Aperture Neuro

Karakuzu A. Toward a woven literature: Open-source infrastructure for networked scientific publishing. *Aperture Neuro*. 2025;5(SI 3). doi:10.52294/001c.146072

Review Articles and Meta-Analyses

Toward a woven literature: Open-source infrastructure for networked scientific publishing

Agah Karakuzu, Ph.D.[1,2] [a]

[1] Department of Electrical Engineering, Polytechnique Montréal, [2] Montreal Heart Institute, University of Montreal

Keywords: reproducible research, open science, next-generation scholarly communication, executable notebooks in publishing, interactive articles, reproducible publishing, open access

https://doi.org/10.52294/001c.146072

Aperture Neuro

Vol. 5, Issue SI 3, 2025

Scientific publishing must evolve to reflect the computational nature of modern research. While code, data, and analysis increasingly shape scientific findings, current publishing formats often flatten these elements into static narratives. In response, a new model is emerging: the woven literature. Inspired by principles of literate programming, this approach integrates code, data, narrative, and interactive environments to enable reproducibility and transparency within the article itself. This review explores the technical foundations, cultural barriers, and infrastructural innovations shaping this transition. Drawing on case studies from the NeuroLibre platform, it illustrates how open-source tools can support next-generation publications that are executable, sustainable, and adaptable across disciplines. The discussion highlights how workflows of varying complexity can be modularized and surfaced through reproducibility-focused platforms, while also addressing limitations imposed by current incentive structures and legacy systems. By embedding computation directly within the scholarly article, woven literature transforms the role of publishing from passive documentation into an active and verifiable part of the research process, allowing readers to engage with and extend the work itself.

INTRODUCTION

The trajectory of scholarly communication is marked by transformative technological shifts. It began with the meticulously hand-copied manuscripts of medieval scholars. This was followed by the advent of the first academic journals in the 17th century.[1] In the 19th and early 20th centuries, typewritten manuscripts were circulated via postal services. By the late 20th century, the digital age ushered in online platforms that enabled widespread electronic dissemination and archival of research articles. While these developments dramatically increased accessibility, they did not fundamentally alter the structure or format of scientific publishing. Most articles remained static PDFs, detached from the underlying data, code, and computational environments that produced their findings.

Today, open-source technologies and reproducibility-driven initiatives create an opportunity to reimagine scientific articles as living documents, encompassing qualities such as interactivity, executability, and true computational depth. This review situates that opportunity within a broader technical landscape. It argues that scholarly publishing is on the brink of a transformation, one that could re-align publication with the ethos of open science, computational reproducibility, and collaborative progress.

The review proceeds in three movements:

1. A diagnostic analysis, exposing the technical limitations and economic asymmetries of the legacy publishing complex. We trace how the infrastructure meant to serve science has instead come to extract disproportionate value from it, and how this dynamic has delayed the adoption of more open, interactive modalities.
2. A survey of enabling technologies, charting the open-source tools and cloud platforms that allow authors to create interactive, executable "living publica-

[a] Corresponding Author:
Agah Karakuzu
Polytechnique Montreal,
Electrical Engineering L-5626, 2500
Chem. de Polytechnique,
H3T 0A3, QC, Montreal, Canada.
agahkarakuzu@gmail.com

tions." We highlight NeuroLibre as a case study for end-to-end reproducibility at scale. It demonstrates how code, data, and computational environments can be integrated into a cohesive and shareable research object.

3. A procedural outlook, exploring how distributed peer review, workflow modularization, and new incentive models can support a more transparent and dynamic publishing ecosystem. We introduce the concept of woven literature, building upon Knuth's original notion of literate programming, to describe publications where prose, computation, and verification not only coexist but evolve together as integrated components.

Through comprehensive analysis and practical recommendations, this review provides researchers, developers, and infrastructure providers with a strategic roadmap for creating, curating, and sustaining the next generation of scientific literature. The practical considerations address critical questions that determine implementation success: What are the resource boundaries of browser-based execution environments? How can large datasets be reproducibly integrated without overwhelming readers or computational platforms? When should upstream, resource-intensive workflows transition to lighter, reader-facing phases of analysis? How can emerging tools ensure complete traceability between raw inputs and final results? How do novel reuse patterns become possible through these advanced publishing technologies?

By addressing these questions, this review bridges the gap between theoretical potential and practical implementation of next-generation scholarly communication systems.

WHEN THE ELM STRANGLES THE VINE: A BROKEN EQUATION

The role of academic publishers in advancing scientific progress is undeniable. For centuries, researchers and publishers have shared a mutually beneficial relationship. This partnership is symbolically captured in Elsevier's logo, which features an old man, representing the scientist, harvesting grapes, a metaphor for knowledge, from a vine entwined around an elm tree, which represents the publisher. The accompanying motto, Non Solus ("not alone"), emphasizes the collaborative nature of scientific discovery.[2]

The symbiosis between the elm tree and the grapevine is multi-layered and has been cultivated since Ancient Greece.[3] The elm provides structural support for the vine, while its canopy helps regulate the microclimate, improving grape quality. During droughts, the tree may even share its sap with the vine, helping it survive under adverse conditions. Elsevier's logo, though a poetic representation of what science expects from publishers, turns out to be botanically inaccurate: without pruning, the depicted elm would reduce grape yield and undermine its potential to support the vine through sap sharing.[4]

More critically, this symbolism has become practically reversed: what began as collaboration has tilted toward commodification. Instead of publishers sustaining the growth of science, it is now the scientific enterprise at scale that fuels the commercial success of a few dominant publishers. The primary beneficiaries of the advances in research dissemination have been a few dominant publishers, whose profit margins (up to 40-50%) have outpaced even those of the most successful high-tech companies,[5] revealing a disproportion between their economic gains and the actual value they add to scholarly communication.

To begin, this discrepancy can be examined through its numerical dimensions: The real cost of hosting a PDF and registering it as scholarly content is estimated to be around $2.71,[6] assuming an open-access online journal that publishes around 300 articles annually (e.g., the Journal of Open Source Software, JOSS[7]). On the other hand, article processing charges (APCs) of a major publisher can reach a staggering $12,000, with an average of $3.3k across nearly 14,000 publications per year.[8] Notably, one such publisher has claimed that the internal cost of publishing an open-access article ranges between $30,000 and $40,000.[8] This nearly 4,500-fold disparity on average, which rises to as much as 14,500-fold according to some claims, between the actual cost of $2.71 and the brand-inflated cost of making an article publicly accessible does not correspond to a proportional increase in scientific impact.[9]

If not delivering impact, is this level of spending at least contributing to the evolution of scientific communication? Unfortunately, an equally pressing concern lies in the underutilization of even basic technological advancements. A telling example is the so-called Continuous Article Publishing model (CAP), introduced in the early 2000s. Here, the term *continuous* refers to the practice of releasing accepted manuscripts online as they are ready, rather than waiting for complete issues to be compiled. Remarkably, some journals are still in the process of transitioning to CAP more than two decades later. This slow adoption underscores the inertia of legacy publishing systems, which struggle to implement even modest procedural changes, let alone the infrastructural transformation needed to support truly modern and integrated forms of scholarly communication.[10]

As a result of the inertia embedded in legacy publishing systems, digital publication often remains structurally tied to the logic of print. Publications remain static, disconnected from the data, code, and computational environments that would allow them to serve as dynamic components of an evolving research ecosystem. In signal processing terms, *the body of knowledge is sampled but never reconstructed*: research exists as discrete outputs without the connective infrastructure for synthesis, independent replication, or reuse.

Encased in this digital shell modeled on print-era conventions, academic papers continue to function as isolated artifacts. The literature becomes increasingly bibliographic, where citations serve more as symbolic gestures than as functional links between interoperable knowledge objects. Sticking with the signal processing analogy, *CAP is only about increasing the sampling rate* of a bibliographic literature, primarily serving to accelerate publishers' revenue

streams, falling far short of the infrastructural leap required for Continuous Science.[11]

The concept of Continuous Science represents a fundamental modernization of research dissemination. It promotes a truly continuous digital medium in which articles are no longer isolated relics but integrated research objects.[10] Returning to the signal processing analogy, this is not merely about increasing the sampling rate. It is about *reconstructing a coherent and interoperable knowledge system*. Such a system would enhance the signal-to-noise ratio of scientific communication by enabling reproducibility, synthesis, and meaningful connectivity across articles.

Before turning to the elements of modern solutions that can enable true continuity in open and reproducible scientific knowledge, it is worth considering another perspective on legacy publishing in light of recent advances in generative artificial intelligence (GenAI).

LEGACY PUBLISHING CAN'T KEEP ITS FOOTING IN A GENAI WORLD

Even a decade before large language models (LLMs) made their sweeping entrance into scientific prose,[12] the oligopolistic nature of the publishing ecosystem had already been described as the most profitable obsolete technology in history.[13]

Against this backdrop of stagnation, the pace of innovation in AI has been nothing short of astonishing. The second quarter of 2025, just four days before the writing of this manuscript, witnessed the introduction of ARC-AGI-2, a new benchmark in the Abstraction and Reasoning Corpus for Artificial General Intelligence.[14] This update was prompted by a rapid rise in top performance on the previous benchmark, which increased from 34% in early 2024 to 88% by the end of the year (for the latest leaderboard see[15]). For comparison, members of the general public typically score around 77%, while STEM graduates score approximately 98%.

As LLMs approach expert-level proficiency in generating coherent and convincing prose so rapidly that we need to move the goalpost in favor of humans, it becomes increasingly important to ask: **what remains of a publication when we remove the prose?**[16]

Even twelve years before the initial public release of advanced AI agents, this challenge had already been articulated by[17]: "*An article about computational results is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result.*" The legacy publishing system is far too outdated to be retrofitted into an ecosystem where papers serve as more than advertisements.

It is only fair to retire the legacy publishing system with a new acronym: **BOOMER** (Barely Operational and Obsolete Manuscript Evaluation Rituals). Perhaps it is time to move toward something more relevant and forward-looking for the century we are in.

FROM BOOMER TO NEXT-GEN PUBLISHING

> *Science is what we understand well enough to explain to a computer. Art is everything else we do. [...] People think that science is the art of geniuses, but the actual reality is the opposite, just many people doing things that build on each other, like a wall of mini stones.*
>
> • Donald Knuth[18]

Scientific progress is often portrayed as the product of isolated breakthroughs by exceptional individuals. Yet as Knuth reminds us, its true foundation is incremental and collective, where people do small, explainable contributions that build on each other. The BOOMER paradigm stands in the way of this accumulative process, perpetuating a format in which knowledge is trapped in static, monolithic documents, divorced from the computational processes and materials that generated it.

Moving beyond this paradigm first requires a return to a more foundational principle that science is what we understand well enough to explain to a computer. This is not merely a technical assertion. It is a claim about findability, accessibility, interoperability, and reproducibility, i.e., the FAIRness of ideas.[19] It is also the central premise of Knuth's concept of literate programming,[20] in which the construction of a program is guided not just by what it does, but by how clearly its purpose and logic can be conveyed to others.

Literate programming proposes that *narrative and computation should not simply coexist, but be intricately woven together*[20]. It allows a program to function both as a computational artifact and as a document that communicates scientific intent. Even though it has not been proven effective for professional software engineering,[21] it has become one of the most preferred approaches for end users who code as part of scientific analysis, also known as end-user software engineering.[22] For example, a neuroimaging researcher using literate programming tools, such as Jupyter Notebooks,[23] to preprocess MRI data is engaging in end-user programming. In the 21st century where computation has become an indispensable part of almost any scientific discipline, literate programming tools boost exploratory analyses by binding prose, code, data, inputs, and outputs into a single space of reasoning.[24] Electronic lab notebooks offer a powerful example of such documentation, capturing the narrative of a researcher's data exploration to serve as a valuable asset for reproducibility.[25]

This approach to reproducible documentation naturally informs the design of a modern and federated publishing network.[26] Drawing inspiration from Knuth's concept of literate programming, we can envision **woven literature** as the natural outcome of such networked scientific publishing. In woven literature, narrative, data, code, and computational environments are stitched together from the outset, not retrofitted after publication. Each thread preserves its own integrity, yet contributes to a fabric that is more than the sum of its parts.

This shift is illustrated in Figure 1. A traditional static article (Figure 1a) presents a self-contained and static document, with a narrative frozen at the time of publication.

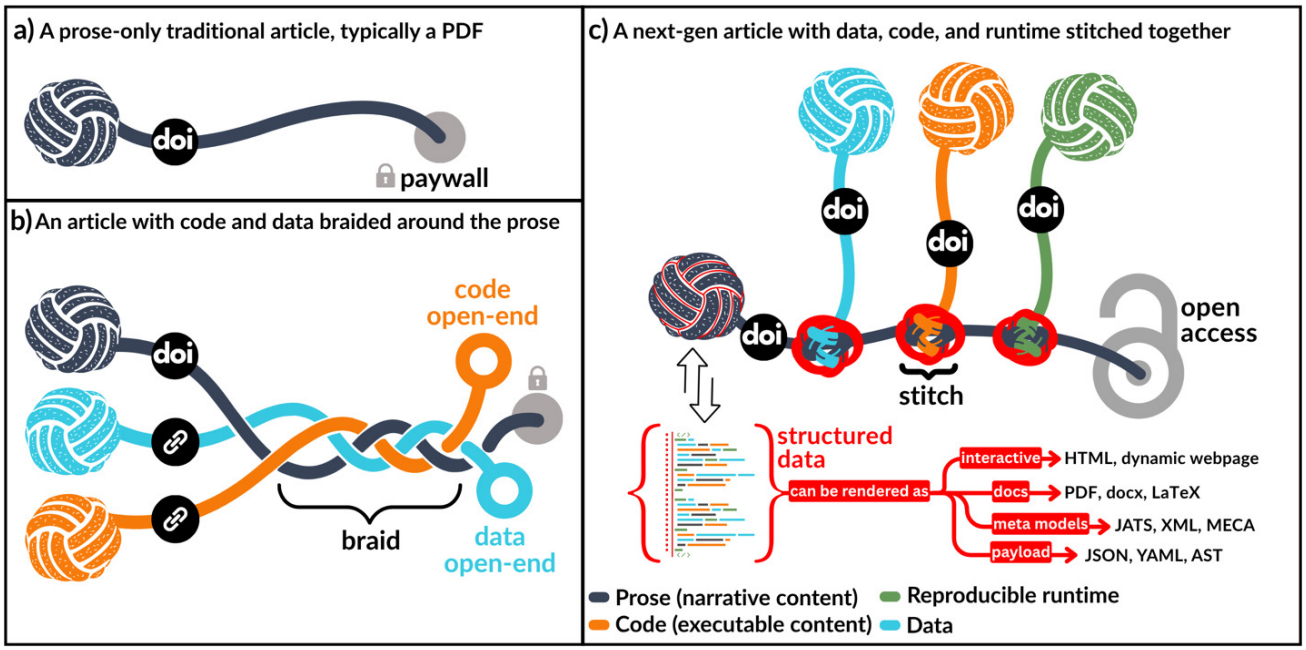

Figure 1. Evolution of article architecture.

(a) Traditional prose-only articles are static documents, typically behind paywalls, containing only narrative content. (b) Current articles with code and data availability statements connect to external resources through fragile hyperlinks that are vulnerable to reference rot, shown as braided strands with open ends. (c) Next-generation articles integrate code, data, and runtime environments as core structural components, creating a unified scholarly object. The structured data foundation enables multi-format rendering and serves as the building block for interconnected research networks, i.e., networked scientific publishing.

While it includes references and a DOI that link it to the broader literature, it lacks open and embedded access to the data, code, or computational context that produced its findings. Even when such artifacts are available (Figure 1b), they are typically hosted externally and linked as afterthoughts, which 20% of the time leads to reference rot.[27] A next-gen article (Figure 1c), by contrast, embodies the principles of literate programming. It stitches together code, data, runtime, and narrative into a cohesive structure. Moreover, representing the article's content as structured data enables seamless transclusion of its components into other documents across different platforms. This way, the article becomes not just a description of research but a functional expression of it.

Figure 2 shows how this transformation scales beyond individual articles to the scholarly ecosystem. Bibliographic literature (Figure 2a) supports credit and citation but not actionable reuse. Even when reproducibility artifacts are linked (Figure 2b), they remain external and structurally isolated. Woven literature (Figure 2c) introduces formal, machine-readable relationships among the elements of research.

The conceptualized relationships in Figure 2c, such as *reuse data* or *transclude a figure*, unlock several reuse patterns: i) transclusion, when one article directly inserts content from another article such as a paragraph or a figure, ii) longitudinal extension, when a living article re-executes itself with new data points to extend its findings, iii) horizontal adaptation, when an article uses another as a methodological template to apply computational frameworks to independent datasets for validation across different research contexts, and iv) cross-pollination, when articles aggregate and synthesize data or methods from multiple sources to generate novel insights that transcend individual publications.

The reuse patterns transform fragmented scientific publishing into what Knuth envisioned as true composability, where small, well-understood computational and narrative components combine to form increasingly sophisticated knowledge structures.[28] Such interlacing enables researchers working transparently and methodically to construct robust foundations for discovery, with each publication serving as both standalone contribution and building block for future work.

In conclusion, woven literature converts the culmination of scientific record from a static collection of isolated PDFs into a dynamic, interconnected (i.e., federated) infrastructure where cumulative knowledge emerges through systematic reuse and extension of validated components. The following subsection introduces open-source solutions and community efforts making interconnected, executable scholarship a practical reality.

OPEN-SOURCE PATHWAYS TO REALIZING WOVEN LITERATURE

NeuroLibre[29] represents one of the earliest fully open-source implementations aligned with the principles of woven literature: https://neurolibre.org. Developed under the Canadian Open Neuroscience Platform,[30] it started as a full-fledged reproducible publishing platform for neuroscience by integrating the open-source applications by JOSS[31] with the documentation engines of the Jupyter Project.[32] It currently hosts 19 next-gen preprints mostly in the fields of neuroimaging and MRI engineering, with ad-

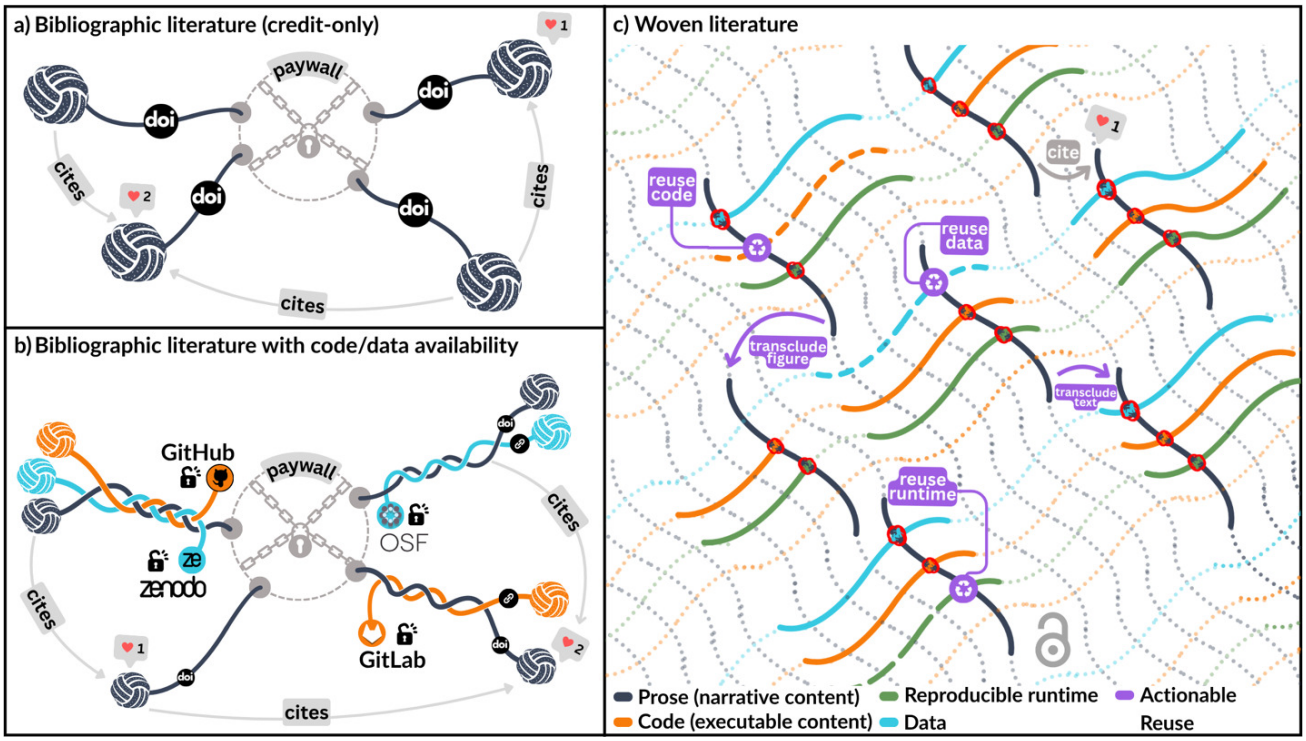

Figure 2. From citation networks to research webs.

(a) Traditional bibliographic literature forms simple citation chains between narrative publications, where papers function primarily as units of academic credit. (b) Enhanced bibliographic approaches add peripheral connections to code and data repositories through platforms like GitHub, GitLab, OSF, and Zenodo, but these remain vulnerable to link decay and disconnection. (c) Woven literature creates dense, multi-dimensional networks where research components form actionable connections that enable various reuse patterns such as transclusion (direct insertion of content from one article into another), longitudinal extension as new data becomes available, horizontal adaptation across different datasets, and cross-pollination of research elements throughout the scholarly ecosystem.

ditional submissions actively coming in as the platform moves toward its second iteration: https://evidencepub.io.

Each NeuroLibre publication is deeply interwoven with its reproducibility artifacts, which are automatically archived on Zenodo (https://zenodo.org) and served on demand through interactive computational environments. Nearly 60GB of input data is made readily accessible within these containerized reproducible runtimes, enabling real-time exploration and interactivity.

Other notable efforts toward modernizing scientific publishing include eLife's *Executable Research Articles*[33] and *Distill* (https://distill.pub),[34] both of which are no longer actively maintained. *Wholetale* (https://wholetale.org)[35] offered another model for stitching together reproducibility assets into executable articles, providing a limited set of example publications. *Quarto* (https://quarto.org) is an open-source scientific and technical publishing system that supports R, Python, Julia, and JavaScript, allowing users to produce documents, websites, and presentations that combine narrative with executable code. *Observable* (https://observablehq.com) provides a reactive notebook environment based in JavaScript, well-suited for building rich, interactive data visualizations and exploratory narratives, though its primary use has been outside formal scientific publishing. In parallel, *Curvenote* (https://curvenote.com)[36] has made substantial contributions to next-gen publishing workflows, particularly through its development of Markedly Structured Text (MyST) markdown language,[37] a foundational technology for bringing about the woven literature. Another open-source demonstration of MyST's potential came with you-only-write-thrice, a proof-of-concept workflow that enabled publishing the same content across multiple formats.[38]

While these initiatives differ in openness, scope, and long-term sustainability, NeuroLibre stands out as an actively growing, fully open, and community-driven infrastructure grounded in the principles of open science and reproducibility. Its publication workflow has been made possible by a thriving ecosystem of open-source projects, whose ongoing innovations have not only enabled its core functionality but also steadily expanded its capabilities.

The sections below provide a brief review of existing tools that support each thread of a next-gen article, including reproducible runtimes, code and data integration, and editorial management systems that orchestrate the publication process.

REPRODUCIBLE RUNTIME

One of the main building blocks of NeuroLibre is *BinderHub*, a cloud-native platform that enables users to share reproducible, interactive computing environments directly from code repositories.[39,40] Binder streamlines the process of creating containerized environments by reducing complex *Dockerfile* specifications to a small set of configuration files interpreted by language-agnostic buildpacks. This simplicity, combined with its seamless integration with Jupyter Notebooks, has made Binder one of the most widely

adopted tools for executable research. As of 2025, the public Binder infrastructure (https://mybinder.org) serves up to 12,000 daily sessions across thousands of repositories (https://archive.analytics.mybinder.org), demonstrating its pivotal role in lowering the barrier to reproducible, in-browser scientific computing.

Google Colab is another widely used platform for interactive computing, particularly popular in data science and machine learning education.[41] It allows users to write and execute Python code in a browser-based Jupyter Notebook interface backed by Google's cloud infrastructure. Colab's ease of use, integration with Google Drive, and free access to GPU and TPU resources have made it especially attractive for rapid prototyping, tutorials, and sharing notebooks across diverse user groups.

However, while both BinderHub and Colab offer interactive computational environments, they differ significantly in design philosophy and sustainability model. BinderHub is grounded in open science principles and supports community-led infrastructure. A notable example is 2i2c (https://2i2c.org), a nonprofit organization that operates scalable BinderHub services for research and education, exemplifying how interactive computing can be delivered sustainably through a community-driven approach. In contrast, Colab is a freemium product provided by a commercial entity, with priorities shaped by business models rather than academic interests.

BinderHub is fully open source and purpose-built to support reproducible science, emphasizing long-term preservation and transparency. It leverages the Reproducible Execution Environment Specification (REES),[42] which uses configuration files like *environment.yml, requirements.txt,* or *install.R* to explicitly declare and version-control runtime dependencies. These specifications are used to build Docker images that can be hosted in private registries and archived, providing a robust foundation for sustained reproducibility. This model aligns strongly with the goals of scientific publishing, where preserving the integrity and reusability of computational environments is essential.

In contrast, Colab's environments are transient and subject to changes in the underlying base images, which may lead to inconsistencies in results over time. While Colab offers generous access to GPU hardware out of the box (at least as of 2025), BinderHub deployments can be configured to integrate with GPU-enabled *Kubernetes* clusters or JupyterHub spawners, offering comparable computational capabilities within an infrastructure designed for reproducibility and archival.

An emerging alternative leverages *WebAssembly*,[43] allowing code execution directly in the browser without relying on a remote server. *Pyodide*, a Python distribution compiled to WebAssembly, enables this shift by running Python code client-side in a manner similar to JavaScript.[44] JupyterLite builds on Pyodide to offer a lightweight, serverless Jupyter environment that launches instantly in the browser.[45] Here, the user interacts with notebooks whose code is executed locally, eliminating the need for backend infrastructure.

This approach is particularly exciting for the future of interactive scientific publishing. It offers a scalable solution for delivering interactive papers without being constrained by centralized compute resources, making it ideal for publications with modest computational demands and minimal data dependencies. It opens the door to creating rich, Python-based interactive articles akin to those pioneered by https://distill.pub, without requiring expertise in web-native programming. However, ensuring long-term preservation of these browser-executed environments remains an open challenge that warrants further exploration in the context of scholarly publishing.

CODE AND PROSE

A wide range of open-source tools now support the creation of executable scientific documents, enabling researchers to combine code and scientific prose in a single environment for reproducible and interactive outputs. *R Markdown*, introduced in 2012, laid the groundwork for integrating analysis and narrative, allowing for dynamic document generation across various formats.[46] Two years later, *GitBook* (https://www.gitbook.com) emerged, popularizing structured, book-like web documentation, albeit not specifically tailored for scientific communication and currently evolved into a hosted commercial platform. In 2016, *Bookdown* extended R Markdown's capabilities, facilitating the creation of books and long-form reports with features like cross-referencing and multi-format output.[47]

The landscape of structured, git-backed web content was bridged with the Jupyter ecosystem by *JupyterBook*,[37] introduced in 2019 as part of the *Executable Books* (EB) (https://executablebooks.org). Leveraging Sphinx (https://www.sphinx-doc.org), a Python-based documentation engine, and MyST markdown language, it renders Jupyter Notebooks and *Markdown* files into interactive, publication-grade documents as static *HTML* files. Features such as collapsible code cells, interactive plotting, and live code execution via *Thebe*[48] enhance its suitability for reproducible and reader-friendly content.

The MyST markdown language has been a major driver of the EB's evolution, enabling its documentation engine to expand beyond a Sphinx-based static site generator toward a framework capable of serving documents as dynamic web pages. This transition (see more here) was supported through collaboration with Curvenote, which contributed a TypeScript library that leverages the *mdast* package from the *UnifiedJS* project (https://unifiedjs.com) to represent the entire content of Jupyter Notebooks and Markdown as abstract syntax trees. Building on this development, MyST-MD (https://mystmd.org) enables the full content of an article to be expressed as structured data (Figure 1c), supporting programmatic access and seamless transformation into a wide range of themes (e.g., single-page or book-like web layouts) and formats (e.g., PDF, XML, JATS, MECA, DOCX, or LaTeX) for publishing.

This structured data representation not only eliminates the need for complex format conversions in the publishing pipeline but also unlocks richer online functionality. It enables actionable content reuse, transclusion, and semantic

understanding of elements such as figures, narrative text, and code cells. As such, MyST-MD represents a significant milestone in the realization of woven literature, and is currently supported by NeuroLibre for dynamically publishing living preprints.

DATA

One of the most commonly missing pieces in building a next-gen article is the input data required for executable content to generate expected outputs. During its early development, the Binder project recognized this challenge and initially recommended storing datasets directly within GitHub repositories.[39] However, this was acknowledged as a stopgap measure rather than a viable long-term strategy, given GitHub's limitations as a platform designed primarily for version-controlling source code. Instead, the Binder team advocated for more robust approaches like Dat,[49] a distributed protocol for data synchronization and versioning.

To understand the importance of thoughtful data distribution in scientific publishing, it is useful to reflect on how digital media became widely shareable. Technologies like BitTorrent[50] revolutionized peer-to-peer file sharing by allowing users to download large files in small pieces from multiple sources, making the process more efficient and resilient. However, while BitTorrent excels at distributing static content, it offers limited support for collaboration or tracking how files evolve, capabilities that are essential for reproducible science.

Version control systems such as Git introduced structured mechanisms for tracking changes in source code across distributed teams. These same ideas have inspired tools for managing datasets, especially in contexts where traceability and reproducibility are essential. For example, git-annex extends Git to support versioning of large files without storing them directly in the repository. DataLad (https://www.datalad.org),[51] built on top of git and git-annex, is widely used in neuroscience for jointly managing code, data, and the links between them. It enables reproducible research workflows that track both inputs and outputs across time and collaborators.

Complementing these solutions, NeuroBagel (https://neurobagel.org) tackles a different but equally critical problem: metadata standardization and dataset interoperability. By harmonizing dataset descriptors, NeuroBagel enables federated queries across distributed datasets—making it easier for researchers to discover and access compatible data resources.[52]

A more general-purpose approach to distributed, versioned data is the InterPlanetary File System (IPFS).[53] Like Dat,[49] IPFS identifies files using content hashes, ensuring that a given address always resolves to the same version of a file. Based on peer-to-peer protocols similar to BitTorrent, IPFS can serve as a decentralized backbone for hosting static datasets and artifacts. This content-addressable architecture is particularly valuable in reproducible science, where exact snapshots of datasets must be preserved and referenced across time.

Another approach, particularly relevant to scientific computing infrastructures, is the CERN Virtual Machine File System (CVMFS).[54] Originally developed to distribute software in high-energy physics, CVMFS is well-suited to delivering large-scale datasets efficiently across distributed systems, with support for secure repositories. Its caching mechanisms and content-addressable design complement reproducibility goals by ensuring users access consistent versions of data, even at scale.

For platforms like NeuroLibre, which deploys executable scientific articles on JupyterHub instances running atop Kubernetes, mounting datasets to individual user sessions introduces unique challenges. Each reader session requires read-only access to the same inputs to ensure consistent reproducibility, while also being isolated and ephemeral.

To address this, NeuroLibre recognizes and supports data hosted on repositories like Zenodo, OSF, and DataLad, using a tool called repo2data to automatically fetch and cache these inputs in a reproducible manner. This makes it possible to reproduce results from linked datasets without requiring manual setup. Looking ahead, the next phase of NeuroLibre's infrastructure will adopt more scalable, efficient solutions such as CVMFS, paving the way for a more robust and larger ecosystem of executable research objects.

EDITORIAL MANAGEMENT AND PEER REVIEW

While advances in tools for code, data, and runtime integration have made it more technically feasible than ever to move beyond the BOOMER paradigm, infrastructure alone does not ensure reproducibility. Without careful curation and modern editorial oversight, executable content written in notebooks may still fall short of yielding truly reproducible outputs.[55]

To address this gap, NeuroLibre pioneered the integration of a technical screening process into the living preprint publication workflow. This process supports curation by establishing an iterative interaction between the submitting author, a technical screener, and an editorial bot (RoboNeuro) on a GitHub issue. Within this workflow, the runtime environment is verified, reproducibility checks are logged, and any necessary changes are communicated and resolved until the submission passes all technical criteria (see examples here).

The platform and tooling that support this next-gen technical screening workflow were built from a fork of the Open Journals' applications, *Buffy* and *JOSS*,[7] and were extended with additional components designed to interface with NeuroLibre APIs. Unlike the JOSS workflow, which involves peer review of both the scientific content and the functionality of the associated research software, NeuroLibre's technical screening focuses solely on ensuring that computational outputs are generated as expected. It does not involve scientific evaluation, similar to commonly used preprint services such as arXiv.org that has posted more than 1 million preprints over the last three decades.[56]

Interestingly, preprints have not only made publicly funded research more accessible, but have also spearheaded the development of new approaches to peer-review, such as eLife's "publish, then review" model.[57] This shift reflects

a broader recognition that review should not be a one-time gatekeeping event as in the BOOMER paradigm, but an ongoing, distributed process that involves diverse perspectives and evolves alongside the work itself. It is on this ground that new models and infrastructures are beginning to take shape, aiming to support more transparent, inclusive, and sustained forms of scholarly evaluation.

A key contributor to advancing this procedural transformation has been the Confederation of Open Access Repositories (COAR), which supports the development of aligned open access infrastructure through a global network of universities, funders, and research institutions.[58,59] One of its notable initiatives, COAR Notify, proposes a decentralized mechanism to connect research outputs hosted in distributed repositories with external services (such as overlay journals and open peer review platforms) using linked-data notifications.[60] This effort stands at the intersection of numerous preprint servers (e.g., arXiv, bioRxiv, medRxiv, Zenodo, Research Square, SciFLO, and Center of Open Science, to name a few), and more importantly, preprint review initiatives such as PreReview (https://prereview.org), Peer Community in (PCI, https://peercommunityin.org), eLife (https://elifesciences.org), Plaudit (https://plaudit.pub), SciPost (https://scipost.org), and Peeref (https://www.peeref.com). Additionally, platforms like ResearchHub (https://researchhub.com) are pioneering novel approaches to post-publication peer review through decentralized evaluation mechanisms and introducing ResearchCoin as an innovative initiative to shift incentive structures in favor of researchers. Collectively, these initiatives represent a growing movement to build a more open, interoperable, and participatory scholarly communication ecosystem where peer review and evaluation are no longer tied to the traditional boundaries of legacy publishers.[61]

RESULTS

Figure 3 highlights the structural shift from conventional publishing to a reproducible research article. On the left, the static PDF presents frozen text, figures, and code snippets, which offer limited transparency or reusability.[62] On the right, the same study is rendered as a next-gen article on NeuroLibre, where narrative prose is stitched together with executable code, openly accessible data, and an embedded runtime.[62] This format allows readers to interact with the actual software package and learn about its applications in real-time with zero installation. The NiMARE paper is one of 19 living preprints hosted on NeuroLibre that illustrate this paradigm in practice: https://neurolibre.org.

Figure 4 showcases how next-gen publishing on NeuroLibre enables the seamless integration of interactive dashboards as first-class figures. The MRShiny Brain dashboard (https://shinybrain.db.neurolibre.org) offers an interface for exploring structural, perfusion, and metabolic brain imaging data from a normative cohort.[63] Built using R Shiny, it empowers readers to navigate multidimensional outcome measures, such as metabolite concentrations and MR thermometry, without being constrained by static layouts or figure limits.

On the right, a Plotly Dash dashboard supports the ISMRM T1 Mapping Reproducibility Challenge, visualizing inter- and intra-site variability across phantom and human scans (https://rrsg2020.db.neurolibre.org). This web-native figure offers toggles, filters, and drill-down visualizations that would require dozens or even hundreds of static panels in a conventional PDF. These dashboards are deployed on NeuroLibre's dedicated Dokku (https://dokku.com) deployment for serving interactive data applications, underscoring the platform's commitment to supporting researchers motivated to communicate complex results through novel, reader-driven interfaces.

Reuse patterns enabled by federated publishing networks demonstrate their transformative potential for scientific validation and knowledge synthesis. Longitudinal extension is exemplified by,[64] where the publication evolved continuously through data submissions from multiple sites participating in a reproducibility challenge. The meta-analysis template provided in[65] illustrates the horizontal adaptation pattern, as it was subsequently reused by[66] with an entirely different dataset, enabling systematic validation of methodological robustness across research contexts. Cross-pollination, as defined earlier, can be seen in a proof of concept implementation at https://agahkarakuzu.github.io/paperception, yet no real-world examples currently exist.

DISCUSSION

ADOPTION OF NEXT-GEN PUBLISHING FALTERS FROM INCENTIVES, NOT INTERFACES

Despite the maturity and accessibility of technologies enabling next-gen articles, their adoption in mainstream scientific publishing remains limited. The culprit is not an insurmountable learning curve that prevents the broader adoption of such tools. If anything, for the next generation researchers, using tools like Binder and Jupyter Notebooks is as natural as word processors were for those trained in the BOOMER ecosystem. It is worth bearing in mind that this new cohort of scientists has grown up with smartphones and now navigates information fluently using LLMs. For them, the tools of woven literature feel intuitive in a landscape where coding for end-user programming is becoming less a specialized skill and more a facet of AI literacy. In this context, a gen-alpha researcher would aptly describe their fluency as "*vibe coding straight through those interactive plots into something reproducible, no sweat*".

The real obstacles lie elsewhere. For publishers, legacy business models tied to static formats continue to dominate, offering little incentive to adopt infrastructures that would undercut their profit margins. On the researchers' side, entrenched incentive structures and network effects still reward publication in "high-impact" journals, regardless of the reproducibility or interactivity of the work. Addressing these issues requires not just advocacy, but the development and visibility of alternative venues that demonstrate what modern scholarly communication can look like. Founded in 2025, the Continuous Science Foun-

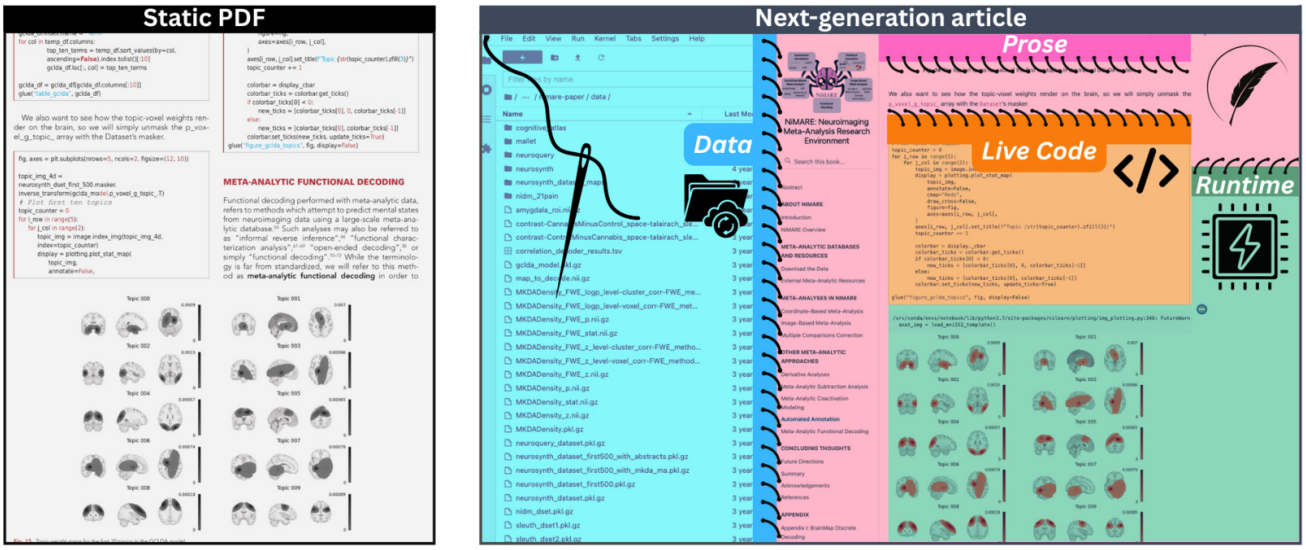

Figure 3. Comparison between a traditional static PDF article (left) and a next-gen reproducible article (right), illustrated using the *Neuroimaging Meta-Analysis Research Environment (NiMARE)* paper.

The next-gen article interlaces prose, live executable code, linked datasets, and computational runtime environments, supporting interactive and transparent scientific communication.

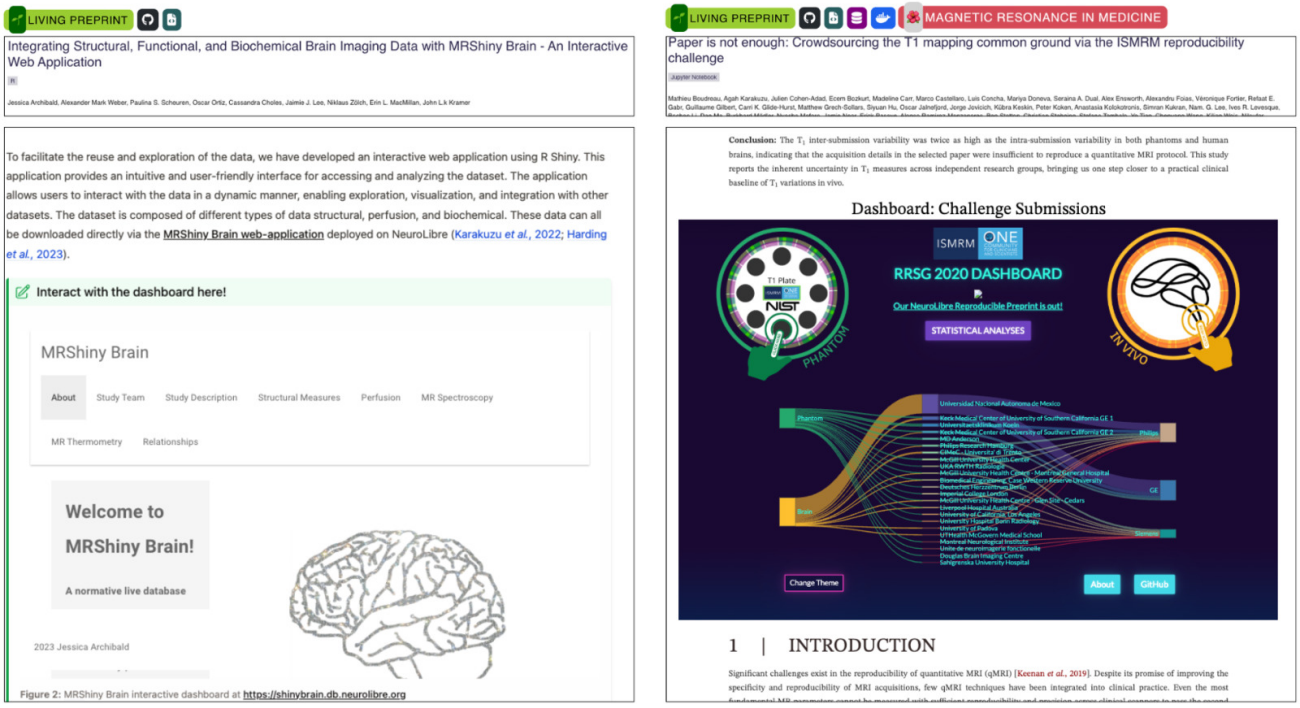

Figure 4. Representative examples of interactive dashboards embedded directly within living preprints hosted by NeuroLibre.

The left panel shows MRShiny Brain, an R Shiny dashboard presenting a normative neuroimaging dataset. The right panel features a Plotly Dash-based visualization from the ISMRM T1 Mapping Reproducibility Challenge. Both dashboards are integrated as figures within their respective articles, enabling dynamic exploration of complex results that exceed the limitations of traditional static figures.

dation (https://continuousfoundation.org) exemplifies this shift through a community-driven approach that places sustainability at its core.

An additional and often overlooked barrier is the difficulty of sustaining platforms like NeuroLibre through traditional scientific funding mechanisms. Despite their potential to transform research communication and reproducibility at scale, infrastructure projects of this nature often fall outside the scope of conventional grant criteria. While programs such as the Chan Zuckerberg Initiative's open science funding (https://chanzuckerberg.com) provide crucial support, there remains a need for broader awareness and recognition within public and national funding systems.

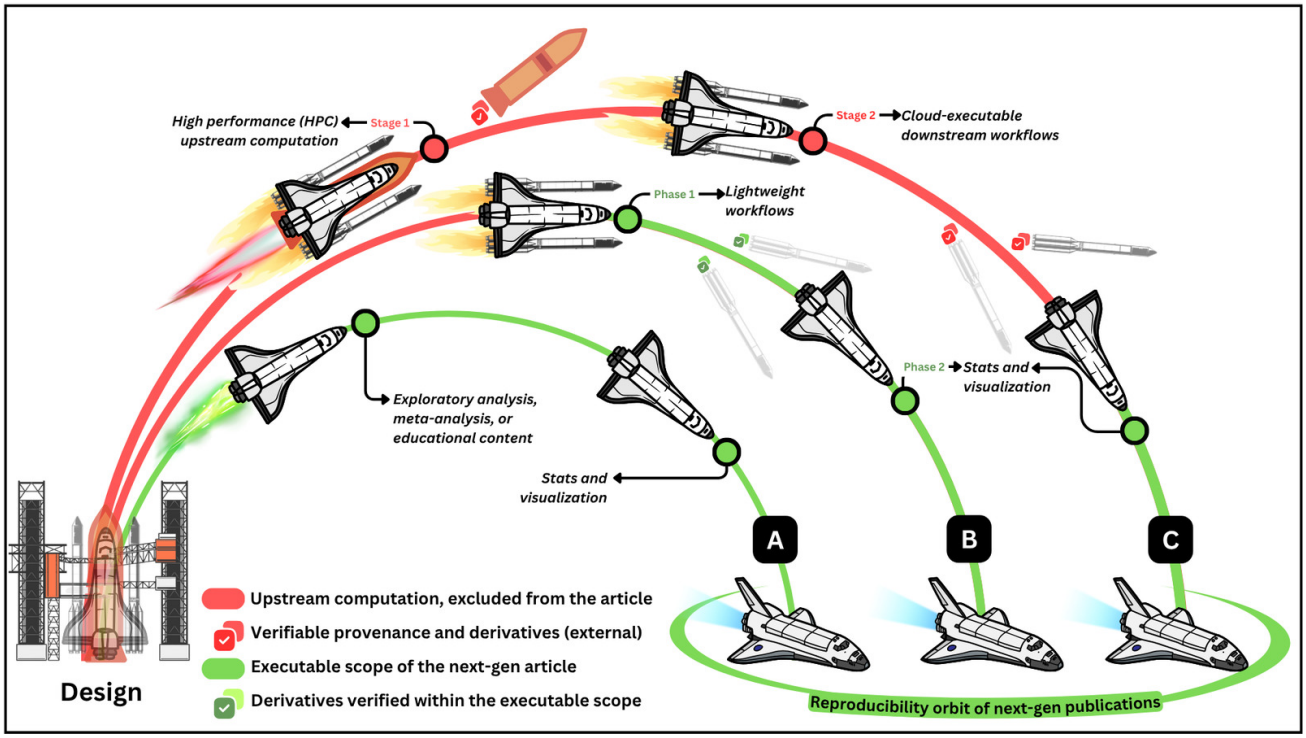

Figure 5. Scientific workflows entering the reproducibility orbit of next-gen publications.

Multi-stage tasks are visualized as space missions, where upstream *external stages* (red arcs) are executed externally and not reproduced in the article, while *internal phases* (green arcs) fall within the executable scope of the respective next-gen article. Checkmarks indicate verified provenance and derivable outputs from external stages (red) or internal phases (green).

Investments in open, reproducible publishing infrastructure may appear modest in size, but their long-term return in terms of saved time, avoided duplication, and enhanced scientific rigor is substantial. Building sustainable pathways for funding these efforts is essential to realizing the full promise of networked science publishing.[67]

DEFINING THE EDGE WHERE A SCIENTIFIC WORKFLOW TRANSITIONS INTO A NEXT-GEN PUBLICATION

The computational and data demands of next-gen articles vary significantly across research domains. For instance, neuroimaging studies often involve processing large volumes of high-resolution MRI data combined with advanced statistical and machine learning methods. Meanwhile, theoretical papers typically have lower data needs but may require intensive symbolic computation. In contrast, fields such as qualitative social sciences often involve relatively small datasets and minimal computational processing, focusing instead on interpretive analysis.

Figure 5 illustrates how scientific workflows enter the reproducibility orbit of next-gen publications by transitioning from external computation stages to internal executable phases. The degree to which a workflow enters this orbit depends on its complexity and structure. Some workflows are fully self-contained, with all operations reproducible in a single session, while others involve modular segments that depend on externally computed stages or pre-generated artifacts that are generated internally.

- **Trajectory A** represents lightweight use cases that do not require batch processing or multi-phase workflow execution logic. All computations, from data ingestion to analysis and visualization, are included within the article and can be executed independently, without strict ordering or dependencies between notebooks. Readers can interact with any part of the content asynchronously, reproducing results in one go without needing to trace interdependent phases.
- **Trajectory B** represents multiple interdependent *phases* that are all technically reproducible within the available computational limits, but differ in practicality for reader interaction. The earlier phases, such as data cleaning, simple model training, or transformation etc., may discourage readers from re-running them interactively. However, these phases are still included in the publication and can be executed if desired. To streamline the reading experience, pre-executed outputs from these initial phases are verified (green checkmarks) and provided as derivatives. Readers are encouraged to engage with the later, more interactive phases, such as analysis, visualization, or interpretation, without being required to reproduce the full pipeline from scratch. This model maintains full transparency and reproducibility, while accommodating asynchronous entry points aligned with reader expectations.
- **Trajectory C** reflects computationally intensive workflows whose upstream *stages* exceed the runtime or resource constraints of next-gen publishing plat-

forms. These workflows often involve multi-step pipelines, such as large-scale HPC preprocessing, simulation, or training procedures, that must be executed externally before publication. Only the final phase, typically involving analysis, visualization, or interpretive synthesis, is included within the executable scope of the next-gen article.

For example, NeuroLibre currently allocates up to 3GB of RAM and 1 CPU per living print user session, with support for storing up to 4GB of input data. These limits are often sufficient to meet the computational requirements of Trajectory A (Figure 5). Representative examples include: i) data-driven exploratory analyses, such as a living preprint by Bellec et al.[68] showcasing a word feature analysis using scikit-learn,[69] ii) a science communication analysis for interactive exploration of patent history,[70] iii) interactive tutorials covering neuroimaging meta-analyses[62] and quantitative MRI,[71,72] and iv) interactive meta-analyses performed focused on myelin imaging[65] and brain imaging demographics in Quebec.[66]

An example representative of Trajectory B is the living preprint resulting from a multi-center reproducibility challenge organized by the International Society for Magnetic Resonance in Medicine (ISMRM), which evaluated the inter-site reliability of T1 mapping.[64,73] In this case, the initial computational phase, such as aligning region-of-interest (ROI) masks and fitting T1 maps, was decoupled from the subsequent statistical analyses, which could be executed within a few minutes.

An example of Trajectory C can be found in a living preprint by Wang et al.,[74] which benchmarks the reproducibility of resting-state fMRI denoising strategies across varying preprocessing pipelines. The upstream processing required to generate these results involved running multiple fMRI pipelines under different parameter configurations, an operation that demanded high-performance computing resources beyond the scope of a typical interactive article. As such, only the downstream derivatives, including statistical analysis and visualization of the precomputed outputs, were integrated into the living preprint. This allowed readers to interact with and explore the results reproducibly, even though the computationally intensive preprocessing steps had to be completed externally.

SEAMLESS CONTINUATION FROM UPSTREAM WORKFLOWS TO REPRODUCIBLE PUBLISHING

Integrating outputs from upstream stages into next-gen articles in a reliable and reproducible way remains an open challenge. Fortunately, platforms like NeuroDesk have begun to address this gap by providing portable, containerized environments capable of running complex neuroimaging workflows.[75] NeuroDesk enables reproducible execution and captures the provenance between inputs and outputs, ensuring traceability across computational stages. Similarly, CBRAIN provides a web-based interface to high-performance computing resources, enabling researchers to run large-scale data analyses while preserving metadata and provenance.[76] Another example is BrainLife, focusing on reproducible neuroscience pipelines, offering cloud-based, interactive execution of modular workflows.[77] More recently, O^2S^2PARC was introduced as a cloud-based platform with a partial pay-per-use model to support the execution and sharing of multiscale computational models in life sciences.[78]

An integration between NeuroLibre and such platforms could offer a seamless continuation from heavy upstream processing to interactive, downstream communication, bridging the full research workflow that can extend from the study inception to publication.[79,80] Emerging infrastructures, such as Orvium (https://orvium.io), and DeSci Nodes (https://nodes.desci.com), further open possibilities for verifiable, tamper-proof records of these computational steps on blockchain, encouraging new models of trust and attribution in scholarly publishing.[81,82]

CONCLUSION

The ultimate convergence of technical and procedural publishing innovations reviewed in this article will mark a critical turning point in scholarly communication. By integrating reproducibility-aware open infrastructure with distributed models of peer review, scholarly publishing is approaching the replacement of BOOMER's static, one-shot logic with the continuous, collaborative, and verifiable attributes of the woven literature.

A Turkish expression, *"ilmek ilmek dokumak"* literally means "*to weave stitch by stitch*" but it is used metaphorically to describe doing something with great care, patience, and attention to detail. This captures the spirit of woven literature: a vision of scholarly publishing built meticulously by interlacing code, data, prose, and peer review into a coherent whole, one next-gen article at a time.

As we move forward, this stitch-by-stitch approach offers not just a more reproducible and transparent future, but also one that values the craft of knowledge creation itself, a quality more vital than ever in an era shaped by the rapid rise of generative AI and automated content.

The threads are already in motion; what lies ahead is ours to weave. We can choose to remain entangled in the static logic of the BOOMER paradigm, or co-create a knowledge network that evolves through novel reuse patterns, modularity, and reproducibility.

ACKNOWLEDGEMENTS

I thank Nikola Stikov for insightful discussions, for reading the initial draft of this manuscript, and for providing thoughtful feedback. I am also grateful to the members of the Canadian Open Neuroscience Platform (CONP) Publishing Committee and to the researchers who contributed to the funding and development of NeuroLibre, including Elizabeth DuPre, Lune Bellec, Jean-Baptiste Poline, Mathieu Boudreau, Patrick Bermudez, Samir Das, and Alan C. Evans. Special thanks to Rowan Cockett and Chris Holdgraf, as well as all members of the Executable Books and Project

Jupyter communities for their unwavering commitment to improving the tools that unlocked the networked science publishing.

Funding was received from the CONP (https://conp.ca), Brain Canada (https://braincanada.ca), Courtois Neuromod (https://www.cneuromod.ca), Quebec Bio-imaging Network (QBIN), and the Canada First Research Excellence Fund (TransMedTech).

CONFLICTS OF INTEREST

The author declares no conflicts of interest.

Submitted: June 16, 2025 CDT. Accepted: October 13, 2025 CDT. Published: November 04, 2025 CDT.

REFERENCES

1. da C. Andrade EN. The birth and early days of the Philosophical Transactions. *Notes Rec R Soc Lond*. 1965;20:9-27. doi:10.1098/rsnr.1965.0002

2. Schlüter L, Vinken PJ. *The Elsevier Non Solus Imprint*. Elsevier Science; 1997.

3. Fuentes-Utrilla P, López-Rodríguez RA, Gil L. The historical relationship of elms and vines. *Forest Systems*. 2004;13:7-15. doi:10.5424/808

4. Heybroek HM. The elm, tree of milk and wine. *IForest*. 2015;8:181-186. doi:10.3832/ifor1244-007

5. Larivière V, Haustein S, Mongeon P. The oligopoly of academic publishers in the digital era. *PLoS One*. 2015;10:e0127502. doi:10.1371/journal.pone.0127502. PMID:26061978

6. Katz DS, Barba LA, Niemeyer K, Smith AM. Cost models for running an online open journal. *Journal of Open Source Software Blog*. Published online 2019. doi:10.59349/g4fz2-1cr36

7. Smith AM, Niemeyer KE, Katz DS, Barba LA, Githinji G, Gymrek M, et al. Journal of Open Source Software (JOSS): design and first-year review. *PeerJ Prepr*. 2018;4:e147. doi:10.7717/peerj-cs.147. PMID:32704456

8. Van Noorden R. Open access: The true cost of science publishing. *Nature*. 2013;495:426-429. doi:10.1038/495426a

9. Maddi A, Sapinho D. Article Processing Charges based publications: to which extent the price explains scientific impact? *Proceedings of the ISSI 2021*. Published online 2021.

10. DuPre E, Holdgraf C, Karakuzu A, Tetrel L, Bellec P, Stikov N, et al. Beyond advertising: New infrastructures for publishing integrated research objects. *PLoS Comput Biol*. 2022;18:e1009651. doi:10.1371/journal.pcbi.1009651. PMID:34990466

11. Cockett R. Continuous Science: Accelerating the speed of scientific discoveries through changes in collaboration. Published online 2024. doi:10.62329/wjck7598

12. Meyer JG, Urbanowicz RJ, Martin PCN, O'Connor K, Li R, Peng PC, et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Min*. 2023;16:20. doi:10.1186/s13040-023-00339-9. PMID:37443040

13. Schmitt J. Academic Journals: The Most Profitable Obsolete Technology in History. Huffpost. 2014. Accessed May 24, 2025. https://www.huffpost.com/entry/academic-journals-the-mos_b_6368204

14. Chollet F, Knoop M, Kamradt G, Landers B, Pinkard H. ARC-AGI-2: A new challenge for frontier AI reasoning systems. *arXiv*. Published online 2025. doi:10.48550/ARXIV.2505.11831

15. ARC-AGI Community. ARCPRIZE leaderboard. 2025. Accessed May 24, 2025. https://arcprize.org/leaderboard

16. Stikov N. ChatGPT and Galactica are taking scientific papers to their logical conclusion. Kantarot [Internet]. 2022. Accessed June 9, 2025. https://qantarot.substack.com/p/chatgpt-and-galactica-are-taking

17. Donoho DL. An invitation to reproducible computational research. *Biostatistics*. 2010;11:385-388. doi:10.1093/biostatistics/kxq028

18. Knuth DE, Shustek L. Let's not dumb down the history of computer science. *Commun ACM*. 2021;64:33-35. doi:10.1145/3442377

19. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. doi:10.1038/sdata.2016.18. PMID:26978244

20. Knuth DE. Literate Programming. *Comput J*. 1984;27:97-111. doi:10.1093/comjnl/27.2.97

21. Lethbridge TC, Singer J, Forward A. How software engineers use documentation: the state of the practice. *IEEE Softw*. 2003;20:35-39. doi:10.1109/MS.2003.1241364

22. Ko AJ, Abraham R, Beckwith L, Blackwell A, Burnett M, Erwig M, et al. The state of the art in end-user software engineering. *ACM Comput Surv*. 2011;43:1-44.

23. Perez F, Granger BE. IPython: A system for interactive scientific computing. *Comput Sci Eng*. 2007;9:21-29. doi:10.1109/MCSE.2007.53

24. Kery MB, Radensky M, Arya M, John BE, Myers BA. The story in the notebook: Exploratory data science using a literate programming tool. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM; 2018. doi:10.1145/3173574.3173748

25. Klokmose CN, Zander PO. Rethinking Laboratory Notebooks. In: *Proceedings of COOP 2010*. Springer London; 2010:119-139. doi:10.1007/978-1-84996-211-7_8

26. Project Jupyter, Bolyen E, Gregory Caporaso J, Cockett R, Garside D, Holdgraf C, et al. Jupyter book 2 and the MyST document stack. In: *Python in Science Conference, 2025*. ; 2025:173-193. doi:10.25080/hwcj9957

27. Klein M, Van de Sompel H, Sanderson R, Shankar H, Balakireva L, Zhou K, et al. Scholarly context not found: one in five articles suffers from reference rot. *PLoS One*. 2014;9:e115253. doi:10.1371/journal.pone.0115253. PMID:25541969

28. Knuth DE. Theory and practice. *arXiv [csGL]*. Published online 1993. doi:10.48550/ARXIV.CS/9301114

29. Karakuzu A, DuPre E, Tetrel L, Bermudez P, Boudreau M, Chin M, et al. NeuroLibre: A preprint server for full-fledged reproducible neuroscience. Published online 2022. doi:10.31219/osf.io/h89js

30. Harding RJ, Bermudez P, Bernier A, Beauvais M, Bellec P, Hill S, et al. The Canadian Open Neuroscience Platform-An open science framework for the neuroscience community. *PLoS Comput Biol*. 2023;19:e1011230. doi:10.1371/journal.pcbi.1011230. PMID:37498959

31. Katz DS, Niemeyer KE, Smith AM. Publish your software: Introducing the Journal of Open Source Software (JOSS). *Comput Sci Eng*. 2018;20:84-88. doi:10.1109/MCSE.2018.03221930

32. Perez F, Granger BE. Project Jupyter: Computational narratives as the engine of collaborative data science. 2015;11:108.

33. Maciocci G, Tsang E, Bentley N, Aufreiter M. Reproducible Document Stack: Towards a scalable solution for reproducible articles. eLife. 2018. Accessed June 5, 2025. https://elifesciences.org/labs/b521cf4d/reproducible-document-stack-towards-a-scalable-solution-for-reproducible-articles

34. Hohman F, Conlen M, Heer J, Chau DHP. Communicating with interactive articles. *Distill*. 2020;5:e28. doi:10.23915/distill.00028

35. Brinckman A, Chard K, Gaffney N, Hategan M, Jones MB, Kowalik K, et al. Computing environments for reproducibility: Capturing the "Whole Tale." *Future Generation Computer Systems*. 2019;94:854-867. doi:10.1016/j.future.2017.12.029

36. Cockett R, Purves S, Koch F, Morrison M. Continuous Tools for Scientific Publishing: Using MyST Markdown and Curvenote to encourage continuous science practices. In: *Proceedings of the Python in Science Conference*. SciPy; 2024:121-136. doi:10.25080/NKVC9349

37. Holdgraf C. Extending Open Standards for Scientific Communication with Jupyter Book and MyST Markdown. *AGU Fall Meeting Abstracts*. Published online 2021:IN55D-02.

38. Sokol K, Flach P. You only write thrice: Creating documents, computational notebooks and presentations from a single source. *arXiv [csPL]*. Published online 2021. doi:10.48550/ARXIV.2107.06639

39. Freeman J, Osheroff A. Toward publishing reproducible computation with Binder. *eLife*. 2016;13.

40. Ragan-Kelley B, Willing C, Akici F, Lippa D, Niederhut D, Pacer M. Binder 2.0-Reproducible, interactive, sharable environments for science at scale. In: Akici F, Lippa D, Niederhut D, Pacer M, eds. *Proceedings of the 17th Python in Science Conference*. ; 2018:113-120. doi:10.25080/Majora-4af1f417-011

41. Bisong E. Google Colaboratory. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress; 2019:59-64. doi:10.1007/978-1-4842-4470-8_7

42. Project Jupyter Contributors. The Reproducible Execution Environment Specification. 2025. Accessed June 6, 2025. https://repo2docker.readthedocs.io/en/latest/specification.html

43. Haas A, Rossberg A, Schuff DL, Titzer BL, Holman M, Gohman D, et al. Bringing the web up to speed with WebAssembly. In: *Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation*. ACM; 2017. doi:10.1145/3062341.3062363

44. Jefferson T, Gregg C, Piech C. PyodideU: Unlocking python entirely in a browser for CS1. In: *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V 1*. ACM; 2024:583-589. doi:10.1145/3626252.3630913

45. Ochkov VF, Stevens A, Tikhonov AI. Jupyter notebook, JupyterLab – integrated environment for STEM education. In: *2022 VI International Conference on Information Technologies in Engineering Education (Inforino)*. IEEE; 2022. doi:10.1109/inforino53888.2022.9782924

46. Xie Y, Allaire JJ, Grolemund G. *R Markdown: The Definitive Guide*. Chapman and Hall/CRC; 2018. doi:10.1201/9781138359444

47. Xie Y. *Bookdown: Authoring Books and Technical Documents with R Markdown*. Chapman and Hall; 2016. doi:10.1201/9781315204963

48. Dehaye PO, Iancu M, Kohlhase M, Konovalov A, Lelièvre S, Müller D, et al. Interoperability in the OpenDreamKit project: The math-in-the-middle approach. In: *Lecture Notes in Computer Science*. Springer International Publishing; 2016:117-131. doi:10.1007/978-3-319-42547-4_9

49. Ogden M, McKelvey K, Madsen MB, et al. Dat-distributed dataset synchronization and versioning. *Open Science Framework*. 2017;10. doi:10.31219/osf.io/nsv2c

50. Pouwelse J, Garbacki P, Epema D, Sips H. The bittorrent P2P file-sharing system: Measurements and analysis. In: *Peer-to-Peer Systems IV*. Springer Berlin Heidelberg; 2005:205-216. doi:10.1007/11558989_19

51. Halchenko YO, Meyer K, Poldrack B, Solanky DS, Wagner AS, Gors J, et al. DataLad: distributed system for joint management of code, data, and their relationship. *J Open Source Softw*. 2021;6:3262. doi:10.21105/joss.03262. PMID:39469147

52. Urchs S, Dai A, Armoza J, Jahanpour A, Bhagwat N, Poline JB. neurobagel_api. Zenodo. doi:10.5281/ZENODO.8088140

53. Benet J. IPFS - Content Addressed, Versioned, P2P File System. *CoRR*. 2014;abs/1407.3561. http://arxiv.org/abs/1407.3561

54. Bocchi E, Blomer J, Mosciatti S, Valenzuela A. CernVM-FS powered container hub. *EPJ Web Conf*. 2021;251:02033. doi:10.1051/epjconf/202125102033

55. Samuel S, Mietchen D. Computational reproducibility of Jupyter notebooks from biomedical publications. *Gigascience*. 2024;13. doi:10.1093/gigascience/giad113

56. Bourne PE, Polka JK, Vale RD, Kiley R. Ten simple rules to consider regarding preprint submission. *PLoS Comput Biol*. 2017;13:e1005473. doi:10.1371/journal.pcbi.1005473. PMID:28472041

57. Eisen MB, Akhmanova A, Behrens TE, Harper DM, Weigel D, Zaidi M. Implementing a "publish, then review" model of publishing. *Elife*. 2020;9. doi:10.7554/eLife.64910

58. Shearer K. Aligning Repository Networks and the Confederation of Open Access Repositories (COAR). Published online 2015.

59. Shearer K. *Publish, Review, Curate A Better Model for Scholarly Publishing*. Digital.CSIC; 2024. doi:10.20350/DIGITALCSIC/16767

60. Walk P, Klein M, Van de Sompel H, Shearer K. Modeling Overlay Peer Review Processes with Linked Data Notifications. *Confederation of Open Access Repositories*. Published online 2020.

61. Henriques SO, Rzayeva N, Pinfield S, Waltman L. Preprint review services: Disrupting the scholarly communication landscape? *SocArXiv*. Published online 2023. doi:10.31235/osf.io/8c6xm

62. Salo T, Yarkoni T, Nichols T, Poline JB, Bilgel M, Bottenhorn K, et al. NiMARE: Neuroimaging meta-analysis research environment. *NeuroLibre*. 2022;1:7. doi:10.55458/neurolibre.00007

63. Archibald J, Weber AM, Scheuren PS, Ortiz O, Choles C, Lee JJ, et al. Integrating Structural, Functional, and Biochemical Brain Imaging Data with MRShiny Brain - An Interactive Web Application. *NeuroLibre Reproducible Preprints*. Published online 2024:29. doi:10.55458/neurolibre.00029

64. Boudreau M, Karakuzu A, Cohen-Adad J, Bozkurt E, Carr M, Castellaro M, et al. Repeat it without me: Crowdsourcing the T1 mapping common ground via the ISMRM reproducibility challenge. *Magn Reson Med*. Published online 2024. doi:10.1002/mrm.30111

65. Mancini M, Karakuzu A, Cohen-Adad J, Cercignani M, Nichols TE, Stikov N. An interactive meta-analysis of MRI biomarkers of myelin. *Elife*. 2020;9. doi:10.7554/eLife.61523

66. Sintu T, Oyeniran O, Anazodo U, Leener BD, Alonso-Ortiz E. Representation in Brain Imaging Research: A Quebec demographic overview. *NeuroLibre Reproducible Preprints*. Published online 2025:35. doi:10.55458/neurolibre.00035

67. Nielsen M. *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press; 2011.

68. Bellec P, Jbabdi S, Craddock RC. Parcellating the parcellation issue - a proof of concept for reproducible analyses using Neurolibre. *NeuroLibre Reproducible Preprints*. 2023;10. doi:10.55458/neurolibre.00010

69. Kramer O. Scikit-Learn. In: *Studies in Big Data*. Springer International Publishing; 2016:45-53. doi:10.1007/978-3-319-33383-0_5

70. McLean E, Abdo J, Blostein N, Stikov N. Little Science, Big Science, and Beyond: How Amateurs Shape the Scientific Landscape. *NeuroLibre Reproducible Preprints*. Published online 2024:31. doi:10.55458/neurolibre.00031

71. Karakuzu A, Boudreau M, Duval T, Boshkovski T, Leppert I, Cabana JF, et al. qMRLab: Quantitative MRI analysis, under one umbrella. *J Open Source Softw*. 2020;5:2343. doi:10.21105/joss.02343

72. Boudreau M, Keenan KE, Stikov N. Quantitative T1 MRI. *NeuroLibre Reproducible Preprints*. Published online 2023:19. doi:10.55458/neurolibre.00019

73. Karakuzu A, Biswas L, Cohen-Adad J, Stikov N. Vendor-neutral sequences and fully transparent workflows improve inter-vendor reproducibility of quantitative MRI. *Magn Reson Med*. 2022;88:1212-1228. doi:10.1002/mrm.29292

74. Wang HT, Meisler SL, Sharmarke H, Clarke N, Paugam F, Gensollen N, et al. A reproducible benchmark of resting-state fMRI denoising strategies using fMRIPrep and Nilearn. Published online 2023. doi:10.55458/neurolibre.00012

75. Renton AI, Dao TT, Johnstone T, Civier O, Sullivan RP, White DJ, et al. Neurodesk: an accessible, flexible and portable data analysis environment for reproducible neuroimaging. *Nat Methods*. 2024;21:804-808. doi:10.1038/s41592-023-02145-x. PMID:38191935

76. Sherif T, Rioux P, Rousseau ME, Kassis N, Beck N, Adalat R, et al. CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Front Neuroinform*. 2014;8:54. doi:10.3389/fninf.2014.00054. PMID:24904400

77. Hayashi S, Caron BA, Heinsfeld AS, Vinci-Booher S, McPherson B, Bullock DN, et al. Brainlife.Io: A decentralized and open-source cloud platform to support neuroscience research. *Nat Methods*. 2024;21:809-813. doi:10.1038/s41592-024-02237-2. PMID:38605111

78. Guidon M, Kaiser D, Iavarone E, Anderegg S, Bisgaard M, Bujard C, et al. Shaping a collaborative, sustainable, accessible, and reproducible future for computational modeling. *bioRxiv*. Published online 2025. doi:10.1101/2025.06.12.656959

79. Karakuzu A, Blostein N, Caron AV, Boré A, Rheault F, Descoteaux M, et al. Rethinking MRI as a measurement device through modular and portable pipelines. *MAGMA*. Published online 2025. doi:10.1007/s10334-025-01245-3

80. Niso G, Botvinik-Nezer R, Appelhoff S, De La Vega A, Esteban O, Etzel JA, et al. Open and reproducible neuroimaging: From study inception to publication. *Neuroimage*. 2022;263:119623. doi:10.1016/j.neuroimage.2022.119623. PMID:36100172

81. Niya SR, Pelloni L, Wullschleger S, Schaufelbuhl A, Bocek T, Rajendran L, et al. A blockchain-based scientific publishing platform. In: *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE; 2019. doi:10.1109/bloc.2019.8751379

82. van Rossum J. The blockchain and its potential for science and academic publishing. *Information Services and Use*. 2018;38:95-98. doi:10.3233/ISU-180003