



LOTS OF COPIES KEEP STUFF SAFE

Distributed Digital Preservation with LOCKSS

Nicholas Taylor ([@nullhandle](#))
Program Manager, [LOCKSS](#) and [Web Archiving](#)
[Stanford Libraries](#)

[Massive Storage Systems and Technology](#)
14 May 2018

overview

- LOCKSS background
- preservation principles
- distributed preservation
- what's next for LOCKSS?



"LAX on take off" by Doug under CC BY-NC-ND 2.0



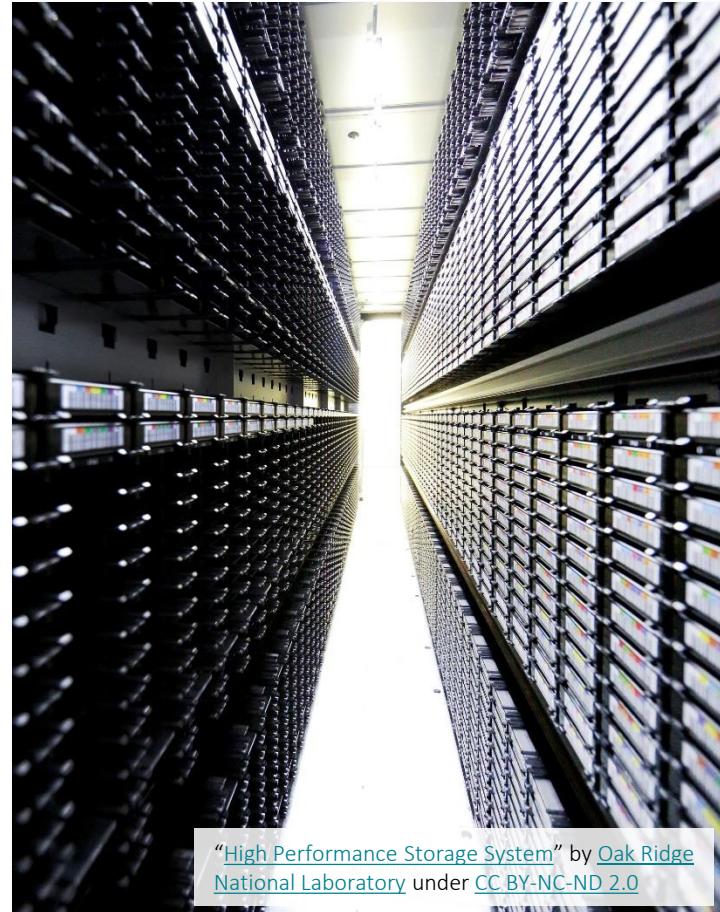
LOCKSS Background

"[Padlocks I](#)" by [Domiriel](#) under [CC BY-NC 2.0](#)

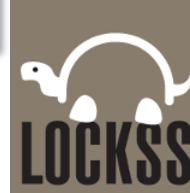
(digital) libraries



[“Boston University - Mugar Library”](#) by
[Gino Roncaglia](#) under [CC BY-NC-SA 2.0](#)



[“High Performance Storage System”](#) by [Oak Ridge National Laboratory](#) under [CC BY-NC-ND 2.0](#)



lots of copies (were already) keeping stuff safe

- print journal holdings
incidentally resilient:
 - distributed
 - decentralized
 - irrevocable
 - tamper-evident
 - publisher-independent



"[serials](#)" by [Timothy Vollmer](#) under [CC BY 2.0](#)

move to online publishing

- own → lease
- local custody → contingent access
- lots of distributed copies → fewer, centralized copies
- net effect:
 - endanger scholarly record
 - obviate library role

POSTMODERN
CULTURE
an interdisciplinary journal
of interdisciplinary
criticism

Volume 1, Number 1 Fall, 1990

Editor: Greg Amerson
Sarah Orr
John Unsworth, Issue Editor

Editorial Assistant: Gloria Maxwell

Editorial Board:

Mathy Acker	Chivalum Nwankwo
Shane Bennett	Patricia O'Donnell
Michael Beraha	Susan Oliver
Marc Chenevier	John O'Farrell
Gregory Curran	Gregory Curran
R. Serge Denisoff	Marjorie Perloff
Robert Dohler	Carl Raschke
Jim English	David Rasmussen
Georges Gartner, Jr.	Ariatal Russell
Robert Hodges	Andrea Schepers
Ivan Illich	George Ruffinelli
Susan Howe	Susan M. Schultz
E. Ann Homann	Mark Stetson
Neil Larsen	Tony Stewart
Jameson Lycan	Gregg Strawhecker
Layne Mykynen	Chris Strayayes
	Greg Ulmer

CONTENTS

Harthead & Table of Contents	CONTENTS 990
Preface	PREFACE 990
Andrew Ross, "Sucking Away at the Counter-culture"	ROSS-1 990
	ROSS-2 990
Bell Hooks, "Postmodern Blackness"	HOOKS 990
Luca Kipnis, "Mars: The Video (A Politics of Political Violence)"	KIPNIS 990
George Yudice, "Feeding the Transcendent Body"	YUDICE 990
Kathy Acker, "Dead Doll Humility"	ACKER 990
John Beverley, "The ideology of postmodern music and left politics"	BEVERLEY 990
Neil Larsen, "Postmodernism and Imperialism: Theory and Politics in Latin America"	LARSEN 990

Features:

Susan M. Schultz, "Voicing the Menace" [a review of "Postmodern Poetics"]	REVIEW 990
J. Richard Green, "The Politics of Language: Roots From North America and the United Kingdom," _Notes_ 1-1 (Spring, 1990): 6-19.	REVIEWS 990
Jim English, "Vacation Notes: Haute-Tech in the Haute-Montagnes" [Popular Culture Column]	POP-CULT 990
The Editors, "Surfaces: Positions on Postmodernism"	POSTFACE 990

Announcements & Advertisements [WWW Version only]

Special thanks to Greg Ulmer for his assistance in starting the journal and preparing the first issue, and to Chuck Reiter, systems programmer, for technical support. We would also like to thank the NCSS Social Sciences Computing Lab, the NCSS Social Sciences Computing Lab, and the Compaq Computer Corporation for their support.

Unless otherwise noted, copyrights for the texts which comprise this issue of *Postmodern Culture*, are held by their authors. The compilation as a whole is Copyright (c) 1990 by *Postmodern Culture*. All rights reserved.

 IATH WWW Server Postmodern Culture: Volume 1, Number 1 under Fair Use
Last Modified:



LOCKSS but for e-journals

- p2p software for e-journal preservation
- restore preservation features of print journal holdings for digital
- re-empower libraries, individually + communally
- improve durability of digital scholarly record



Permanent Publishing on the Web

NEWS: October, 2002
Stanford University Libraries have received a planning grant from the National Science Foundation to explore the potential applicability of the LOCKSS technology to preservation of US Federal government information. LOCKSS (Lots of Copies Keeps Stuff Safe) software allows individual libraries to take custody of stable content in all formats delivered via HTTP and safeguard their community's access. This NSF Small Grant for Exploratory Research provides \$50,000 to assist the LOCKSS team and eight content partners drawn from the government documents community explore the technical, economic, social and legal viability of various LOCKSS architecture models for their GPO depository program. The LOCKSS partner group includes librarians in the following organizations: US Government Printing Office, Stanford University, University of Minnesota, California Digital Library, US National Agricultural Library, University of Colorado, California State University, San Bernardino and Saint Charles [Missouri] City-County Library District. For further information contact Chuck Eckman (eckman "at" stanford "dot" edu) or Vicki Reich (vreich "at" stanford "dot" edu).

[Demo of the User Interface](#) | [Live Map of Worldwide Beta Sites](#)

Project Descriptions

[Brief/Medium](#)
[Frequently Asked Questions](#)
[Press Releases & News Articles](#)
[Presentation](#)

Technical Specifications

[Published Papers](#)
[Security](#)

Project Status

[Beta Test Participants - Libraries and Endorsing Publishers](#)
[Beta Test Release & Testing Status](#)

Software

[License](#)

Credits

[Funders & Technical Wizards](#)

Related Work

[Other Archive Projects](#)
[References](#)

Contact

[Vicky Reich](#), vreich "at" stanford "dot" edu

LOTS OF COPIES KEEP STUFF SAFE™

Internet Archive: "LOCKSS | Lots of Copies Keep Stuff Safe"



LOCKSS for more than e-journals

- set out to build e-journal preservation system
- ended up building generic digital preservation core
- growing number of communities use to preserve other digital materials



THE ALABAMA DIGITAL PRESERVATION NETWORK
PRESERVING ALABAMA'S DIGITAL RESOURCES



community + digital preservation

- communities complement LOCKSS:
 - resilience against organizational failure
 - native heterogeneity
- preservation is an active community effort
- lots of communities keep stuff safe



"Redwood Canopy" by Floyd Stewart under CC BY-NC-SA 2.0



Preservation Principles

"Verbeke Foundation" by [Expectmehr](#) under [CC BY-NC-SA 2.0](#)

lots of copies

- intuitive **best practice**
- LOCKSS typically operates w/ **4+ copies**
- **enlist copies to attest** to expected integrity value
- lots of copies **enables**:
 - majority **votes w/ minority** of participating copies
 - **higher-confidence attestations** via landslide agreement



"Untitled" by Craig Donachy under CC BY-NC-ND 2.0

routine audit + repair

- ensuring **long-term bit integrity**
 - must **read data** to know it's good
 - easier to **repair data** sooner
- network nodes **conduct polls** to validate integrity of distributed copies
- more nodes = **more security**
 - more nodes can be **down**
 - more copies can be **corrupted**
 - ...and polls will **still** conclude
- nonces **force re-hashing**
- peers are **mutually-distrusting**



"DSC_4346" by Dennis Jarvis under BY-SA 2.0

fail slowly

- fast-operating, tightly-coupled systems **fail quickly**
- LOCKSS is **conservative + sophisticated** about repairs
- polls run slow to enable **detection + mitigation** of cause of damage



[“Galápagos giant tortoise on Santa Cruz Island”](#)
by [John Solaro \(sooolaro\)](#) under [CC BY-ND 2.0](#)

threat model

- familiar threats:
 - media + hardware obsolescence
 - software obsolescence
 - natural disaster
- more typical threats:
 - economic failure
 - organizational failure
 - operator error
- security threats:
 - internal attack
 - external attack



distributed + decentralized

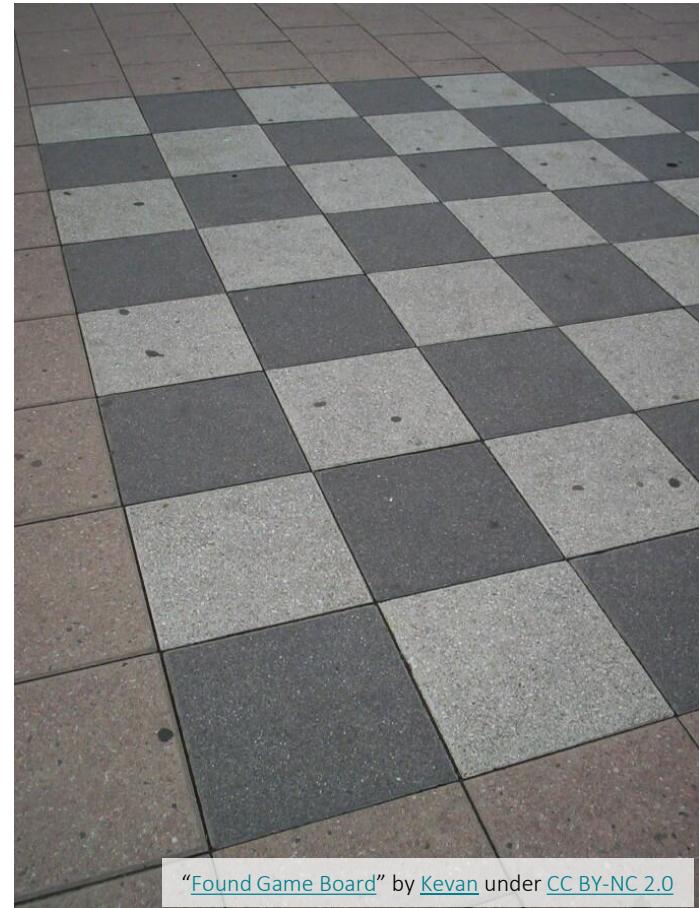
- no monopoly on copy-making
- more copies doesn't mitigate **correlated risk**
- independent, **de-correlated copies**
- minimize central points of failure or vulnerability



“Domino’s” by [david pacey](#) under [CC BY 2.0](#)

no centralized fixity store

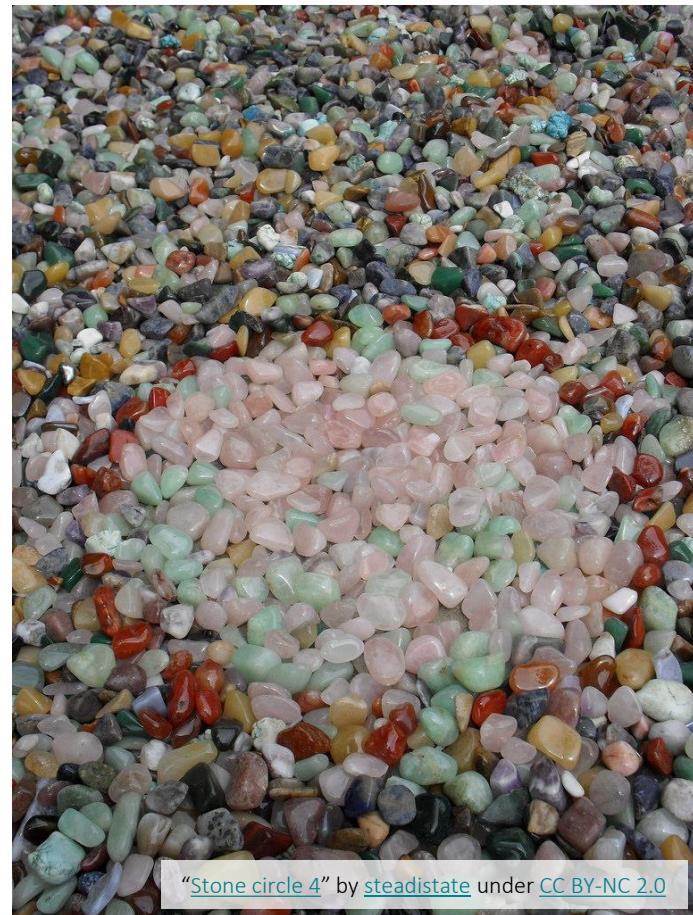
- fixity data **subject to same threats** as data whose integrity it assures
- fixity data is **more vulnerable**, in fact, since more valuable + more centralized
- LOCKSS uses fixity data in **limited ways**



["Found Game Board"](#) by [Kevan](#) under CC BY-NC 2.0

local custody

- if preserving data is **core to mission**, LOCKSS helps maintain that competency + commitment in-house
- **unencumbered access** for use by designated community
- **conserving autonomy** + leverage w/ content + service providers
- **jurisdictional transparency** + control



"Stone circle 4" by [steadystate](#) under CC BY-NC 2.0

Distributed Preservation

[“Catho longtime \[explored\]”](#) by Bill Collison under [CC BY-NC 2.0](#)

where does distributed preservation fit?

- may be **integrated into own infrastructure** (e.g., offsite replication)
- as a **wholly hosted service**:
 - for some, may be **main preservation solution**
 - for others, may **supplement local preservation**



[“Stone stacks”](#) by [Jack Malvern](#) under [CC BY-NC 2.0](#)

use cases

- scholarly record
- government documents
- web archives
- collaborative collections
- any types of content
valued in common by a community



"147-Smartie Colour Wheel" by Michael Scott under CC BY-NC-ND 2.0

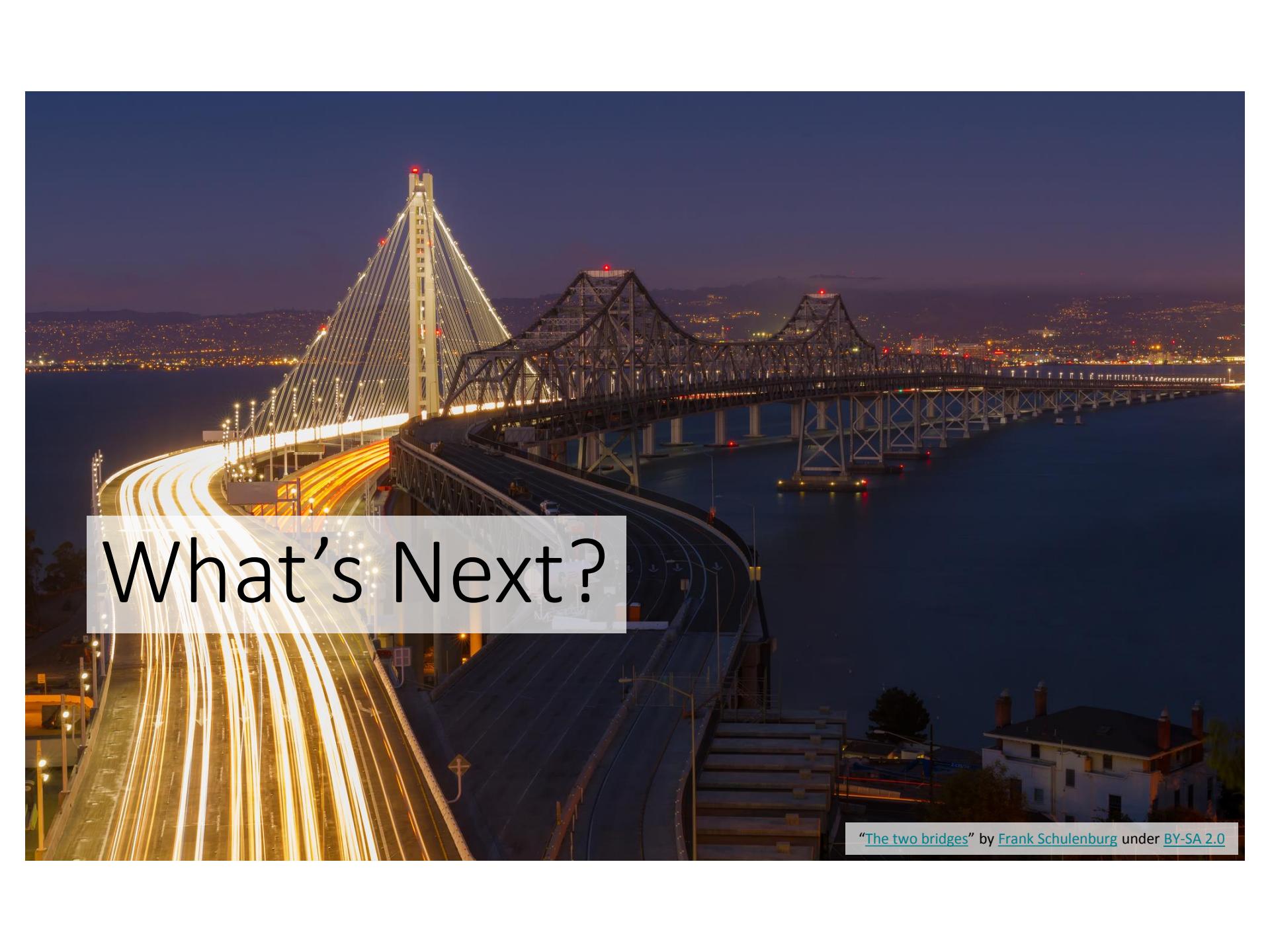
distributed preservation providers

- hosted services w/ varied architectures, service tiers, levels of assurance, replication factors
- replication nodes include memory orgs + cloud
- none (including LOCKSS) require local preservation infrastructure
- LOCKSS provides opportunity for co-preservation



ACADEMIC
PRESERVATION TRUST



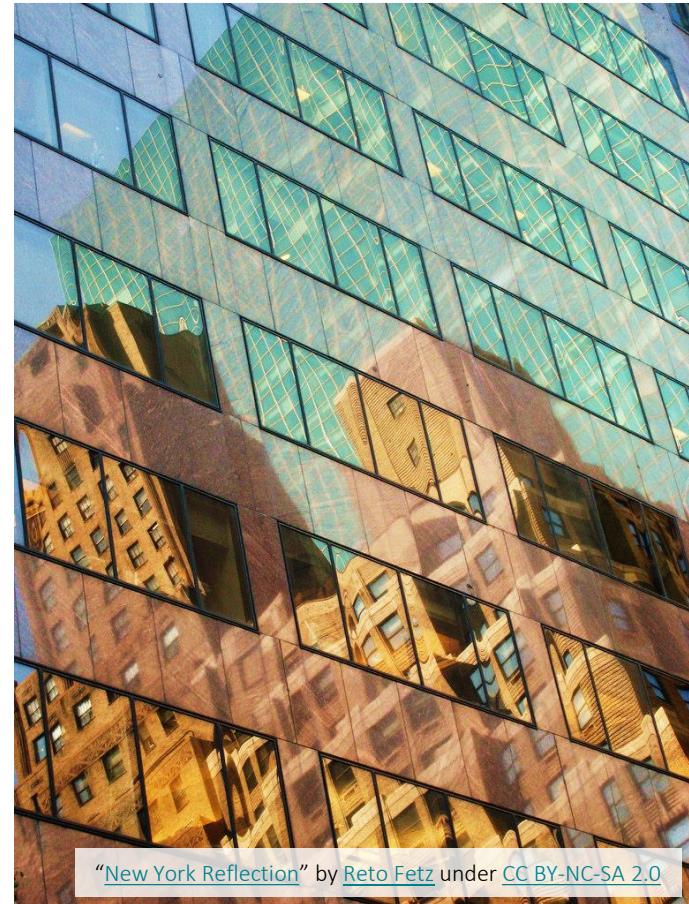


What's Next?

["The two bridges"](#) by [Frank Schulenburg](#) under [BY-SA 2.0](#)

re-architecture rationale

- de-silo + enable **external integrations**
- foster **developer community**
- capitalize on **work of broader communities**
- create space for **system enhancements**
- evolve w/ **web + digital preservation ecosystem**



"New York Reflection" by [Reto Fetz](#) under [CC BY-NC-SA 2.0](#)

anticipated outcomes

- functional parity + backward compatibility
- components providing value outside of end-to-end system
- better integration + data hand-offs w/ other apps
- increased use to preserve repository content
- increased use to preserve content managed by non-memory institutions



“ Bloom” by [Jak WI](#) under [CC BY-NC 2.0](#)

Questions

["Any Questions?"](#) by Matthias Ripp under [CC BY 2.0](#)