

# SHARE in 2016: Technology, Community, Pedagogy

Cynthia Hudson-Vitale  
Judy Ruttenberg  
Texas Digital Library Webinar  
Thursday, May 6, 2016

## **Mission**

SHARE is a higher education initiative whose mission is to **maximize research impact** by making research widely accessible, discoverable, and reusable.

# Mission

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**.

# Thank you to our sponsors

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



**Alfred P. Sloan  
FOUNDATION**

# SHARE Phase II objectives

Expand providers

Enhance metadata collected from existing providers, (100+) including linking related research objects

Look for opportunities for institutional and other systems use and integration

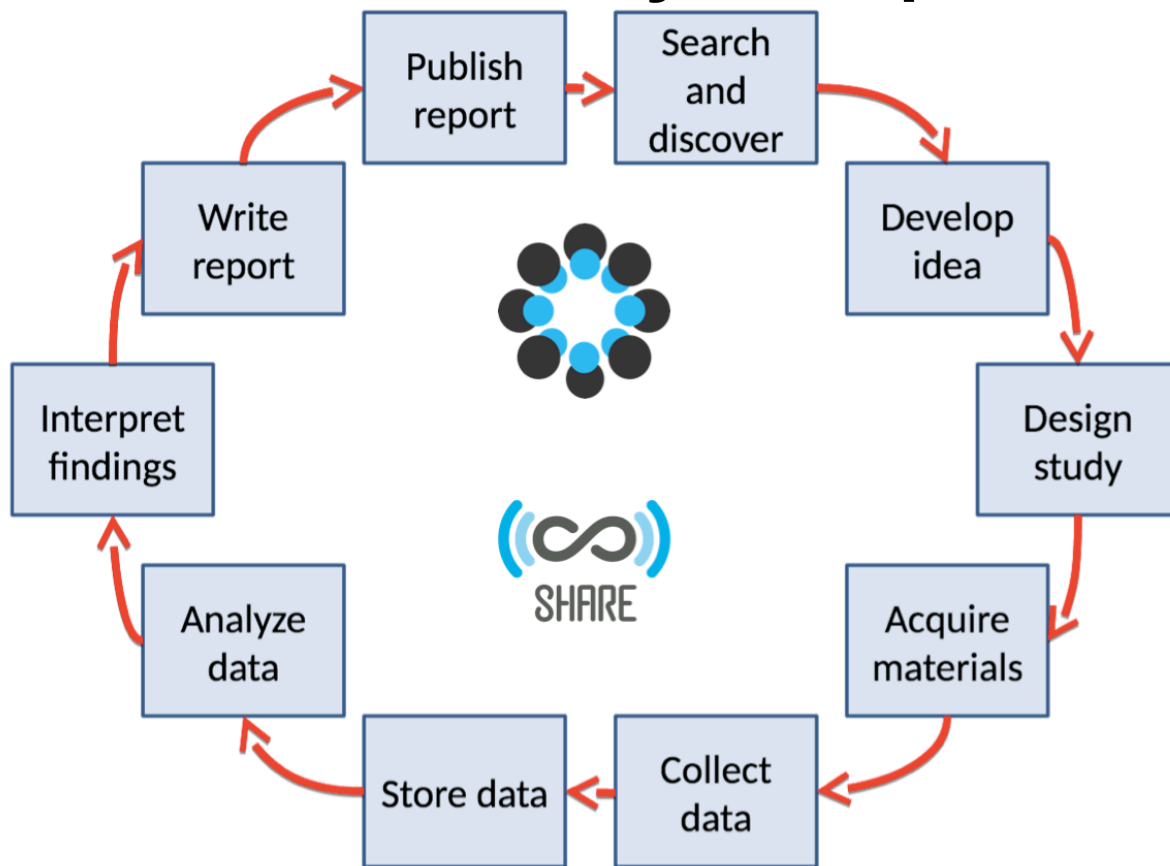
# Open infrastructure for open access

OA advocacy & policy is more mature than OA infrastructure & implementation

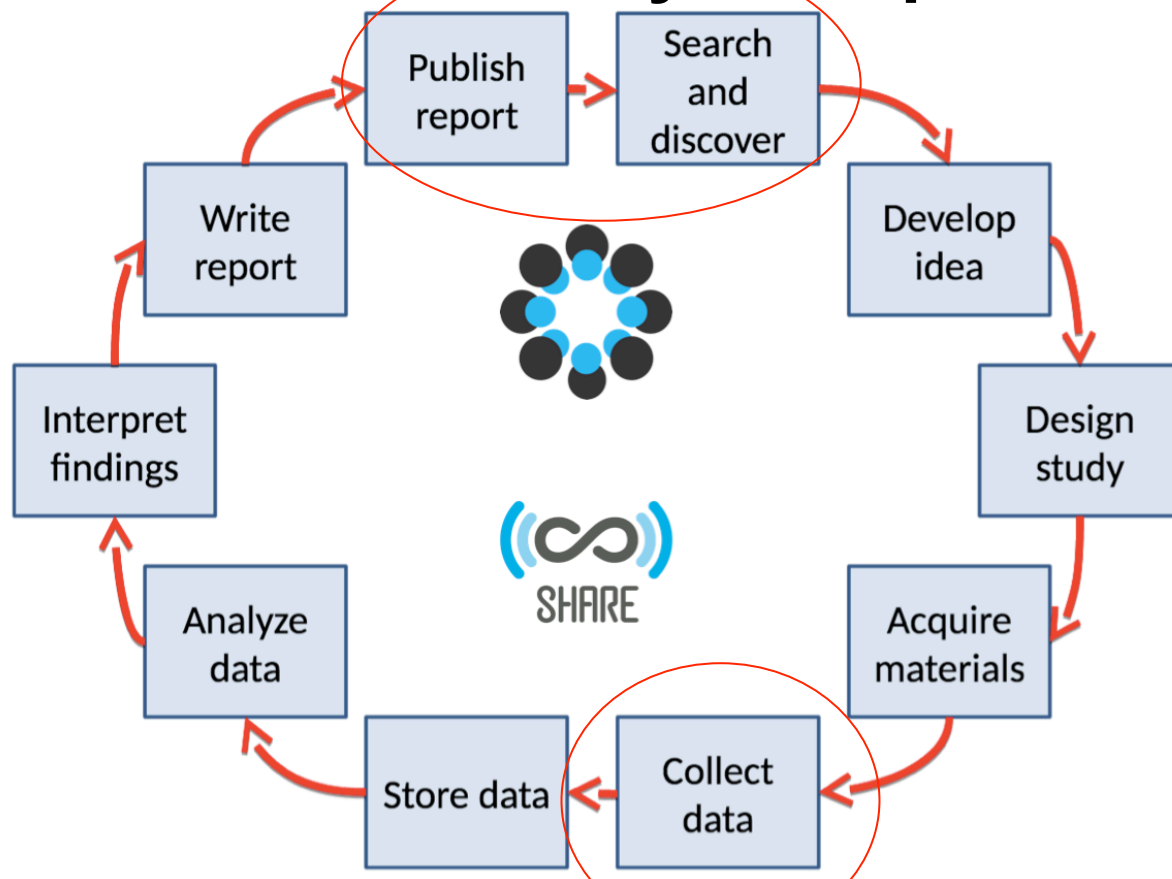
OA implementation infrastructure should be **open**

- Metadata
- Standards and identifiers
- Code, platforms, and APIs

# Why is the research life cycle important?



# Why is the research life cycle important?





# The future of open scholarship

“We find ourselves two decades into the transition of research communication from a paper-based endeavor to a web-based digital enterprise, and well into the transition of the research process itself **from a largely hidden activity to one that becomes plainly visible on the global network.**”

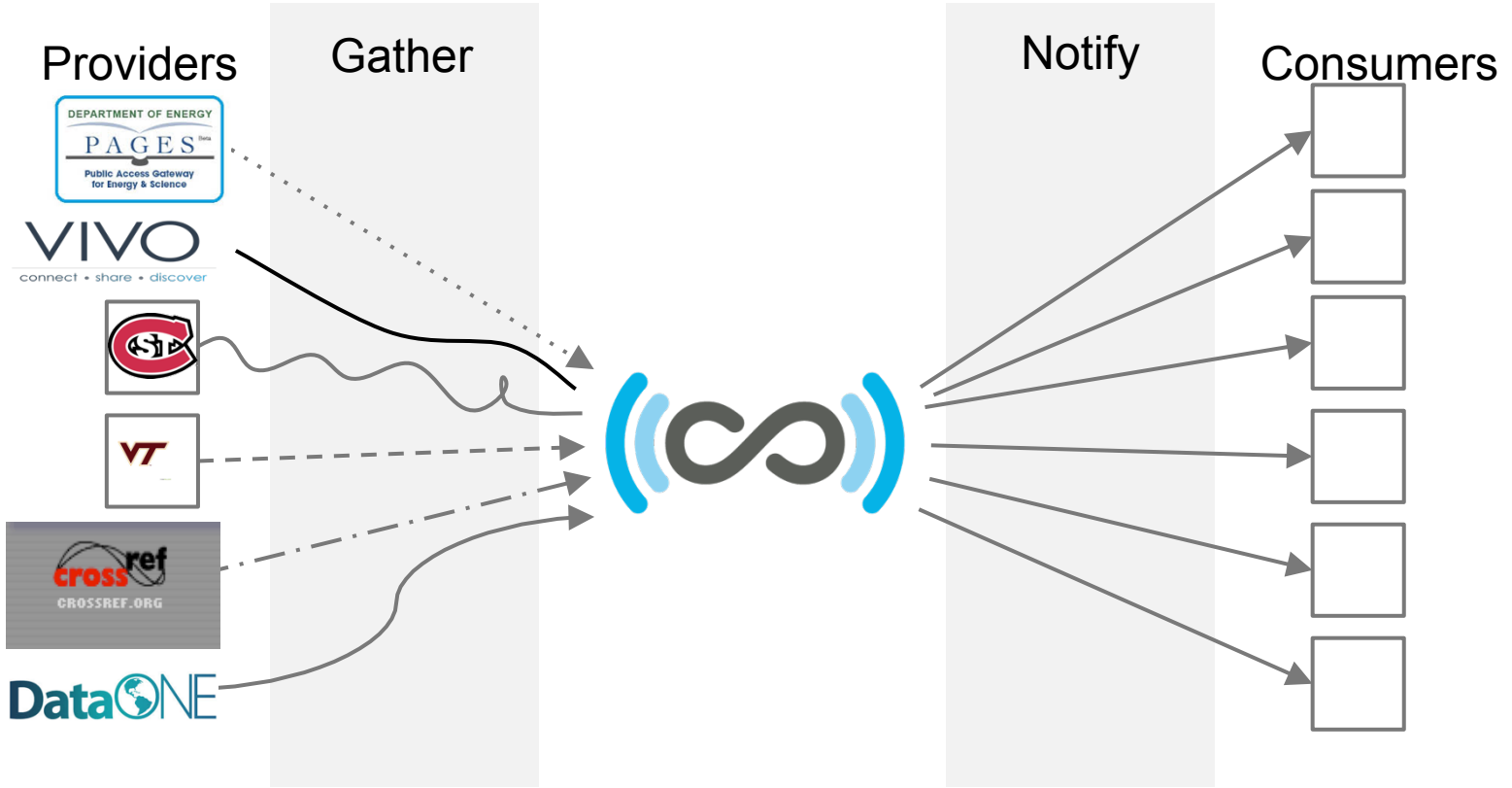
Herbert Van de Sompel

“Reminiscing About 15 Years of Interoperability Efforts” (DLib Magazine, Nov./Dec. 2015)

# Access is restricted/hidden by

- Licenses
- Terms of service
- Lack of programmatic access
- Pay walls
- Unavailability
- Inconsistency
- Low quality

SHARE is assembling and enhancing a free, open dataset of scholarly research activity so others can avoid these challenges.





# Example discovery interface <http://osf.io/share>

Notice: this is a public beta release

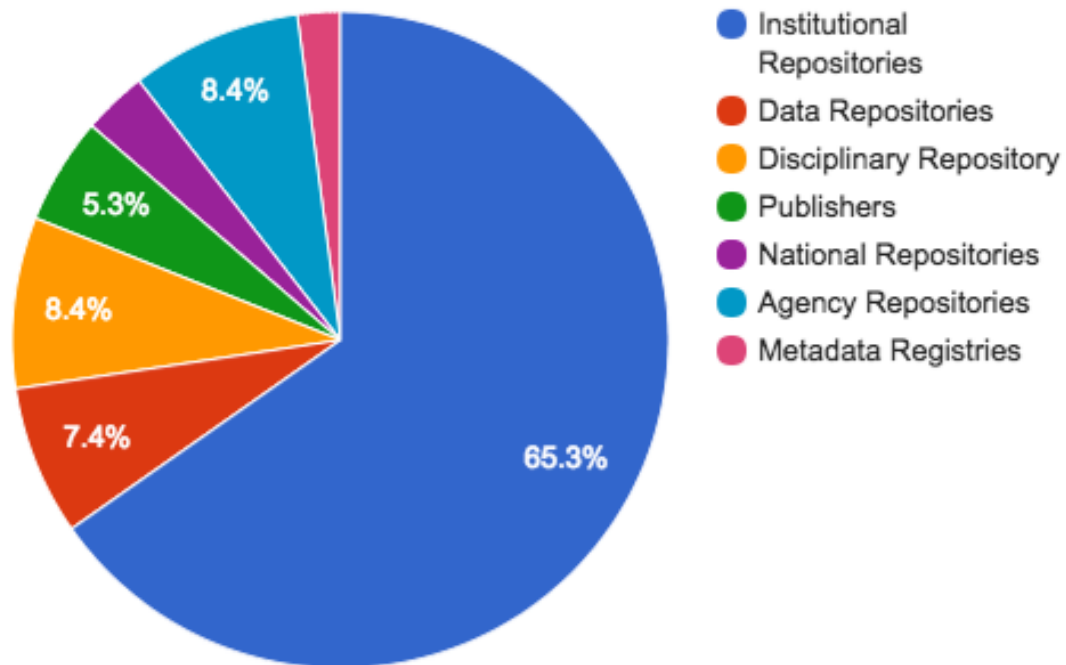


5,819,886 events as of Wed Apr 06 2016

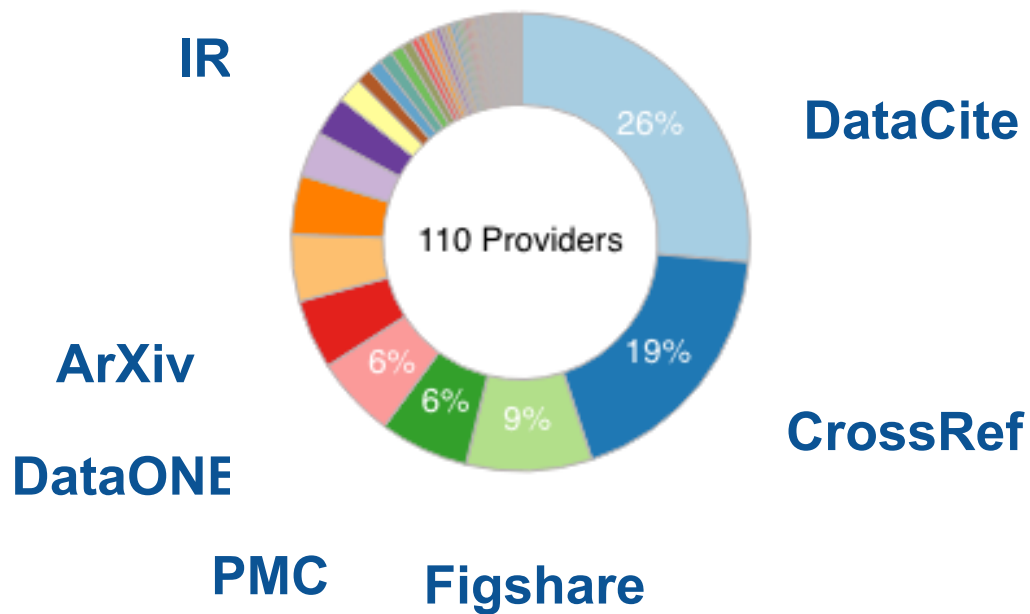
104 content providers

- Addis Ababa University Institutional Repository
- Apollo @ University of Cambridge
- Arizona State University Digital Repository
- ArXiv
- Biodiversity Heritage Library OAI Repository
- BioMed Central
- Calhoun: Institutional Archive of the Naval Postgraduate School
- CaltechAUTHORS
- Lake Winnipeg Basin Information Network
- London School of Hygiene and Tropical Medicine Research Online
- Mason Archival Repository Service
- MOspace: University of Missouri
- National Oceanographic Data Center
- Nature Publishing Group
- NeuroVault.org
- New Prairie Press at Kansas State University

# SHARE Notify Providers by Type



# SHARE Providers by % of 6 Million Records



# Normalization of metadata

- Arizona State University Digital Repository
- ArXiv
- Biodiversity Heritage Library OAI Repository
- BioMed Central
- Calhoun: Institutional Archive of the Naval Postgraduate School
- CaltechAUTHORS
- Carnegie Mellon University Research Showcase
- Chapman University Digital Commons
- CiteSeerX Scientific Literature Digital Library and Search Engine
- Cleveland State University's EngagedScholarship@CSU
- ClinicalTrials.gov
- Cognitive Sciences ePrint Archive
- Columbia Academic Commons
- CrossRef
- CU Scholar University of Colorado Boulder
- CyberLeninka - Russian open access scientific library
- DataCite MDS
- DataONE: Data Observation Network for Earth
- Deep Blue @ University of Michigan
- Department of Energy Pages

## Autism and tuberous sclerosis.

Autism is a behavior disorder with genetic influences indicated from twin and family studies and from the co-occurrence of autism with known genetic disorders. Tuberous sclerosis complex (TSC) is a known genetic disorder with behavioral manifestations including autism. A literature review of these two disorders substantiates a significant association...

Smalley, SL - Tanguay, PE - Smith, M - Gutierrez, G

Released on 2015-03-30 08:00 PM | Data

male tuberous sclerosis family child humans See All

eScholarship @ University of California

Raw Normalized

```
<record
  xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <header>
    <identifier>q0k7463tx</identifier>
    <timestamp>2015-08-21</timestamp>
  </header>
  <metadata>
    <oai_dc:dc
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
        <dc:identifier>q0k7463tx</dc:identifier>
        <dc:rights>Attribution (CC BY): http://creativecommons.org/licenses/by/3.0/</dc:rights>
        <dc:creator>Smalley, SL</dc:creator>
        <dc:creator>Tanguay, PE</dc:creator>
        <dc:creator>Smith, M</dc:creator>
        <dc:creator>Gutierrez, G</dc:creator>
        <dc:title>Autism and tuberous sclerosis.</dc:title>
        <dc:description>Autism is a behavior disorder with genetic influences indicated from twin and family studies and from the co-occurrence of autism with known genetic disorders. Tuberous sclerosis complex (TSC) is a known genetic disorder with behavioral manifestations including autism. A literature review of these two disorders substantiates a significant association of autism and TSC with 17-58% of TSC subjects manifesting autism and 0.4-3% of autistic subjects having TSC. In initial data collected on 13 TSC probands and 14 autistic probands in our family study of autism and TSC, we identified 7 TSC subjects with autism. The seven TSC autistic probands are similar to non-TSC autistic probands on the Social and Communication domains of the Autism Diagnostic Inventory (ADI) (Le Couteur et al., 1989), but show fewer Repetitive Rituals. There are more male TSC probands with autism than female, despite a 1 equal sex ratio among TSC probands. The TSC probands with autism have significantly more seizures and mental retardation than those without autism; however, the extent and etiology of associations require further study. Our preliminary findings suggest that a fruitful approach for delineating genetic influences in autism may come from further investigation of possible mechanisms underlying the association of autism and TSC.</dc:description>
        <dc:coverage>339 - 355</dc:coverage>
        <dc:subject>Humans</dc:subject>
        <dc:subject>Tuberous Sclerosis</dc:subject>
        <dc:subject>Family</dc:subject>
        <dc:subject>Autistic Disorder</dc:subject>
        <dc:subject>Genetics, Behavioral</dc:subject>
        <dc:subject>Interview, Psychological</dc:subject>
        <dc:subject>Sex Factors</dc:subject>
        <dc:subject>Child</dc:subject>
        <dc:subject>Child, Preschool</dc:subject>
        <dc:subject>Female</dc:subject>
        <dc:subject>Male</dc:subject>
      </oai_dc:dc>
    </metadata>
  </record>
```

- Arizona State University Digital Repository
- ArXiv
- Biodiversity Heritage Library OAI Repository
- BioMed Central
- Calhoun: Institutional Archive of the Naval Postgraduate School
- CaltechAUTHORS
- Carnegie Mellon University Research Showcase
- Chapman University Digital Commons
- CiteSeerX Scientific Literature Digital Library and Search Engine
- Cleveland State University's EngagedScholarship@CSU
- ClinicalTrials.gov
- Cognitive Sciences ePrint Archive
- Columbia Academic Commons
- CrossRef
- CU Scholar University of Colorado Boulder
- CyberLeninka - Russian open access scientific library
- DataCite MDS
- DataONE: Data Observation Network for Earth
- Deep Blue @ University of Michigan
- Department of Energy Pages

## Autism and tuberous sclerosis.

Autism is a behavior disorder with genetic influences indicated from twin and family studies and from the co-occurrence of autism with known genetic disorders. Tuberous sclerosis complex (TSC) is a known genetic disorder with behavioral manifestations including autism. A literature review of these two disorders substantiates a significant association...

Smalley, SL - Tanguay, PE - Smith, M - Gutierrez, G

Released on 2015-03-30 08:00 PM | Data

male tuberous sclerosis family child humans See All

eScholarship @ University of California

Raw Normalized

```
{
  "versions": [
    {
      "341cbf4b-9e95-49d6-94d6-dd38017e79e5",
      "b4f64043-df2a-46a1-913f-29c29e8b1fca",
      "e3184cba-7660-4538-99fe-4ef124b4ea5a",
      "67f28d5f-22f7-4ee6-93a9-5c6e38dd9a9d",
      "8f787696-08b8-436a-bd6b-50c17538a64c",
      "3669cc6e-e8c1-4c0a-8beb-055496b0845c"
    }
  ],
  "licenses": [],
  "publisher": {
    "name": "eScholarship, University of California"
  },
  "title": "Autism and tuberous sclerosis.",
  "shareProperties": {
    "source": "ucescholarship"
  },
  "otherProperties": [
    {
      "properties": {
        "type": "article"
      }
    },
    {
      "name": "type"
    },
    {
      "properties": {
        "format": "application/pdf"
      }
    },
    {
      "name": "format"
    },
    {
      "properties": {
        "date": "1992-01-01"
      }
    },
    {
      "name": "date"
    },
    {
      "properties": {
        "identifier": [
          "q0k7463tx",
          "http://www.escholarship.org/uc/item/q0k7463tx",
          "q0k7463tx"
        ]
      }
    }
  ]
}
```

## Schema

Toggle view: **View** Compare

## Menu

## Project Wiki Pages

- Home
- API Documentation
- Code Review
- Creating a Harvester
- Error Recovery
- Feed Options
- Glossary of Terms
- Metadata Analysis
- OSF
- Overview
- Running the Harvesters
- Schema**
- scrAPI
- Setting up scrAPI

## View

(Current) Erin Braswell: 2015-11-17 20:56:27 UTC

## Table of Contents

- Beta SHARE Metadata Schema
  - Current Schema
  - Example

<https://osf.io/wur56/wiki/schema/>

## Beta SHARE Metadata Schema

The beta SHARE metadata schema is based on a [json-schema](#). Details, code generating schemas, examples and more information available on the [SHARE Schema repo on github](#).

Many of the fields in this schema are not required. The only required fields include:

- title
- contributors
- uris
- providerUpdatedDateTime

See the schema below for details on each of the fields.

## Current Schema



Here is the current schema:

```
definitions:
  otherProperties:
    type: "object"
  properties:
    name:
      description: A name that describes this collection of properties
```



# Open SHARE API

← → ↻ <https://osf.io/dajtq/wiki/SHARE%20Search/> ☆ 🍷 🗄️ ☰

 Open Science Framework My Dashboard Browse ▾ Help ▾ 🔍  Judy Ruttenberg ⚙️ ↻

↑ [SHARE API Documentation](#) Files **Wiki** Analytics Registrations Forks

– 📁 Project Wiki Pages

📄 Home

📄 Push API

📄 **SHARE Search**

## SHARE Search

Visit the SHARE search API endpoint [here](#)

For security reasons, unencrypted HTTP access is not permitted.

The `/api/v1/share/search/` endpoint supports both GET and POST requests. The data returned by both endpoints is JSON, by default in the format:

```
{
  "count": # of metadata documents found,
  "results": list of documents corresponding to the SHARE schema,
  "time": Amount of time the query took
}
```

Both endpoints have the following query parameters in common:

count

If the `count` parameter is defined, all other parameters except for `q` are ignored. The `results` field of the response is always empty, and the `count` field corresponds to the number of matches for the query.

raw

If the `raw` parameter is defined, the results of the request will be the raw Elasticsearch results, rather than the processed version normally returned. The Elasticsearch results will usually be of following form (Note: for most use cases, the filtered form of the API will be adequate



Public

0

0

SHARE /

# SHARE Examples and Tutorials

Contributors: [Erin Braswell](#)

Date created: 2016-02-01 09:44 AM | Last Updated: 2016-02-22 01:09 PM

Category: Uncategorized

## <https://osf.io/bygau/>

Wiki



## SHARE Examples and Tutorials

Here you'll find links to three Jupyter notebooks that explore three areas of using the SHARE API:

- [Basic API Queries](#)
- [Complex Queries and Basic Visualization](#)
- [Using External Data](#)

Condensed notebook from the SHARE Webinar on Feb 22:

- [Querying the SHARE API](#)

Files



Filter



Name ^ ▾

Citation

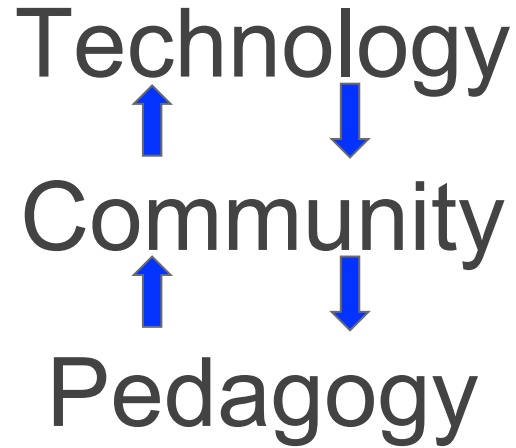
osf.io/bygau ▾

Recent activity

All times displayed at -0500 UTC offset.

- 2016-02-22 01:09 PM** [Erin Braswell](#) updated wiki page [home](#) to version 3 of [SHARE Examples and Tutorials](#)
- 2016-02-22 10:57 AM** [erinspace](#) added file [SHARE API Webinar.ipynb](#) to GitHub repo [erinspace / share\\_tutorials](#) to [SHARE Examples and Tutorials](#)
- 2016-02-08 01:47 PM** [Erin Braswell](#) updated wiki page [home](#) to version 2 of [SHARE Examples and Tutorials](#)
- 2016-02-01 02:23 PM** [Erin Braswell](#) updated file [README.md](#) to GitHub repo [erinspace / share\\_tutorials](#) in [SHARE Examples and Tutorials](#)
- 2016-02-01 02:19 PM** [erinspace](#) removed file [3\\_SHARE\\_Data\\_in\\_the\\_Wide\\_World.ipynb](#) from GitHub repo [erinspace / share\\_tutorials](#) from [SHARE Examples and Tutorials](#)

**There is *a lot* of work to do**



# Technology

Data processing pipeline

Provenance modeling

Harvesters for metadata providers

Automated metadata enhancement techniques

Curation interface

Curation pipeline (e.g., validation rules)

Discovery interface

APIs

# SHARE Community

Joint Working Group and Task Groups

Users

Data providers

Code contributors

Hackers

Pilot Institutions

Consortia

Curation Associates

Related systems  
(e.g. VIVO)

# Users and Use Cases

- Tracking research by particular people or institutions
- Inter-repository linking: data and publications
- Intervention in sponsored research workflow
- Meta-scholarship



## SHARE to Offer Webinar for Developers on Querying SHARE Data Set of Research Events

Judy Ruttenberg | [202-296-2296](tel:202-296-2296) | [judy@arl.org](mailto:judy@arl.org) | March 3, 2016

The SHARE team is hosting a [free webinar](#) for the higher education and library developer community on Thursday, March 17, 2016, at 2:00–3:30 p.m. EDT. Erin Braswell from the Center for Open Science—and SHARE’s primary developer—will introduce the SHARE application programming interface (API), present how to perform queries and visualizations with the data, and discuss potential uses for the data in universities and libraries, such as querying the data set for new research from your institution.



## SHARE to Offer Webinar for Developers on Querying SHARE Data Set of Research Events

Judy Ruttenberg | [202-296-2296](tel:202-296-2296) | [judy@arl.org](mailto:judy@arl.org) | March 3, 2016

The SHARE team is hosting a [free webinar](#) for the higher education and library developer community on Thursday, March 17, 2016, 10:00 AM EDT. The webinar, led by Judy Ruttenberg from the Center for Open Science—and SHARE's primary developer—will introduce the SHARE application programming interface (API), present how to perform queries and visualizations with the data, and discuss potential uses for the data in universities and libraries, such as querying the data set for new research from your institution.

<http://bit.ly/1ToHJt8>



## **Pedagogy: SHARE Curation Associates**

An opportunity for library professionals to develop digital curation and computational thinking skills to enhance local institutional repositories in a service-learning setting. Together, we can advance library digital competencies and move scholarly communications forward.

# **Pedagogy: SHARE Curation Associates**

Information extraction through application programming interfaces (API) and OAI-PMH feeds

Building content harvesters

Basic programming in Python

Methods and tools to automate data cleaning and metadata enhancements

**All tools used will be open source.**

# Pedagogy: Demand for Open APIs

- Digital scholarship workshops
- Computer science students
- Internships

SHARE provides an opportunity to advance an academy-friendly scholarly communication agenda in a service learning context

# Participating TDL sources

- Texas A&M
- Texas State University
- University of Texas -Austin

# Registering as a new provider



To include your organization's publications, data, repository holdings, or other research in [SHARE](#), please register here.

[Register as a Provider](#)

Watch a [90-second video](#) that shows how easy it is to register to become a SHARE metadata provider.

# Thank you!

[www.share-research.org/updates](http://www.share-research.org/updates)



info@share-research.org

@SHARE\_research