

Progress in including research software in the scholarly communication ecosystem

22 March 2024

Daniel S. Katz, University of Illinois at Urbana-Champaign, d.katz@ieee.org
Morane Otilia Gruenpeter, Software Heritage, morane@softwareheritage.org
Kristina Vrouwenvelder, AGU, kvrouwenvelder@agu.org
Daniel Garijo, Universidad Politécnica de Madrid, daniel.garijo@upm.es



<https://doi.org/10.5281/zenodo.10849689>



Outline

1. Introduction - Dan
2. Software & metadata - Morane
3. Software & publications - Kristina
4. Software & repositories & registries - Daniel
5. Questions and discussion - all

1 - Introduction

Daniel S. Katz

Importance of research software

- Studies [1, 2, 3] show
 - >90% of researchers use research software in their work
 - >50% of researchers develop software as part of the research
 - ~50% of research grants produce software
- Software is as important as a paper or dataset for full understanding and dissemination of research
 - It must be properly entered in the scholarly record
- Finding and accessing software helps further research & provides a means for other researchers to access software in order to
 - support proper attribution and credit (similar to that of papers, data, etc.)
 - enable peer-review, validation, and reproducibility of findings
 - support collaboration and reuse
 - encourage building on the work of others

Research software citation

Two FORCE11 working groups:

- Software Citation Working Group (2015-16): Developed [Software Citation Principles](#)
- Software Citation Implementation Working Group (2017-2023): Developed [best practices](#) for software citation
 - [Software citation implementation challenges](#)
 - Checklists for [\(paper\) authors](#) and [\(software\) developers](#)
 - [Best practices for software repositories and registries](#)
 - [Guidance for journals](#)

Today:

- Citing software is developing as a common practice
- But journals and editors still need consistent guidance, and to provide this to authors

We will discuss the types of software citation, challenges, and recommended approaches

FAIR for research software

- [FAIR for Research Software \(FAIR4RS\)](#) convened in 2021 as joint Research Data Alliance (RDA) Working Group, FORCE11 Working Group, and Research Software Alliance (ReSA) Task Force
- Via a community-intensive (>500 participants) iterative process, over 18 months developed
 - Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A-L, Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2022). FAIR Principles for Research Software version 1.0. (FAIR4RS Principles v1.0). Research Data Alliance. DOI: <https://doi.org/10.15497/RDA00068>
 - Introduced and discussed in <https://doi.org/10.1038/s41597-022-01710-x>
- Revises and extends original FAIR (data) principles; needed because of [inherent differences](#) between research software and research data (software can be stored as data, but it's not just data)

Common challenges remain

- Technical and social
- Particularly related to metadata, archiving, versions
 - Creator/publisher sequence for data doesn't typically apply to software
 - Where is software metadata stored? (in code repository for open source?, for closed source?, in archival repository?, in registry?)
 - Where is software archived? (GitHub/Gitlab are not archival, registries are not archival, repositories? Software Heritage?)
 - Different use cases need specific version, latest version, all versions
- Getting best practices actually into practice
- Lots of work needed beyond citation and FAIR
 - Quality, correctness, reproducibility, openness, community support, ...

2. Software & metadata

Morane Gruenpeter

<https://doi.org/10.5281/zenodo.10849689>

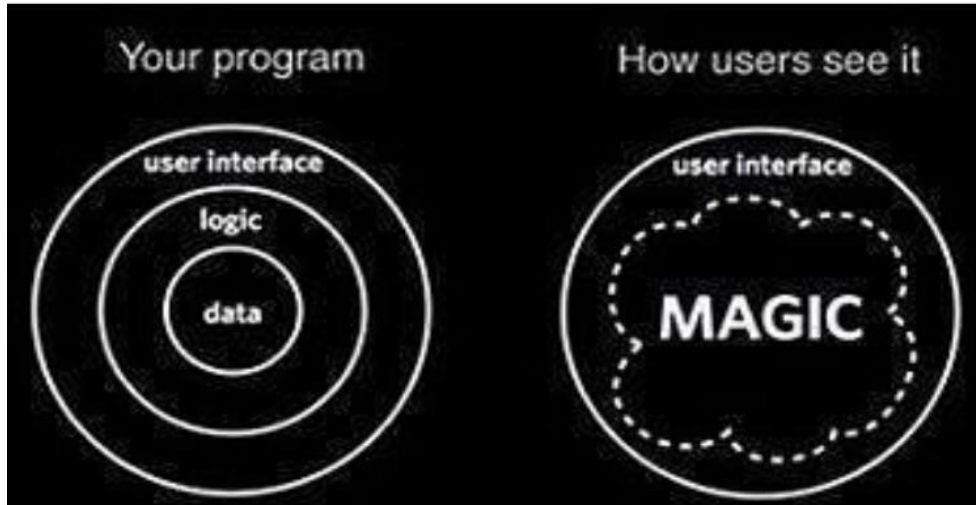


What is software? Clarifying the magic...



worldofprogrammers

...



https://www.reddit.com/r/ProgrammerHumor/comments/70fuamp/programming_is_magic/

Software as a concept

- **project** or entity
- the **community** around the project
- the software **idea** / algorithms / solutions

Not a digital artifact

Software artifacts

- Executables
 - For multiple environments
- Source code

A very large collection of digital artifacts

The Knowledge is in the Source Code

"Source code provides a view into the mind of the designer."

Len Shustek, 2006

Computer History Museum

Apollo 11 source code (excerpt)

```
P63SPOT3      CA   BIT6      # IS THE LR ANTENNA IN POSITION 1 YET
              EXTEND
              RAND   CHAN33
              EXTEND
              BZF    P63SPOT4      # BRANCH IF ANTENNA ALREADY IN POSITION 1

              CAF    CODE500      # ASTRONAUT:  PLEASE CRANK THE
              TC     BANKCALL      #              SILLY THING AROUND
              CADR    GOPERF1
              TCF     GOTOP00H     # TERMINATE
              TCF     P63SPOT3     # PROCEED      SEE IF HE'S LYING

P63SPOT4      TC     BANKCALL      # ENTER        INITIALIZE LANDING RADAR
              CADR    SETPOS1

              TC     POSTJUMP      # OFF TO SEE THE WIZARD ...
              CADR    BURNBABY
```



René Magritte, 1933, *The Human Condition*

*"Programs must be written for people to read,
and only incidentally for machines to execute."*

Harold Abelson, 1985

Structure and Interpretation of Computer Programs (1st ed.),

<https://doi.org/10.5281/zenodo.10849689>

Software source code is not “just” data

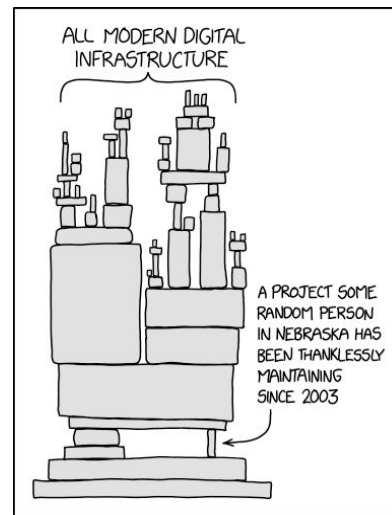
Software evolves over time

- projects may last decades
- the development history is key to its understanding

Complexity

- millions of lines of code
- large web of dependencies
- easy to break, difficult to maintain
- sophisticated developer communities

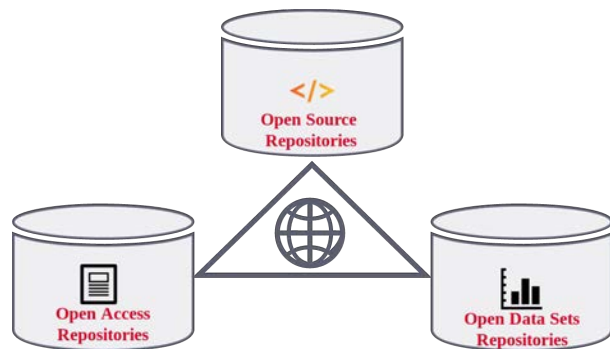
**research software is a thin layer
on top of the global software stack**



https://www.reddit.com/r/ProgrammerHumor/comments/c1zmc/dependency_xkcd/

Software is a pillar of Open Science

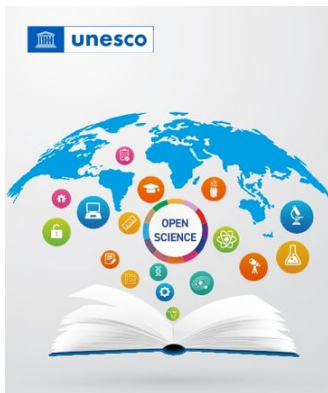
A key pillar: software (source code)



Three pillars of Open Science
Software Heritage CC-BY 4.0 2019

Software has multiple facets:

- a **tool**
- a research **outcome** or result
- the **object** of research



UNESCO Recommendation
on Open Science

“Open Scientific knowledge [includes] open source software: source code must be included in the software release and made available on openly accessible repositories”.

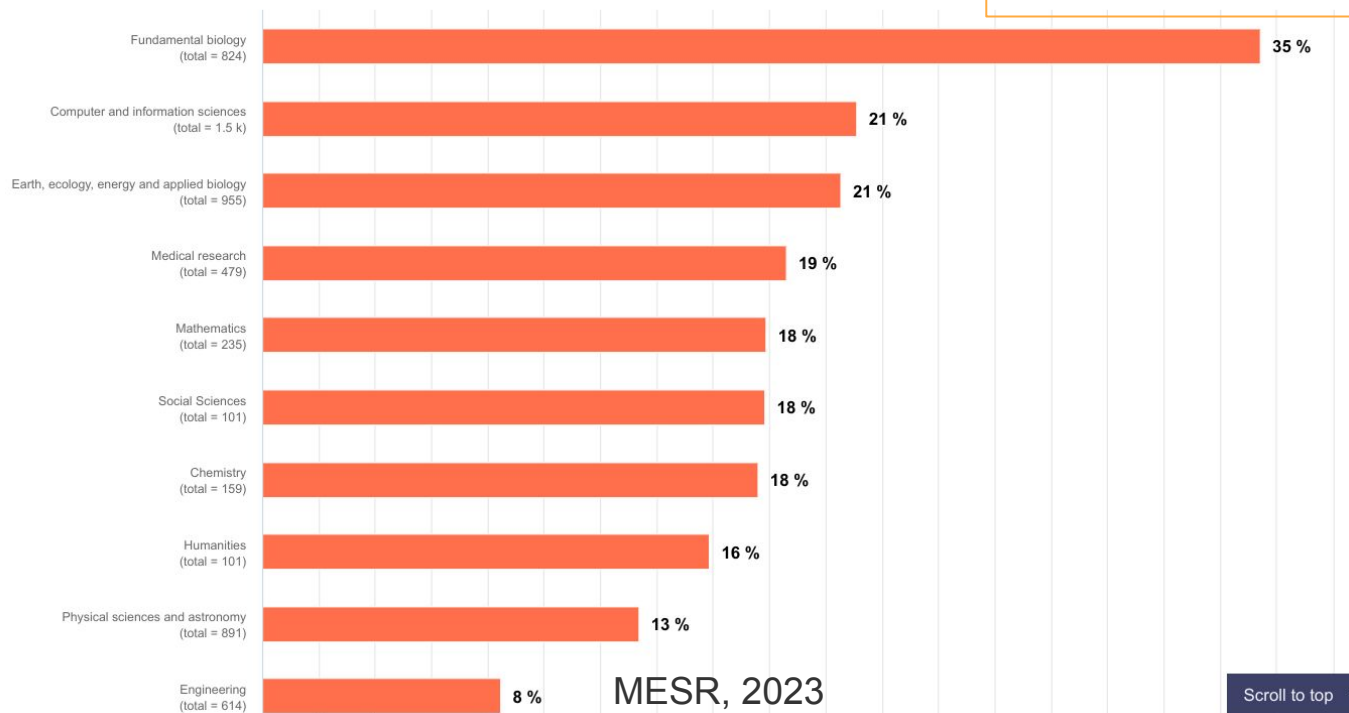
Open Source in [UNESCO recommendations for Open Science](#), November 2021

<https://doi.org/10.5281/zenodo.10849689>

Proportion of publications in France that mention code or software sharing by discipline

Sort by:

☐ Highest volume ☒ Highest sharing rate



MESR, 2023

Scroll to top ↑

Research, across all disciplines, heavily depends on software!

Proportion of French publications mentioning use of code or software, by discipline (2023 data from <https://barometredelascienceouverte.esr.gouv.fr/>)

<https://doi.org/10.5281/zenodo.10849689>

What is at stake?

Archive

- make sure we can access to retrieve the software (reproducibility)

Reference

- make sure we can identify the software artifacts (reproducibility)

Describe

- make it easy to discover the software projects (visibility)

Cite (for credit)

- make it rewarding to create software by giving credit to authors (evaluation!)

Where is the metadata available ?

Software development platforms

(on platform page)

- GitHub
- Bitbucket
- SourceForge
- ...

Catalogs and registries

- ASCL
- swMath
- OpenAire
- libraries.io
- ...

Scholarly repositories

- Zenodo (InvenioRDM)
- HAL
- Institutional repositories
- ...

Package managers

- PyPI
- NPM
- ...

Scholarly publishers

- IPOL
- eLife
- Dagstuhl
- Episciences
- ...



In the source code (as a file)

- README
- LICENSE
- AUTHORS
- Package manager file
- codemeta.json
- CFF
- ...

<https://doi.org/10.5281/zenodo.10849689>

Mapping metadata with CodeMeta

→ A subset of schema.org, an academic community & a crosswalk table

→ In 2023

- ◆ New [governance model](#)
- ◆ [version 3.0](#) released 🚀
- ◆ A new open [w3id](#) identifier
- ◆ [Open source tool](#)

CodeMeta generator

Most fields are optional. Mandatory fields will be highlighted when generating Codemeta.

The software itself


Name
My Software
the software title

Description
My Software computes ephemerides and orbit propagation. It has been developed from early '99.

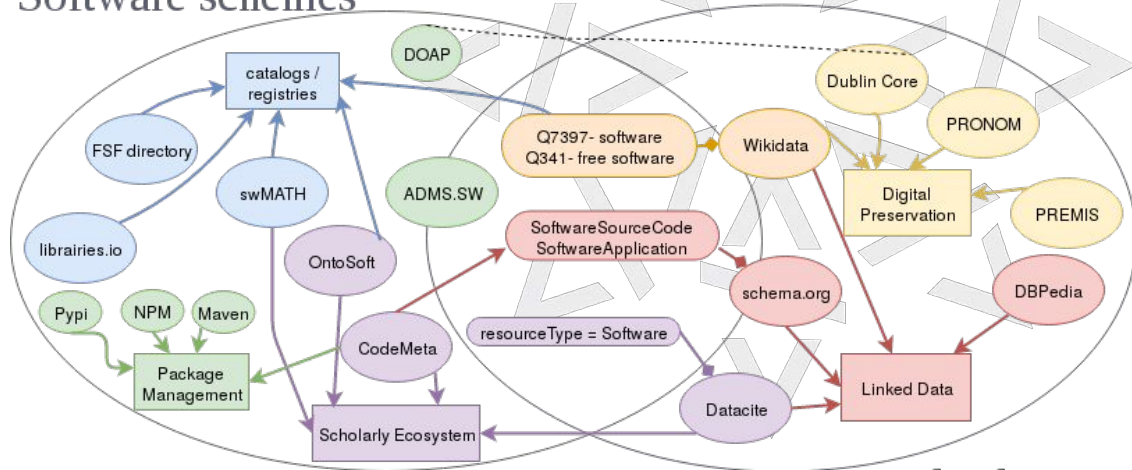
Creation date
YYYY-MM-DD

First release date
YYYY-MM-DD

Contributed to the community by

 Software Heritage
THE GREAT LIBRARY OF SOURCE CODE

Software schemes



General schemes

Gruenpeter M. and Thornton K. (2018) Pathways for Discovery of Free Software (slide deck from LibrePlanet 2018). <https://en.wikipedia.org/wiki/File:Pathways-discovery-free.pdf> accessed on 6.11.2020.

Next steps - Improve the CodeMeta mappings:
A Simple Standard for Sharing Ontology Mappings [by N. Matentzoglou](#)

<https://doi.org/10.5281/zenodo.10849689>

Keeping humans in the loop: moderation on [HAL](#)

Software deposits in the French national archive, HAL, are moderated by archivists who verify the metadata quality.

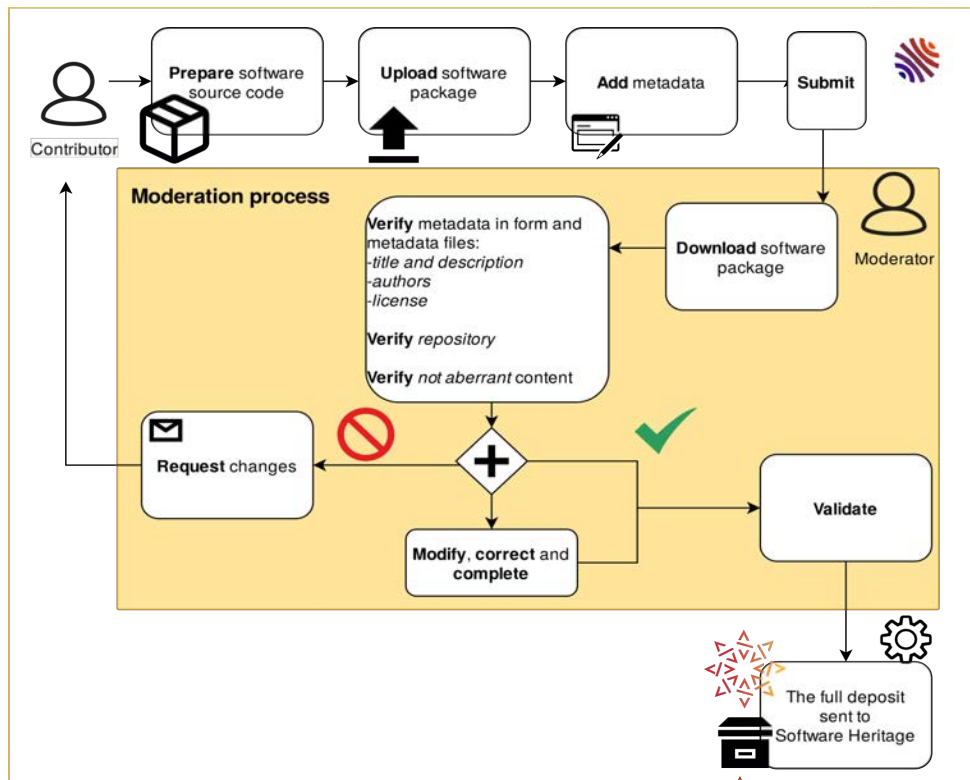
When credit is at stake, automation/crowdsourcing is not enough!

- Humans are needed to get quality information

Just Metadata is not enough!




HAL's citation format

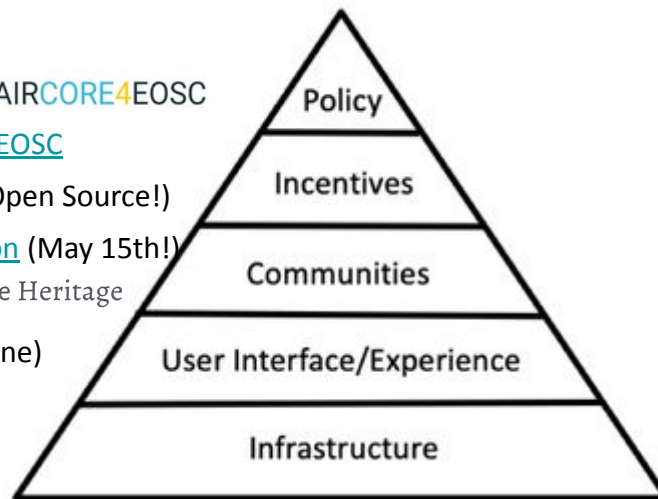
Matteo Frigo, Mauro Zucchelli, Rachid Deriche, Samuel Deslauriers-Gauthier. TALON: Tractograms As Linear Operators in Neuroimaging. 2021.
<https://hal.archives-ouvertes.fr/hal-03116143:visit=sw:1.snp:465d89956196578717f4cb5155e456c279aa6a22:anchor=sw:1.rev:10247a14640a280b9140a27ce003d382d70cccac:path=/hal-03116143>



<https://doi.org/10.5281/zenodo.10849689>

Road ahead

- Research Software Metadata Guidelines:  **FAIR-IMPACT**
Expanding FAIR solutions across EOSC
 - Join the Open Call by EU project FAIR-IMPACT to implement the guidelines:
<https://fair-impact.eu/support-offer-1-assessment-and-improvement-research-software>
- Join the community discussing software metadata:
<https://github.com/codemeta/codemeta/discussions>
- Research Software APIs and connectors:  **FAIRCORE4EOSC**
 - Implemented as part of the EU project [FAIRCORE4EOSC](#)
 - You can apply to [test the components](#) (which are Open Source!)
 - Come to the [RDA P22 Software Source Code session](#) (May 15th!)
- Archive any software on Software Heritage:  Software Heritage
 - save.softwareheritage.org (like the wayback machine)
- Create Metadata files using the CodeMeta generator
 - <https://codemeta.github.io/codemeta-generator/>



Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible

Pyramid from Strategy for Culture Change: Brian Nosek (2019)
<https://www.cos.io/blog/strategy-for-culture-change>

Thank you!

morane@softwareheritage.org

<https://doi.org/10.5281/zenodo.10849689>

3. Software & publishing

Kristina Vrouwenfelder

<https://doi.org/10.5281/zenodo.10849689>

Strategies for Sharing Research Software: A Publisher Perspective

Kristina Vrouwenvelder, Ph.D
Assistant Director, Publications, &
AGU Open Science
ORCID: 0000-0002-5862-2502
kvrouwenvelder@agu.org



24 journals

60,000 members across 137
countries

Annual meeting of ~25k
researchers

Nonprofit scientific society governed by Board of Directors
and Council.

Open Science at AGU includes:

Shelley Stall, VP of Open Science Leadership

Brian Sedora, Sr Program Manager, Pubs and Open Science

Sophie Hanson, Program Specialist, Data and Publications

<https://doi.org/10.5281/zenodo.10849689>

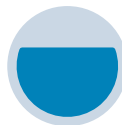


Open Science at AGU



Inclusion and Equity

We're lowering barriers to participation in science through diverse efforts:



Open Books and Preprints

We publish open books through a partnership with **Geological Society of London** and host a free preprint community server, **ESS Open Archive**.



Open Access Journals

All AGU journals offer open access options; 12 are fully open and transition to full OA is ongoing!

Community
Science
Exchange

Global
Research
Inclusion
Policy

Co-Reviewer
and
Mentoring
programs

& more!

<https://doi.org/10.5281/zenodo.10849689>

Open Data and Software at AGU



2017: AGU partners with community, COPDESS on enabling FAIR Data project (copdess.org).

2021: Data and Software Sharing Guidance for Authors Submitting to AGU Journals published.



2022: First AGU journal (*GeoHealth*) enters **Data Citation Pilot**.

Aug 2023: All AGU journals entered in Data Citation Pilot.

 **OpenTopography**
High-Resolution Topography Data and Tools

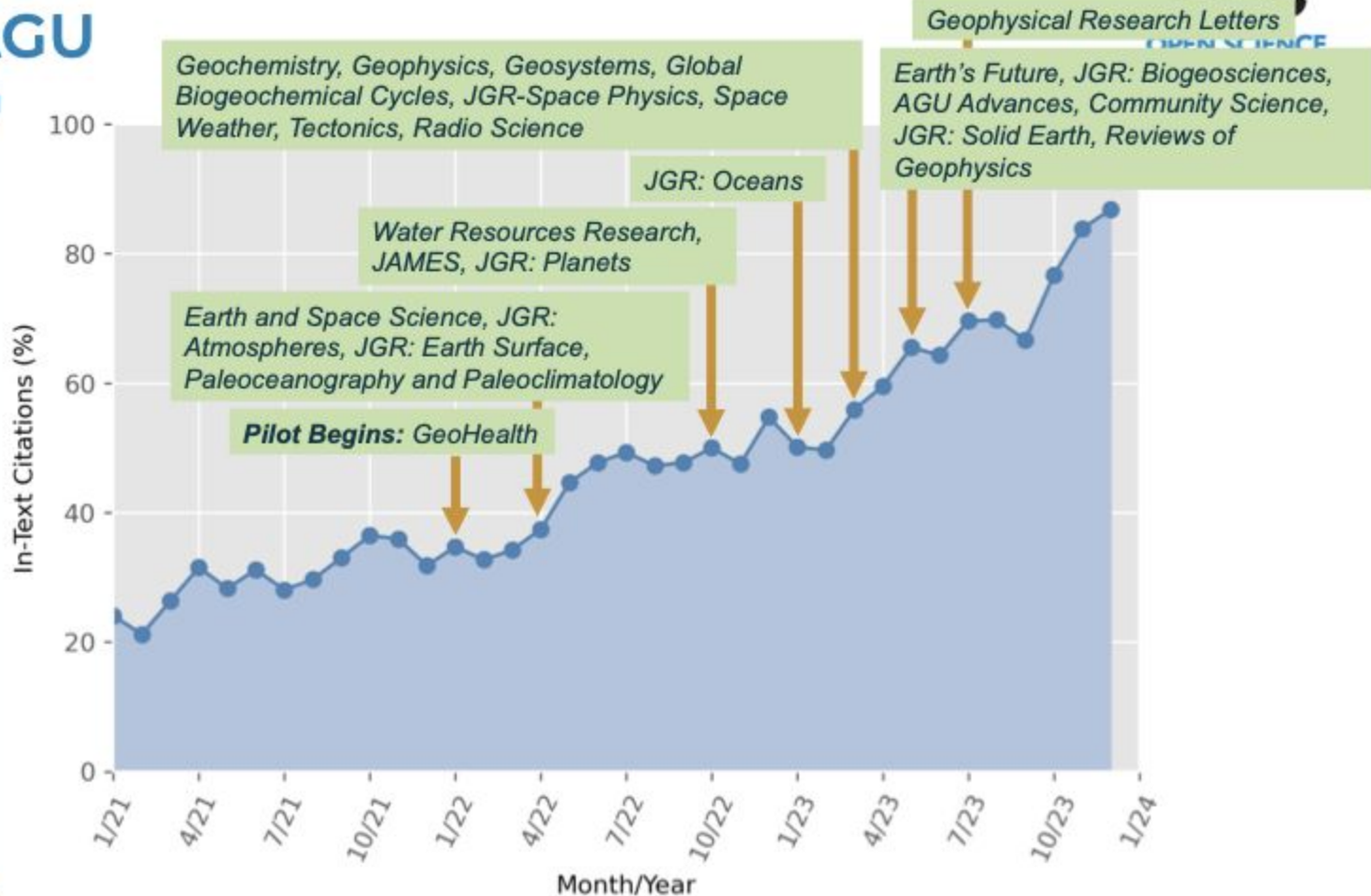
Dec 2023: First **Open Science Recognition Prizes** are awarded at AGU23.



Data & Software Citation at AGU

Authors are asked to:

1. Deposit data/software in a community accepted, trusted repository, preferably with a DOI
2. Include an Availability Statement: where and how to access the data and software
3. Include citation(s) to data/software in the Reference Section



What does successful citation look like?

The article might...

- contain data/code in SI
- describe how to access authors' data/code in text, with link
- above, + description of how to find any reused data/code, with link
- Cite authors' data/code in References section with DOI
- Above, + references for reused data/code!

<https://doi.org/10.5281/zenodo.10849689>



What does successful citation look like?

The article might...

- contain data/code in SI
- describe how to access authors' data/code in text, with link
- above, + description of how to find any reused data/code, with link
- Cite authors' data/code in References section with DOI
- Above, + references for reused data/code!

Data:

Edmunds, P. J., Didden, C., & Frank, K. (2021). Mean percentage cover of corals and Porites astreoides at each site by year at St. John, VI from 1992 to 2019 (Version 1) [Dataset]. Biological and Chemical Oceanography Data Management Office (BCO-DMO). <https://doi.org/10.26008/1912/BCO-DMO.843284.1>

Software:

Shobe, C. (2023). Code and data for "The uncertain future of mountaintop-removal-mined landscapes 1: How mining changes erosion processes and variables" (v1.0) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.10059514>

Include bracketed description!

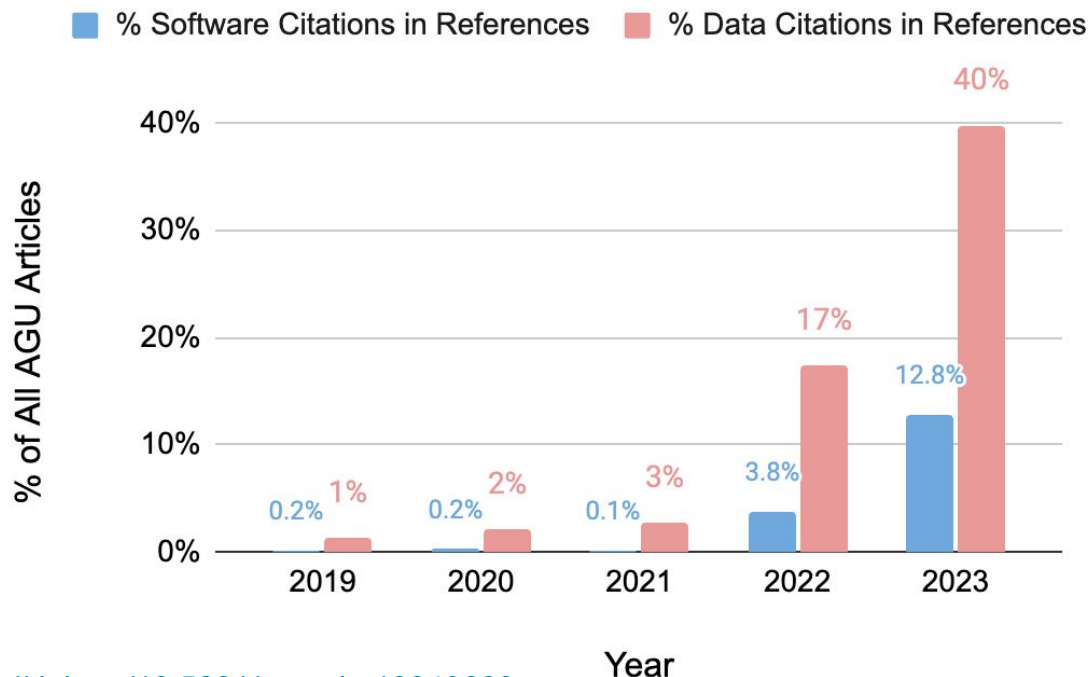
See Stall et al, 2023, "Journal Production Guidance for Software and Data Citations". DOI: [10.1038/s41597-023-02491-7](https://doi.org/10.1038/s41597-023-02491-7)



What does successful citation look like?

The article might...

- contain data/code in SI
- describe how to access authors' data/code in text, with link
- above, + description of how to find any reused data/code, with link
- Cite authors' data/code in References section with DOI
- Above, + references for reused data/code!



<https://doi.org/10.5281/zenodo.10849689>

Author Resources and Education

*Workshops,
presentations, &
more in our Zenodo
community:*



*Introduction to Open
Science course
materials*



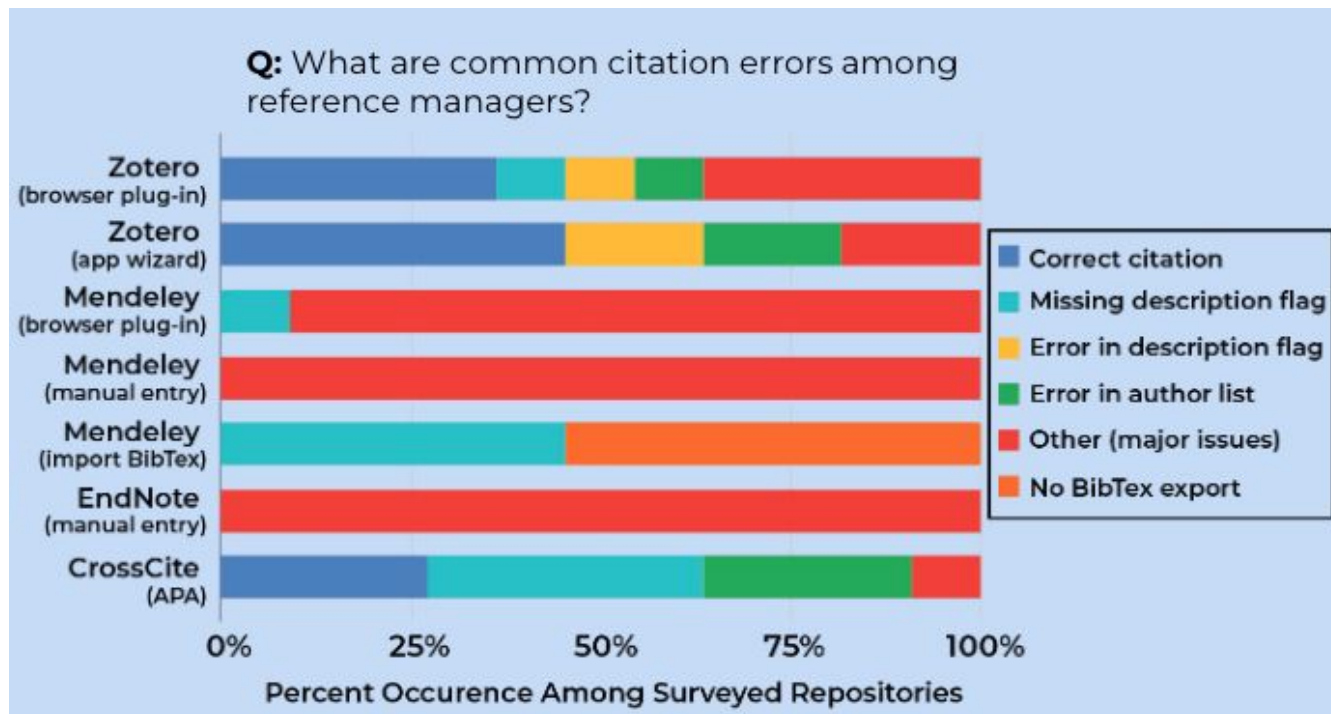
*AGU's full author
guidance for data
and software
citation*

*Postcards: Skills for
Open Science*



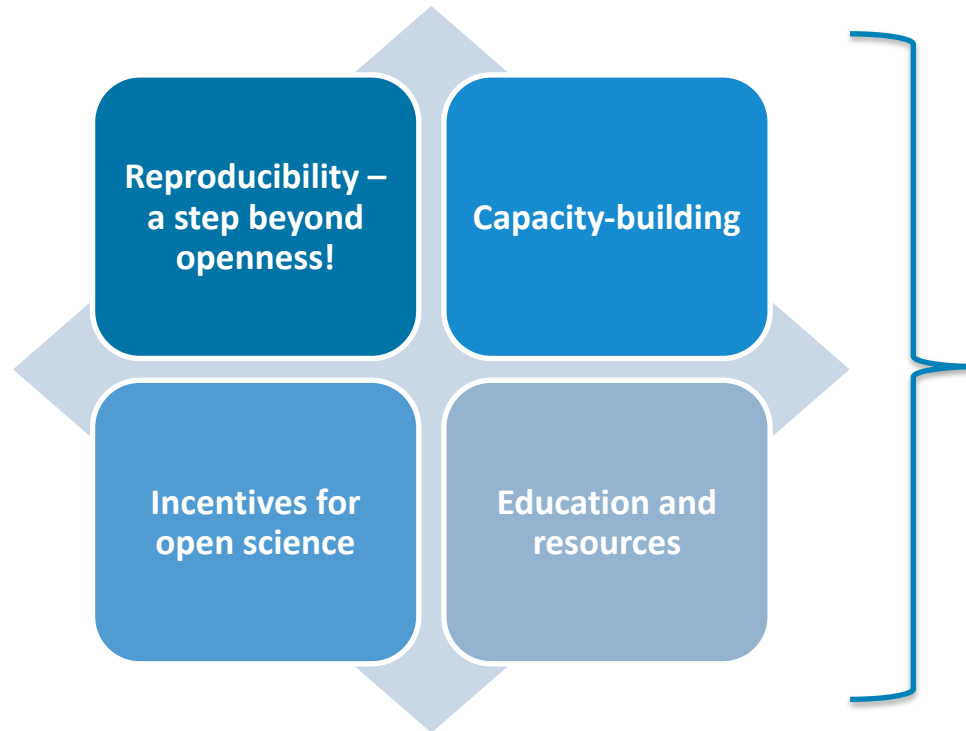
Please feel free to send us comments and suggestions for our resources!

Barriers to Data and Software Sharing



See Raia, Natalie; Vrouwenvelder, Kristina (2024). Barriers to Data and Software Sharing: The Reference Manager Gap. ESIP. Poster. <https://doi.org/10.6084/m9.figshare.25012655.v1>

Challenges in open publishing



*For researchers;
editors, peer
reviewers – **and** for
publishing staff,
infrastructure
providers, and
repositories!*

4. Software registries & repositories

Daniel Garijo

Consortium of scientific software registries and repositories

<https://scicodes.net/>

- Open community of **registries and repositories** in different domains
 - 37 resources (registries/repositories)
- Forum to discuss **best practices**
- Collaborative work towards **Codemeta** adoption
- Monthly meetings & presentations
 - Interested? Join us! <https://scicodes.net/meetings/>



<https://doi.org/10.5281/zenodo.10849689>

Scicodes: The story so far

<https://asclnet.github.io/SWRegistryWorkshop/>

“This workshop’s goals are to demonstrate unique aspects of our respective services, discuss challenges and share solutions to **common issues** that arise in managing our resources, **finalize a list of best practices for our resources**, and work cooperatively to **speed adoption of the CodeMeta and/or CFF standards**”



<https://doi.org/10.5281/zenodo.10849689>

Representation of resources in the original workshop

14 participating resources (different disciplines)

| Question | #Yes | #No | #Other |
|---|------|-----|--------|
| Is the resource discipline-specific? | 6 | 8 | 0 |
| Does the resource accept software only? | 8 | 6 | 0 |
| Does the resource require a software deposit? | 2 | 12 | 0 |
| Does the resource accept software deposits? | 10 | 4 | 0 |
| Can the resource mint DOIs? | 6 | 8 | 0 |
| Is the resource actively curated? | 10 | 1 | 3 |
| Can the resource be used to cite software? | 11 | 2 | 1 |

Creating best practices



Workshop



Arxiv report



PeerJ



 **FORCE11**

<https://github.com/force11/force11-sciwg>

<https://scicodes.net/>

<https://doi.org/10.5281/zenodo.10849689>

Consider these practices when creating a new resource

1. Provide a **public scope** statement
2. Provide guidance for **users**
3. Provide guidance to software **contributors**
4. Establish an **authorship policy**
5. Document and share your **metadata schema**
6. Stipulate **conditions of use**
7. State a **privacy policy**
8. Provide a **retention policy**
9. Disclose **end-of-life policy**



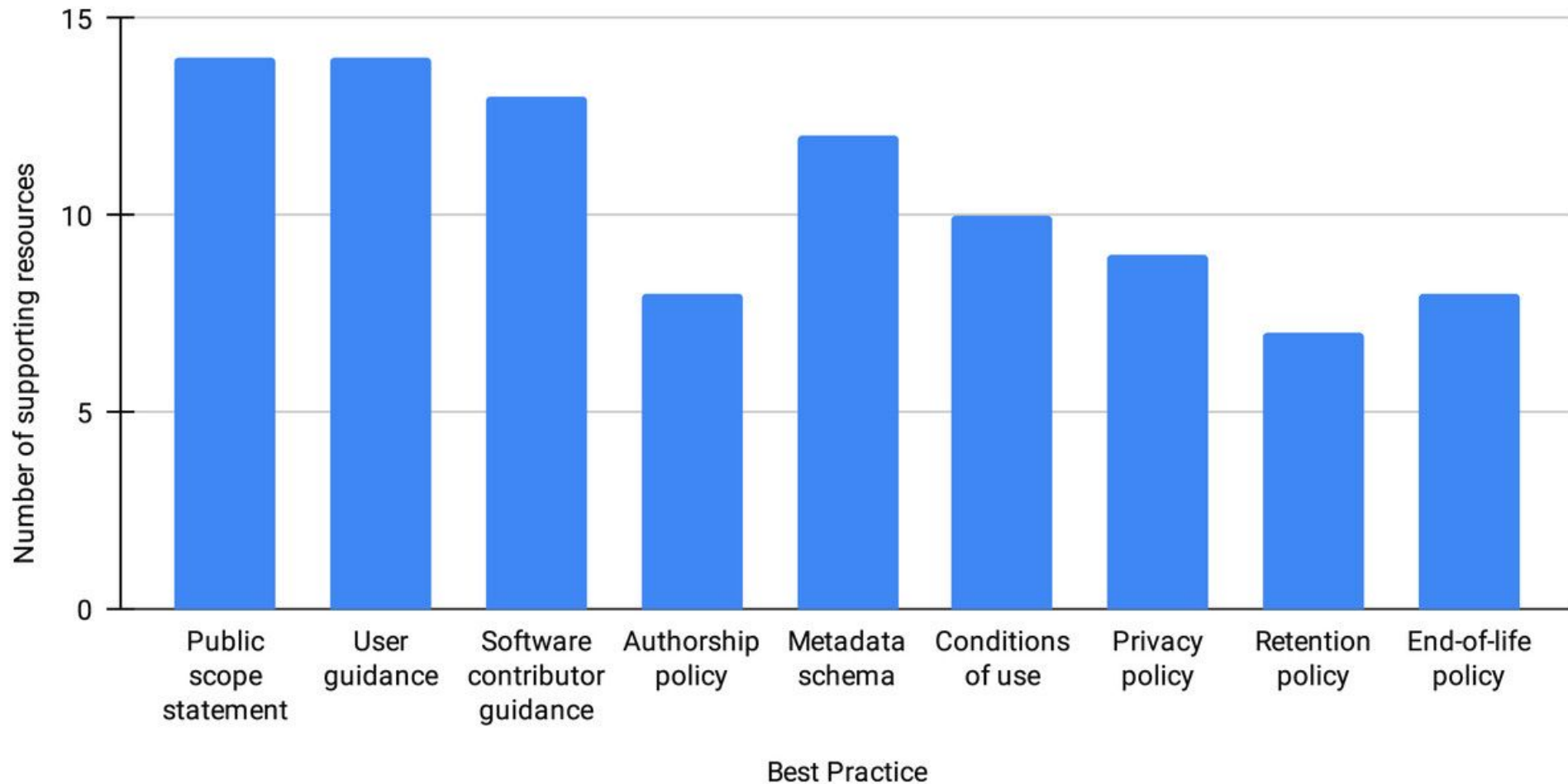
Join the discussion! <https://scicodes.net/>

Garijo D, Ménager H, Hwang L, Trisovic A, Hucka M, Morrell T, Allen A, Task Force on Best Practices for Software Registries, SciCodes Consortium. 2022. Nine best practices for research software registries and repositories. PeerJ Computer Science 8:e1023

<https://doi.org/10.7717/peerj-cs.1023>

<https://doi.org/10.5281/zenodo.10849689>

Tracking the adoption of best practices



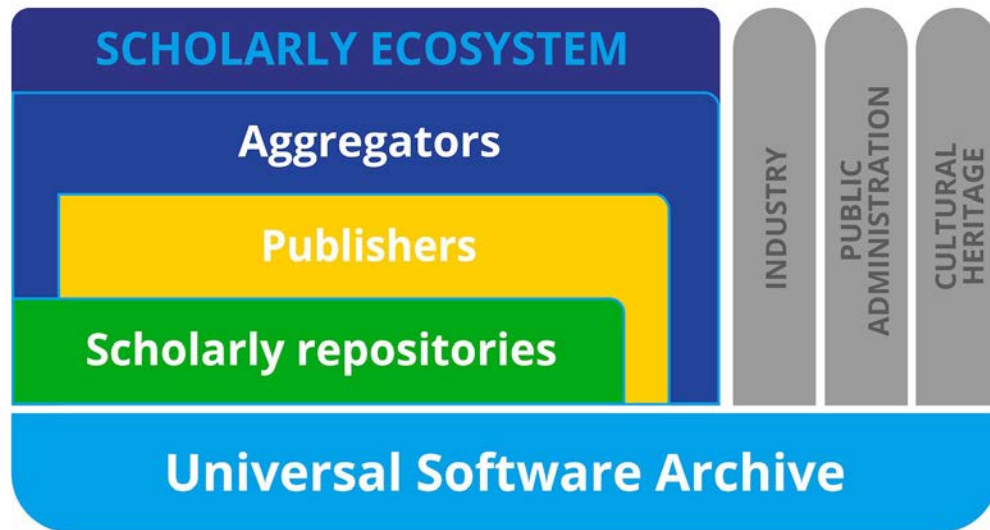
<https://doi.org/10.3201/261000.10049000>

Evolving to the Scicodes consortium

The consortium is growing! Up to 30 resources <https://scicodes.net/>

| Question | #Yes | #No | #Other |
|---|------|-----|--------|
| Is the resource discipline-specific? | 18 | 12 | 0 |
| Does the resource accept software only? | 17 | 13 | 0 |
| Does the resource require a software deposit? | 5 | 25 | 0 |
| Does the resource accept a software deposit | 22 | 8 | 0 |
| Can the resource mint DOIs? | 16 | 14 | 0 |
| Is the resource actively curated? | 21 | 3 | 6 |
| Can the resource be used to cite software? | 21 | 6 | 3 |

Scholarly Infrastructures for Research Software



SIRS report: European Commission, Directorate-General for Research and Innovation, *Scholarly infrastructures for research software : report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS*, Publications Office, 2020, <https://data.europa.eu/doi/10.2777/28598>

Video: [EOSC Software Infrastructures for Research Software: J. B. Gonzalez Lopez \(CERN\)](#)

SIRS GAP Analysis report : <https://zenodo.org/records/10376006>

<https://doi.org/10.5281/zenodo.10849689>

5. Questions and discussion