Supporting Open Data with SHARE

Cynthia Hudson-Vitale
Data Services Coordinator
Washington University in St. Louis

SHARE

SHARE is a free, open catalog of research and scholarly metadata across the research life-cycle.

Mission and Values

SHARE's mission is to maximize research impact by making research widely accessible, discoverable, and reusable.

SHARE is developing services to gather and freely share information about research and scholarly activities across their life cycle.

Making research and scholarship freely and openly available encourages innovation and increases the diversity of innovators.
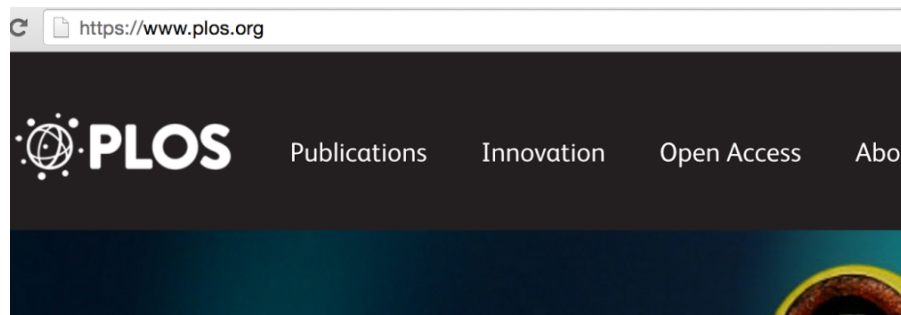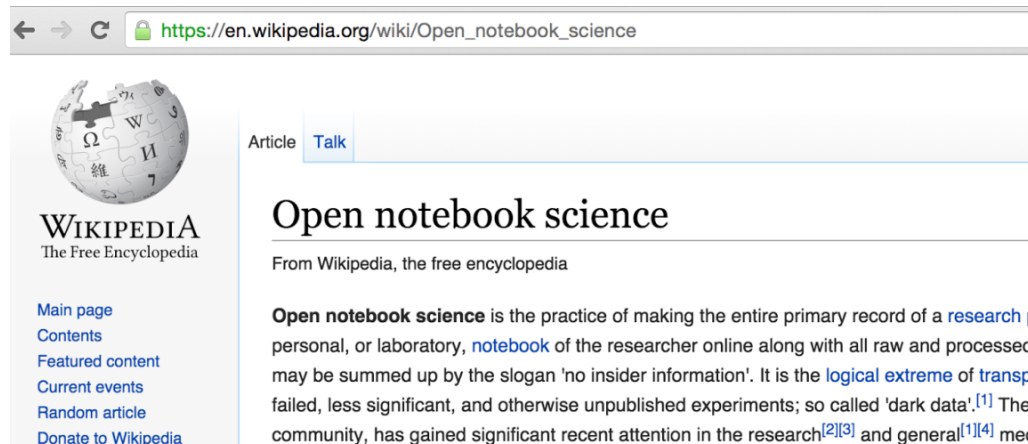
Open science and open scholarship
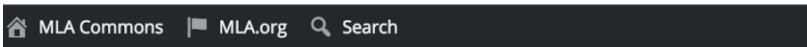
Open, public goods vs. closed, proprietary applications

Open Source

Depends on collaborative community dedicated to transparency and openness

Facilitates interoperability and sustainability

Freely available on Internet. Don't pay licensing fee
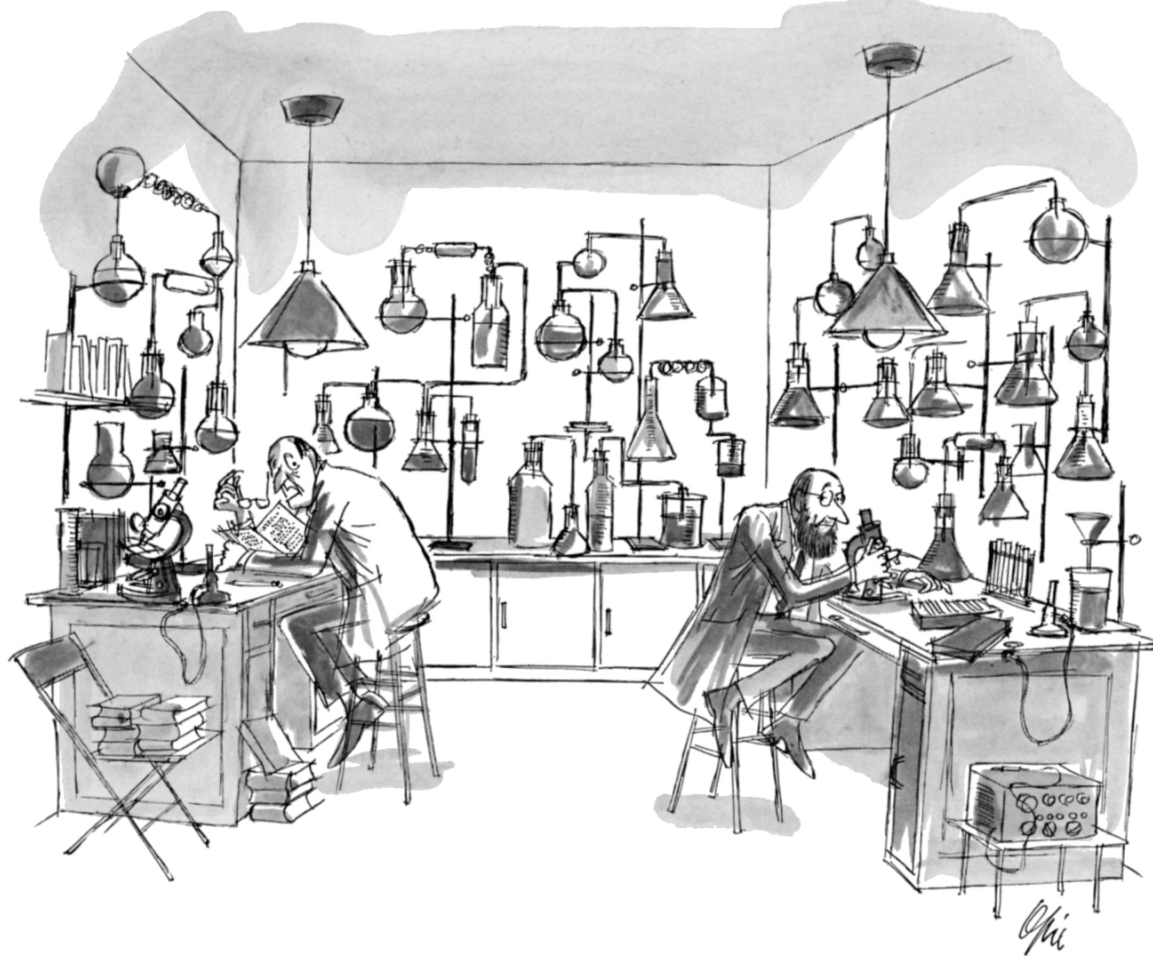
Can modify at will

Proprietary Software

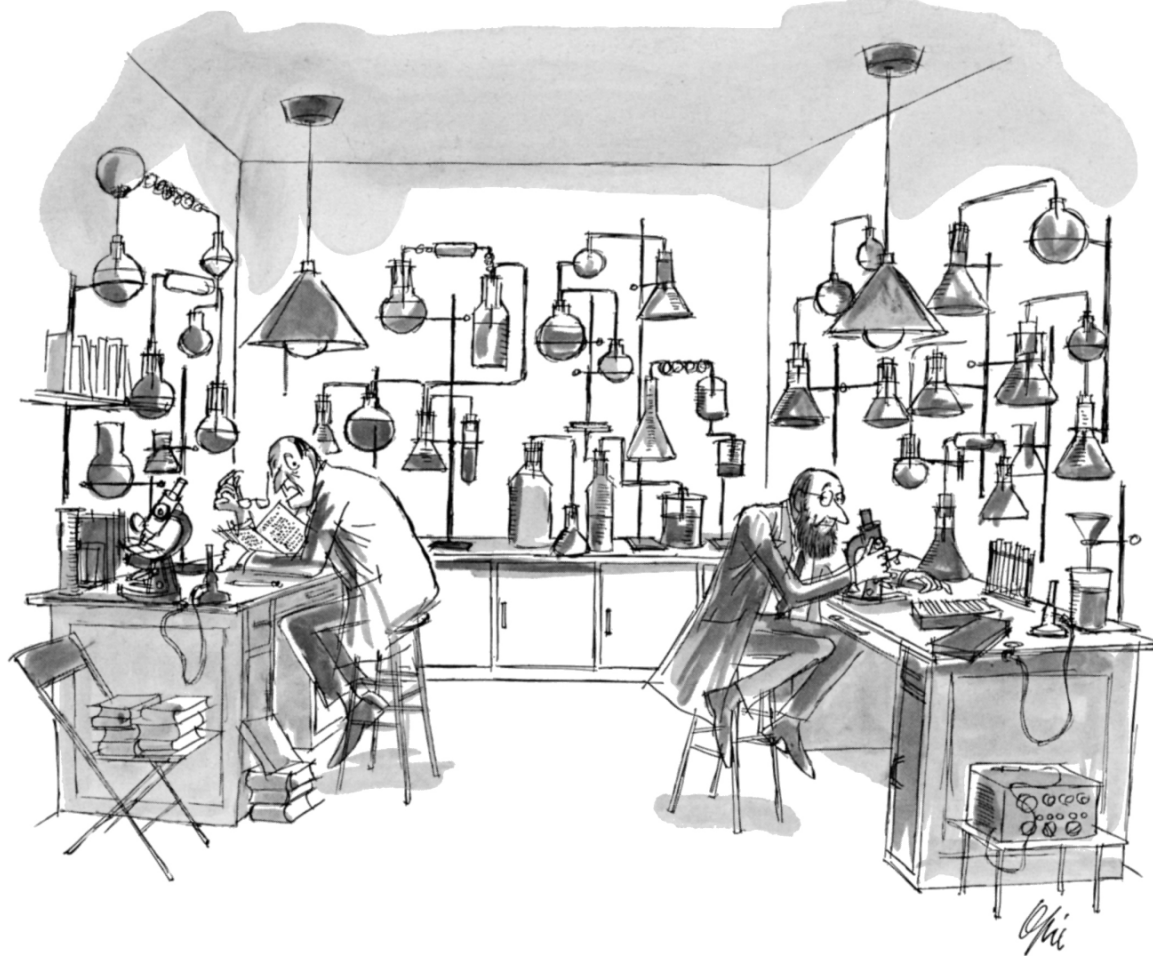Have to rely on vendor for updates, problem fixes

Pay licensing fees

Owned by the individual/ company that developed it

Harder to share research data

"I see by the current issue of 'Lab News,' Ridgeway, that you've been working for the last twenty years on the same problem I've been working on for the last twenty years."

Everett Opie The New Yorker Collection/The Cartoon Bank, **1976**

"I see by the current issue of 'Lab News,' Ridgeway, that you've been working for the last twenty years on the same problem <u>I've</u> been working on for the last twenty years."

SHARE catalog

150 data sources
- Registries (e.g. CrossRef, DataCite)
- Disciplinary repositories and preprint services
- Data repositories
- Institutional repositories
- Agency repositories (e.g. DOE SciTech Connect)

17.8 million metadata records

https://share.osf.io/

Collecting the Research Lifecycle
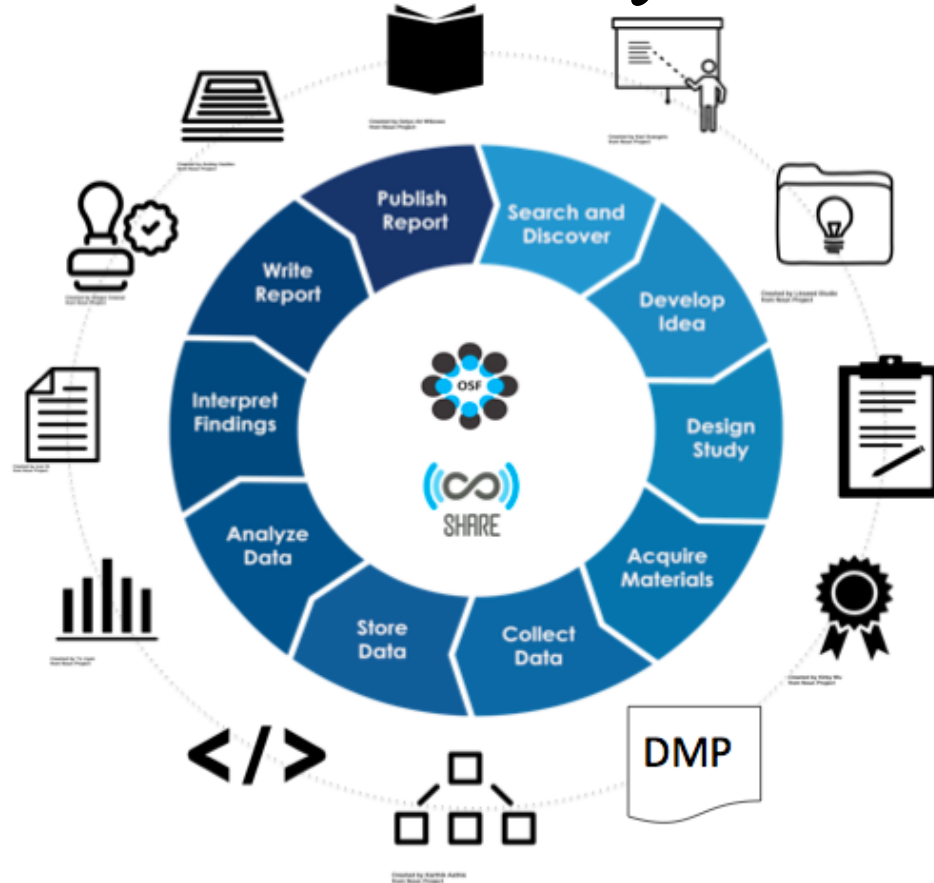
Metadata records by type

Data Set

Patent

Poster

Presentation

Publication
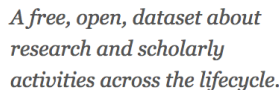
Repository

Retraction

Software

The Lifecycle Approach

"All scholarly digital research objects—from data to analytical pipelines—benefit from application of these [FAIR data] principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability."
Nature. Scientific Data Comment. Wilkinson, et al. "The FAIR Guiding Principles for scientific data management and stewardship" March 15, 2016.

"Rather than focusing on acquiring the products of scholarship, the library is now an engaged agent supporting and embedded within the processes of scholarship."
"21st Century Library Collections: Calibration of Investment and Collective Action," (ARL 2012)

SHARE¶

SHARE is a higher education initiative whose mission is to maximize research impact by making research widely accessible, discoverable, and reusable. To fulfill this mission SHARE is building a free, open, data set about research and scholarly activities across their life cycle.

SHARE harvests metadata nightly from 100+ repositories, transforms that metadata into one format, and makes it accessable via a web API.

The technical side of SHARE has many pieces that you can interact with:

- A search endpoint powered by elasticsearch that indexes the transformed data allowing:
    - Thorough search of creative works:
        - Creative works: /api/v2/search/creativeworks/_search
        - more info on the elasticsearch docs page
    - Data aggregations across fields

- An Ember application using the SHARE API for:
    - Searching the SHARE database
    - Discovering new Projects
    - Corrections and Updates

- API Endpoints for accessing transformed metadata
    - https://share.osf.io/api/v2

Guide

Table Of Contents

SHARE
- Guide
- Contribute
- Get In Touch

Quick search

Go

A free, open, dataset about research and scholarly activities across the lifecycle.

v: latest

((∞)) S H A R E

Discover   FAQ   Sign in with OSF

"washington University" NOT george

29,903,653 events as of May 25, 2017   Found 17,866 events in 0.464 seconds   Aggregating 151 sources
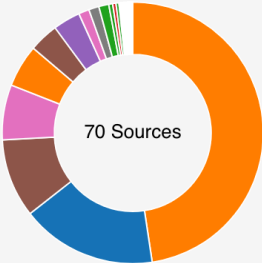
View query body ▲

Clear filters

◄ **1** 2 3 ••• 1,000 ►

787 additional pages not shown

Sort by: Relevance

Source

Add Source filter ▼

70 Sources

Date

📅 All time ▼

Effect of neoadjuvant chemotherapy on bone marrow micrometastases in women with locally advanced breast cance

R. Aft - L. Ylagan - S. Sadeghi - T. Eberlein - V. Herrmann - J. Dietz +3

Article   CrossRef

oncology   cancer research

Last updated Apr 05, 2017   Published by American Society of Clinical Oncology (ASCO)

Effect of neoadjuvant chemotherapy on bone marrow micrometastases in women with locally advanced breast cance
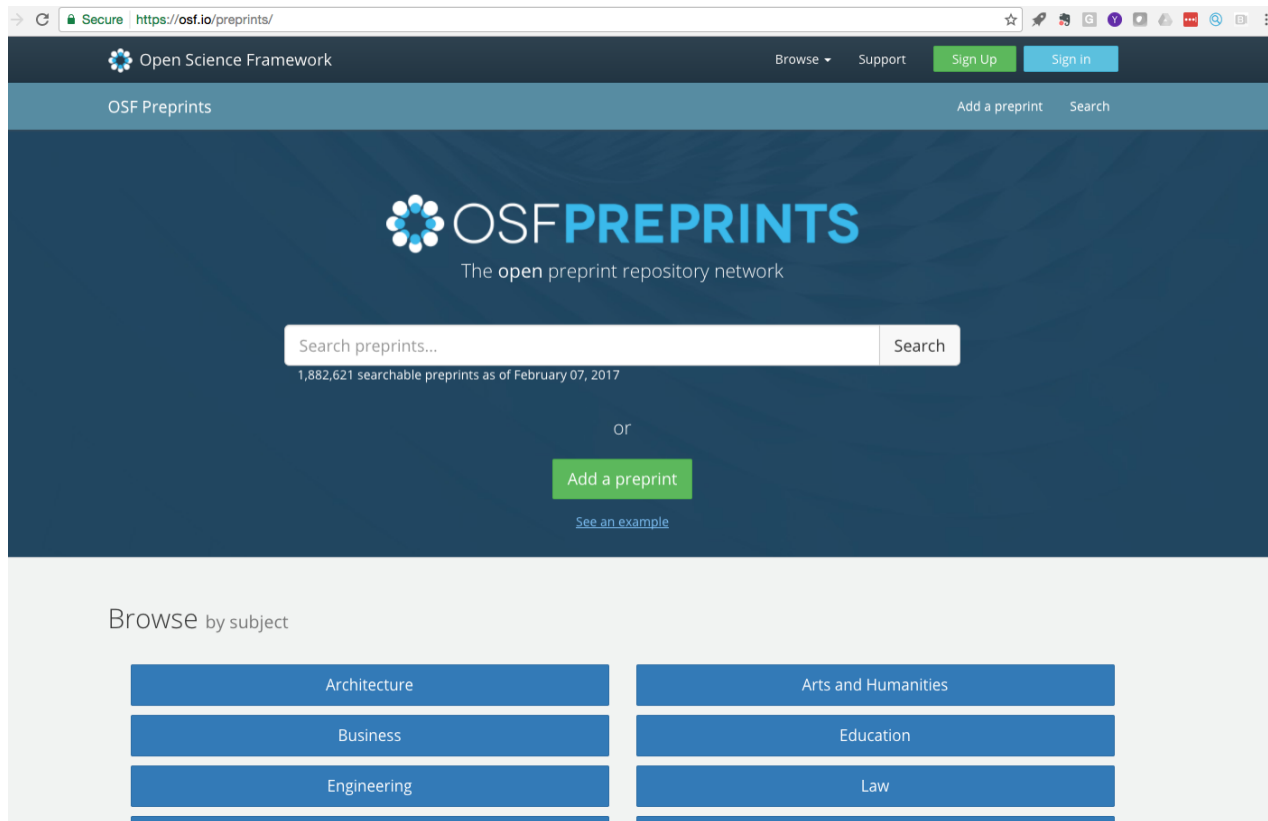
R. Aft - L. Ylagan - S. Sadeghi - T. Eberlein - V. Herrmann - J. Dietz +3

Article   CrossRef

oncology   cancer research

SHARE API powers new discovery services

SHARE API powers institutional analytics

With institutionally-contributed data on affiliation, SHARE API can be used to visualize:

Awards

Highlighted collections

Top contributors

Top subject tags

Prototype in development by team from UCSD Research Data Curation Program and the Center for Open Science

Check it out at: http://tritonshare.ucsd.edu

Types

project: 28.90%	article: 22.18%	creative work: 21.04%	data set: 10.18%

Top Contributors

1767	Scripps Institution of Oceanography
590	SIO Geological Data Center (GDC)
495	University of California, San Diego
323	Global Change Data Center
252	DOC/NOAA/NESDIS/NODC > National Oceanographic Data Center, NESDIS
98	Scripps Institution of Oceanography (SIO) ; University of California [San Diego] (UC San Diego)
56	University of California [San Diego] (UC San Diego)
39	Argonne
30	University of California

html + js + css + SHARE API
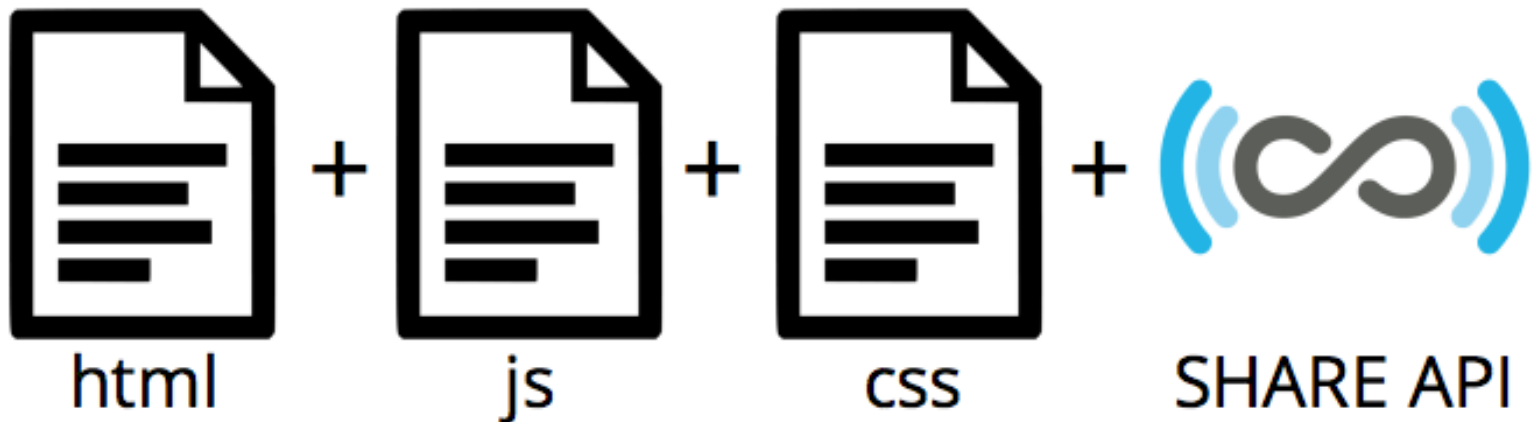
hosted statically on Github sites.

Note: we did put a caching proxy in between the client and SHARE to be nice.

Code: https://github.com/cos-labs/share-analytics

Get more information

www.share-research.org/updates

@SHARE_Research

Get involved!

info@share-research.org

Questions

1. What existing services and resources do we offer that support openness? What new services or expertise might we need to develop?
2. What are the best methods to provide outreach for open tools?
3. What are some practical incentives for researchers to start using open tools?
4. What is a reasonable level of service to provide faculty for OSF, ReproZip, and SHARE?
5. What opportunities and challenges do librarians face when engaging with openness?
6. Next thing on the horizon - what will people want next?
7. What are your libraries currently doing to address open scholarship?