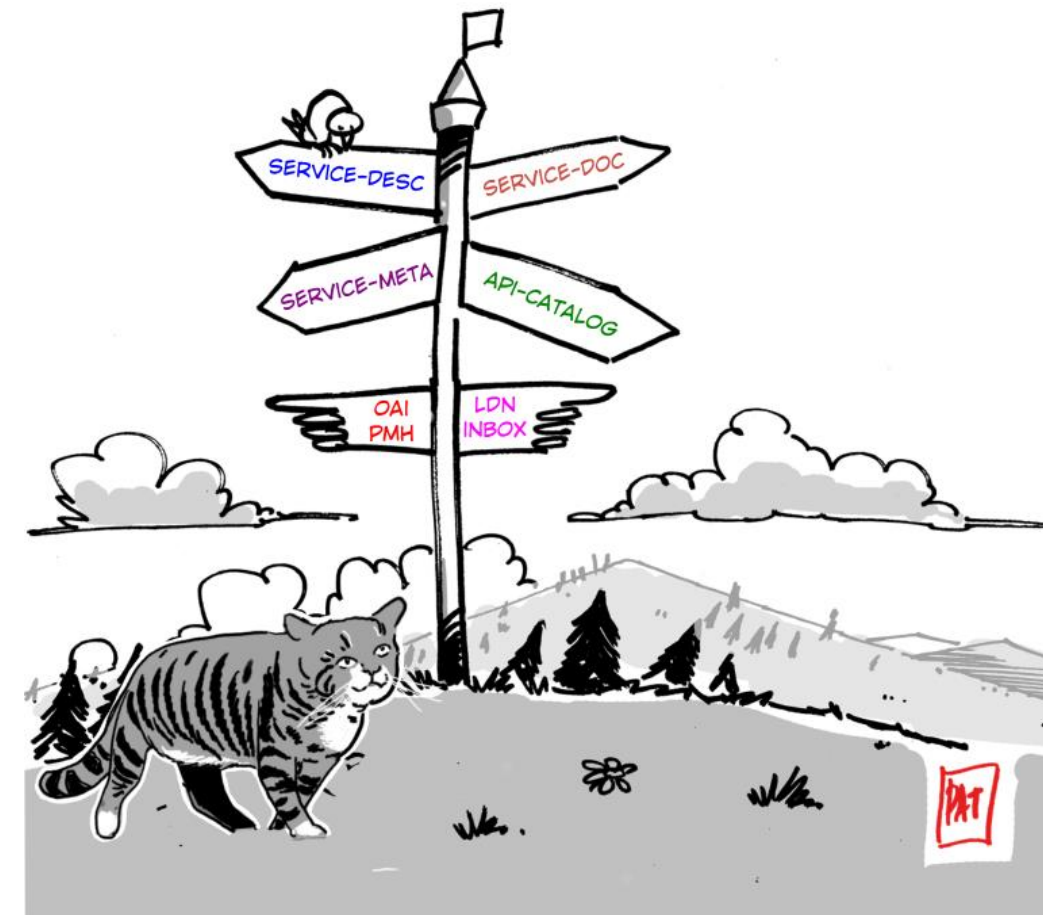


Image courtesy of Patrick Hochstenbach

Interoperability Approach #4

- How can we learn about all contributions by researcher hosted in a repository?
 - By unique identifier as key
 - e.g., ORCID, ISNI
- **authorIDy**
 - Machine-readable interface listing contributions
 - "authorIDy: a so simple protocol that could actually work." - Andrea Bollini

<https://signposting.org/authorIDy/>

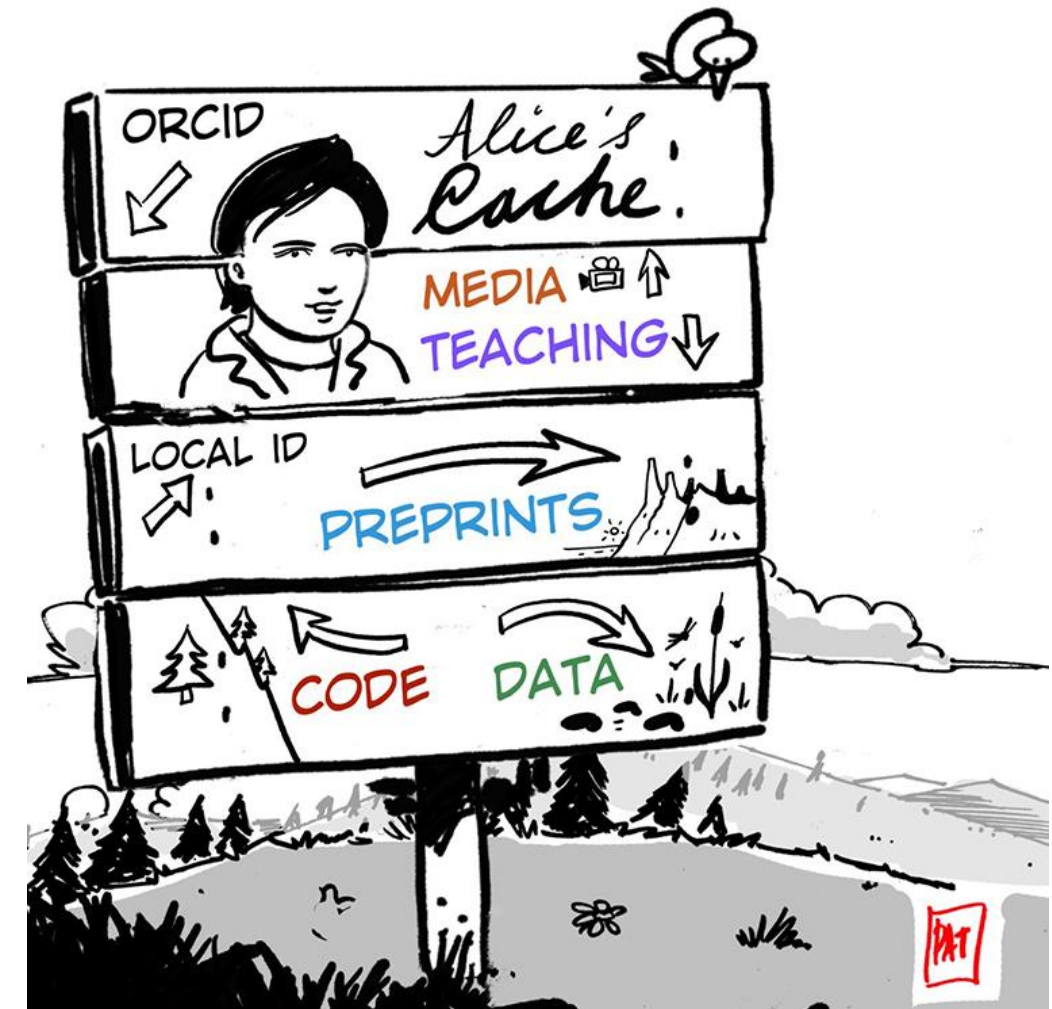
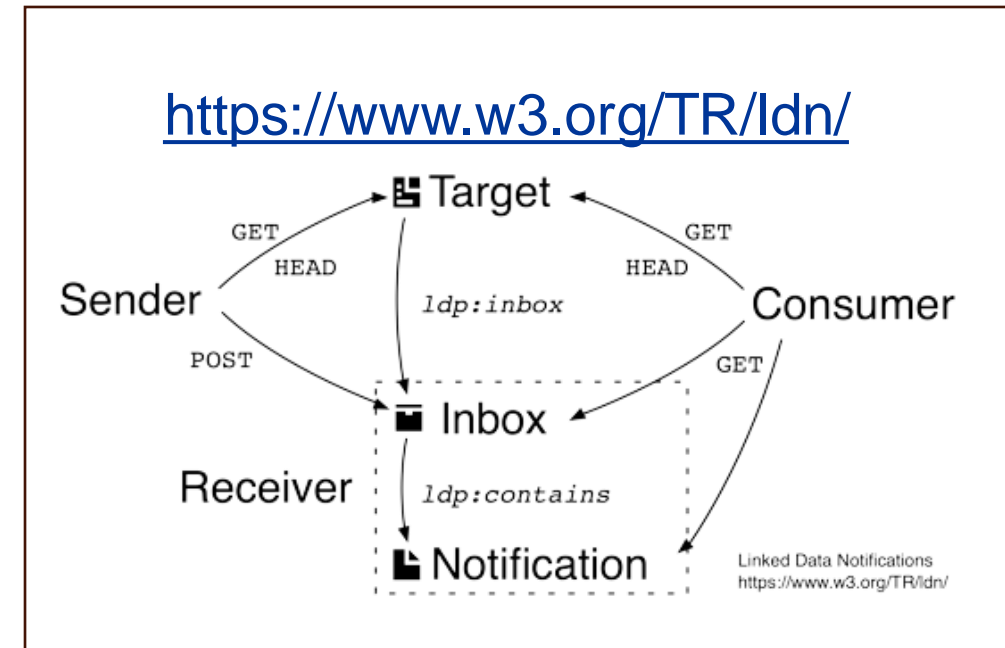


Image courtesy of Patrick Hochstenbach

New Ways to Achieve Interoperability



+




<https://coar-notify.net/>

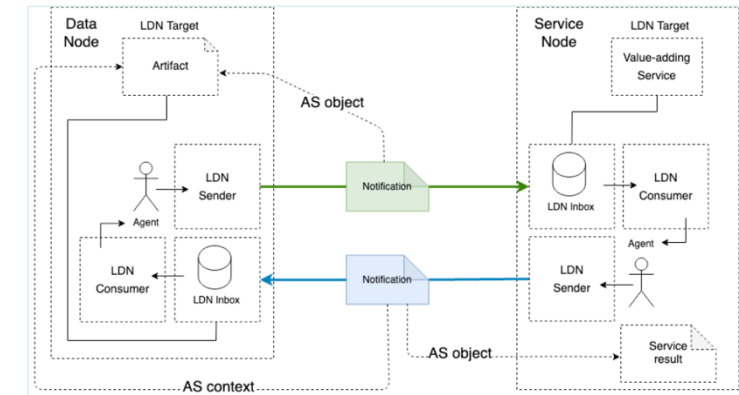


<https://www.eventnotifications.net/>

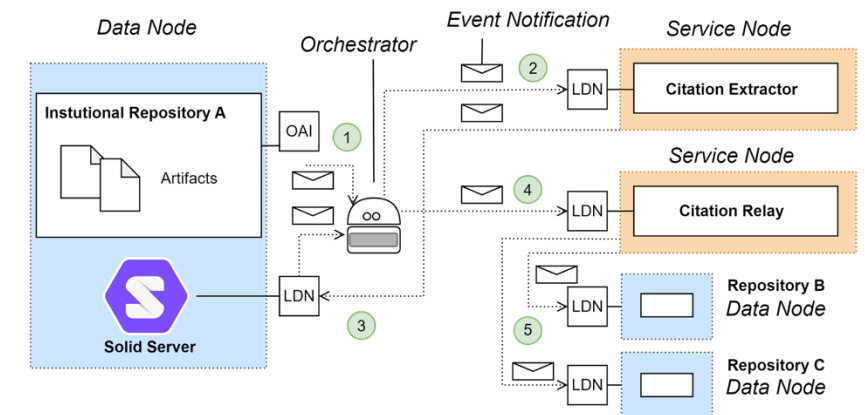
New Levels of Interoperability 1/2

- Repository-to-repository linking
 - Citation notifications
 - Paper \Leftrightarrow data notifications
 - Event notifications

- Status quo: reverse engineering linkage
 - Compute intensive
 - Imprecise due to heuristics
 - Significant delays (publishing, aggregation)



<https://doi.org/10.5281/zenodo.8076843>

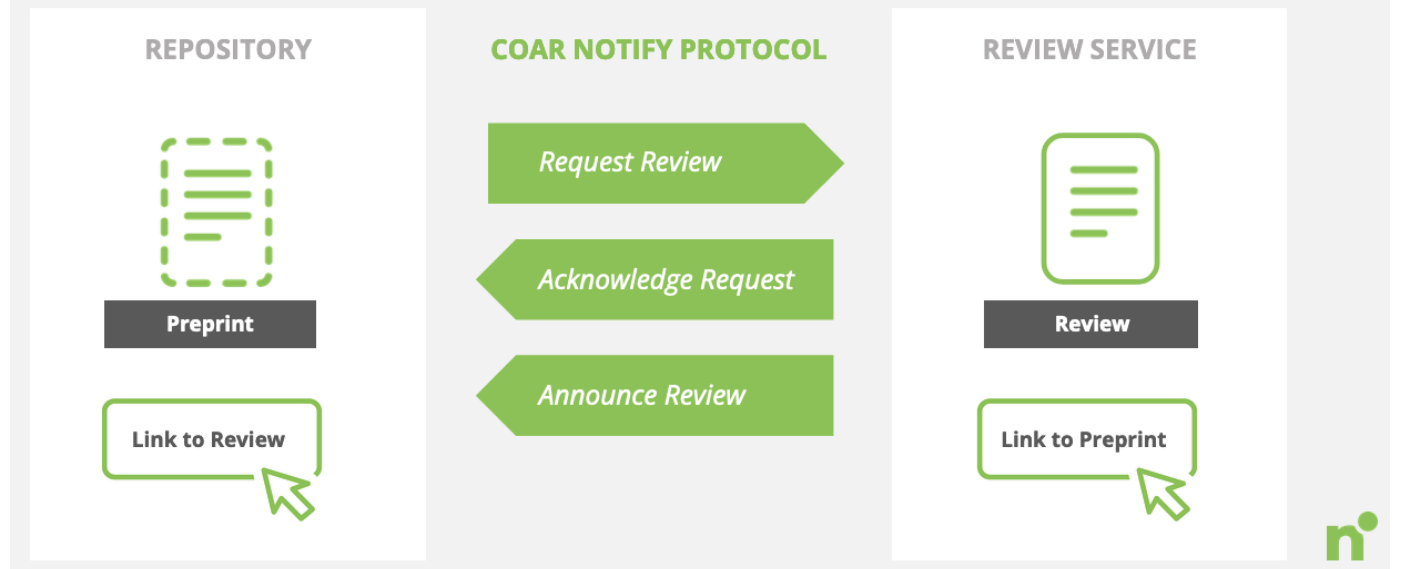


<https://journal.code4lib.org/articles/17823>

New Levels of Interoperability 2/2

- Repository-to-service linking
 - Peer review request
 - Relationship announcement
 - Bi-directional linking
- Status quo: siloed resources
 - Manual linking
 - Significant delays (publishing, aggregation)

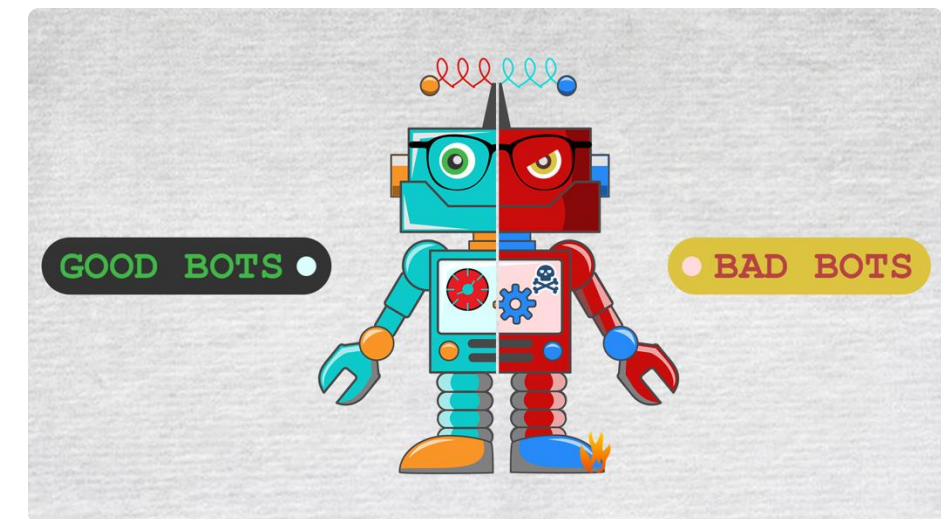
Facilitating communications across repository and review service



<https://coar-notify.net/>

Threat Model

- If/when these interoperability approaches become successful...
 - Lots of bots, notifications, value-added services
 - Research assessment based on derivative knowledge graphs
- ... they will become a target for unethical behavior
 - Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure"
 - Risk of repositories "closing their doors"
- Need for an infrastructure of trusted bots
 - Can't be solved by individual repository action (Cloudflare example)



Summary

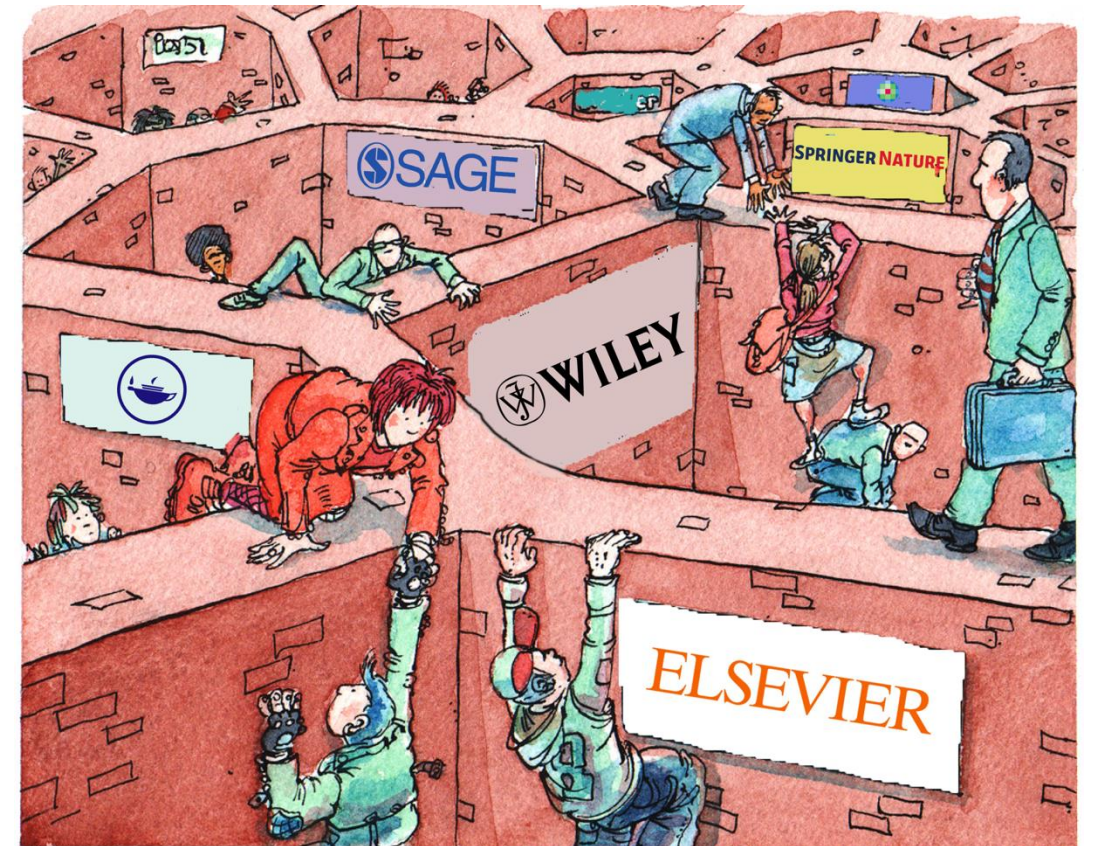
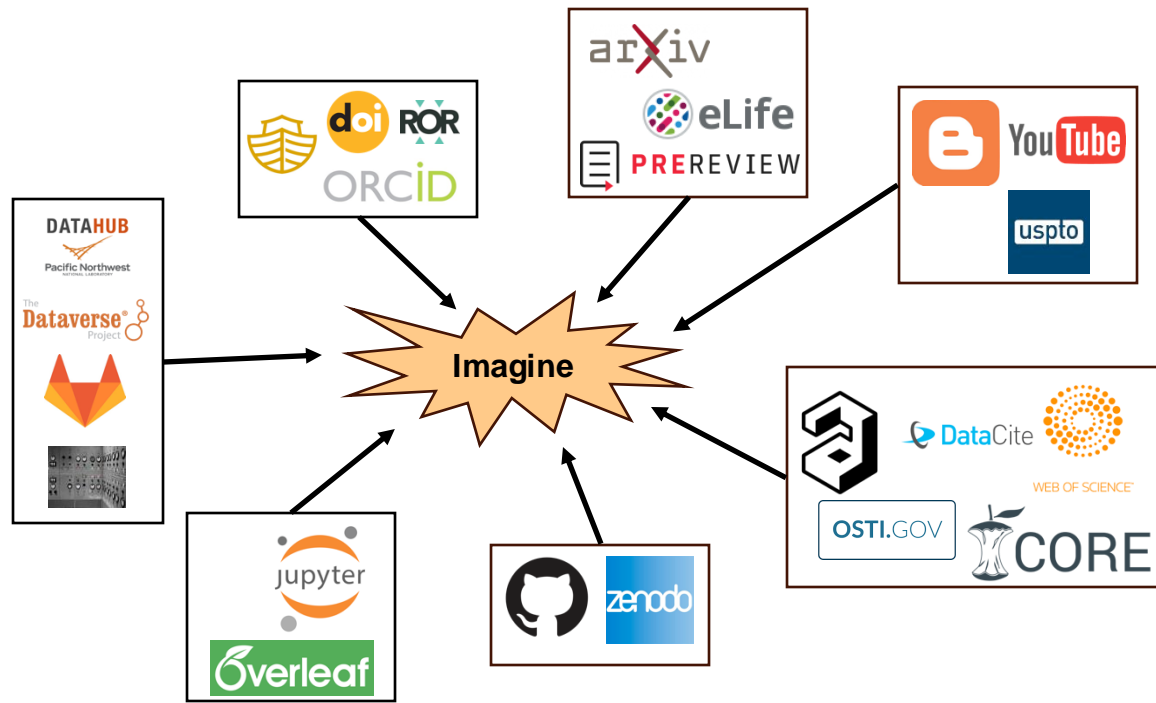


Image courtesy of Herbert Van de Sompel

