The Repository Ecosystem: Vision and Action

GREI Workshop Keynote, January 24, 2023

Martha Whitehead,
Vice-President for the Harvard Library;
Co-Chair, USRN Steering Group;
Chair, COAR Executive Board

Key Takeaways

We need a diverse, distributed network of repositories

Let's collaborate nationally

And also act locally

My perspectives

- *Librarian:* values of equitable, perpetual access to diverse information and protection of privacy

- *Administrator:* values-based sustainable practices

- *Global and national:* Confederation of Open Access Repositories, Digital Research Alliance of Canada, U.S. Repository Network

Dare to Know

"Unlike paintings, prints and drawings traveled easily, offering more room for experimentation and for new and unconventional subject matter. As multiples, prints circulated the same visual information to widely dispersed viewers. They vastly expanded access to new knowledge put forth by anatomists, missionaries, and colonists, among many others, and they helped form a cohesive public out of heterogeneous and anonymous audiences."

LONG—TSING—YU *ou* LES YEUX DE DRAGONS.

14.

25.

2.

1.

1. *Le Quintebande*. 2. *Le Norimembre*.
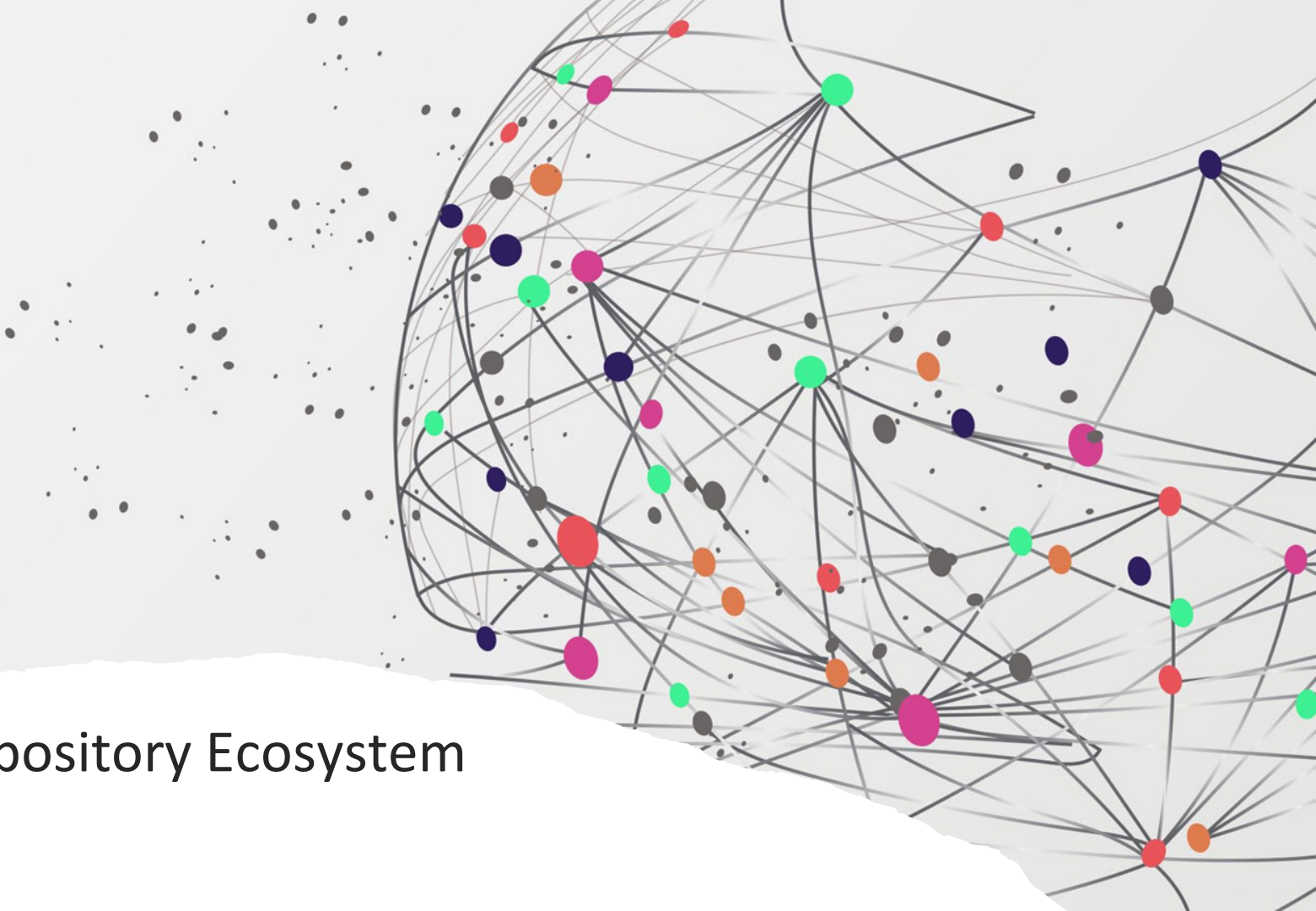
What researchers want:

A system of interoperable digital research objects

- Dynamic, version controlled
- Can be searched and text-mined
- Near-immediate publication
- Different models for post-publication review exist
- Contributorship in diverse roles
- Open to everyone

Data Code Interpretations Preregistration Materials DMP Review/Quality control
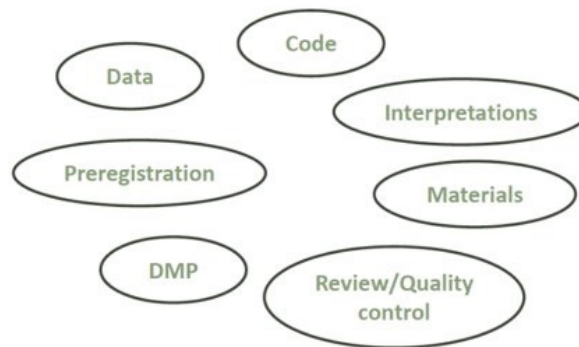
Dr. Bjoern Brembs, neurobiologist, in his keynote at the COAR Annual Meeting 2022
https://www.coar-repositories.org/the-role-of-repositories-in-radical-transformation-of-scholarly-communications/

Ideal Repository Ecosystem

What: open, inclusive and equitable

All valuable research outputs are as openly available as possible, as early as possible, and included in research evaluation processes

Trust – peer review that's preferably open and transparent

No costs for readers to access or authors to publish, and reasonable, transparent costs for services and infrastructure

Support diversity of perspectives and inclusion of all regions, and their local control of assets

Representative community governance – not out of the scholarly community's control

FAIR and CARE principles

How: diverse, distributed network of repositories

- A diverse ecosystem is necessary so that there is **no single point of failure**.

- Local infrastructure and services promote **equity**.

- Multiple types of repositories **meet varying needs** (some discipline-specific, some generalist, some managed by government agencies, some by non-profit organizations, some by institutions).

- **Institutional repositories** are generalist repositories. They represent a compliance option for public access to federally funded research results that is **fast and affordable**.

- Because institutional repositories are mainly hosted by universities and research centers (and community supported), they are aligned with the **research mission.**

- Investment in open infrastructure and expertise at local institutions can be leveraged to **ensure curation and adoption of good practices**.

Repositories are critical components of open science infrastructures

Open science infrastructures are often the result of community-building efforts which are crucial for their long-term sustainability and therefore should be not-for-profit and guarantee permanent and unrestricted access to all public to the largest extent possible.
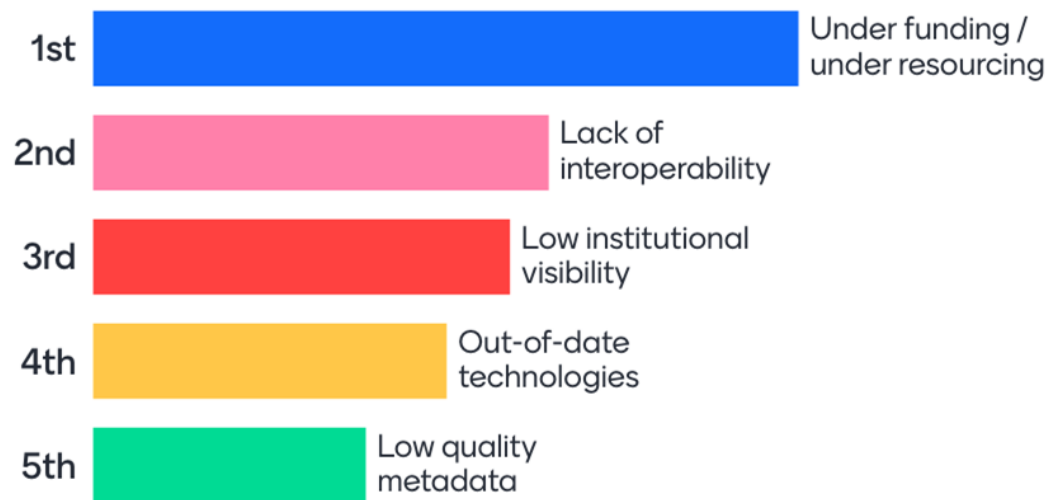
What is the most important barrier related to consolidating the role of the global network of repositories?

COAR Confederation of Open Access Repositories

COAR International Survey of Repositories 2021

Rank	Barrier
1st	Under funding / under resourcing
2nd	Lack of interoperability
3rd	Low institutional visibility
4th	Out-of-date technologies
5th	Low quality metadata

Lessening the burden for researchers

Making data sharing more than a compliance checkbox

Choosing a repository

Improving metadata processes

Improving curation and preservation

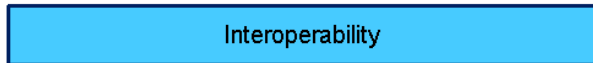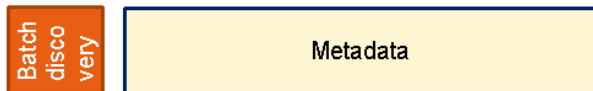Invest in and promote best practices in all repositories

Some key aspects of the ecosystem

COAR Next Generation Repositories (2016-17)

Current repositories

Services we can develop with repositories today

Conceptual layer
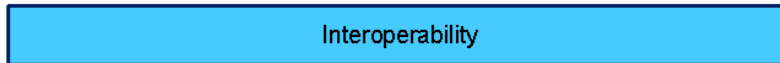
Batch discovery | Metadata

Interoperability
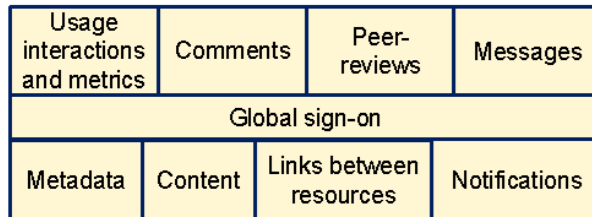
Persistence layer

Next generation repositories

Services we can develop with the next generation repositories

Conceptual layer

Discovery mechanisms
- Batch
- Navigation
- Notification

| Usage interactions and metrics | Comments | Peer-reviews | Messages |

Global sign-on

| Metadata | Content | Links between resources | Notifications |

Interoperability

Persistence layer

Pubfair:
a distributed framework for open publishing services
(November 2019)

COAR Confederation of Open Access Repositories
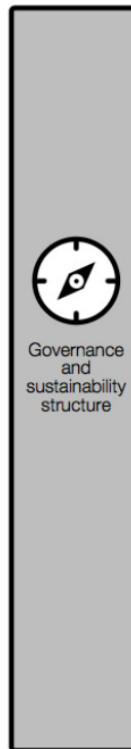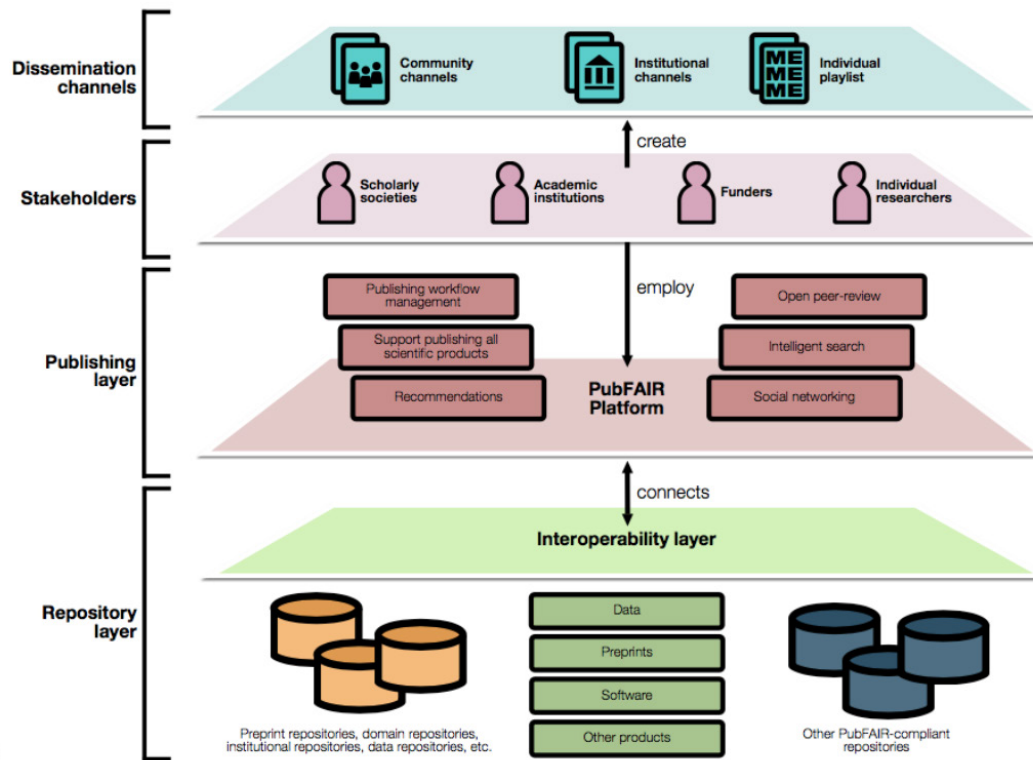
The COAR Notify Initiative

COAR Notify Project
Example use case: linking an article to a dataset
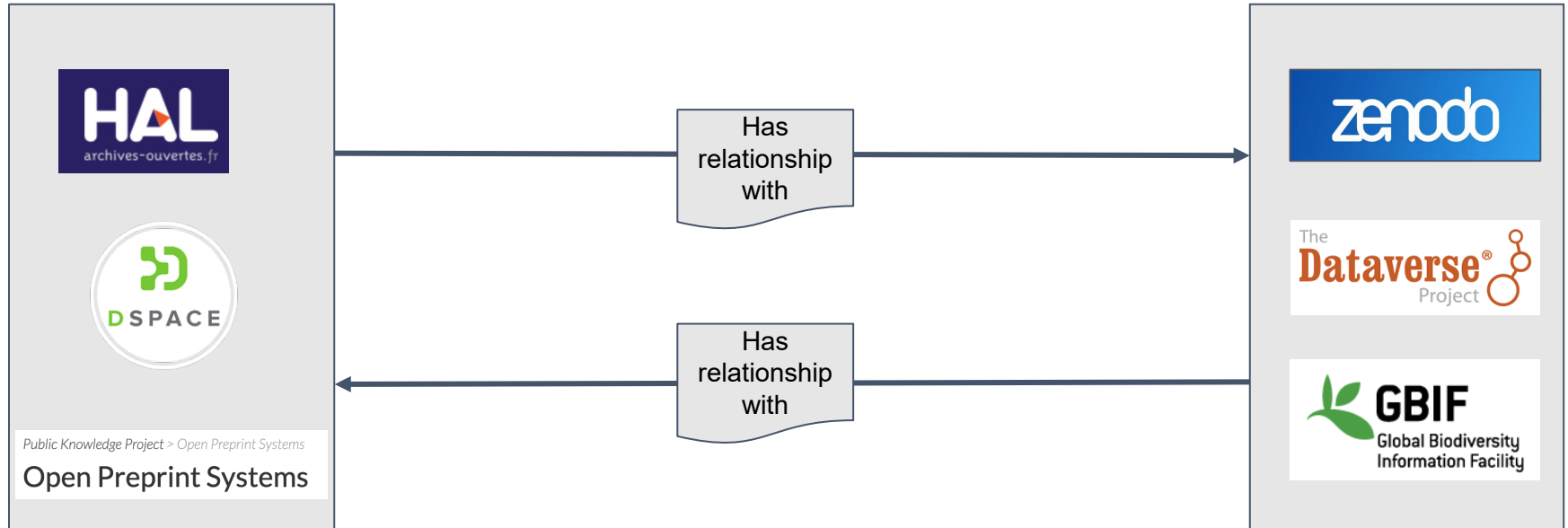
COAR Community Framework for Good Practices in Repositories

———

+ Purpose

+ Process

+ Relationship to other frameworks

+ Definitions

+ COAR Repository Assessment Working Group

+ Previous versions

+ Translations of Version 1

+ Translations of Version 2

VERSION 2 – PDF

COAR Confederation of
Open Access Repositories

https://www.coar-repositories.org/coar-community-framework-for-good-practices-in-repositories

Multilingual and Non-English Content

Ten Recommended Practices for Managing Preprints in Generalist and Institutional Repositories

COAR launches strategy to modernize the global repository network

COAR Confederation of Open Access Repositories

US in this global network: USRN

U.S. Repository Network is a partnership

Scholarly Publishing and Academic Resources Coalition

SPARC membership includes more than 200 libraries and academic organizations across North America. SPARC is a non-profit advocacy organization that supports systems for research and education that are open by default and equitable by design.

Confederation of Open Access Repositories

COAR is an international association with 155 members and partners representing libraries, universities, research institutions, government funders, and others. COAR brings together individual repositories and repository networks in order to build capacity, align policies and practices, and act as a global voice for the repository community.

SPARC✳

C⦿AR Confederation of Open Access Repositories
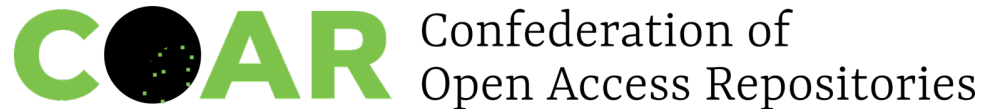
Good Practice Principles for Scholarly Communication Services

COAR and SPARC have developed seven good practice principles to ensure that scholarly communication services are transparent, open, and support the aims of scholarship. These principles can be used by users to make decisions about which services they will contract with, and by service providers to improve their practices and governance

GOOD GOVERNANCE

The service has strategic governance that allows community input on the direction of the service and operational governance with community representation and decision making power.

OPEN STANDARDS

The service uses open APIs to enable interoperability, and adheres to open standards. Ideally, the platform is based on open-source software, but in cases where it is not, user-owned content is managed according to well-established, international standards.

FAIR DATA COLLECTION

Only data necessary for the service's provision are collected from users and the type of the data collected and how they are used is clearly and publicly articulated.

TRANSPARENT PRICING AND CONTRACTS

The service's contract conditions and pricing are transparent and equitable, with no non-disclosure agreements included.

EASY MIGRATION

User-owned or generated content can be easily migrated to another platform or service upon termination of contract, without any additional fee from the service provider.

SUCCESSION PLANNING

If the service is a nonprofit the organization's bylaws state the conditions and terms governing how the organization may be transferred or wound down. If the service is provided by a for-profit entity, the contract/agreement should not be assignable to another entity without the client's express permission.

OPEN CONTENT

Content, metadata and usage data are immediately, openly and freely available in machine-readable format via open standards, and using licenses (like CC0 or similar) which facilitate reuse.

U.S. Repository Network is an inclusive network

The USRN is envisioned as an inclusive community committed to advancing all open repositories in the U.S., including:

- content such as publications, data, code, gray literature, emerging forms of scholarship, etc.

- hosted by higher education institutions, research centers, or other nonprofit organizations,

- using open source or vended platforms.

Developing a Vision

- 63-member Expert Group
    - U.S. COAR members
    - SPARC Steering Committee members
    - Library leaders
    - Repository managers
    - Consortium leaders
- Coalition of Open Access Policy Institutions (COAPI)
- OpenCon Librarians
- Open call for public comment

Strategic Vision for U.S. Repositories

An interoperable network of repositories is an essential component of our national research infrastructure, offering rapid and open access to research, and plays a crucial role in collective efforts to transform global research communications, leading to a more open, inclusive, and equitable system.

USRN Foundational Characteristics

Accountability	Community	Equity and Inclusivity	Local and Global	Sustainability
Collaboration	Discoverability	Interoperability	Preservation	User-centered

USRN Steering Group

Vicki Coleman, North Carolina A&T State University (co-chair)

Dan Coughlin, Penn State University

Greg Eow, Center for Research Libraries

Harriett Green, Washington University in St. Louis

Kristi Holmes, Northwestern University

Cynthia Hudson Vitale, Association of Research Libraries

Amanda Hurford, Private Academic Library Network of Indiana

Heather Joseph, SPARC (ex-officio)

David Minor, University of California San Diego

Kristi Park, Texas Digital Library

Anali Maughan Perry, Arizona State University

Torsten Reimer, University of Chicago
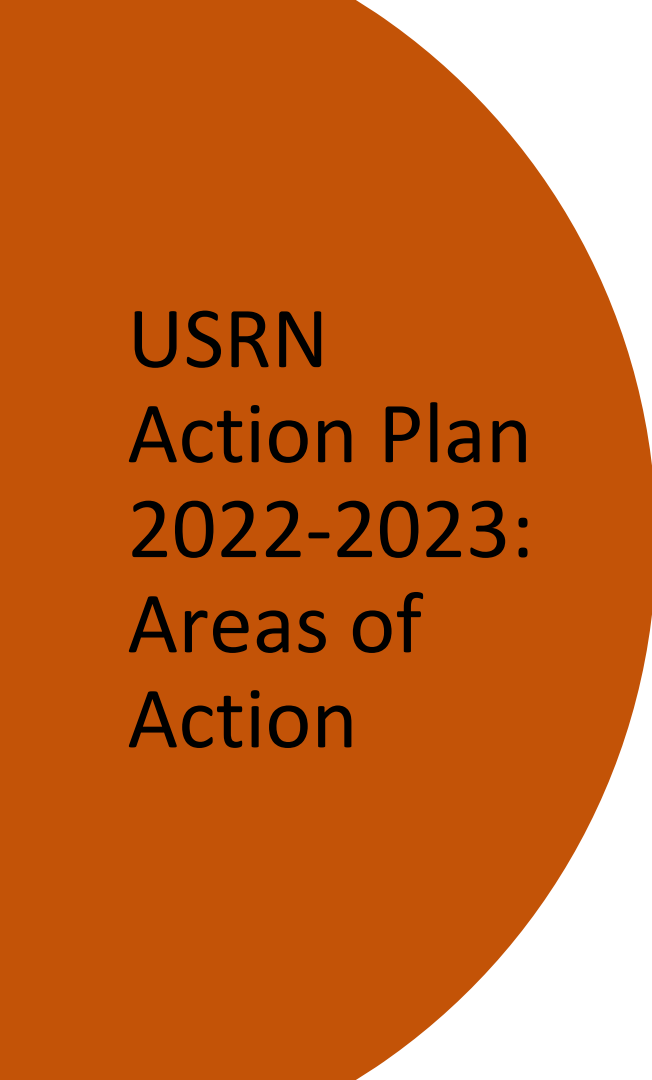
Kathleen Shearer, COAR (ex-officio)

Sarah Shreeves, University of Arizona

John Unsworth, University of Virginia

Meg Wacha, City University of New York

Martha Whitehead, Harvard University (co-chair)

USRN Action Plan 2022-2023: Areas of Action

To begin coalescing and strengthening repositories, sharing best practices, and developing value added services, our first steps are to

1. Engage with OSTP and federal funding agencies on implementation of public access guidance

2. Develop network governance model

3. Build community and external relationships

https://sparcopen.org/our-work/us-repository-network/action-plan/

Acting locally

Advancing Open Knowledge: A Strategic Framework

OUR PURPOSE

We champion curiosity for the betterment of the world.

OUR VISION

We aspire to be global leaders in expanding world knowledge

OUR MISSION

To advance the learning, research and pursuit of truth that are at the heart of Harvard.

OUR VALUES

- Embrace Diverse Perspectives
- Champion Access
- Aim for the Extraordinary
- Seek Collaboration
- Lead with Curiosity

A DISCUSSION OF HARVARD LIBRARY'S STRATEGIC DIRECTIONS
Advancing Open Knowledge
November 2020

HARVARD LIBRARY AND HARVARD COLLEGE LIBRARY
Multi-Year Goals and Objectives and FY21 Priorities
JULY 2021

HARVARD LIBRARY AND HARVARD COLLEGE LIBRARY
Multi-Year Goals and Objectives and FY22 Priorities
JULY 2021

HARVARD LIBRARY

library.harvard.edu/advancing-open-knowledge

HARVARD LIBRARY

EDIBA

Advance Equity, Diversity, Inclusion, Belonging, and Antiracism (EDIBA) in our workforce, services, collections, and spaces

"In all we do, we are seeking to become a truly antiracist research library that addresses issues of bias and minimizes all forms of oppression."

- Advancing Open Knowledge

In addition to stewarding vast collections and open digital content, we at Harvard Library want to see communities in all parts of the world empowered to produce, share, and retain ownership of local research outputs and cultural resources.

We believe this will benefit scholars at Harvard as well as people around the world.

Harvard Library News, March 15, 2021

Research Libraries in a Global Knowledge Commons

Our shared goal

Collaboratively develop a global knowledge commons that benefits our local communities and the world.

"Data" at the core, surrounded by enabling layers of:

- policy

- distributed infrastructure

- services that facilitate user (including machine) engagement with data
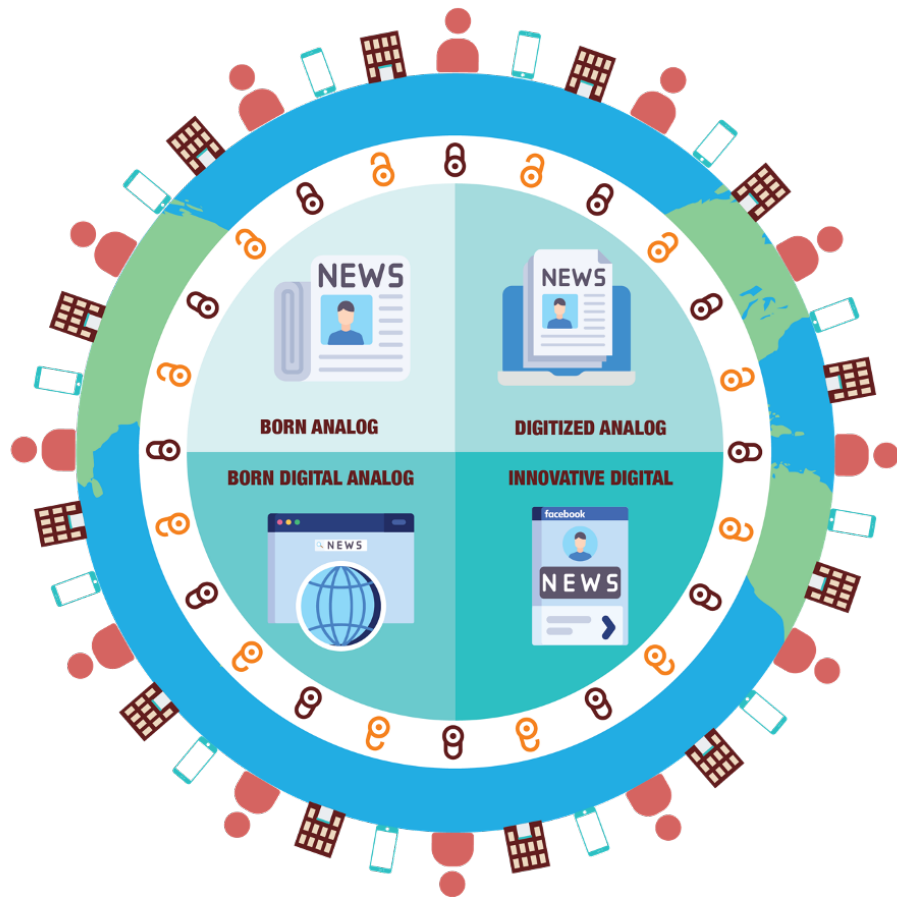
A Key Priority in our Multi-Year Goals and Objectives

Invest in open-access infrastructure and services that support equitable, sustainable models for scholarly communication and open knowledge

Harvard Data Commons

The Harvard Data Commons is a collaboration between Harvard Library, University Research Computing, the Office of the Vice-President for Research and several schools across the university to create an interoperability framework between research, management tools, and Harvard's computing resources and repositories.

Vision: The Data Commons will improve the researcher experience by automating the flow of research data and derivative scholarship from research computing environments to management, publication, and preservation environments. As a result, data integrity, provenance, and reproducibility of research data will be improved in order to meet sponsor requirements.

The Problem (as we see it)

- Lack of data-centric tools throughout the research lifecycle
- Lack of interoperability between tools

PLANNING	STORING	ANALYSIS	MANAGEMENT	DISPOSAL
Creating data Discovery Acquisition	Data transfer Raw data Reference data	Data wrangling Data analysis Data sharing	Data repositories Data preservation	

- Ease the **organization** of large data
- Facilitating the **reproducibility** of data
- **Sharing** of active data internally and externally

- Supporting data **retention** compliance
- Tracking the **provenance** of data
- Providing **metrics** on data access and use

Services as infrastructure

Research Data Management @Harvard

Home   Vision   Data Lifecycle ▾   Policies   Resources   Contacts

A reference guide with information and resources to help you manage your research data

DATA LIFECYCLE

Planning Data Management	Data Acquisition and Collection	Storage, Security, and Analysis	Dissemination and Preservation
How can I best manage my data throughout the lifecycle of my research to save time and money in the future?	How can I acquire data in an efficient and ethical way, and how can I ensure that my data is used appropriately?	What are my options for effectively organizing, storing, securing, computing, and analyzing my research data?	Why is it worthwhile to share my data? What do funders and journals require? Can I get help with data curation?
• Data Management Plans (DMPs) • DMP requirements and tools • What research objects should be tracked and documented	• Data Use Agreements (DUAs) • Institutional Review Boards (IRB and IACUC) • Subscription data	• Data security • Computing, research methods, data science, and viz support • Electronic Lab Notebooks	• Data repositories • Open Access • Data citation, FAIR principles • Data disposal

https://researchdatamanagement.harvard.edu/

Key Takeaways

We need a diverse, distributed network of repositories

Let's collaborate nationally

And also act locally

Acknowledgements and sincere thanks!

- Tina Baich, SPARC Visiting Program Officer for USRN
- Heather Joseph, Executive Director, SPARC
- Kathleen Shearer, Executive Director, COAR
- Harvard research data management partners (many!)

Get USRN updates!
Get involved!

https://sparcopen.org/our-work/us-repository-network/