# The Open Science Toolkit

\*\*A Comprehensive Guide to Open Research & Knowledge Preservation\*\*

\*\*Version 2.2 – October 2025\*\* | Enhanced with community feedback

## 1. Scientific Repositories & Preprints

\*\*arXiv.org\*\* – Free preprint server for STEM fields

<https://arxiv.org/>

\*\*bioRxiv\*\* – Biology preprints

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

\*\*medRxiv\*\* – Health sciences preprints

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

\*\*ChemRxiv\*\* – Chemistry preprints

<https://chemrxiv.org/>

\*\*PsyArXiv\*\* – Psychology preprints

<https://psyarxiv.com/>

\*\*SocArXiv\*\* – Social sciences preprints

<https://osf.io/preprints/socarxiv/>

\*\*Zenodo\*\* – General research data and DOIs

<https://zenodo.org/>

\*\*CORE\*\* – Open access aggregator

<https://core.ac.uk/>

\*\*EarthArXiv\*\* – Geoscience preprints

<https://eartharxiv.org/>

\*\*marXiv\*\* – Materials science preprints

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

\*\*AfricArXiv\*\* – African-focused preprints

<https://info.africarxiv.org/>

\*\*SciELO\*\* – Latin American open access journals

<https://scielo.org/>

## 2. Open Peer Review & Assessment

\*\*OpenReview\*\* – Transparent peer review platform

<https://openreview.net/>

\*\*DORA\*\* – Research assessment reform

<https://sfdora.org/>

\*\*CoARA\*\* – Advancing research assessment

<https://coara.eu/>

\*\*CrossRef\*\* – Citation metadata

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

\*\*Dimensions\*\* – Research metrics

<https://www.dimensions.ai/>

\*\*Lens.org\*\* – Open scholarly data

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

\*\*Altmetric\*\* – Broader impact tracking beyond citations

<https://www.altmetric.com/>

\*\*ORCID\*\* – Persistent researcher identifiers

<https://orcid.org/>

## 3. AI & Computational Tools

\*\*Hugging Face Datasets\*\* – AI/ML datasets

<https://huggingface.co/datasets>

\*\*AllenNLP\*\* – NLP research library

<https://allennlp.org/>

\*\*CatalyzeX\*\* – Links papers to code

<https://catalyzex.com/>

\*\*Papers with Code\*\* – ML papers with implementation code

<https://paperswithcode.com/>

\*\*Semantic Scholar\*\* – AI literature search

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

\*\*Open Science Grid\*\* – Distributed computing

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

\*\*Jupyter / MyBinder\*\* – Executable notebooks

<https://mybinder.org/>

\*\*ExecutableBooks / Jupyter Book\*\* – Interactive books from notebooks

<https://jupyterbook.org/>

\*\*EleutherAI\*\* – Open LLM training data and models

<https://www.eleuther.ai/>

\*\*OMol25 (Open Molecules 2025)\*\* – Large quantum chemistry dataset (Meta, 2025)

<https://omol25.org/> \*(verify launch status)\*

\*\*UMA (Universal Model for Atoms)\*\* – Interatomic potential model (2025)

<https://uma.org/> \*(verify launch status)\*

\*\*Qiskit\*\* – Open quantum SDK (IBM)

<https://qiskit.org/>

\*\*PennyLane\*\* – Quantum ML library (Xanadu)

<https://pennylane.ai/>

\*\*LiveCodeBenchPro\*\* – Advanced coding benchmark for LLM evaluation

\*(check latest releases)\*

\*\*Lock-LLMs\*\* – Framework for secure/controlled large language models

\*(check latest releases)\*

## 4. Earth, Climate & Environmental Data

\*\*NASA Earthdata\*\* – Satellite & climate datasets

<https://earthdata.nasa.gov/>

\*\*NOAA NCEI\*\* – Environmental archives

<https://www.ncei.noaa.gov/>

\*\*Pangaea\*\* – Earth & environmental data

<https://www.pangaea.de/>

\*\*Copernicus Climate Service\*\* – EU climate data

<https://climate.copernicus.eu/>

\*\*GBIF\*\* – Biodiversity data

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

\*\*OBIS\*\* – Ocean biodiversity

<https://obis.org/>

\*\*iNaturalist\*\* – Citizen science observations

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

\*\*REACH (NOAA, 2025)\*\* – Environmental research discovery portal

<https://reach.noaa.gov/> \*(verify launch status)\*

\*\*FAIR² (Frontiers, 2025)\*\* – AI-enhanced biodiversity platform

<https://fair2.frontiers.org/> \*(verify launch status)\*

## 5. Humanities & Arts Repositories

\*\*Humanities Commons\*\* – Open humanities scholarship

<https://hcommons.org/>

\*\*DARIAH-EU\*\* – Digital research infrastructure for arts & humanities

<https://www.dariah.eu/>

## 6. Equity, Diversity & Global South

\*\*OCSDNet\*\* – Open science in low/middle-income countries

<https://ocsdnet.org/>

\*\*CRAFT-OA\*\* – Diamond open access initiative

<https://www.craft-oa.eu/>

\*\*Plan S\*\* – Full open access transition initiative

<https://www.coalition-s.org/>

\*\*CARE Principles / Indigenous Data Sovereignty\*\* – Ethical data guidelines

<https://www.gida-global.org/care>

## 7. Open Data, Code & Hardware

\*\*GitHub\*\* – Open source code hosting

<https://github.com/>

\*\*Dryad\*\* – Research data repository

<https://datadryad.org/>

\*\*Mendeley Data\*\* – Data sharing

<https://data.mendeley.com/>

\*\*Code Ocean\*\* – Reproducible research platform

<https://codeocean.com/>

\*\*Open Hardware Repository (OHWR)\*\* – Instrument designs

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

\*\*Software Heritage\*\* – Archive of open-source code

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

## 8. Digital Preservation & Reproducibility

\*\*Internet Archive\*\* – Wayback Machine & digital library

<https://archive.org/>

\*\*HathiTrust\*\* – Digitized books

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

\*\*Project Gutenberg\*\* – Free ebooks

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

\*\*Dataverse\*\* – Social science data preservation

<https://dataverse.org/>

\*\*Protocols.io\*\* – Experimental protocols

<https://www.protocols.io/>

\*\*OSF Registries\*\* – Study pre-registration

<https://osf.io/registries/>

\*\*CLOCKSS\*\* – Long-term digital archiving

<https://clockss.org/>

\*\*Portico\*\* – E-journal preservation

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

## 9. Education & Training

\*\*MIT OpenCourseWare\*\* – Free MIT course materials

<https://ocw.mit.edu/>

\*\*Khan Academy\*\* – Free K–14 education

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

\*\*OpenStax\*\* – Free college textbooks

<https://openstax.org/>

\*\*Software Carpentry / Data Carpentry\*\* – Reproducible research training

<https://software-carpentry.org/>

\*\*The Turing Way\*\* – Guide to reproducible data science

<https://the-turing-way.netlify.app/>

\*\*Open Science MOOC\*\* – Open science training modules

<https://www.openaire.eu/open-science-mooc>

\*\*Reproducible Science Curriculum\*\* – Comprehensive training materials

<https://reproducible-science-curriculum.readthedocs.io/>

## 10. Practical Workflow & Visualization

### Research Lifecycle Flow

\*\*STEP 1: PLAN & PRE-REGISTER\*\*

→ OSF Registries, Protocols.io

\*\*STEP 2: COLLECT DATA\*\*

→ Assign DOI (Zenodo/Dryad/Dataverse)

\*\*STEP 3: ANALYZE\*\*

→ Jupyter Notebooks, GitHub, Reproducible Code

\*\*STEP 4: PUBLISH PREPRINT\*\*

→ arXiv/bioRxiv/domain-specific preprint servers

\*\*STEP 5: PEER REVIEW\*\*

→ OpenReview, traditional journals with OA options

\*\*STEP 6: PRESERVE & ARCHIVE\*\*

→ Internet Archive, Software Heritage, CLOCKSS

\*\*STEP 7: TRACK IMPACT\*\*

→ Dimensions, Lens.org, Altmetric

\*\*STEP 8: ENGAGE COMMUNITY\*\*

→ Zooniverse, iNaturalist, social media

\*\*Throughout all steps:\*\* Apply FAIR Principles + CARE Ethics + ORCID/CRediT

### Essential Checklist

✓ Follow \*\*FAIR\*\* principles (Findable, Accessible, Interoperable, Reusable)

✓ Apply \*\*CARE\*\* principles (Collective benefit, Authority to control, Responsibility, Ethics)

✓ Assign \*\*DOIs\*\* to all datasets and code releases

✓ Use \*\*ORCID\*\* for persistent researcher identity

✓ Apply \*\*CRediT\*\* taxonomy for author contributions

✓ Create \*\*reproducible notebooks\*\* (Jupyter/R Markdown)

✓ \*\*Pre-register\*\* studies when applicable

✓ Share \*\*code\*\* via GitHub/Software Heritage

✓ Document \*\*protocols\*\* on Protocols.io

✓ Archive \*\*final versions\*\* in preservation systems

## Recent Additions & Updates (2025)

\*\*New Tools:\*\* OMol25 • UMA • REACH • FAIR² • LiveCodeBenchPro • Lock-LLMs • Qiskit • PennyLane

\*\*Enhanced Coverage:\*\* Altmetric • Papers with Code • EleutherAI • SciELO • Plan S • CLOCKSS • Portico • Reproducible Science Curriculum

\*\*Note:\*\* Items marked \*(verify launch status)\* are upcoming/recently announced—check official channels for availability.

## Contributing & Feedback

\*\*GitHub Repository:\*\* Host this toolkit on GitHub for community contributions

- Enable issue tracking for corrections/additions

- Use GitHub Actions for automated version updates

- Assign a Zenodo DOI to the repository for citability

- Consider RSS feeds from listed sites for automated updates

\*\*Feedback welcomed via:\*\* [Insert GitHub repo URL or contact method]

## Version History

- \*\*v2.2 (Oct 2025):\*\* Added Altmetric, Papers with Code, EleutherAI, SciELO, Plan S, CLOCKSS, Portico, workflow visualization, renamed to “The Open Science Toolkit”

- \*\*v2.1 (Oct 2025):\*\* Original comprehensive compilation

- \*\*v2.0 and earlier:\*\* Initial development

## Citation & License

\*\*Suggested Citation:\*\*

The Open Science Toolkit: A Comprehensive Guide to Open Research & Knowledge Preservation (Version 2.2). October 2025. [DOI to be assigned via Zenodo]

\*\*License:\*\* CC BY 4.0 – Free to share and adapt with attribution

\*\*For the latest version and to contribute:\*\* [GitHub repository URL]

\*\*Share using:\*\* #OpenScienceToolkit