

A New Storage Paradigm for Sustainable Digital Stewardship

CNI 2022-12

Washington, D.C.

Andrew Woods - Harvard University

<https://bit.ly/2022-cni-drs>



Our Shared Opportunity

“ We see digital preservation as one of the greatest opportunities for collaboration between institutions that are inherently committed to ensuring that information is not only accessible and usable today, but far into the future.”

(Advancing Open Knowledge | Harvard Library, 2020, p.12)

Harvard Digital Repository Service (DRS)



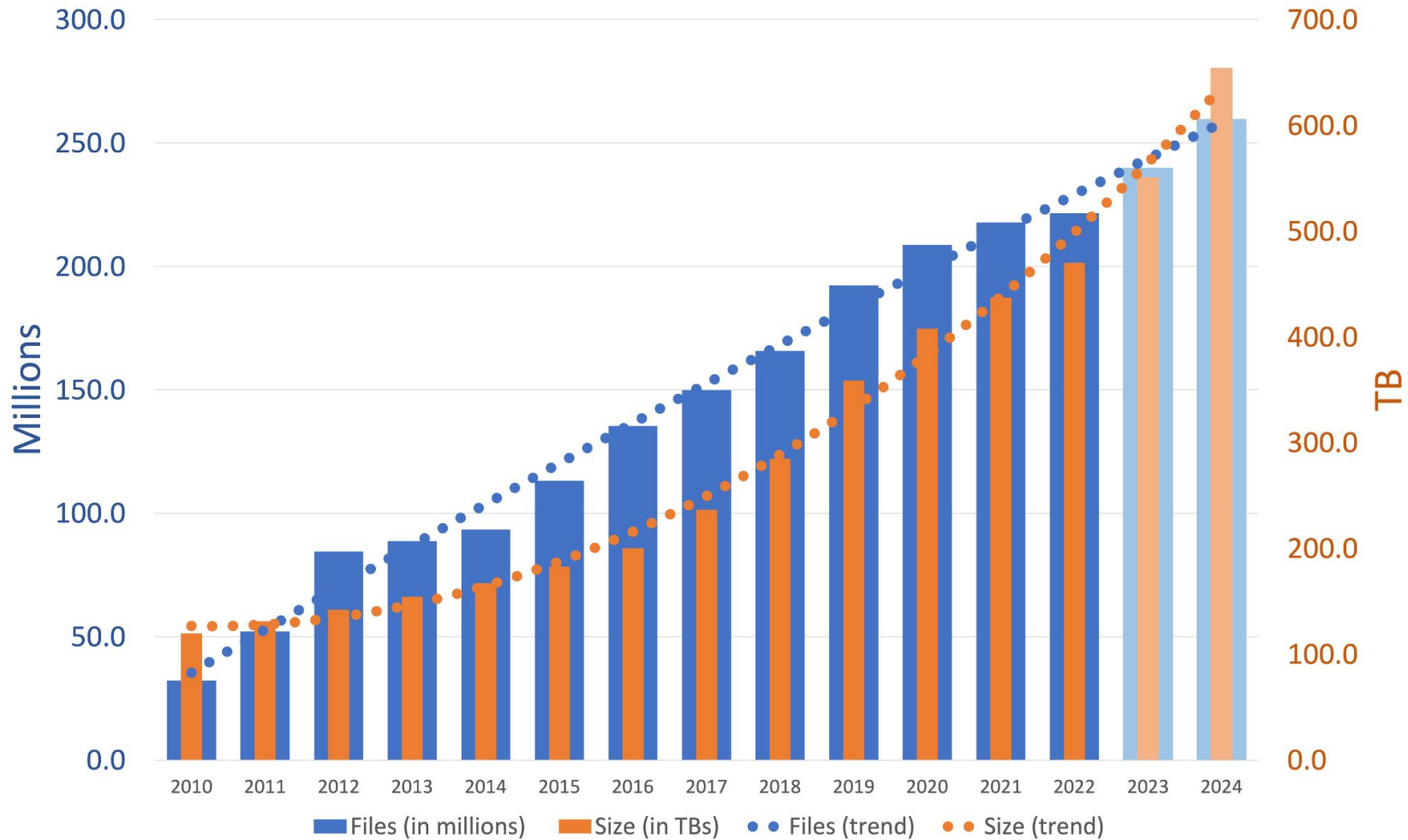
- Scope

Harvard Library's centrally-supported solution for the long-term preservation of, and access to, Harvard's digital collections

- Scale

- 63 contributing libraries, archives, museums, research centers, & admin offices
- 116 formats
- 453 thematic collections
- 10.6M objects
- 221M files (892M fully replicated)
- 470TB storage (2PB fully replicated)

DRS File Count and Size



Harvard & S3

Migration from POSIX to Object Storage (S3-API)

- Lower costs
- Billed by usage
- Leased instead of purchased storage
- On-demand extensibility
- Flexible development, QA, production environments

Harvard & OCFL

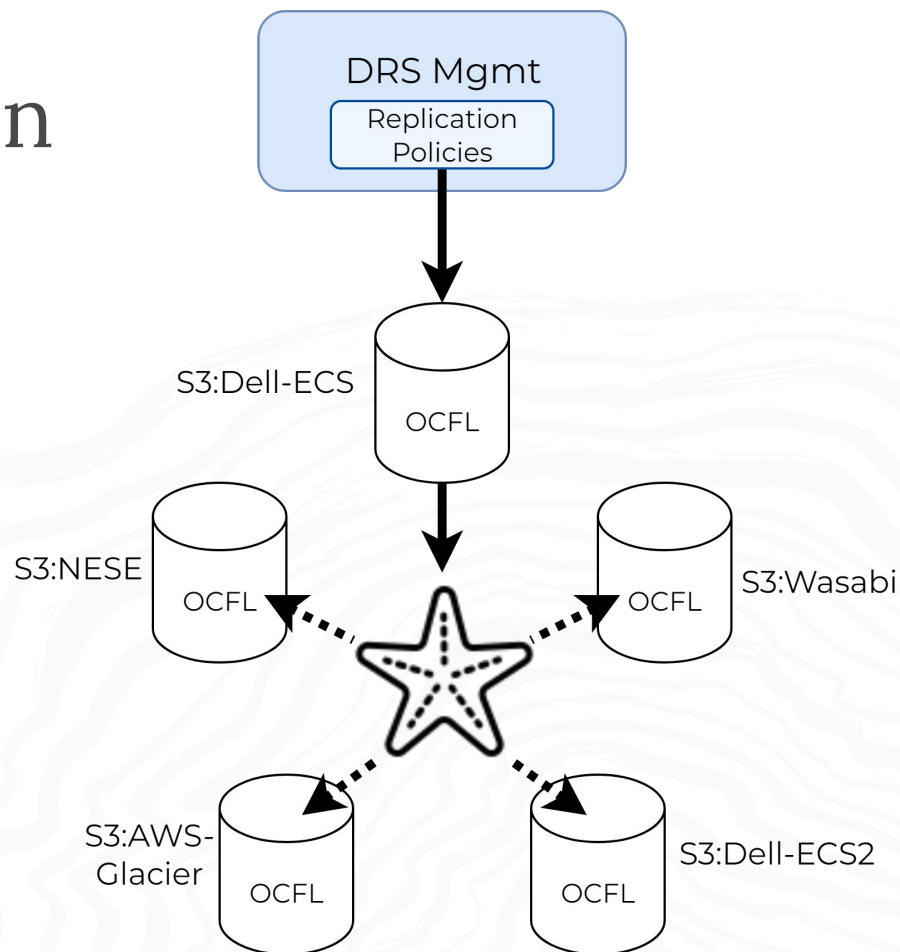
OCFL - Oxford Common File Layout

- <https://ocfl.io>

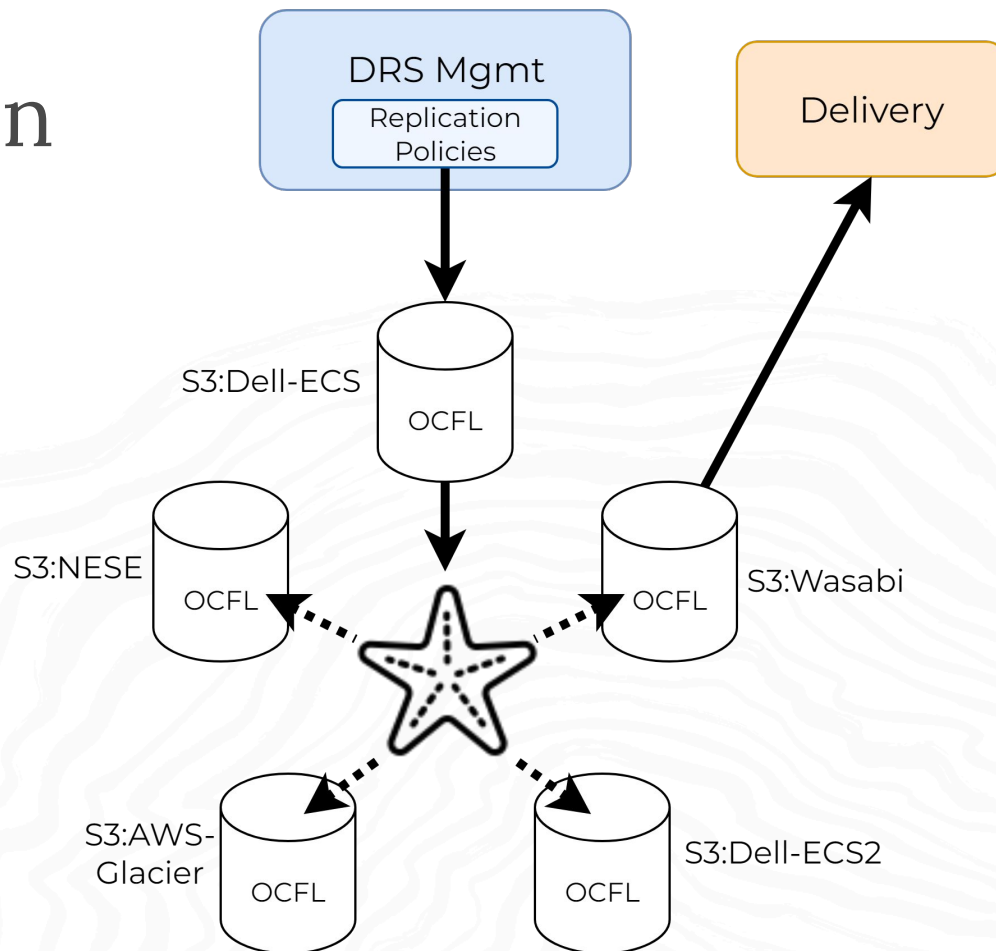
Opportunities with standardized file layout

- All of the benefits of OCFL:
 - completeness, robustness, versioning, parsability, storage diversity
- Shared community tooling
- Decoupled storage layout from repository software
- New use cases!

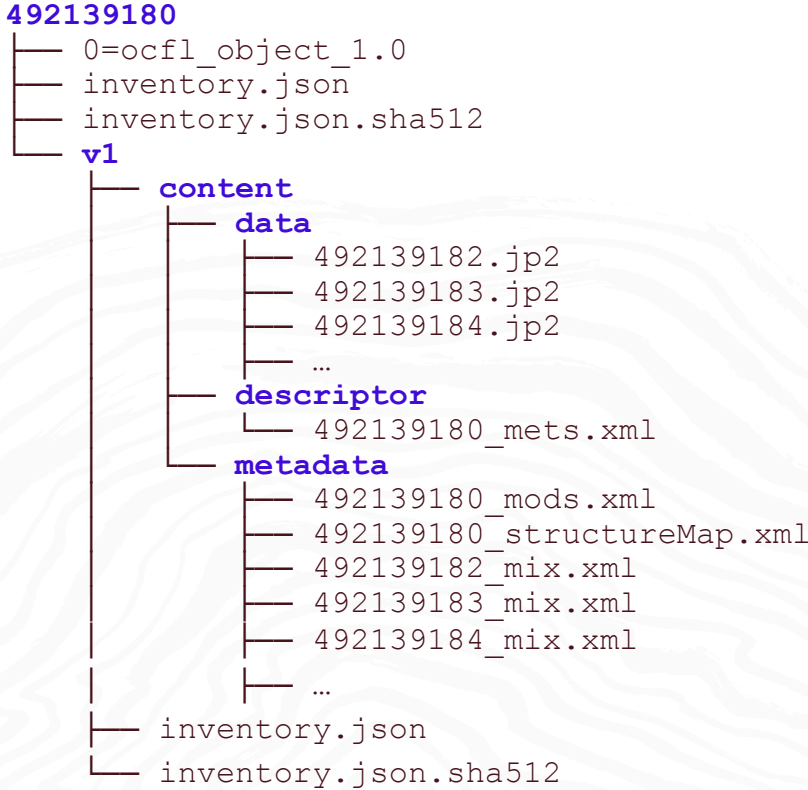
Policy-driven Replication



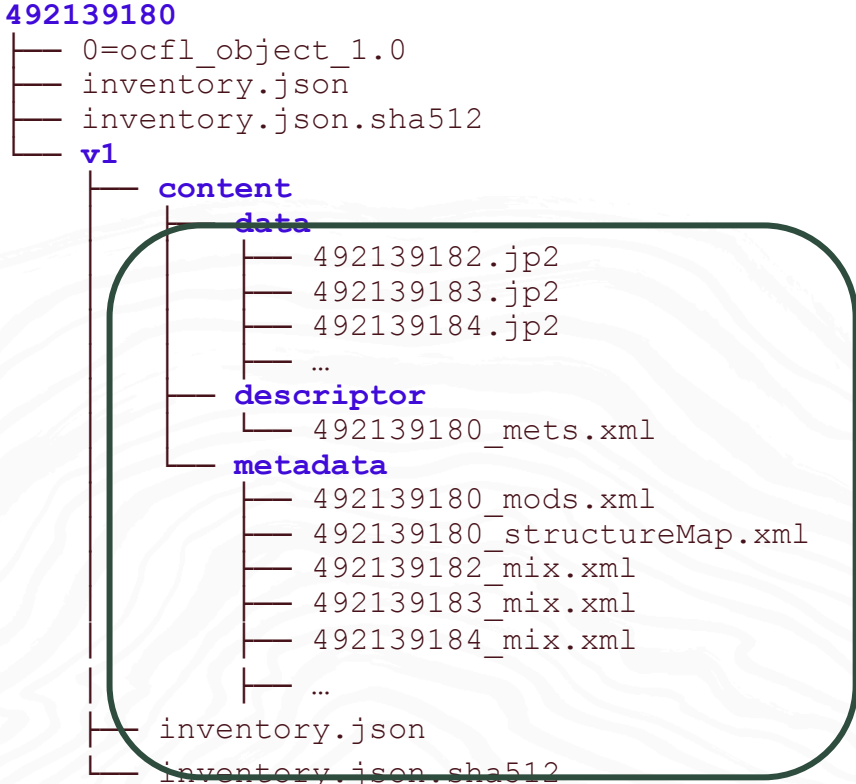
Policy-driven Replication



DRS OCFL Object Structure



DRS OCFL Object Structure



Let's Work Together

- For long-term digital preservation, it is important to have **transparent persistence layout**.
- **Applications come and go**. Our data is the bedrock.
- As a community, we have an opportunity to **standardize interaction with preservation persistence**.

Harvard DRS Futures: The Next Chapter

“We plan to modernize and rationalize our repository approach with a goal of disentangling preservation, asset management, and access while exploring opportunities for interoperability between systems.”

(Advancing Open Knowledge | Harvard Library, 2020, p.11)

Schedule



The Team



Stephen Abrams



Miriam Leigh



Janet Taylor



**Julianna
Barrera-Gomez**



Anthony Moulen



Robin Wendler



Stefano Cossu



Tricia Patterson



Andrew Woods



**Michelle
Gallinger**

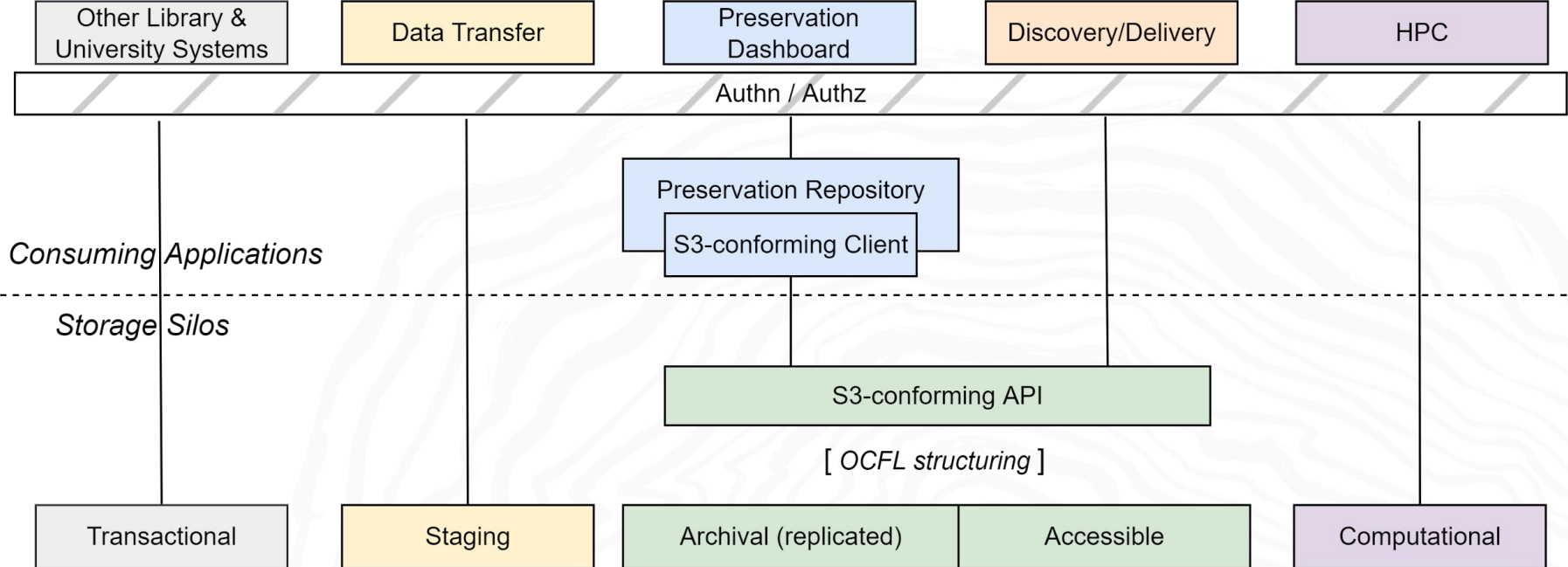


Stu Snyderman

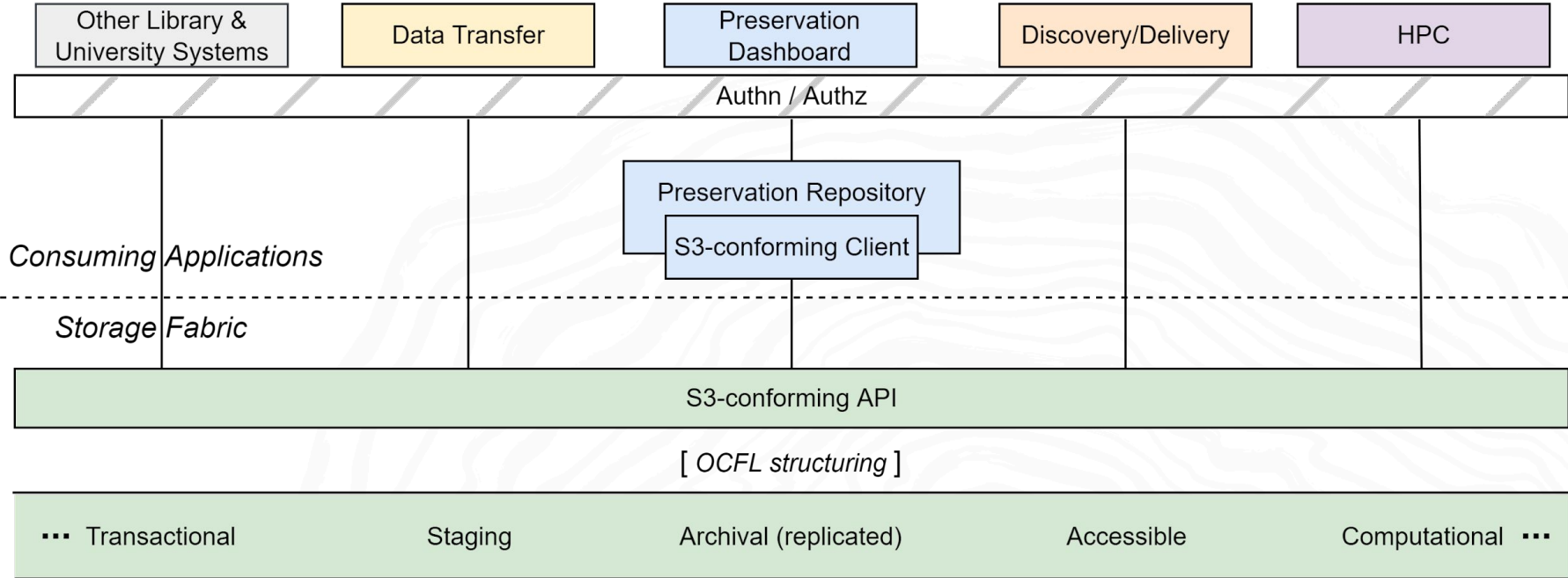


Vitaly Zakuta

Current



Vision



[Online / Nearline / Offline; Onsite / Offsite / Cloud; Institutional / Consortial / Commercial; Purchased / Leased]

Let's Work Together

<https://bit.ly/2022-cni-drs-c>

I (my team) would like to:

- receive project updates
- collaborate on next-generation digital preservation solutioning

Stay Connected

DRS Futures

- Follow: <https://sites.harvard.edu/drs-futures>
- Contact: drs-futures-feedback@HU.onmicrosoft.com

OCFL

- Follow: <https://ocfl.io/>
- Contact: <https://groups.google.com/forum/#!forum/ocfl-community>
- Chat: <https://code4lib.org/slack> (#ocfl)

Andrew Woods

- Contact: andrew_woods@harvard.edu
- Chat: <https://code4lib.org/slack> (@awoods)

Thank you!