# Scholarly Communication: Deconstruct and Decentralize?

A Bold Speculation Featuring







Bob



Carol

Herbert Van de Sompel

<a href="mailto:ohvdsomp">ohvdsomp</a>
Los Alamos National Laboratory









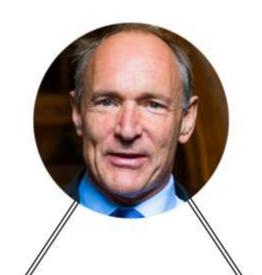
Tim Berners-Lee

Brewster Kahle



Paul Ginsparg









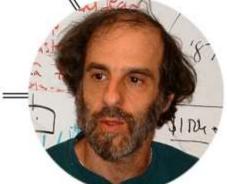


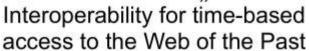


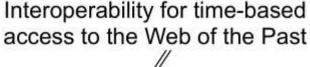














Web archiving paradigm for archiving the scholarly record



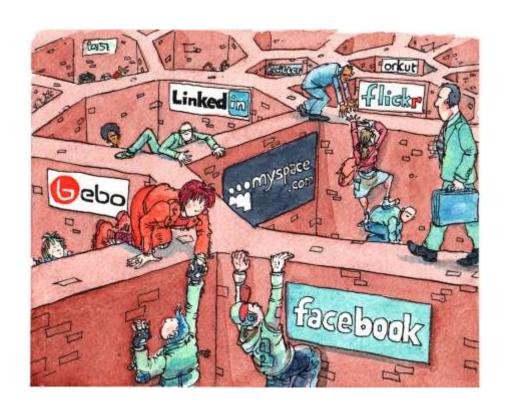
#### In this Talk

- Decentralized Web
- Why consider Decentralized Web for scholarly communication?
- Technical and Organizational inspiration
- Bold speculations





### Decentralized Web - In a Nutshell



#### Decentralized Web: Different Strands

- Distributed file systems
- Blockchain
- W3C Social Web & Linked Data Activities
  - Exemplified by MIT Solid Project
  - The Web as a distributed read/write file system with access controls
  - Enabled by a set of standardized HTTP-based protocols





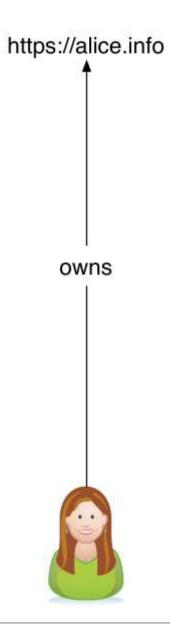
#### Social Network Silos

- Some massive central portals dominate the web
- Service is smooth, free; the user is the product
- No mobility; the user not in control of (moving) her data
- No interoperability; different APIs (if any) for different platforms
- Functionality contained within a portal, can't be reused on content that resides in other portal

Put the user back in the driver's seat by means of user owned "pods" that store the user's resources. Third parties are granted access. Introduce cross-pod applications.

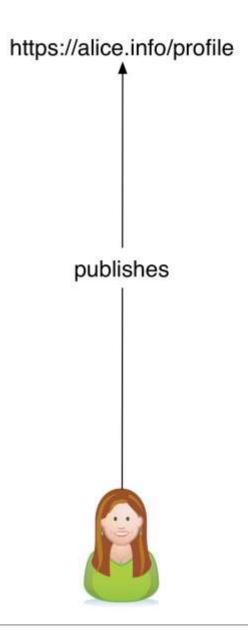
# MIT Solid Architecture





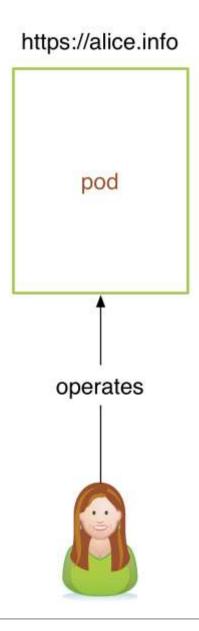






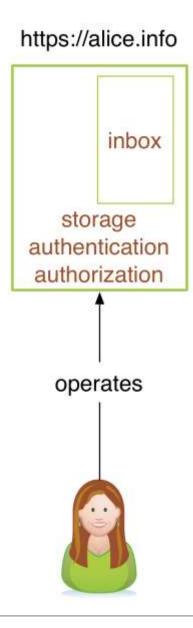






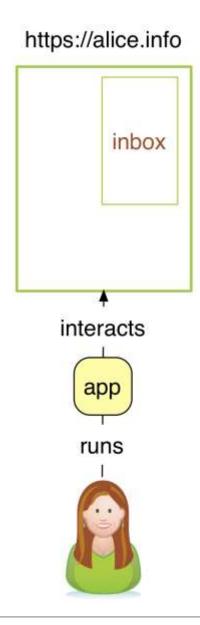






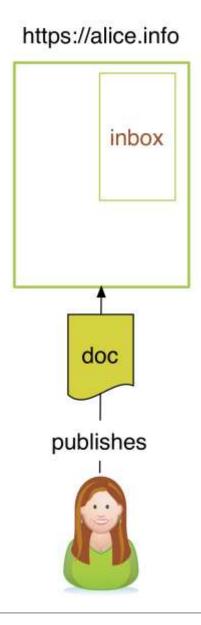








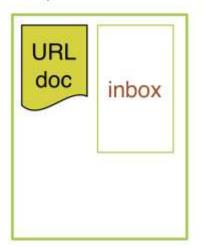








#### https://alice.info

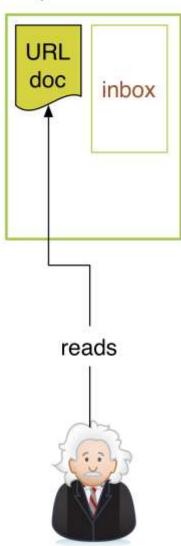






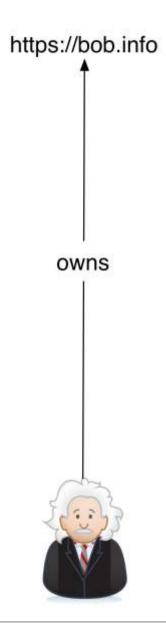


### https://alice.info



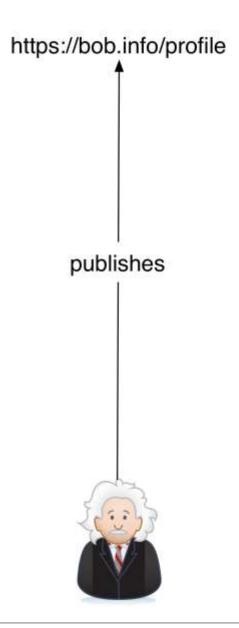






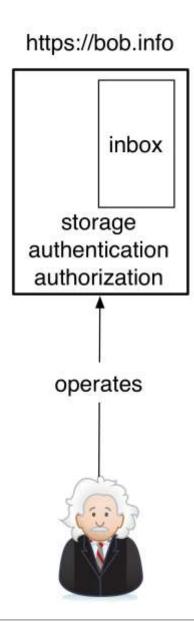






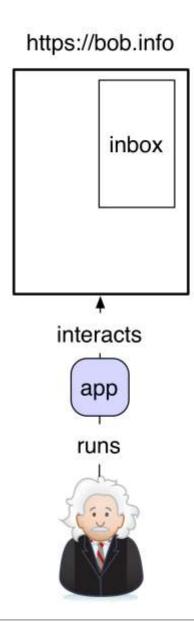






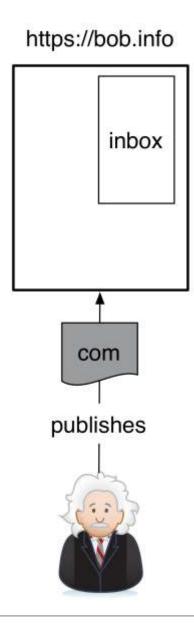








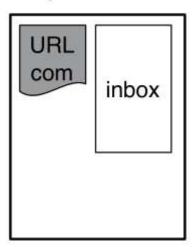








### https://bob.info

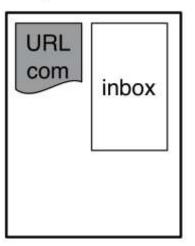








https://bob.info



https://alice.info

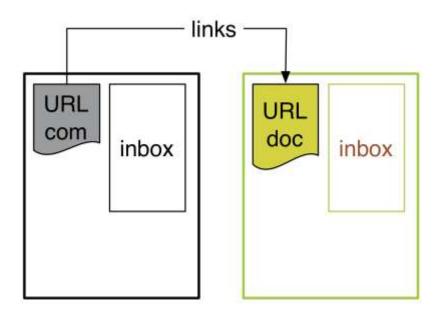










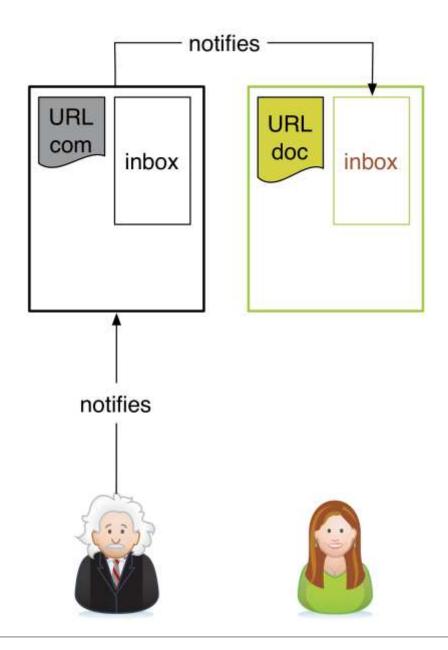






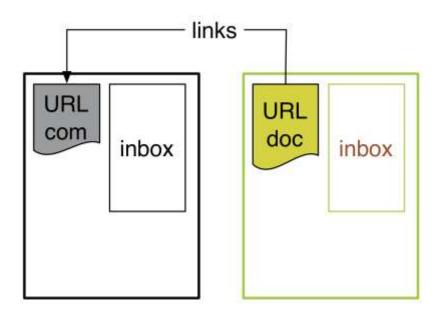










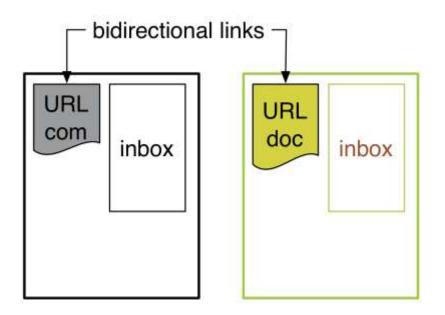










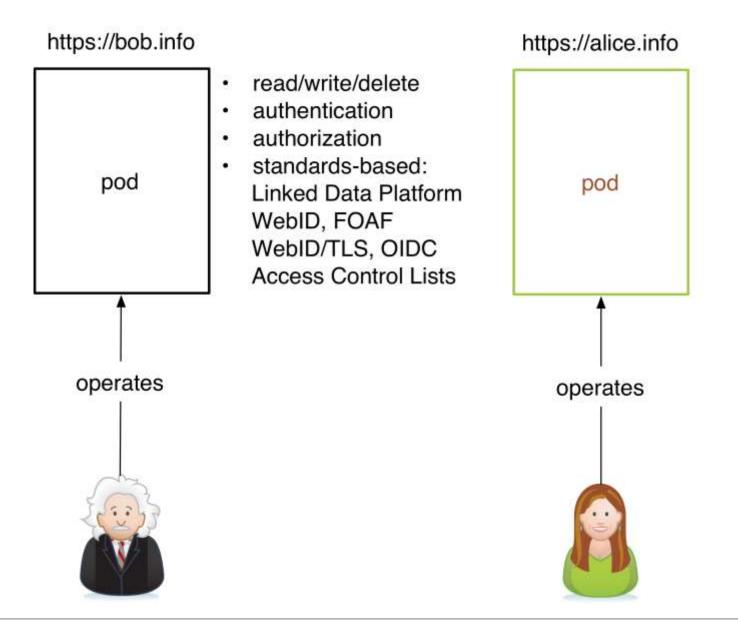






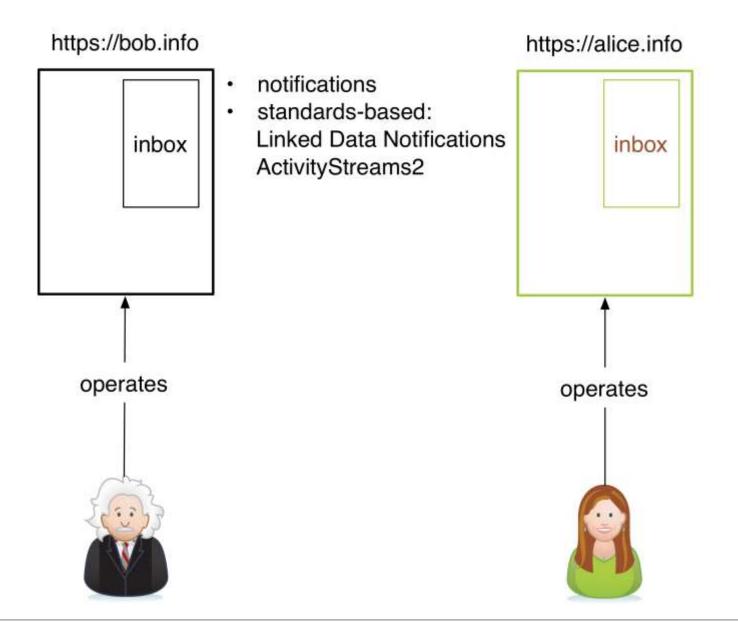






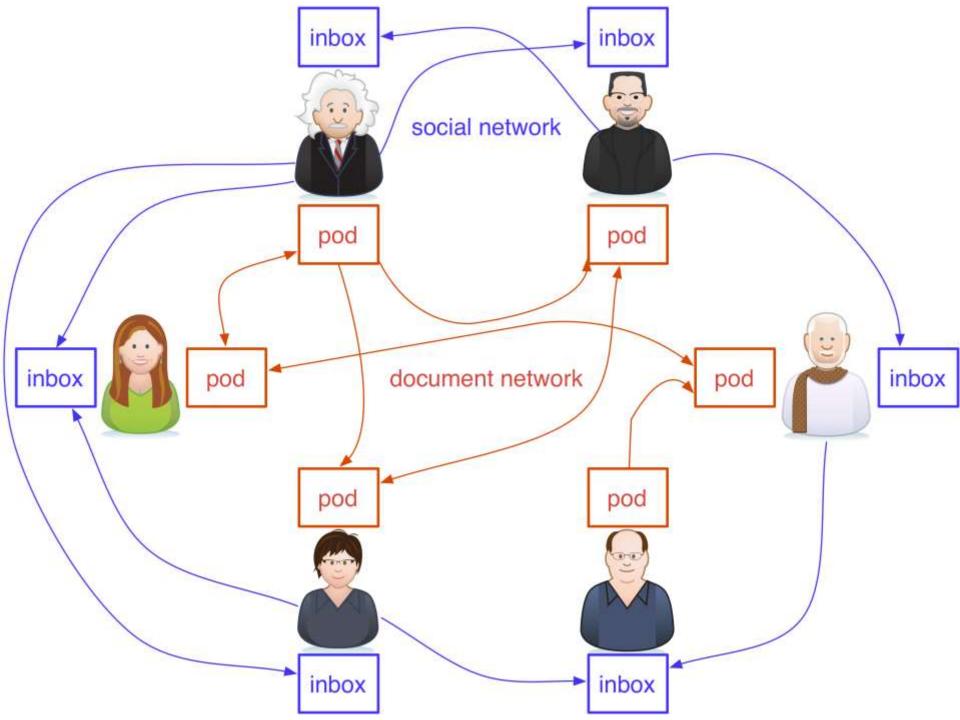




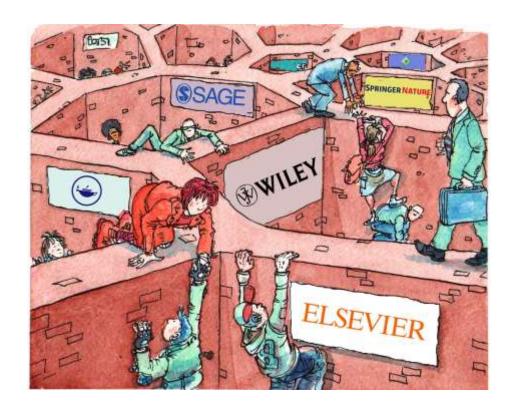








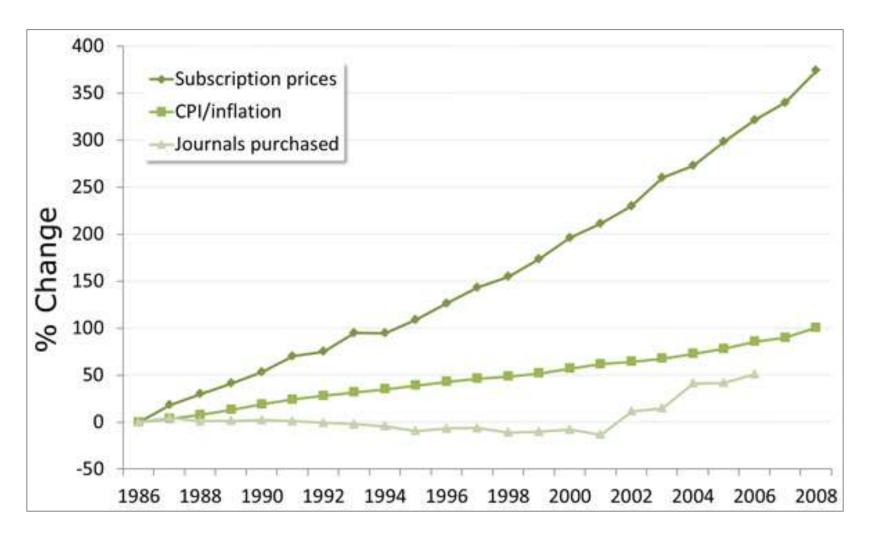
# **Scholarly Communication**



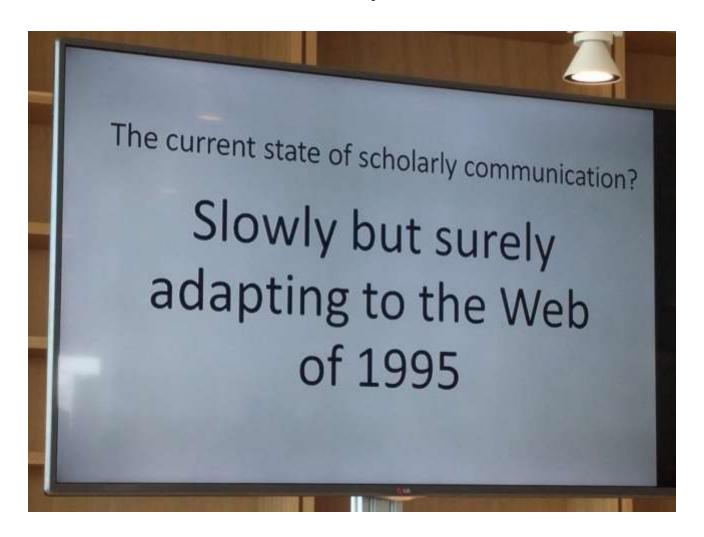




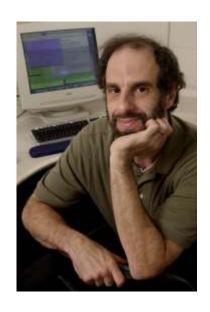
## 36 Years of My Career ...



# 20 Years of My Career ...



#### It's Not Like We Haven't Been Told



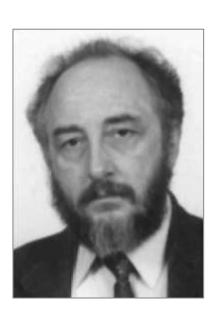
Paul Ginsparg

deconstruct discipline-centric



Stevan Harnad

deconstruct researcher-centric



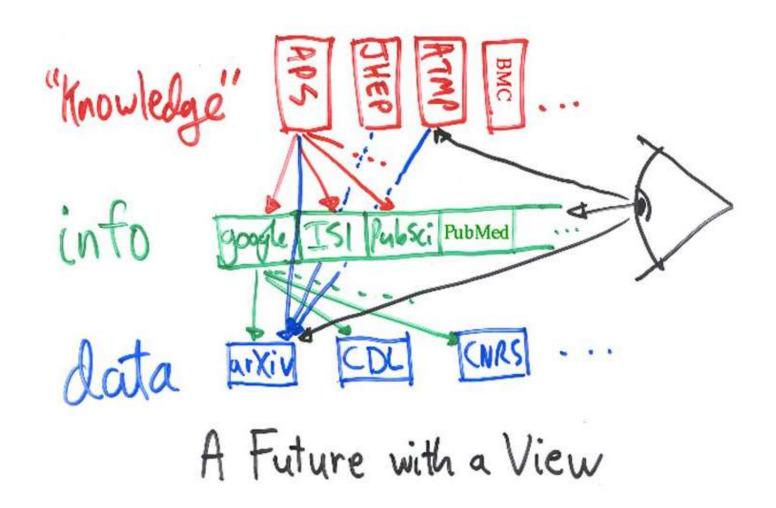
John W.T. Smith

deconstruct researcher/institution-centric





## "Decoupling Registration from Certification" Didn't Happen



### What Happened: Open Access

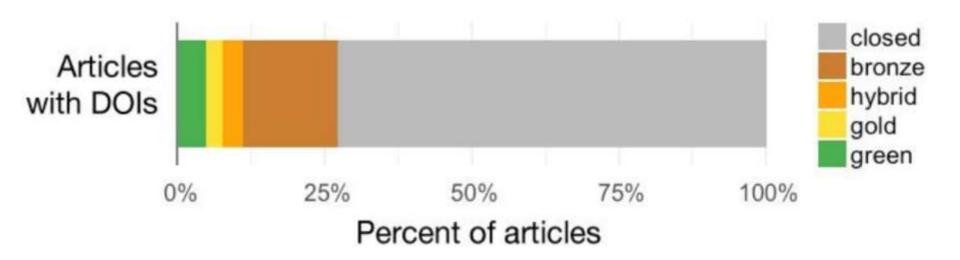
- No actual paradigm shift, but focus on business models to support access to reviewed articles without paywall
- Strategies include:
  - Author publishes article in open access, pays article processing charge
  - Author deposits article in institutional repository, access subject to publisher embargo (conditions vary)



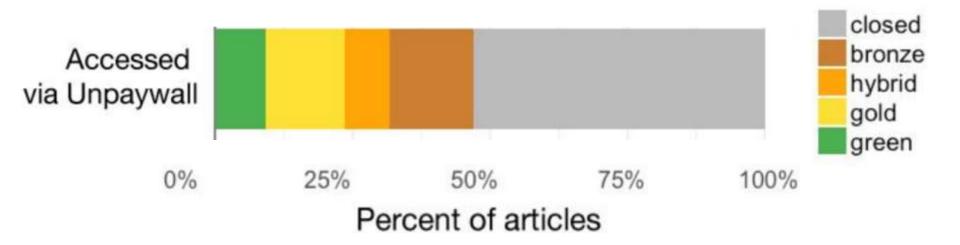




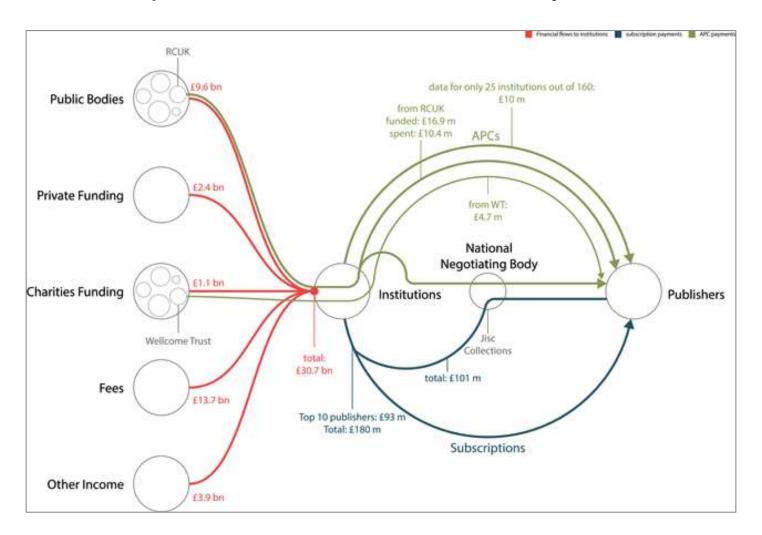
## Open Access: Improved Access



## Open Access: Improved Access

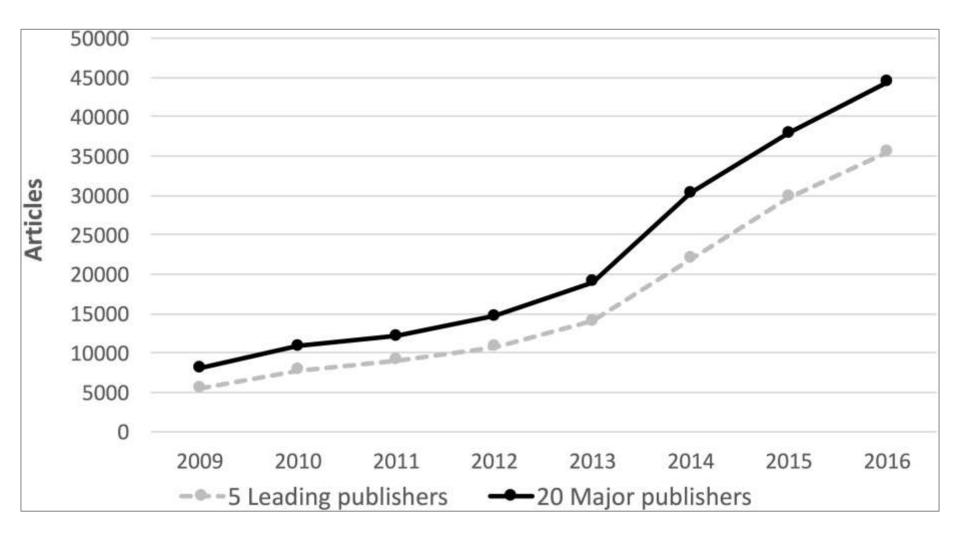


## Open Access: More \$ in the System



Lawson S., Gray, J., Mauri, M. (2016) Opening the Black Box of Scholarly Communication Funding. Open Library of Humanities. https://doi.org/10.16995/olh.72

# Open Access: More \$ in the System



Björk B. (2017) Growth of hybrid open access, 2009–2016. PeerJ. https://doi.org/10.7717/peerj.3878

# What Happened: Increased Control by Established Players

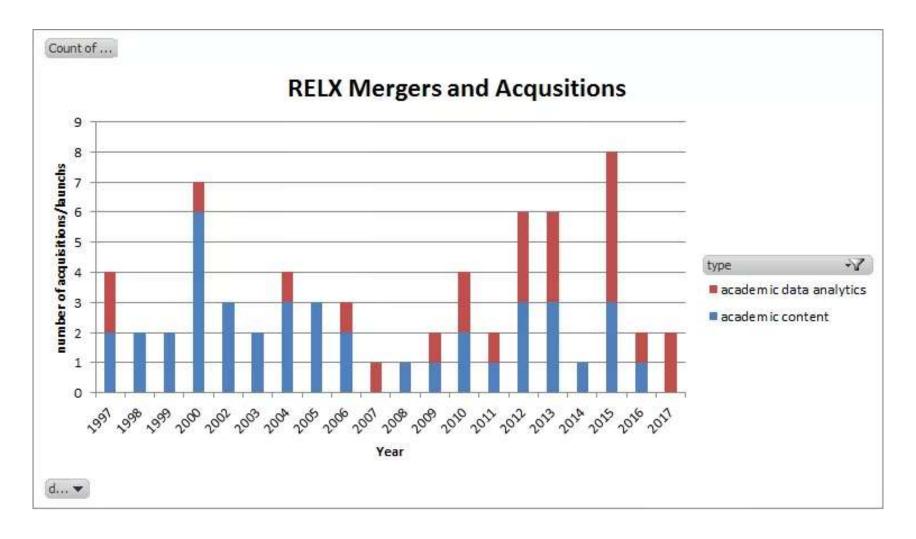
- Acquisition of innovative venues for publishing, analytics, research process (SSRN, bepress, publons, ...)
- Institutional repositories face increased competition from CRIS systems
- Action radius is extending into provision of solutions for the entire research life cycle, not just article publishing





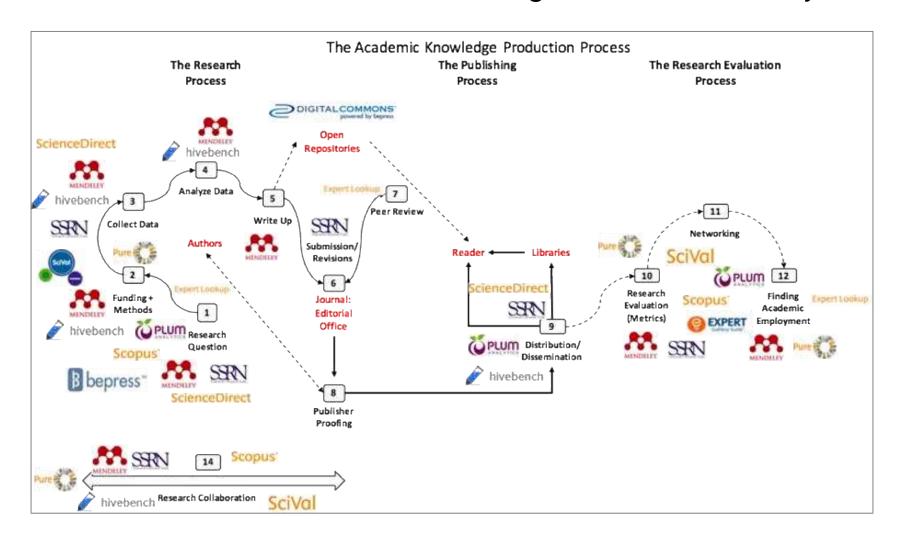


### Increased Control: From Content to Analytics



Posada, A., Chen, G. (2017) Publishers are increasingly in control of scholarly infrastructure and why we should care http://bit.ly/2xKRnSr

# Increased Control: From Publishing to Research Lifecycle

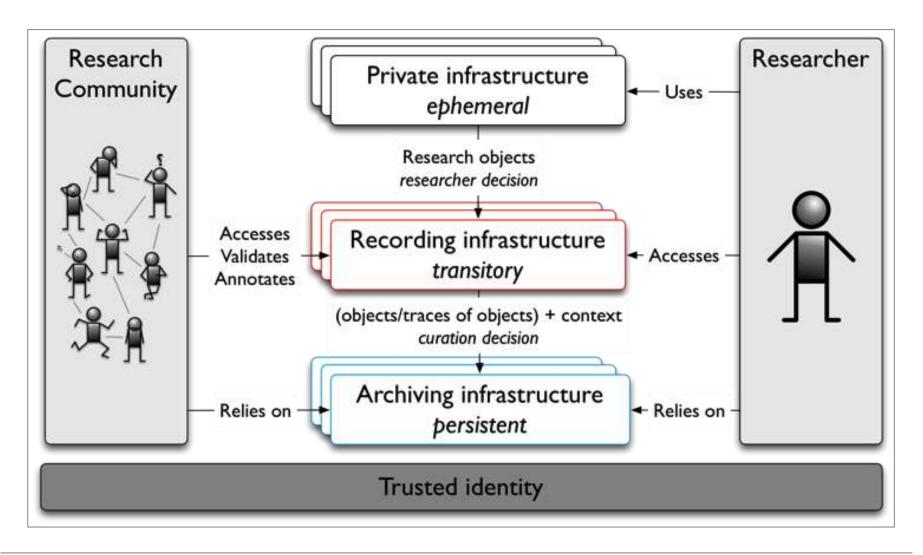


Posada, A., Chen, G. (2017) Publishers are increasingly in control of scholarly infrastructure and why we should care http://bit.ly/2xKRnSr

# What Happened: Research Artifacts Surfacing on the Web

- The research process, not just its outcome, is becoming visible on the web
- Dramatic extension of the scholarly record with a wide variety of artifacts
- These artifacts are hosted across the web, commonly on web platforms that are not dedicated to scholarship
- These platform record artifacts but typically don't archive them

# Recording is not Archiving



Herbert Van de Sompel & Andrew Treloar (2014) A Perspective on Archiving the Scholarly Web http://public.lanl.gov/herbertv/papers/Papers/2014/iPres2014\_Sompel\_Treloar.pdf

# How to Realize FORCE11 Scholarly Commons Principle 1?

P1. The scholarly commons is an agreement among knowledge producers and users

There is global commitment and participation in the commons' long-term viability and preservation

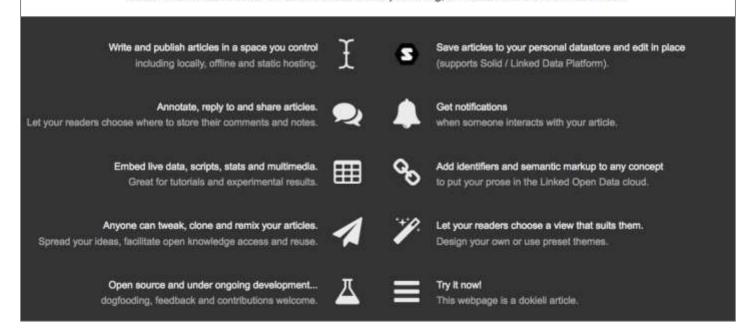
- Technical component
  - I formulate ideas inspired by my explorations of decentralized web approaches
- Organizational component
  - I provide pointers to ongoing thinking about a possible organizational foundation for a scholarly commons

## Technical Inspiration: Sarven Capadisli's dokie.li



### dokieli

dokieli is a clientside editor for decentralised article publishing, annotations and social interactions.



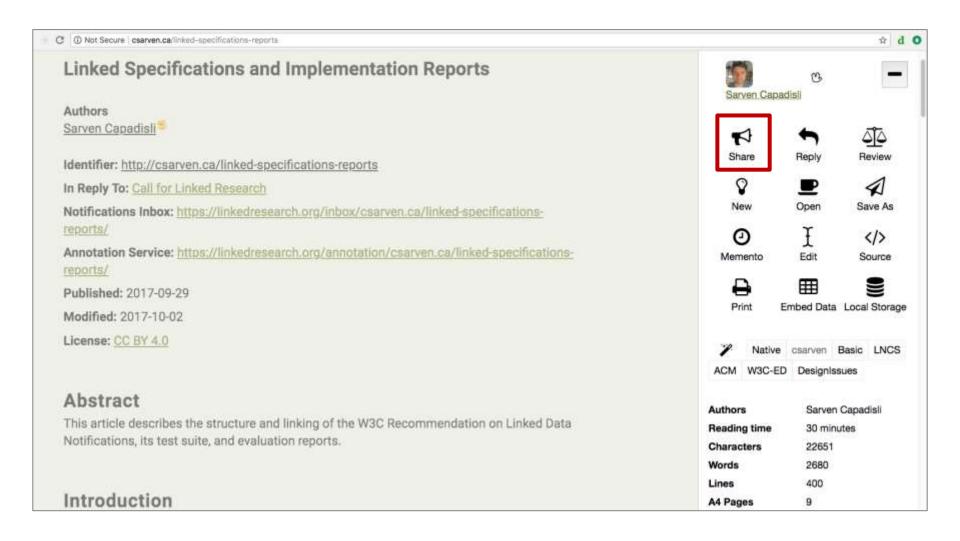
#### dokieli Features

- A client-side application that can perform CRUD operations with pods
  - Browser extension or embedded JavaScript
  - Supports a wide range of standards, e.g. WebID, FOAF, OpenID Connect, ACL, LDP, LDN, Web Annotation, AS2, CC, CiTO, PROV-O, schema.org, ...
- Application functionality:
  - HTML+RDF editor for decentralized article publishing, reviewing, annotating
  - A platform for social interactions, e.g. share, request review
- Application features include:
  - Human-friendly, multi-view documents
  - Machine-readable Linked Data documents
  - Inline semantic citations
  - Fine-grained URI-addressability
  - Authentication & Authorization based on ORCID OpenID Connect

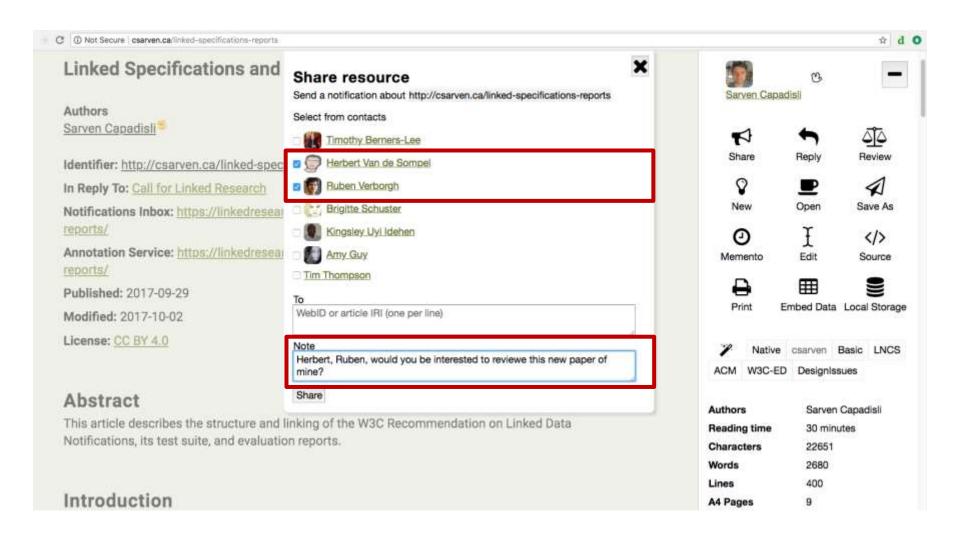




#### Send a Notification to Solicit a Review



#### Send a Notification to Solicit a Review



#### Receive the Notification that Solicits a Review

9b043210 db8b 11e7 a11d d38a32b8da8d.ttl license

type Announce

actor http://csarven.ca/#i
object linked specifications

reports

profile

1.0

summary Herbert, Ruben, would you

be interested to review this

new paper of mine?

to

updated 2017-12-07T20:17:09.017Z

linked specifications reports

license type

4.0 Post

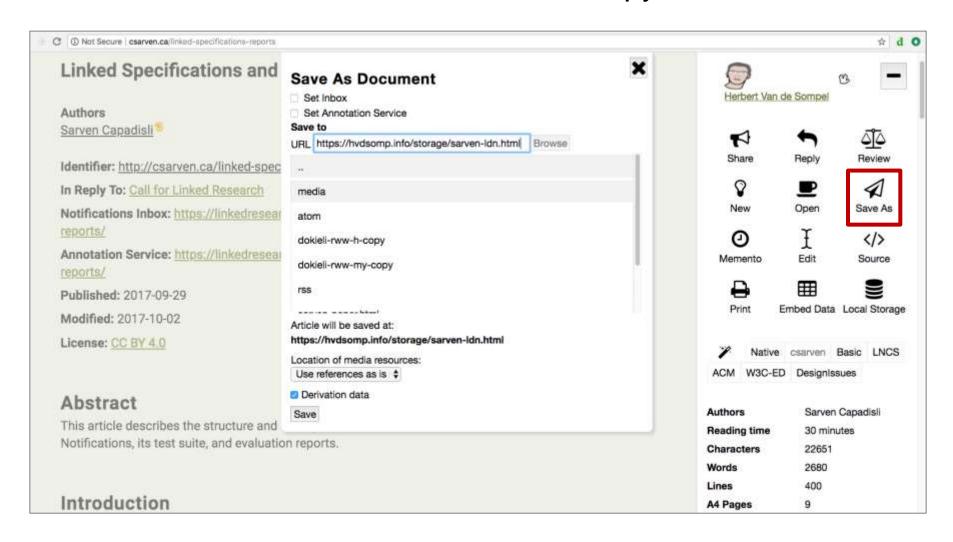
**Creative Work** 

Entity

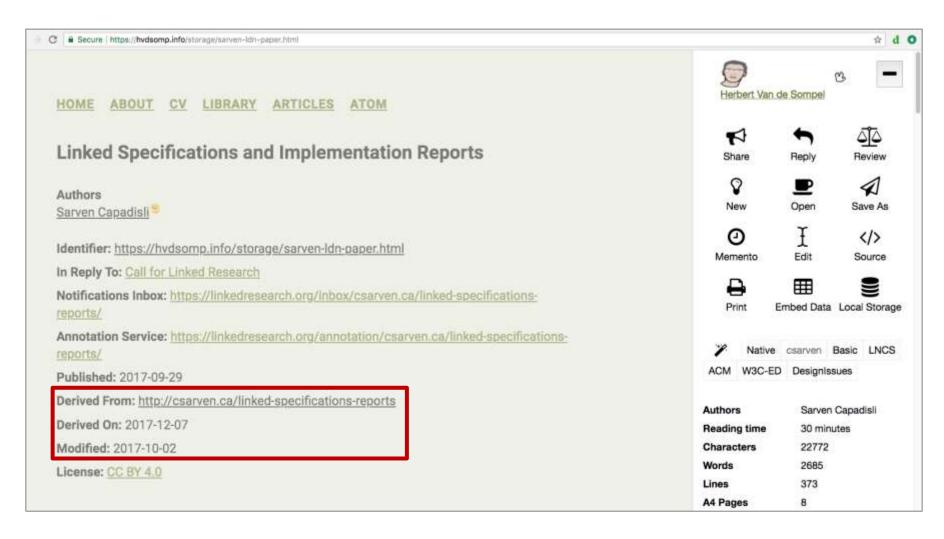
**Document** 

Article

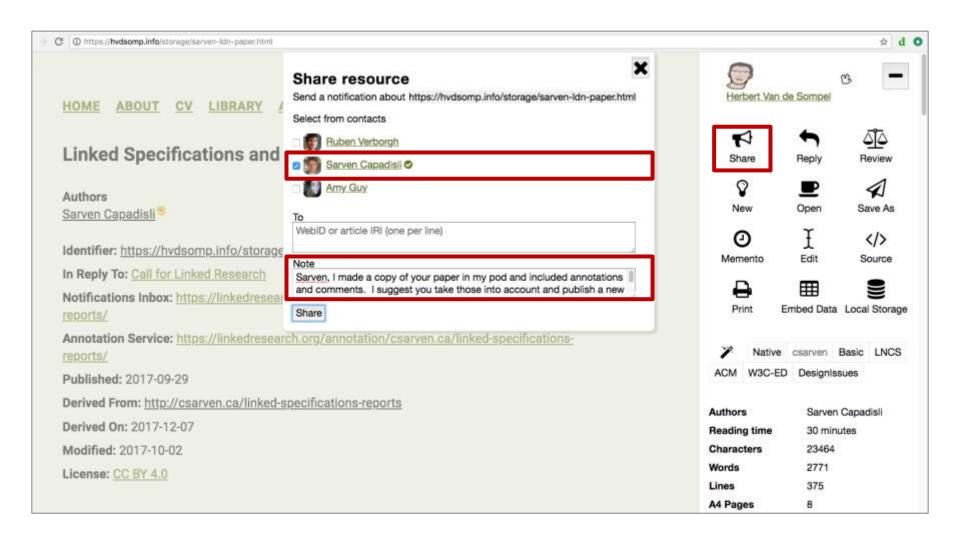
# Create a Derivation Copy



# Review the Derivation Copy



#### Send a Notification to Announce the Review



#### Receive the Notification that Announces the Review

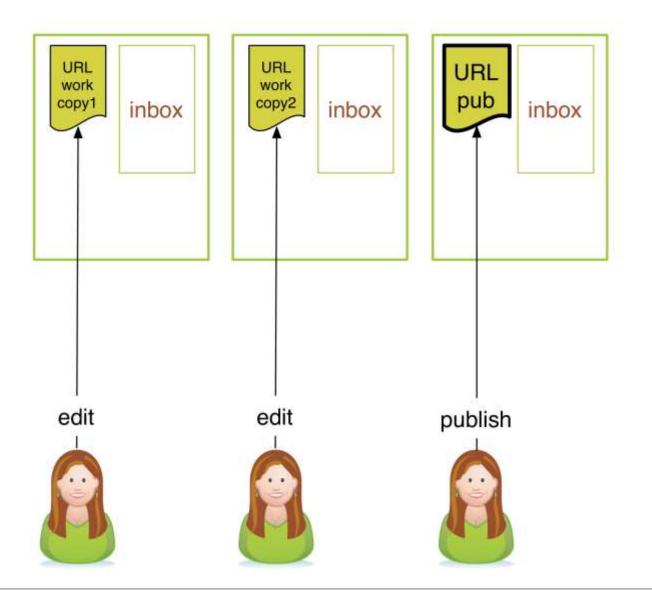
Types Announce Object https://hvdsomp.info/storage/sarven-ldn-paper.html Types http://schema.org/CreativeWork http://rdfs.org/sioc/ns#Post http://www.w3.org/ns/prov#Entity http://xmlns.com/foaf/0.1/Document https://www.w3.org/ns/activitystreams#Article License https://creativecommons.org/licenses/by/4.0/ Updated 2017-12-07 18:49:47 Summary Sarven, I made a copy of your paper in my pod and included annotations and comments. I suggest you take those into account and publish a new version. Actor http://orcid.org/0000-0002-0715-6126 License https://creativecommons.org/publicdomain/zero/1.0/ To http://csarven.ca/#i

## Beyond dokie.li ...

- dokie.li provides a glimpse of a possible technical approach for decentralized scholarly communication
- Concrete technical challenges arise related to aspects that are crucial from a scholarly communication perspective, including:
  - authentication & authorization issues with WebID/TLS, OIDC
  - versioning not supported by Linked Data Platform
  - collaborative writing
- Conceptual questions arise regarding how to fulfill the core functions of scholarly communication using this technical paradigm:
  - registration
  - certification
  - awareness
  - archiving



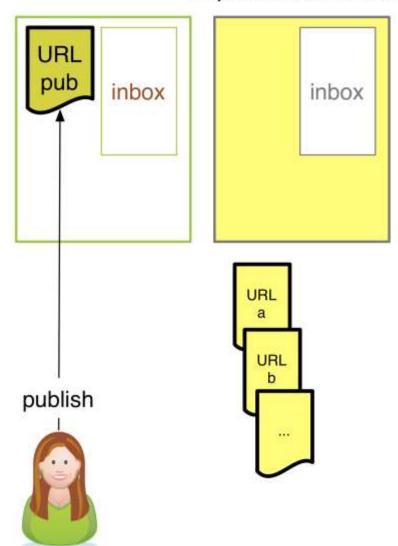






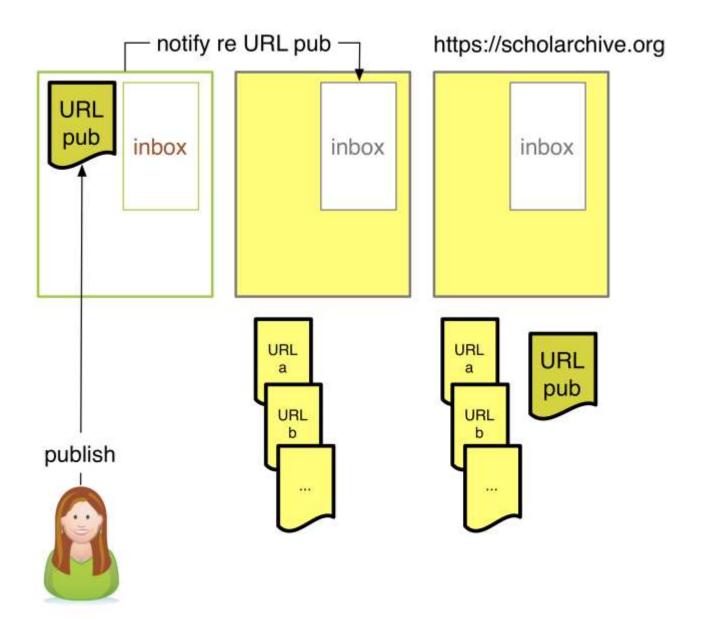


### https://scholarchive.org



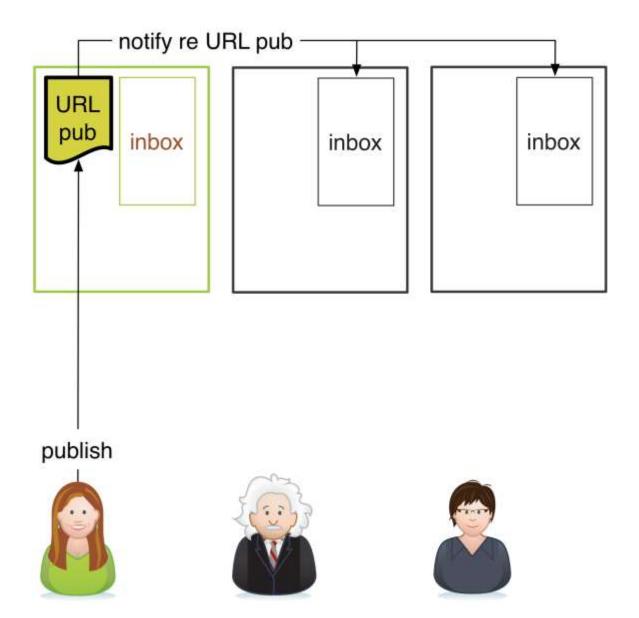






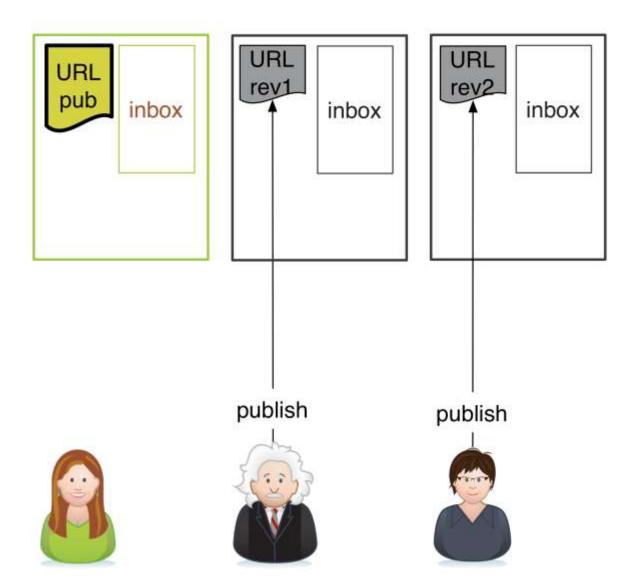






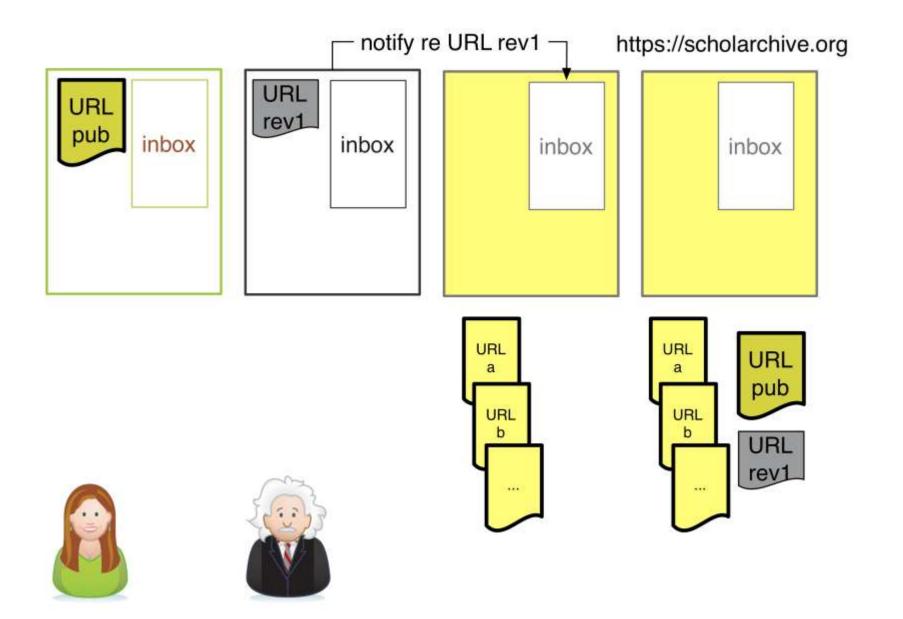






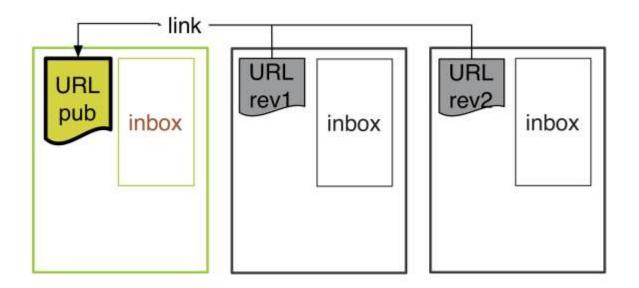












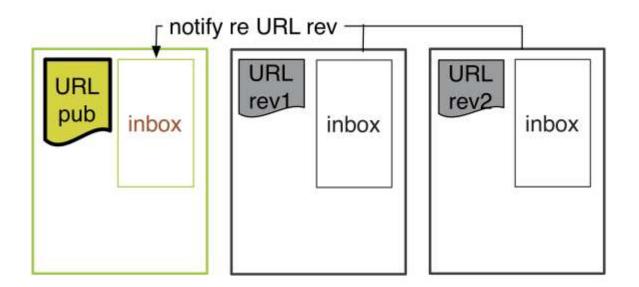












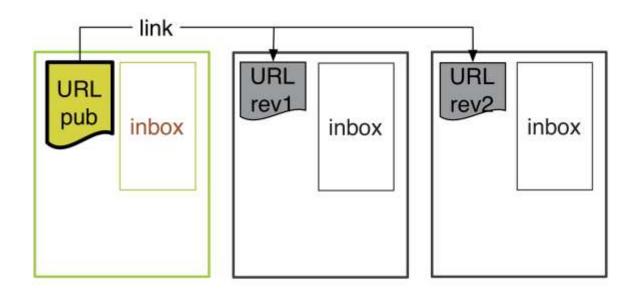












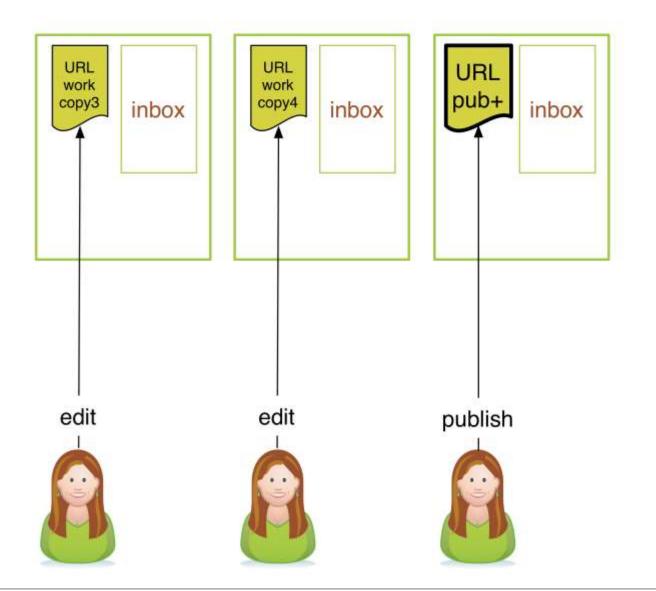








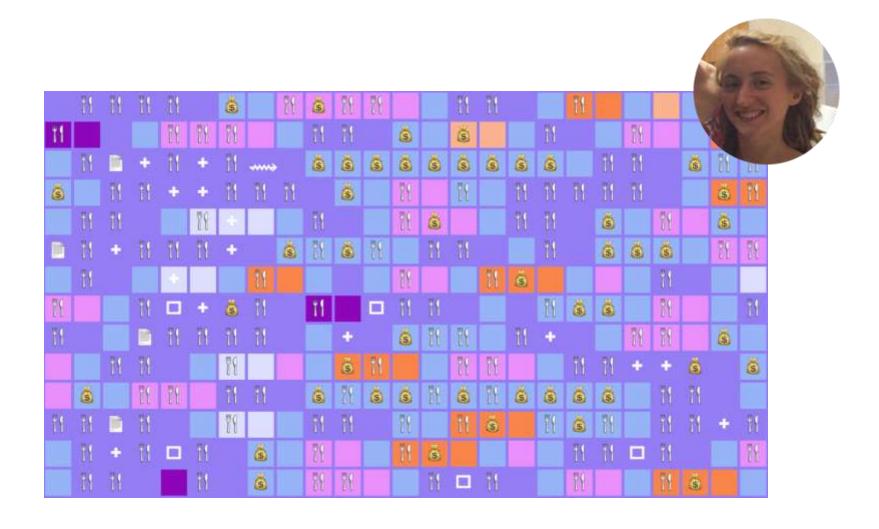




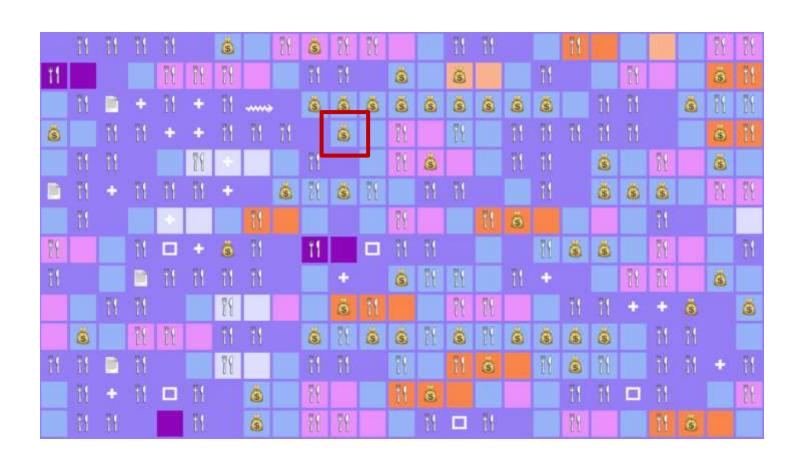




# Technical Inspiration: Amy Guy's sloph



#### Overview of Personal Activities



# A Personal Activity

Tuesday the 5th of September at 7:25pm (+01:00)



### Personal Web Observatory

- Records Amy's activities
  - Some by leveraging web portal APIs using "connectors"
  - Some by manual entry
- Observatory adheres to a variety of W3C standards e.g. webmention, Activity Streams, ActivityPub, Micropub, ...
- Activities recorded in Event Store
- Inspired by:
  - The Quantified Self movement, i.e. lifelogging, self-tracking
  - The Locker Project

that share the notion of aggregating a person's web traces into a personal web environment.

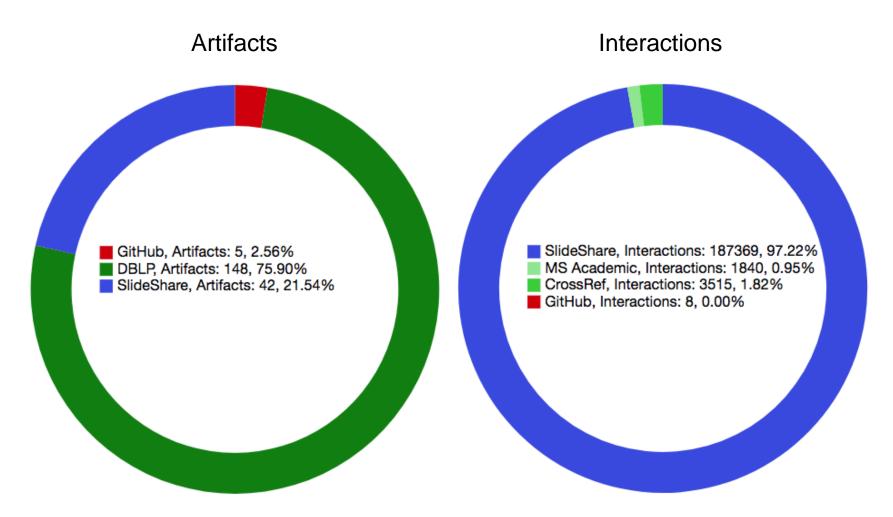
## Personal Scholarly Web Observatory

- Automatically tracks the researcher's web activities, including created artifacts, in a variety of portals
- Records metadata about scholarly artifacts that a researcher deposits in Web portals in a personal pod
- Records metadata about interactions that others have with those artifacts
- Event Store holds the metadata
- Activity metadata contains artifact URI, which is a seed for archiving the actual artifact





## My (utterly experimental) Scholarly Web Observatory







## My (utterly experimental) Scholarly Web Observatory

Interactions via MS Academic		
Artifact	Interactions	
co authorship networks in the digital library research community	440	
a principal component analysis of 39 scientific impact measures	264	
toward alternative metrics of journal impact a comparison of download and citation data	160	
clickstream data yields high resolution maps of science	123	
usage impact factor the effects of sample characteristics on usage based impact metrics	68	

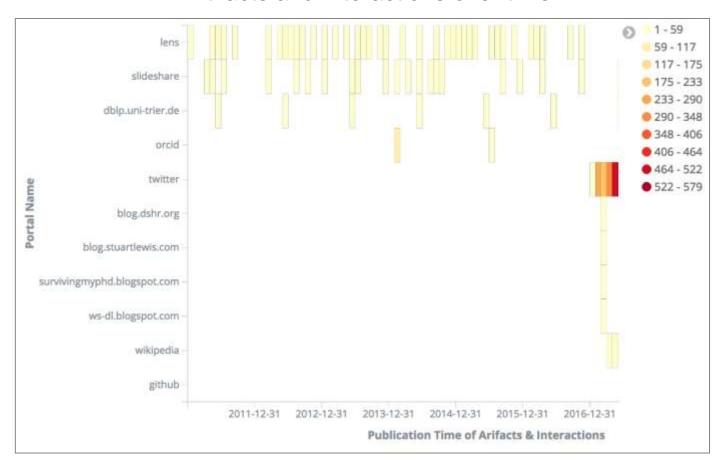
Interactions via SlideShare	
Artifact	Interactions
Memento: Time Travel for the Web	11825
Paint-Yourself-In-The-Corner Infrastructure	11662
An Overview of the OAI Object Reuse and Exchange Interoperability Framework	11209
Hiberlink: Investigating Reference Rot, December 2013	8884
An HTTP-Based Versioning Mechanism for Linked Data	7140





## My (utterly experimental) Scholarly Web Observatory

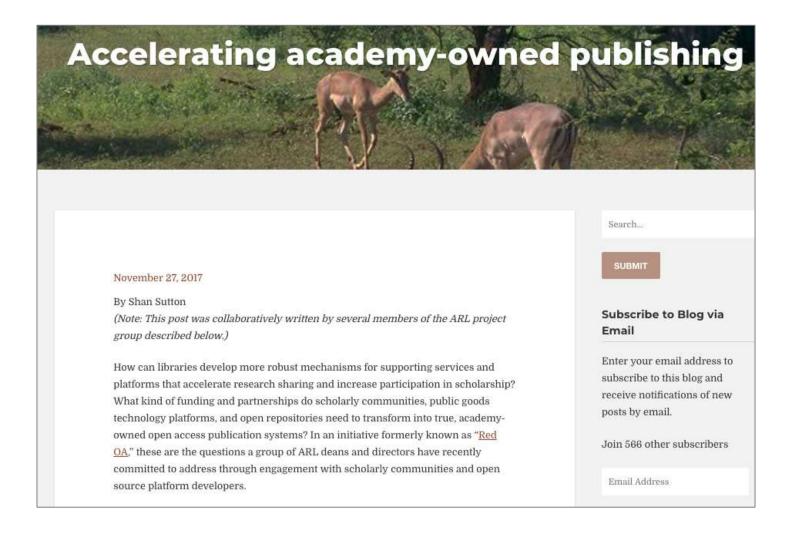
#### Artifacts and Interactions over time







## Organizational Inspiration: Shan Sutton et al.



Shan Sutton et al. (2017) Accelerating Academy-Owned Publishing http://intheopen.net/2017/11/accelerating-academy-owned-publishing/

## Organizational Inspiration: David W. Lewis

#### The 2.5% Commitment

David W. Lewis

Dean of the IUPUI University Library
dlewis@iupui.edu

September 11, 2017

© 2017 David W. Lewis. This work is licensed under a <u>Creative Commons Attribution 4.0</u> International license

**The Commitment:** Every academic library should commit to contribute 2.5% of its total budget to support the common infrastructure needed to create the open scholarly commons.

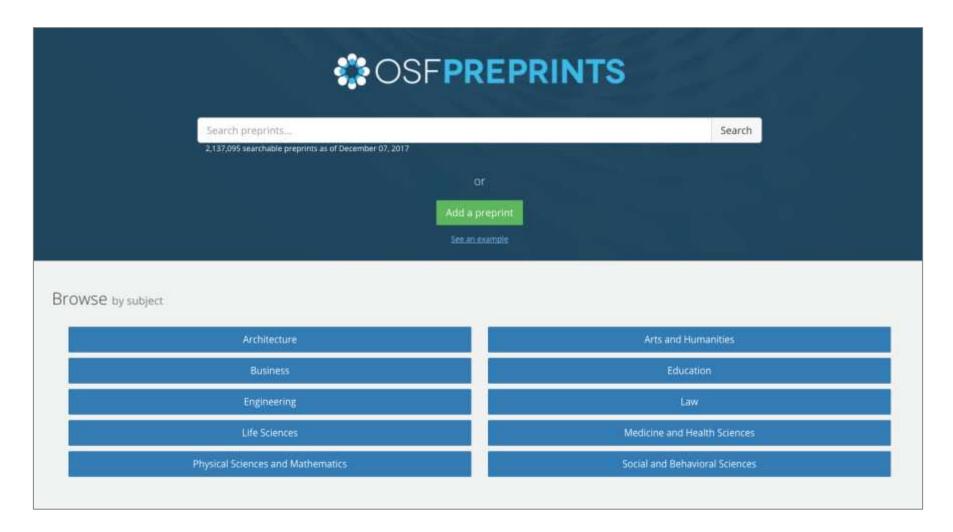
## Organizational Inspiration: Björn Brembs

the American ARPANET took shape thanks to well-managed state subsidies and collaborative research environments and the Soviet network projects stumbled because of unregulated competition among self-interested institutions, bureaucrats, and others.

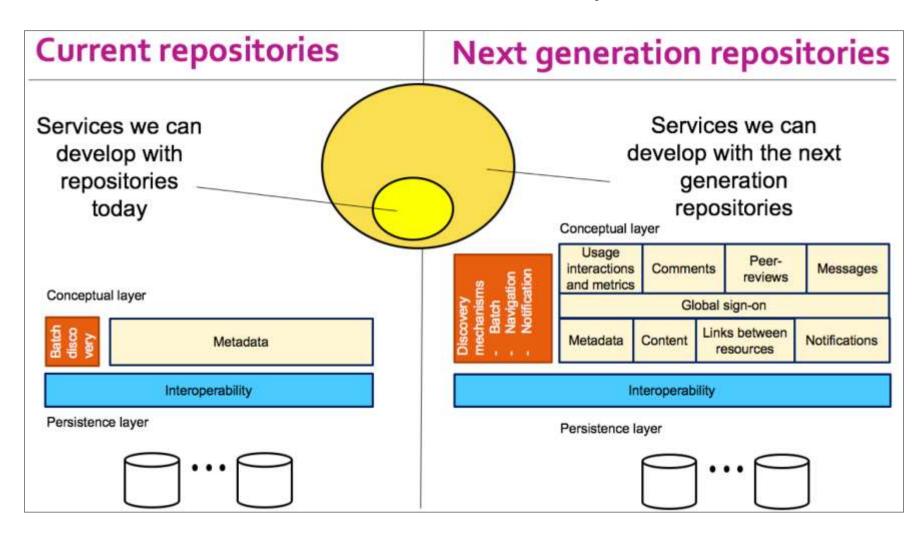
We need the same collaborative spirit if institutions are to abandon subscriptions, just as they cooperated to spend money to draw cables between institutions, even though they were separated by borders and even oceans. If in the 1980s, our institutions collaborated across nations and continents to spend money on a research infrastructure nobody yet knew, can't they collaborate now to save money being wasted on an obviously outdated infrastructure? Has neoliberal worship of competition poisoned our academic institutions to such a degree, that within 25 years they went from cooperating even if it means spending money to never cooperating even if it means saving money? I refuse to believe that, even though that's what some try to tell me.

Instead of trying to tell scholars to behave ethically in spite of the downsides to them personally, maybe we ought to tell institutions that our infrastructure is outdated and that we need the functionalities everybody else but academics are enjoying? We need to get the same mechanisms going that in the 1980s got our universities to invest in networking hardware and cables, despite a functioning telephone and snail mail system. Cancelling subscriptions doesn't mean losing access, so nobody can tell me that canceling subscriptions is more difficult than installing compatible networking hardware across the globe. I'm now paying for my phone bills out of my budget, while my skype calls are provided by my university. Maybe rather than trying to convince scholars to choose the ethical over the advantageous, it would be more effective to ask our institutions to provide us with modern technology and have those who still want to use the legacy version pay for it themselves?

## Path 1: Discipline-Oriented Preprint Servers

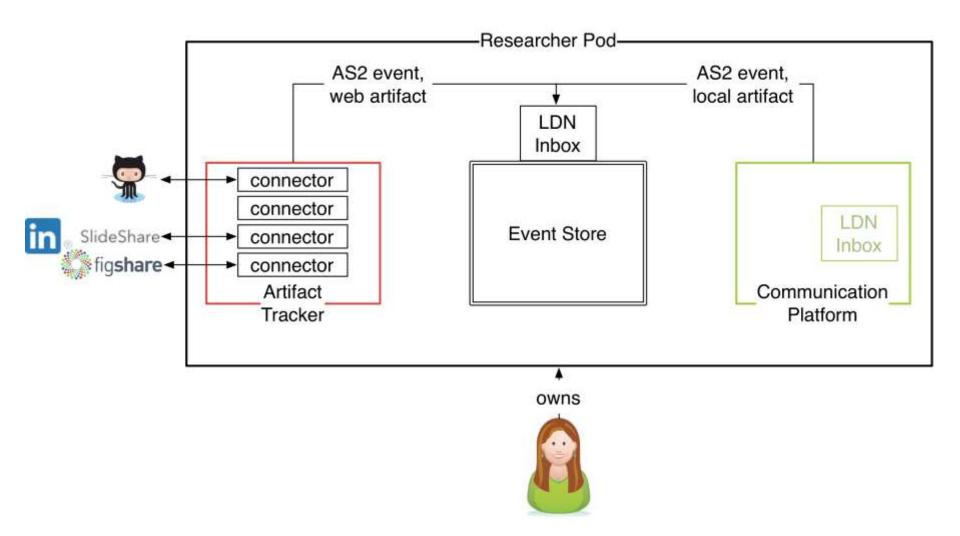


Path 2: Institution-Based Repositories



COAR Next Generations Repositories Working Group (2017) Behaviours and Technical Recommendations https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf

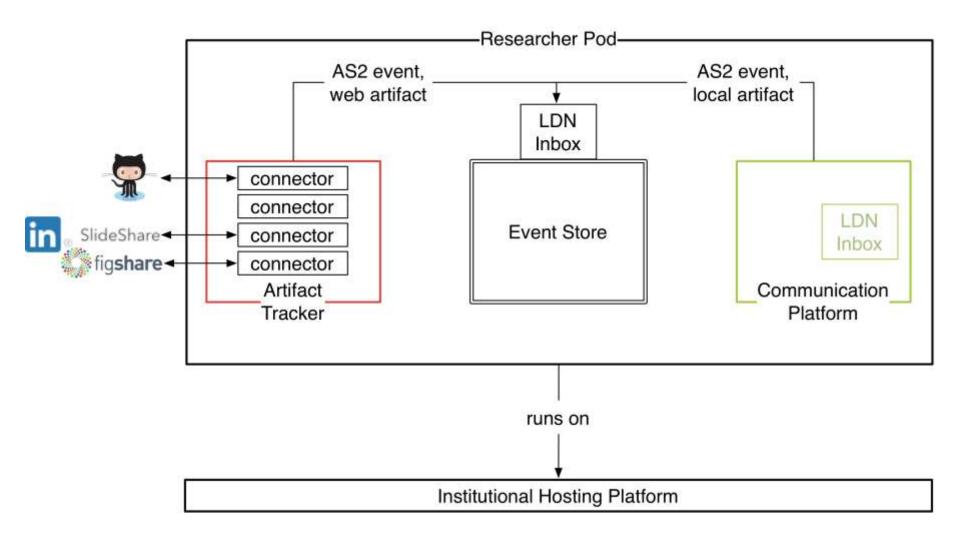
Path 3: Researcher Pod ???







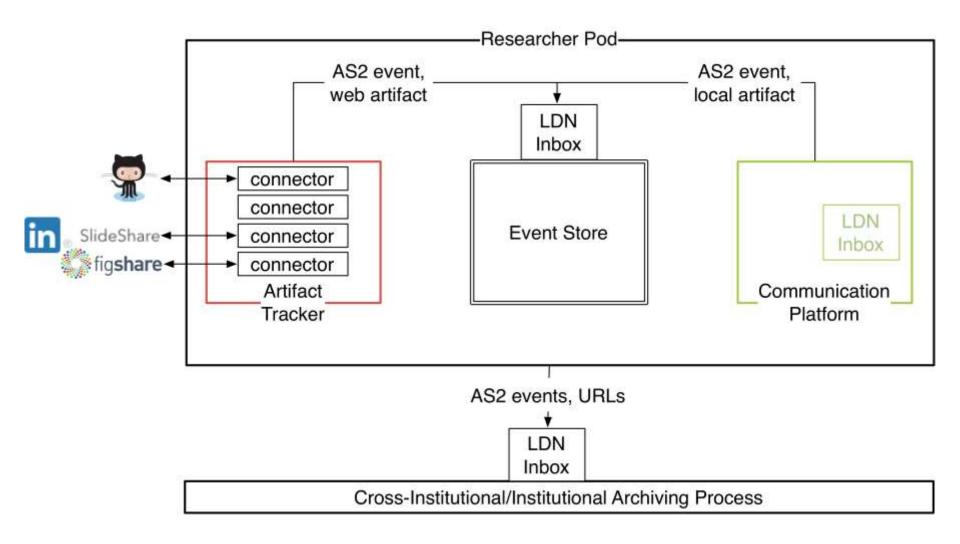
## Researcher Pod: Institutional Hosting







## Researcher Pod: Cross-Institutional/Institutional Archiving







### Contributor Centric

- The researcher has a personal domain for scholarly communication
  - Not institution bound, but truly personal
- A researcher pod for scholarly activities operates in the researcher's domain
  - The researcher takes responsibility for publication and review
  - David Rosenthal: "Persistence through Self-Interest"
- The researcher pod has three functions:
  - self-publishing: the researcher's papers, overlay reviews, comments, annotations, ...
  - notifying: the discipline-oriented social network glue
  - self-tracking: keep track of the researcher's artifacts in web portals





### Contributor Centric

- A researcher pod provides a comprehensive overview of the researcher's contributions and interactions of peers with those contributions
- The researcher pods are globally interoperable via support of a suite of standards
  - the essential standards exist
  - the tools are currently brittle (understatement)





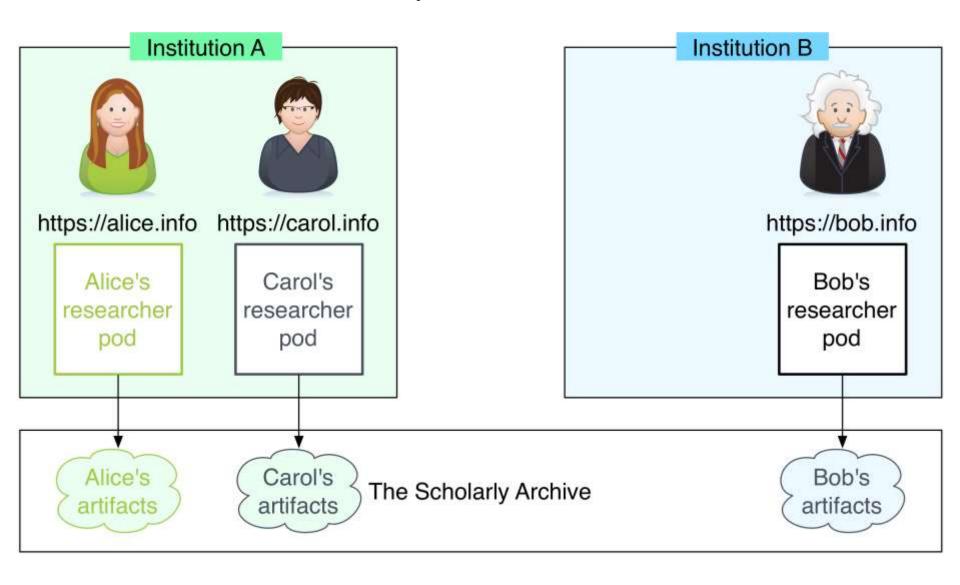
#### Institution Enabled

- Institutions provide a hosting platform and operational support for their researchers' pods
  - a researcher's domain resolves to the institutional platform
- Institutions archive their researchers' contributions by monitoring their pods and capturing new resources
  - Aggregation of institutional scholarly web archives?
  - Cross-institutional web archive, cf. LOCKSS?
  - Cross-institutional IPFS/DAT archive (HTTP bridge essential)?

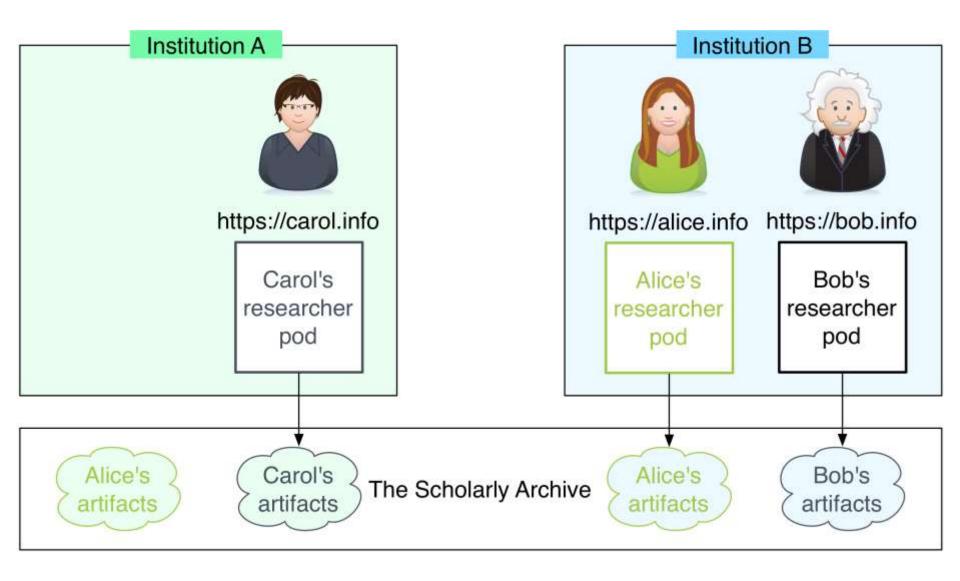




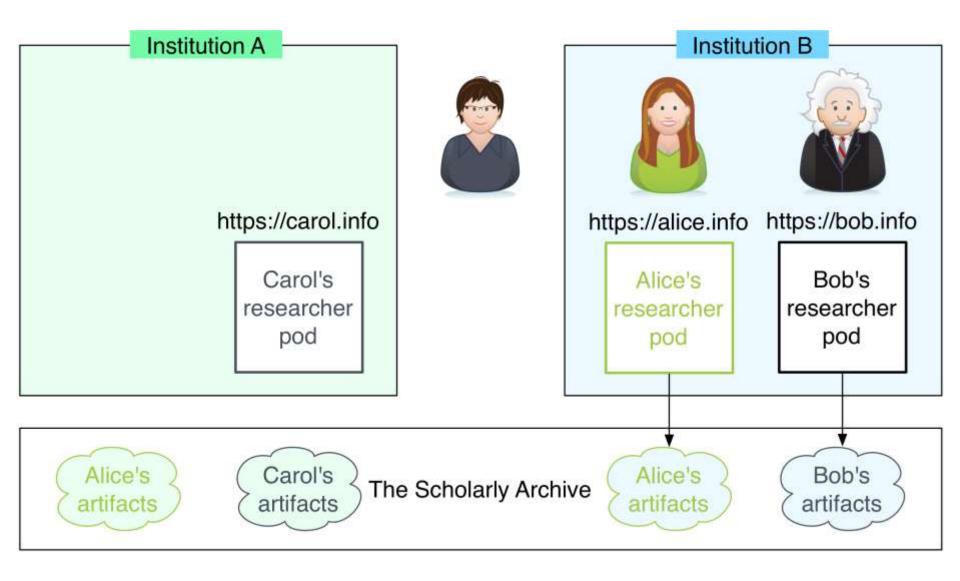
## Mobility & Persistence



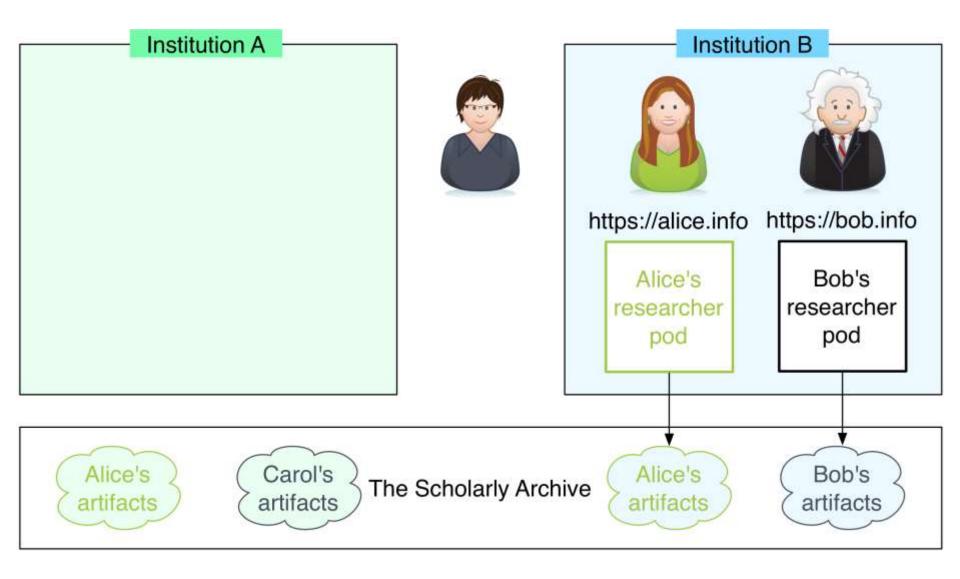
## Mobility & Persistence: Alice Changes Institution



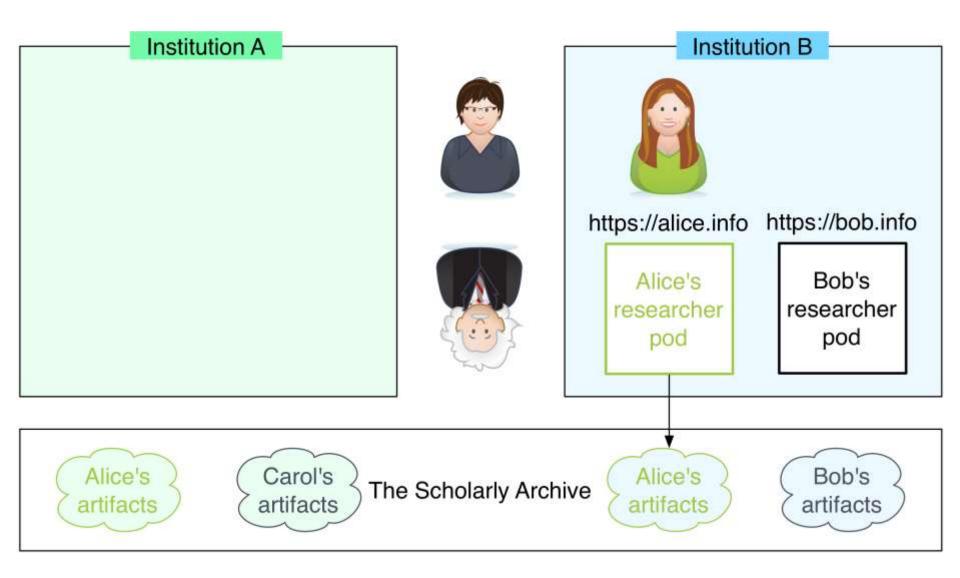
## Mobility & Persistence: Carol Leaves Academia



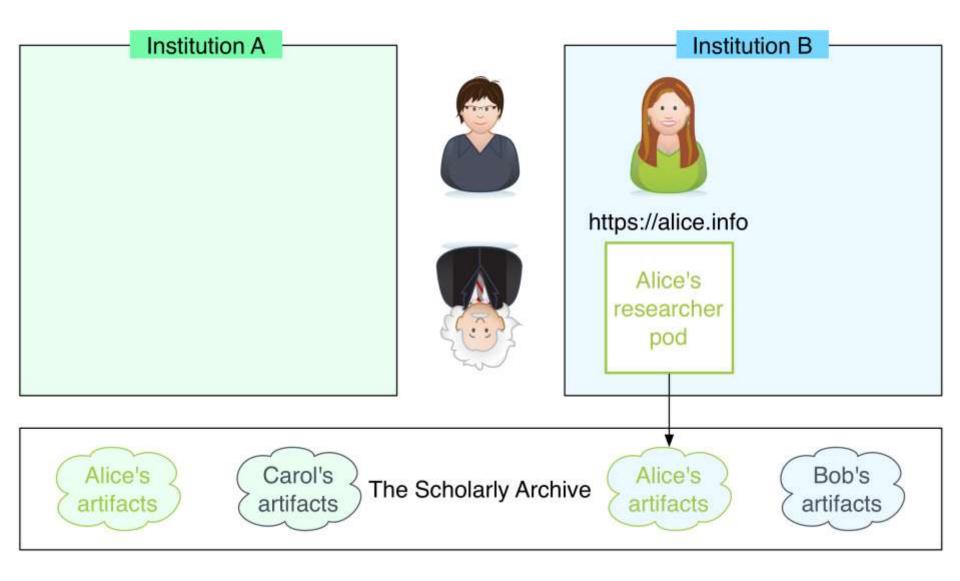
## Mobility & Persistence: Carol Leaves Academia



## Mobility & Persistence: Bob Passes Away



## Mobility & Persistence: Bob Passes Away



## Mobility & Persistence

- Persistently linking to artifacts:
  - Alice moving from A to B: She holds on to her domain, links keep working
  - Carol and Bob give up on their pod: Technologies exist to make links to their artifacts lead directly into the archive, e.g. PIDs, Robust Links





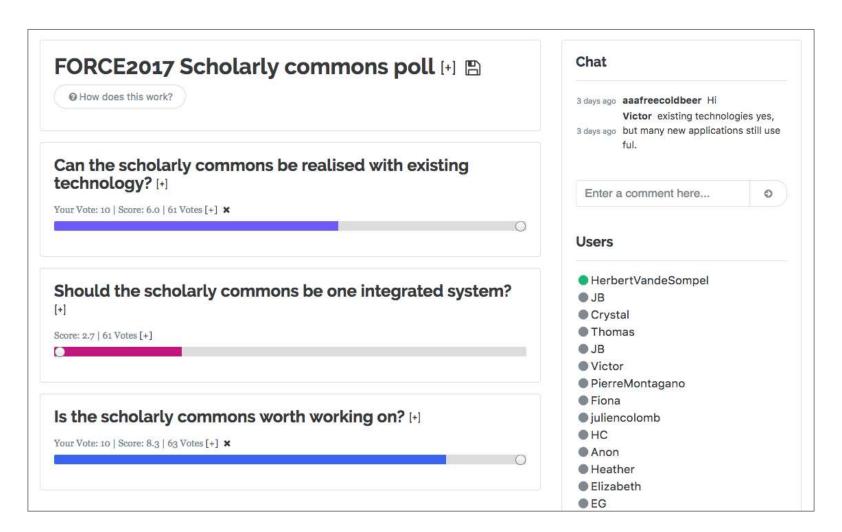
## In Summary ...

- Think of researcher pods as a different kind of institutional repository
  - contributor-centric instead of document-centric
  - disciplinary component provided by social network functionality
  - built-in portability
  - archive-ready





## In Closing ...



# Scholarly Communication: Deconstruct and Decentralize?

A Bold Speculation Featuring







Bob



Carol

Herbert Van de Sompel

<a href="mailto:ohvdsomp">ohvdsomp</a>
Los Alamos National Laboratory



