

Monday 17 June, 2024



The Research Data Alliance

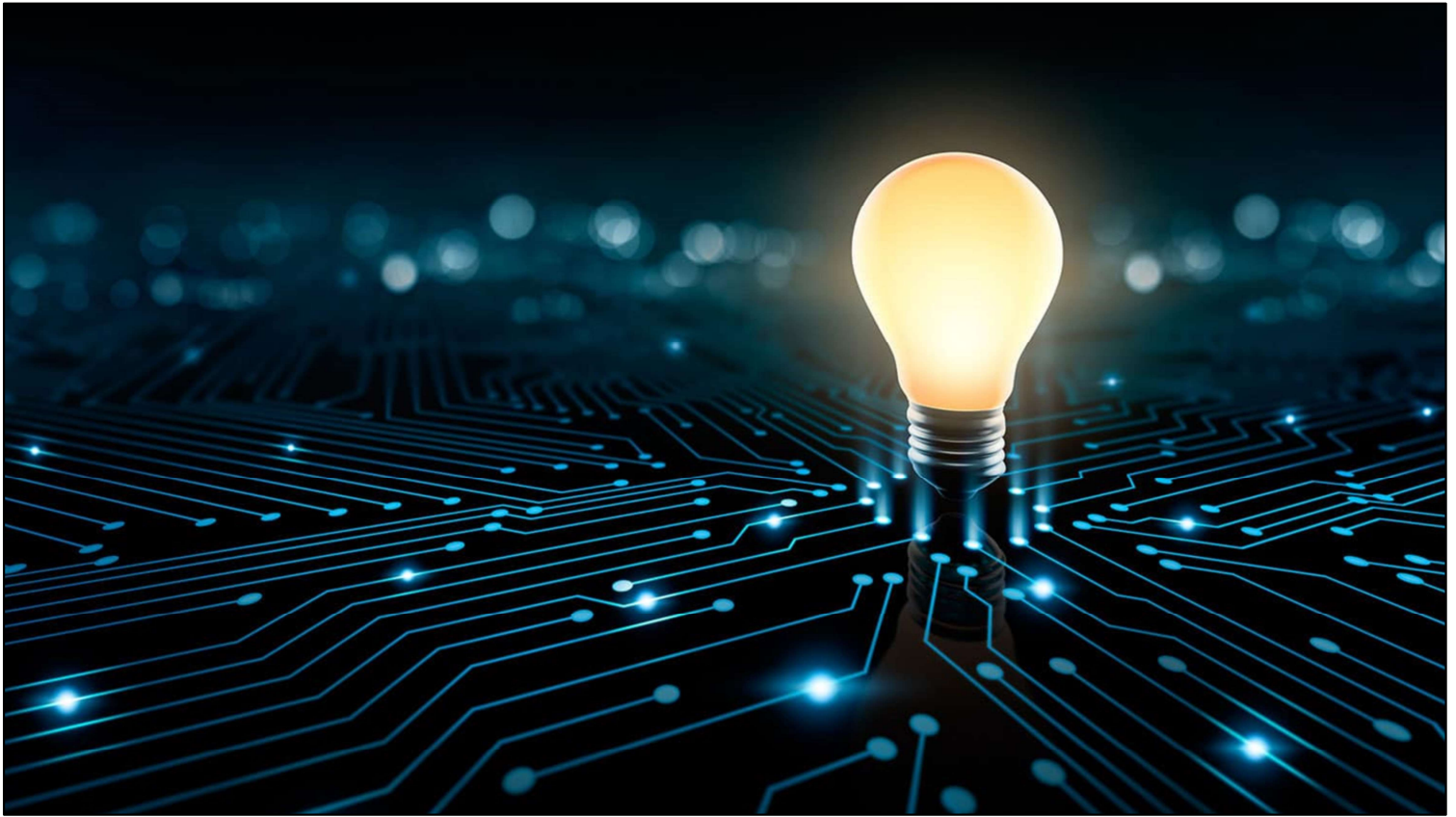
Hilary Hanahoe
Secretary General, RDA
hilary.hanahoe@rda-foundation.org
www.rd-alliance.org



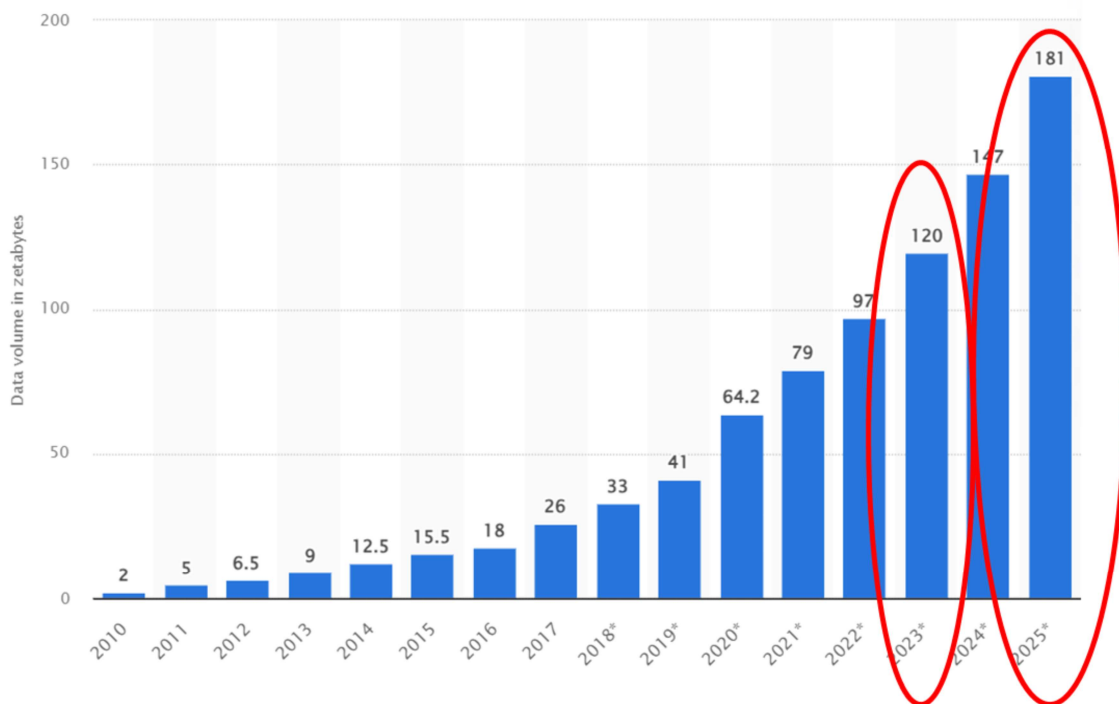
Association of Research Libraries (ARL) | SCOSS Webinar



Imagine a world without electricity. How would you charge your phone? How would you cook your meals? How would you cool and heat the environment you live in?



- Light helps us find our way in the dark. Light creates an ambience and a setting. Light changes our view of things.
- Data is light. Data is essential. But Data needs structure , data needs value, data needs experts. That's why the RDA was set up to make sure there is that structure, value and expertise in managing data across the world.



Source: [Statista 2024](#)

- Nowadays digital data is produced at an exponential rate every single day. In 2023 the world produced 120 zettabytes and by 2025 we are expected to reach a rate of 180 zettabytes. **To put this into perspective,**



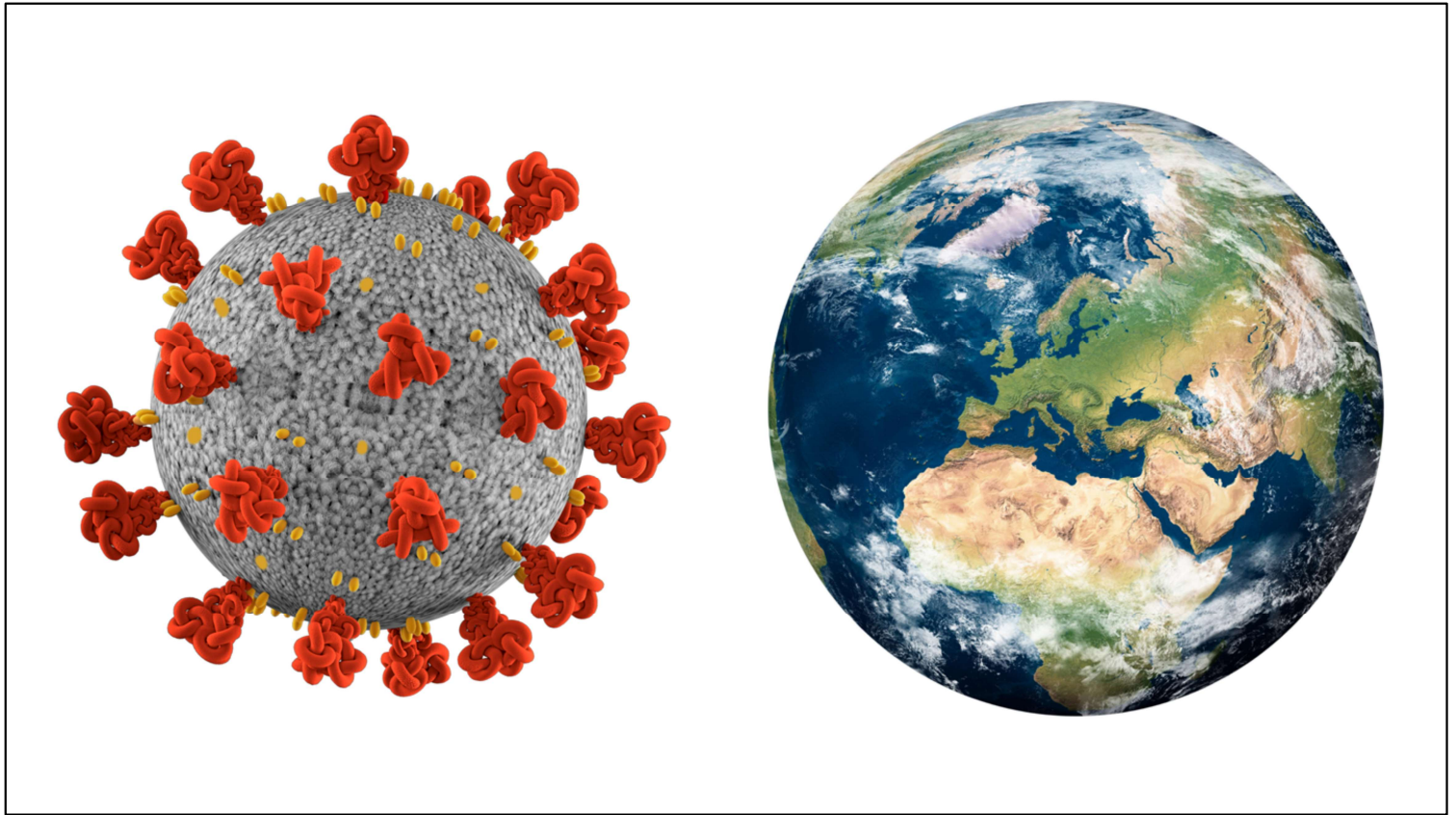
if zettabyte were a length then it would be equivalent to 1,300 round trips to the moon and back (769,000 kilometers). multiply that by 180. That's 235.000 round trips to the moon. Not all digital data produced needs to be stored, accessed, preserved but how do you know what to keep and what to throw away? When you have the answer to that question how do you store, preserve and make findable the data that you keep? How do you add value to that data to discover new ways of addressing societal challenges, for example to understand and slow down climate change ?



Without the RDA work on wheat data, farmers in remote areas in Africa would not have access to essential climate and agricultural data to assist them in making accurate and lifesaving crop management decisions. – The Wheat Data Interoperability guidelines generated by the RDA Agricultural Data Interest Group (IGAD) found this solution.



Without the RDA community activities around digital preservation and curation, libraries and librarians around the world would be struggling to create digital archives and replicas of their artefacts. Remember that libraries are gatekeepers of our history and societal memory.



In 2020, 600 RDA community members from RDA worked tirelessly to create the RDA COVID-19 WG Recommendations and Guidelines on Data Sharing: <https://doi.org/10.15497/rda00052>

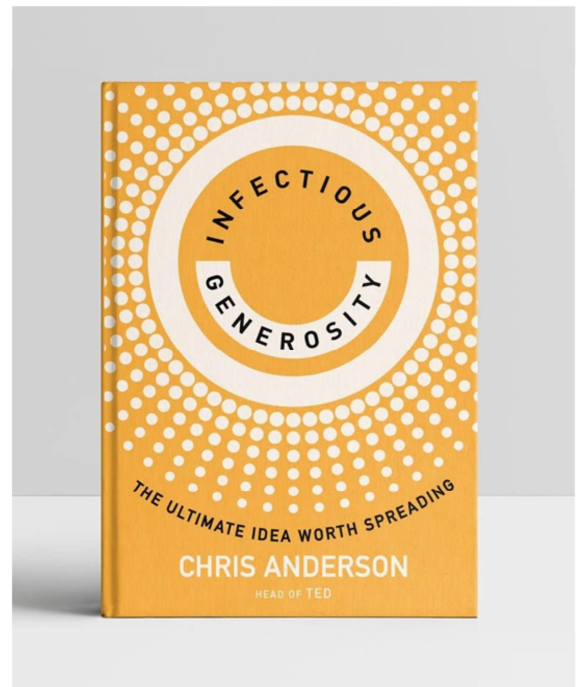
Without this RDA community work on COVID 19 data, funders and researchers would have invested months to understand how to format, save and make available their research data outputs so that others could quickly and effectively benefit from their findings to save lives.



The Research Data Alliance is global community of data experts working on the many complexities of digital data, to give it structure and value. As an open, international, community driven initiative the RDA facilitates the development and delivery of standards to facilitate interoperability of data.

Generosity is at the heart of being human. It's how we've co-operated, innovated and grown as a civilization. It's what makes me into we, net takers into net givers, selfishness into selflessness.

Source: infectiousgenerosity.org



Association of Research Libraries (ARL) | SCOSS Webinar

You will all be familiar with the TED talks and TED concept, and probably also know of Chris Anderson, the CEO / Head of TED who recently published a book entitled «Infectious Generosity» In his own words., Generosity is at the heart of being human. It's how we've co-operated, innovated and grown as a civilization. It's what makes me into we, net takers into net givers, selfishness into selflessness. Infectious Generosity is the idea that through the power of the internet, small acts of thoughtfulness spread to change lives at a scale never experienced before. The Research Data Alliance is a global community of volunteer, passionate, driven data experts who practice infectious generosity every day of their lives.

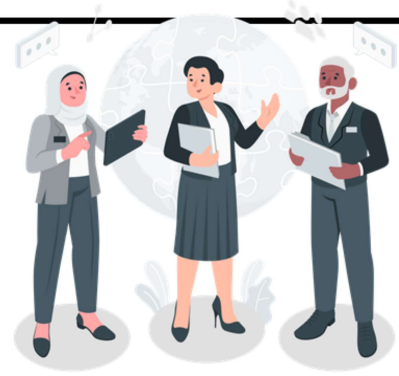
About the RDA

Vision

Researchers and innovators openly share and re-use data across technologies, disciplines, and countries to address the grand challenges of society.

Mission

RDA builds the social and technical bridges that enable open sharing and re-use of data.



Guiding principles

- Openness
- Consensus
- Inclusivity
- Harmonisation
- Community-driven
- Non-profit & technology neutral

RDA produces the adaptors to ensure you can access the electricity you need anywhere in the world.

The heart of the RDA: *Community groups*



Grass-roots, bottom-up, organic creation to address global challenges

Working Groups

- Develop & implement **tools, policy, practices & products** for data that are adopted & used by projects, organisations, & communities
- Life span: 12-18 months

RECOMMENDATIONS:

Concrete deliverables - "Running code", tools, standards, etc.

Interest Groups

- Focus on solving a specific **data sharing problem** and identifying what kind of infrastructure needs to be built
- Life span: As long as group is active

OUTPUT:

Best practices, guidelines, emergence of new working groups (WGs), etc.

Communities of Practice

- **Domain/disciplinary focus**, coordination and awareness raising role - '**umbrella group**'
- Life span: as long as CoP is active (review every 18 months)
- 1 CoP on Agricultural Data

OUTPUT:

New WGs & IGs, bridge building across the RDA and externally.



Data needs structure, data needs value, data needs experts. Data needs the RDA. In short, the data world would be a dark place without the RDA.

Data needs structure, data needs value, data needs experts. Data needs the RDA. In short, the data world would be a dark place without the RDA.

SUPPORTING RDA: SCOSS FUNDING



The overarching goals of the Research Data Alliance SCOSS funding appeal are to:

- 1) support the creation of global standards to enable best practice research data management, and
- 2) build research data management capacity and capability with stakeholders in the Global South.

**You can help us to ensure that science and research is truly seamless, equitable and global by pledging to RDA. Organisations supporting RDA
Contact Hilary Hanahoe**

[SCOSS FUNDING FOR RDA](#)

Association of Research Libraries (ARL) | SCOSS Webinar

To support the RDA, to continue to do fundamental work across the globe, please consider pledging through the SCOSS model to RDA.

Support global research data management: pledge to RDA

SCOSS pledging organisations become RDA Organisational members. A group of global stakeholders supporting Open Science. This group is a forum for interaction and exchange between institutions

DONATION LEVELS

ORGANISATION SIZE

ANNUAL CONTRIBUTION LEVEL

Large organisations from high-income countries	4,000 EUROS annually for three years
Small organisations from high-income countries	2,000 EUROS annually for three years
Funding agencies	8,000 EUROS annually for three years
Organisations from low and middle income countries and others with lower incomes	500 EUROS per year for three years

A 25% discount will be deducted from consortia that provide pledges from 10 or more pledging member institutions.



[SCOSS FUNDING FOR RDA](#)

Association of Research Libraries (ARL) | SCOSS Webinar

All organisations that pledge through SCOSS are invited to join the almost 100 RDA Organisational members. A group of global stakeholders supporting Open Science. This group is a forum for interaction and exchange between institutions where we provide updates to RDA, presentations of the solutions and many benefits for the organisations that support us. See Organisational Membership – <https://www.rd-alliance.org/organisational-membership/>
We also give visibility to our SCOSS pledgers via the web site - <https://www.rd-alliance.org/scoss-funding-for-the-rda/>

Monday 17 June, 2024



Thank you

Hilary Hanahoe
Secretary General | Research Data Alliance
hilary.hanahoe@rda-foundation.org



All Illustrations from [Storyset.com](https://www.storyset.com)

Association of Research Libraries (ARL) | SCOSS Webinar

Please feel free to contact me, Hilary Hanahoe the Secretary General | Research Data Alliance at hilary.hanahoe@rda-foundation.org and I would be very happy to interact with you.

The Source Code Library

By Morane Gruenpeter
(Head of Open Science Operations,
@Software Heritage)



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

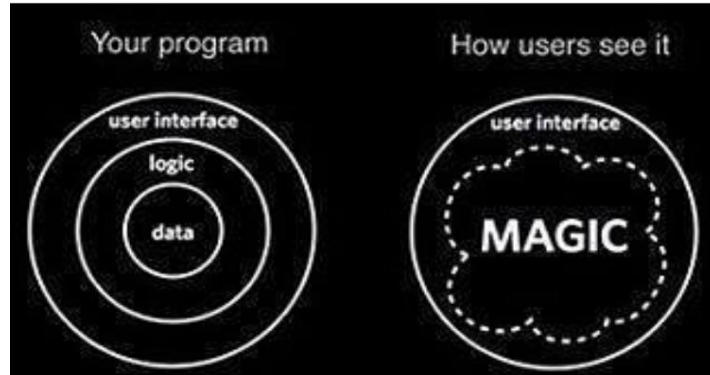
Inria

What to archive? what is software?



worldofprogrammers

...



Software as a concept

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

Not a digital artifact

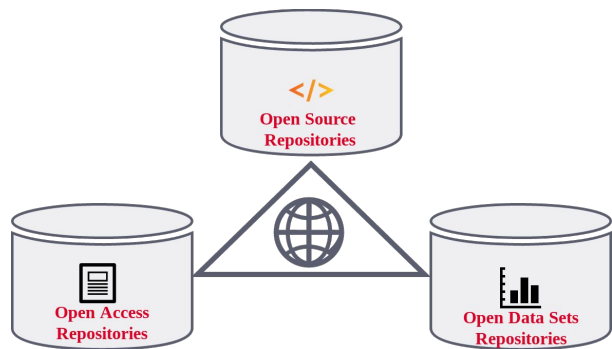
Software artifacts

- Executables
 - For multiple environments
- Source code

A very large collection of digital artifacts

Software is a pillar of Open Science

A key pillar: software (source code)



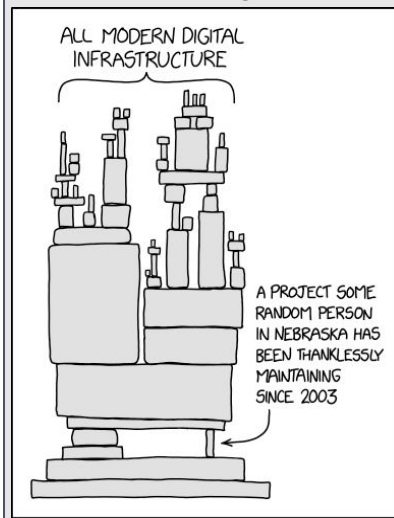
Three pillars of Open Science
Software Heritage CC-BY 4.0 2019

The links in the picture are important!

Source code is fragile, it can disappear...

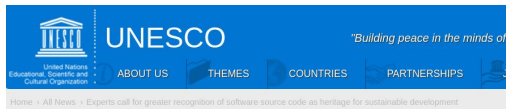
Software has multiple facets:

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



https://www.reddit.com/r/ProgrammerHumor/comments/ic1zmc/dependency_xkcd/

Software is increasingly recognized as a research output



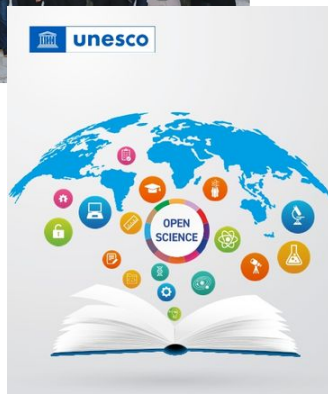
[Paris Call on Software Source Code, 2019](#)

Experts call for greater recognition of software source code as heritage for sustainable development

16 November 2018



[40 international experts call to](#) *“promote software development as a valuable research activity, and research software as a key enabler for Open Science/Open 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

Research libraries have a pivotal role

Software preservation requires a global and coordinated effort.

Picture by Lili Popper on Unsplash



The long road ahead:

- availability,
- findability,
- traceability,
- reproducibility

Have you met software archival requests in
your library or organization?
(please share in the chat)





A special guest deserves a special library:

Meet Software Heritage, the Alexandria library of software source code



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

An international and non-profit initiative launched in 2016



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Inria

Inria
La Fondation

built for the long term

Sharing the vision



Donors, members, sponsors



<http://www.softwareheritage.org/support/testimonials>



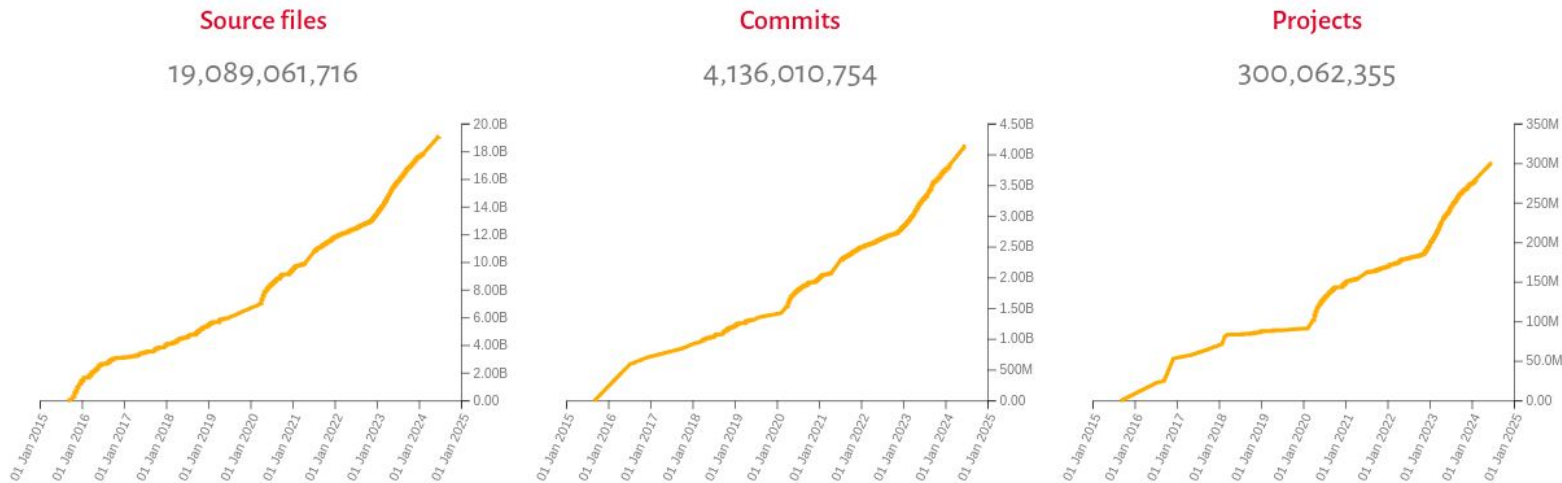
Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

Are you familiar with Software Heritage?

Size

As of today the archive already contains and keeps safe for you the following amount of objects:





An Open infrastructure

to collect, preserve and share *all source code*



DEPOSIT

to submit software source archives and its associated metadata



CRAWLING

Bitbucket 2,036,916 origins	git 7,300 origins	R 21,620 origins
debian 129,507 origins	h 6,463 origins	GitHub 156,095,789 origins
GitLab 3,990,594 origins	Guix 12,466 origins	GNU 354 origins
heptapod 1,098 origins	launchpad 368,181 origins	Maven 93,710 origins
NixOS 12,466 origins	npm 1,799,296 origins	Maven 4,083 origins
Phabricator 184 origins	pypi 432,092 origins	SOURCEFORGE 308,965 origins



SAVE CODE NOW SERVICE

At any time, for free:

- For a project
- For a forge

And also ...

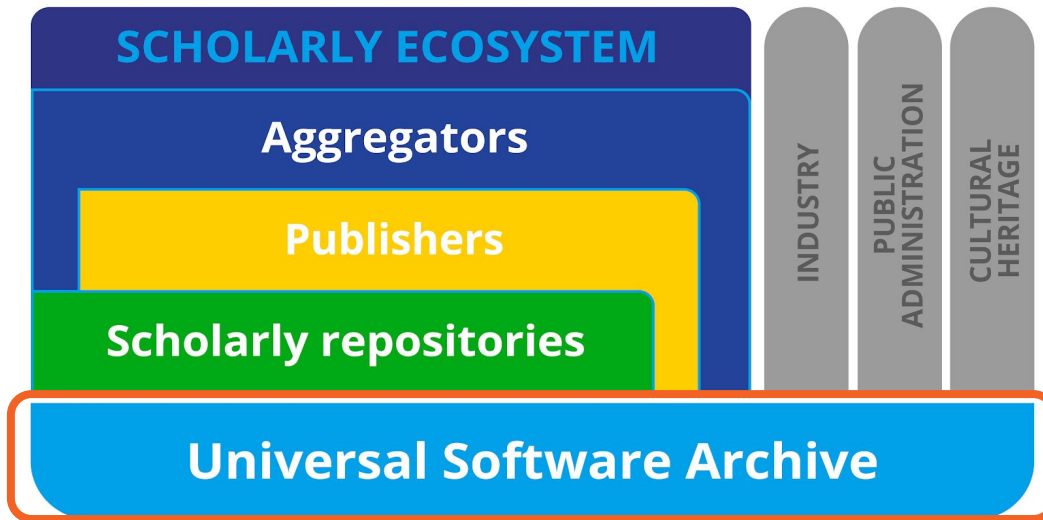


Rescue operations



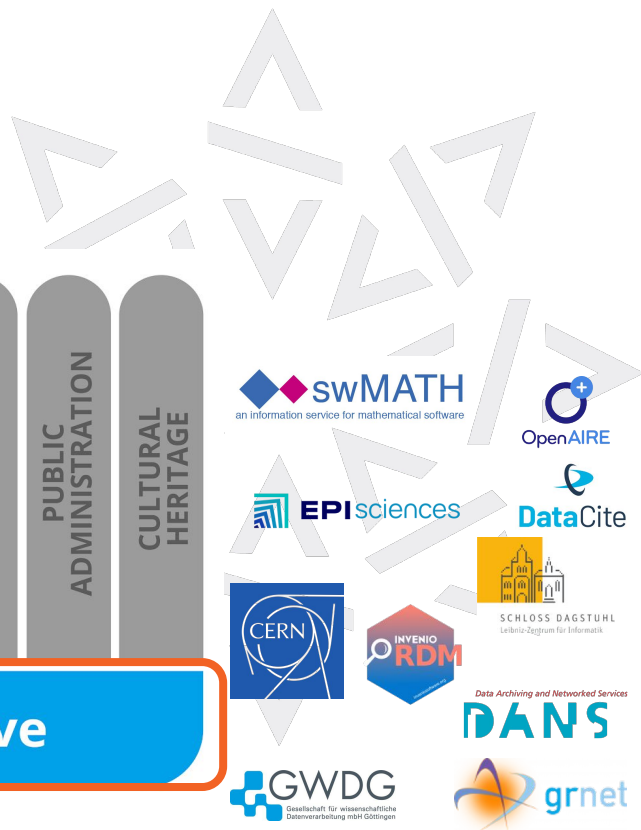
Serving the scholarly ecosystem

Creating connections and mutualizing efforts



Software Heritage

THE GREAT LIBRARY OF SOURCE CODE

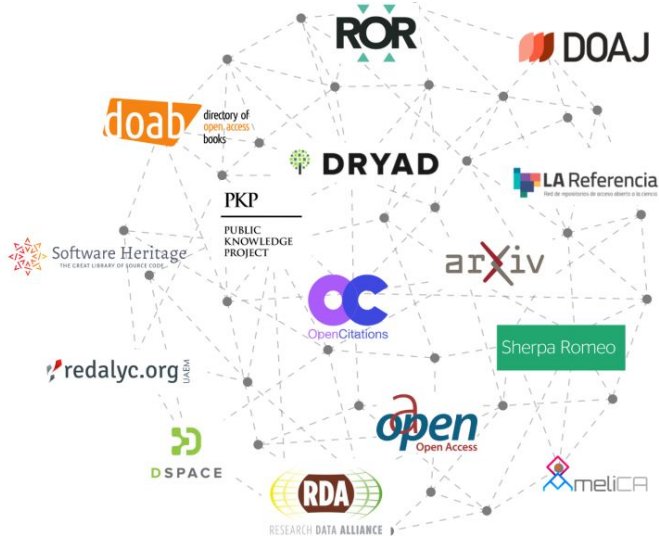


EOSC FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

Joining the SCOSS Family

→ Selected by **The Global Sustainability Coalition for Open Science Services (SCOSS)** for the **5th pledging round** with RDA

◆ Establishing a new membership agreement through the **Archives and Libraries Interest Group (ALIG)**



SOFTWARE HERITAGE

THE LIBRARY OF ALEXANDRIA
OF SOFTWARE SOURCE CODE

Software Heritage is an open non-profit infrastructure for archiving, referencing and sharing software source code, launched by Inria in 2016, in partnership with UNESCO.

Archiving over 260 million software projects already, it is built according to the UNESCO recommendations for Open Science: open, multi-stakeholder, non-profit, using exclusively open source components, it serves as a cornerstone for Open Science.

It simplifies the deposit of research software and associated metadata, amplifying the visibility and impact of scholarly outputs. Researchers take advantage of Software Heritage's vast collection of software projects, that enables citation, referencing and sharing of software artefacts, improving reproducibility and traceability of research. Libraries benefit from Software Heritage's robust infrastructure, which offers long-term archival and unique identification of software, removing the need for custom and in-house archival solutions.

By supporting Software Heritage, you're supporting unfettered access, reference and citation of software produced by academic research, reinforcing the principles of open science.

WHY HAS IT BEEN DEEMED AN ESSENTIAL INFRASTRUCTURE?

The SCOSS Board considers Software Heritage to be an essential open science infrastructure because it provides continued access to the software and code outputs produced by researchers globally.

SCOSS FUNDING TARGET

€ 900,000

The Archives and Libraries Interest Group (ALIG)

- ★ Supporter benefits include:
 - Acknowledgement of individual / consortium pledge
 - [Membership in the ALIG](#): access to a dedicated information and consultation channel through discussions in the Interest Group
 - Assistance in software archival dilemmas

- ★ Sponsorship levels

- Sponsors status depending on contribution level
- See [program](#)

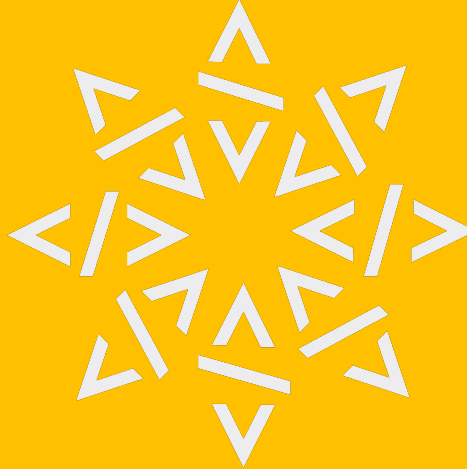
LEVEL	BENEFITS	ANNUAL CONTRIBUTION LEVEL
Paper	Supporter benefits include: Acknowledgement and membership in the Archives and Libraries Interest Group	2,000 EURO
Wood		4,000 EURO
Copper	Supporter benefits plus the Bronze sponsor status	8,000 EURO
Bronze		10,000 EURO
Silver		25,000 EURO
Gold	Supporter benefits plus the Gold sponsor status	50,000 EURO

What are the needs in your organization or community when it comes to software?





Behind the scenes: a dedicated and engaged team!



Thank you for your attention

alig@softwareheritage.org

<https://www.softwareheritage.org/>

Cultural Heritage



Industry



Research



Public Administration



Software Heritage

Time for your questions...

Links to be shared with your users:



Guidelines for researchers:

<https://www.softwareheritage.org/howto-archive-and-reference-your-code/>



Describe a software (intrinsic metadata made easy) :

<https://codemeta.github.io/codemeta-generator/>



The FAQ: <https://www.softwareheritage.org/faq/>



The SWH guided tour:

https://archive.softwareheritage.org/?guided_tour=0&guided_tour_next=https://archive.softwareheritage.org/



"Open Science tutorial: source code deposit" (how to build services upon SWH):

<https://youtube.com/playlist?list=PLD2VqrZz2-u3bOWtoCoBlh5Flt6iYXsq3>



"Book" a SWH ambassador: <https://www.softwareheritage.org/ambassadors/>



Let's keep in touch

<https://twitter.com/SWHeritage>

<https://twitter.com/rdicosmo>

<https://twitter.com/moraneottilia>

<https://www.linkedin.com/company/swheritage/>

<https://sympa.inria.fr/sympa/info/swh-science>

<https://www.softwareheritage.org/newsletter/>

<https://www.softwareheritage.org/blog>



Implementing new services at the library: the SWH documentation treasure island

The Web API: <https://archive.softwareheritage.org/api/>

API terms of use: <https://www.softwareheritage.org/legal/api-terms-of-use/>

API endpoints: <https://archive.softwareheritage.org/api/1/>

The technical roadmap:

<https://docs.softwareheritage.org/devel/roadmap/roadmap-2023.html#roadmap-current>

SWH legacy software services: <https://www.softwareheritage.org/swhap/>

Sources

Albornoz, D., Huang, M., Martin, I. M., Mateus, M., Touré, A. Y., & Chan, L. (2018). Framing Power: Tracing Key Discourses in Open Science Policies. In L. Chan & P. Mounier (Eds.), *ELPUB 2018*. <https://doi.org/10.4000/proceedings.elpub.2018.23>

The Association of Research Libraries (ARL). (2018). *Code of Best Practices in Fair Use for Software Preservation*. <https://www.arl.org/resources/code-of-best-practices-in-fair-use-for-software-preservation/>

Barborini, Y., Cosmo, R. D., Dumont, A. R., Gruenpeter, M., Marmol, B., Monteil, A., Sadowska, J., & Zacchiroli, S. (2018, March 21). The creation of a new type of scientific deposit: Software. RDA Eleventh Plenary Meeting, Berlin, Germany. <https://hal.inria.fr/hal-01738741>

Brinkman, L., Schettino, A., Zurita-Milla, R., Eerland, A., Imming, M., Zormpa, E., Schwamm, H., Heininga, V. E., Tsang, E., & van 't Veer, A. (2020). Open Science Community Starter Kit (A. Blanchard, Trans.). International Network of Open Science & Scholarship Communities (INOSC). <https://osf.io/7vez3>

Di Cosmo, R. (2020). Archiving and Referencing Source Code with Software Heritage. In A. M. Bigatti, J. Carette, J. H. Davenport, M. Joswig, & T. de Wolff (Eds.), *Mathematical Software – ICMS 2020* (Vol. 12097, pp. 362–373). Springer International Publishing. https://doi.org/10.1007/978-3-030-52200-1_36

Di Cosmo, R., Gruenpeter, M., & Zacchiroli, S. (2020). Referencing Source Code Artifacts: A Separate Concern in Software Citation. *Computing in Science & Engineering*, 22(2), 33–43. <https://doi.org/10.1109/MCSE.2019.2963148>

Sources

EOSC Executive Board & EOSC Secretariat. (2020). Scholarly infrastructures for research software. Report from the EOSC Executive Board Working Group (WG) Architecture Task Force (TF) SIRS. European Commission. Directorate General for Research and Innovation.

<https://data.europa.eu/doi/10.2777/28598>

Jetten, M., Grootveld, M., Mordant, A., Jansen, M., Bloemers, M., Miedema, M., & Gelder, C. W. G. van. (2021). *Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship*. Zenodo.

<https://doi.org/10.5281/zenodo.4623713>

LIBER. (2023). *Report on outputs on LIBER Winter Event Workshop “Data management skills and training development for libraries.”*

https://libereurope.eu/wp-content/uploads/2023/04/Report-on-Winter-Event-Workshop_PDF-1.pdf

Mudditt, A. (2020, February 18). Reforming Research Assessment: A Tough Nut to Crack. *The Scholarly Kitchen*.

<https://scholarlykitchen.sspnet.org/2020/02/18/reforming-research-assessment-a-tough-nut-to-crack/>

Poldrack, R. A. (2019). The Costs of Reproducibility. *Neuron*, 101(1), 11–14. <https://doi.org/10.1016/j.neuron.2018.11.030>

Research Data Alliance/FORCE11 Software Source Code Identification WG, Allen, A., Bandrowski, A., Chan, P., di Cosmo, R., Fenner, M., Garcia, L., Gruenpeter, M., Jones, C. M., Katz, D. S., Kunze, J., Schubotz, M., & Todorov, I. T. (2020). *Software Source Code Identification Use cases and identifier schemes for persistent software source code identification*. <https://doi.org/10.15497/RDA00053>

Saby, F. (2013). Approche historique du dépôt légal en France. *Sociétés & Représentations*, 35(1), 15–26. <https://doi.org/10.3917/sr.035.0015>