Open Access | https://doi.org/10.62329/kcep6732

Continuous Science Foundation

From Tools to Adoption

A Path to Modular and Interactive Scientific Publishing

Rowan Cockett[1,2], and **Tracy Teal**[3]

[1]Continuous Science Foundation, [2]Curvenote Inc., [3]openRxiv

Abstract

Scientific content remains trapped in static formats that limit reuse, attribution, and machine readability, despite the availability of modern tooling that makes this possible and better matches how scientists conduct research. This initiative, led by openRxiv and Continuous Science Foundation, addresses this challenge through practical implementation rather than consensus-driven standardization. The project will convene developers of leading modular publishing tools—MyST, Quarto, Curvenote, and Kotahi—to create a working, federated reference architecture using real bioRxiv content and deploying through the newly created openRxiv labs. Through a facilitated virtual meeting followed by a 2.5-day in-person implementation sprint, participants will develop shared interoperability standards and demonstrate live reuse of scientific content. Following the successful ipynb model, the approach prioritizes working software over abstract frameworks. Key deliverables include a reference implementation demonstrating cross-platform interoperability, pilot deployments using bioRxiv articles, progress updates at the CZI Open Science Meeting, and adoption materials for publishers and platforms. This implementation-first approach represents a concrete step toward modular publishing that meets current needs while preparing for the LLM era.

1. From Tools to Adoption: A Path to Modular and Interactive Scientific Publishing

1.1. Workshop Overview

In San Diego November 6-8th, 2025 a group of 25 open-science leaders gathered at the San Diego Made Factory for an in-person implementation workshop to unify modular scientific publishing tools around shared reference implementations. The working meeting, hosted by openRxiv and Continuous Science Foundation convened developers of leading modular publishing tools and licensing to create a working, federated reference architecture using real bioRxiv content, connected authoring tools, and worked towards shared standards. Our goal was to resolve technical gaps, demonstrate live interoperability, and lay the foundation for modular, machine-readable research publishing. The workshop was preceded by a virtual kickoff and included discussions and activities at CZI's Open Science Meeting and JupyterCon. This meeting was designed to unlock adoption by showing, not telling, how modular publishing can work in practice today.

Our approach for the workshop was to convene a small group of 25 leaders and tool makers in open science to align on a collective vision/demonstrations and implementations to drive meaningful change. Given the small size and implementation focus, this meeting was by invitation only.

This meeting is funded by the Meeting Fund of The Navigation Fund Open Science Program. The Navigation Fund [1].

Published Nov 09, 2025

Correspondence to
Rowan Cockett
rowan@continuousfoundation.org

Figure 1. *Participants of the workshop in November 8, 2025 in San Diego, California. Agah Karakuzu, Anton Molina, Carlos Scheidegger, Carol Willing, Chris Wilkinson, Franklin Koch, Jason Priem, John Bohannon, John Kaye, Kevin-John Black, Matt Akamatsu, Michael Markie, Milton Pividori, Monica Granados, Nokome Bentley, Paul Shannon, Raj Palleti, Rose Reatherford, Rowan Cockett, Steinn Sigurdsson, Taylor Campbell, Ted Roeder, Tom Scott, Tracy Teal.*

1.2. The Three Days

Over two and a half days, we moved from setting shared context and vision to hands-on building and prototyping, culminating in demos and concrete next steps. Each day built on the previous one, with structured discussions giving way to parallel work streams and finally public presentations of what we'd accomplished together.

🗺️ Setting Context & Direction (Nov 6)

We kicked off with lightning talks from Quarto, eLife, Creative Commons, arXiv, and others, then dove into ideation sessions to dream up an idealistic future for scientific communication. The group aligned on a powerful metaphor—Bedrock, Soil, and Flowers—to describe the evolving structure of scientific knowledge, and broke into working groups to tackle concrete implementation challenges. Read about the full day's discussions, the case studies we explored, and how we organized ourselves for the work ahead.

/day-1

👩💻 Hacking, Building, Vibe-ing (Nov 7)

Teams worked in parallel on six key initiatives: building the Open Exchange Architecture (OXA) schemas, integrating Creative Commons licensing, extending MECA for interoperability, exploring computational research content with AI, prototyping a "Spotify for Science" application, and identifying model communities for pilot deployments. See what we built, the technical decisions we made, and how the different groups collaborated throughout the day.

/day-2

🚀 Demo Day & Next Steps (Nov 8)

We presented our progress: a new OXA format with working schemas and repositories, a TASLAR framework for modular licensing and attribution, a live "Spotify for Science" prototype, and discussions on computational research content and community adoption. Then we mapped out concrete next steps for moving these initiatives forward. Watch the demos and see where we're heading next.

/day-3

1.3. Participants

Name	Organization
Agah Karakuzu	NeuroLibre
Anton Molina	b.next
Carlos Scheidegger	Quarto/Posit
Carol Willing	Willing Consulting
Chris Wilkinson	PreReview
Franklin Koch	Curvenote, Jupyter
Jason Priem	OpenAlex
John Bohannon	alphaXiv
John Kaye	octopus
Kevin-John Black	bioRxiv/openRxiv
Matt Akamatsu	DiscourseGraphs
Michael Markie	eLife
Milton Pividori	University of Colorado Anschutz Medical Campus
Monica Granados	Creative Commons
Nokome Bentley	Stencila
Paul Shannon	eLife
Raj Palleti	alphaXiv
Rose Reatherford	PLOS
Rowan Cockett	CSF, Curvenote, Jupyter
Steinn Sigurdsson	arXiv
Taylor Campbell	Creative Commons
Ted Roeder	bioRxiv/openRxiv
Tom Scott	PLOS
Tony Alves	HighWire
Tracy Teal	openRxiv

Regrets: Stephan Van der Walt (Jupyter), Chris Hartgerink (ResearchEquals), Gabe Stein (KnowledgeFutures), Chris Holdgraf / Greg Caporaso (MyST Steering Council), Adam Hyde (Coko/Katahi).

2. Day 1

2.1. Setting Context

Before the in-person meeting, we went through the context of where the group is, and what we are all thinking and expecting to get out of the meeting. In this two-hour virtual call we tried to set some of the intention that our role in this workshop was to work “bottom-up” rather than “top-down”; we are aware of many of the incentive problems in the scholarly ecosystem, but in the group we are bringing together have limited direct ability to influence or change them — especially in two days. Our role, however, can be to enable new possibilities through technology, organizational connections, and sparks of new ideas that come from these in-person events. We can show, not tell, what exciting things are possible in scientific communication. As we shared in that meeting, especially our ambitious goals of introducing new formats for scientific communication someone inevitably shared that xkcd comic about “there being 14 competing standards… now 15.” It’s funny because it’s true — ***sometimes***. We tried to identify some of the paradigm shifts at play — AI, cloud computing, distributed & composable research — and why sometimes those efforts need new ways to communicate.

> The medium of communication determines what you can communicate; and professional scientific communication is based on print.

The way we do science has changed, but the way we communicate it has not kept up. To illustrate through an example of standards evolution we introduced a case study on the format evolution of compressed and distributed scientific data; from ZIP –> HDF5 –> Zarr, and how the format changed to suit new capabilities and ultimately has enabled new ecosystems with each evolution. You can read the full case study here [2].

We also shared a case study on Jupyter and JupyterHub, where Carol Willing talked about the beginnings of these conversations that started with a dream, ambition, values, and a community of extremely talented people. Both of these tools and standards a decade later have millions of users and support unique and **fundamentally** new use cases that were not previously possible.

> We invited participants to dream, to choose an open, community approach, and execute.

=== Lightning Talks

To further set the stage, we invited lightning talks from Quarto, eLife, Creative Commons, arXiv, JATS, COAR Notify, Stencila, NueroLibre, openRxiv, Curvenote, OpenAlex, and MECA. These set a baseline understanding of some of these tools and approaches in play and the roles that people in the room are working on that can weave together and interoperate.

Figure 2. *We presented a case study [2] of 30 years of evolution in sharing scientific data, from zip files in the 1989 to HDF5 that introduced structured metadata alongside binary arrays to open up parallel IO on clusters to Zarr introduced in 2015 that enabled distributed data-access in a simplified, cloud-native way. These new formats changed the ecosystem at each stage, and also brought with them the best practices from the standard before.*

Figure 3. *A small sample of some of the brainstorming and ideas from the group.*

2.2. *Ideation and Group Discussion*

We prompted the group to ideate, dream on an idealistic future, and then focus on concrete actions that we could take in the next day and a half that would make progress towards that future. We took 10 minutes and wrote ideas on sticky notes. Then we paired up to discuss, identify themes and create new ideas. Pairs then connected into groups of 5 or 6, and those groups reported out on the themes.

> What do you wish was possible in scientific communication that is not possible or broadly adopted today?

2.2.1. *Common Themes*

Across all groups, several recurring themes and also some concerns and opportunities around artificial intelligence surfaced:

Component-based and modular science Multiple groups emphasized breaking research into smaller, composable units—figures, datasets, methods—that could be recombined for different audiences (policy makers, educators, the public). These “component pieces” would form a new modular infrastructure for knowledge and could come with their own attribution and licensing.

Computational content — built in, not bolted on Participants highlighted that computation should be natively included in the research narrative, not an afterthought. The goal is to treat code, data, and visualizations as first-class objects within the scientific record—interwoven with text and metadata, not hidden in supplements or external notebooks. Projects like Stencila, Curvenote, MyST, and Quarto were cited as leading examples of this integrated approach, where execution, interactivity, and reproducibility are embedded directly into the document itself.

Accessibility and education Many wanted to ensure research outputs can reach younger learners and non-experts, through interactive and modular content rather than static PDFs. Education was recognized as the historical vector of adoption for new scientific tools (e.g., Jupyter, R Markdown, Quarto, MyST).

Interoperability and standardization Participants stressed the need for shared schemas or “record locators” for research objects to make data, code, and media discoverable and reusable across distributed systems. This would remove the friction of incompatible formats between repositories and journals and make it possible to more easily link components together.

Human connection amid automation Despite technical focus, there was an emotional thread—preserving human relationships and trust networks in a landscape increasingly mediated by algorithms and AI filters.

Trust, reputation, and accountability Participants worried about how to maintain human accountability in an era of AI-generated content and papermills. Identity and persistent identifiers for people (not just papers) were seen as one important component to rebuild trust and reputation in the scientific record.

Attribution and licensing When modular components of science - data, code, images - can be more easily distributed or remixed, participants recognized that these components can then come with their own attribution and licensing, and we can build that into our technologies and approaches.

Quality versus quantity in publishing The flood of low-quality or fraudulent content is overwhelming peer review and submission. Some groups discussed using AI-assisted review as "fighting fire with fire," with both potential and pitfalls.

In the group discussion, we came up with a metaphor to describe a "decoupled journal" [3] as an organizing principle to break into groups to get towards more concrete ideas to progress on these ideas in the next few days.

> What are the pathways to getting there that we can make progress on together **tomorrow**?

2.3. The Bedrock–Soil–Flowers Metaphor 🌸🌷🌻

The group aligned on a three-layer ecological metaphor to describe the evolving structure of scientific knowledge:

2.3.1. *Bedrock – The Foundational Content*

Represents the raw scientific materials: datasets, code, preprints, lab notes, and other primary research artifacts. It is the substrate upon which all higher layers depend—analo-

Figure 4. *Working together in small groups on various ideas. Left to right: Franklin Koch, John Kaye, and Rose Reatherford.*

gous to the geological foundation of an ecosystem. Participants described it as "boring but essential": the invisible, structured layer that ensures reliability, provenance, and persistence. Some proposed a "dynamic bedrock"—still foundational but continuously reshaped as new data, formats, and schemas emerge — this is especially true from an adoption perspective of existing tools in the room that want to use this format, but not be locked into something that isn't flexible.

🪨 **Describing the Bedrock**

We took notes on an architecture for the open exchange of scientific ideas that is accessible by developers and includes computational content and data as well as bakes in licensing and attribution in modular ways.

/bedrock

2.3.2. *Soil – The Connecting Layer*

The soil digests, organizes, and connects the bedrock materials—metadata, graphs of relationships, citation networks, peer review, and trust signals. It represents context, curation, and integration: where reputation systems, provenance graphs, and aggregation tools (like Scopus, CORE, or Google Scholar) operate. The soil also filters and moderates quality, providing "nutrients" that enrich the bedrock and allow meaningful reuse of content. Trust was seen as growing from the soil[1]:

> If we can trust the bedrock, we can trust the flowers—but the soil makes it possible for the flowers to grow

=== Flowers (the ecosystem) – The Visible, Dynamic Outputs and Applications

Flowers (and trees and shrubbery) represent applications, interfaces, and experiences built atop the lower layers. These are the journals, educational tools, visualizations, interactive notebooks, and AI-driven summaries. This is the layer that the public, educators, and policymakers engage with—the visible ecosystem that blooms from solid foundations and fertile connections

It encompasses both human-curated and AI-generated outputs—the "living ecosystem" of scientific communication. These applications often have the easiest time attracting new funding or excitement, but ultimately rely on the underlying layers to deliver value.

2.3.3. *Alignment and Insights*

Trust propagates upward A trustworthy and transparent "bedrock" is made possible through identifying the connections between research objects, and ultimately new applications/experiences that can provide different curation, badging and labelling services that also contribute meaningful and reusable data.

Quality can tolerate imperfection The group agreed the bedrock can also contain low quality outputs; open, messy data/resources *are acceptable* as long as filtering and curation (the soil/ecosystem) evolve accordingly and/or communities have the ability to self organize. This mirrors many of the existing cultures decoupling dissemination from review.

Ecosystem thinking Each layer enables the next; no single actor can build all three. The system's resilience comes from modularity and interoperability.

Strategic focus Participants agreed to identify "boring but high-leverage" components—standards, schemas, and identifiers—that quietly enable everything else.

[1] Others also commented that trust is likely in all layers of our metaphor, not just the "soil"!

> The bedrock is the structured scientific content; the soil is the web of connections, curation, and trust; and the flowers are the dynamic, interactive ways science reaches people. We need all three layers for the ecosystem of knowledge to thrive.

== Break into groups to think through the How

2.3.4. *Bedrock*

In this group the overall conversation centered around Scientific content produced in granular form so others can reuse, consume, and remix. The group decided to collaborate on a schema/architecture for the open exchange of scientific content, ensuring it is web-native, cloud ready, and enables new AI workflows. The scope of the effort for the workshop was to create a small library of schemas, examples, and tests. We identified existing tools and ideas in Stencila, MyST, Quarto, Curvenote, Pandoc, JATS, and schema.org that would be the basis for this new format. Starting with authoring tools enables the researcher to be directly involved in the production of something that is standards compliant. We also had advisors from Creative Commons, who helped identify places where licensing and attribution can be embedded directly in the workflows.

The general approach we sketched out for Day 2 was:

- Identify the units of scientific communication (figures, charts, paragraphs)
- Ensure that these include license and attribution at all levels
- Ensure any schema can be mechanically reused in many contexts, starting with adoption in authoring tools and ensuring compatibility with publishing, production, and archiving workflows.

2.3.5. *Soil*

This group came up with concepts for new ways to connect the components of science, these were based on existing schemas (e.g. ActivityPub, JSON-LD, COAR Notify, etc.). High level concepts such as "Federated DataCite" and "Science DNS" that can be resilient to failure as well as include concepts of spam-protection from email services. A concept of **collections and containers** emerged, with new types of research objects that could be composition of other objects, including recognizing the existing role and potential of MECA. There was also the concept of "an index of indexes", a role that OpenAlex is currently fulfilling in the ecosystem, that connects and exposes research objects in many diverse repositories.

The group proposed architecture that could bring together some of the content from the "bedrock" in new ways to enable composable science in containers that are "in-family" with today's preprints and articles. For example, pulling in tables, or figures from other existing research that was published. There were overlaps identified with the group working on an open exchange architecture and participants ended up joining that group effort.

2.3.6. *Ecosystem*

The group discussed the potential for many new types of applications and services that could be built from the new components in the bedrock, especially if they are made more accessible through APIs and services (e.g. OpenAlex, DocMaps) that connect and add meaning through graph-networks (citations, person disambiguation, reviews, etc.). The group talked about new metaphors like "spotify for science" that could be rapidly prototyped to show new curation and trust metrics on the research.

The group proposed to do a sprint using collaborative prototyping and new LLM-based programming tools to quickly pull together concepts and iterate as a group. As not everyone was familiar with these tools, they decided to mostly work together on a single new prototype

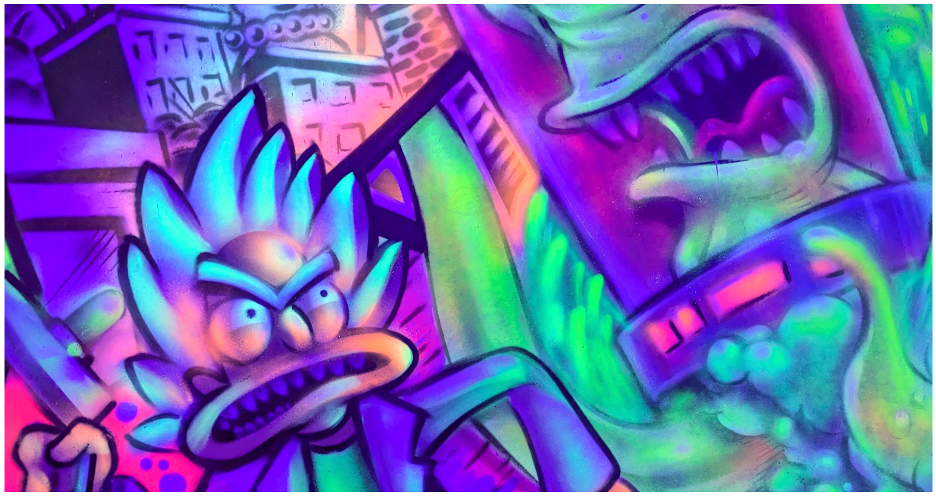

Figure 5. *The sensory experience of the arcade was intense, lots of fun was had at ski-ball, mario, pac-man, and basketball.* 👾🕹️

that brought together aspects of search, discovery, curation, trust signals, and embedded reading experiences from the open exchange architecture. A separate effort was proposed to look at the combination of multiple computational research articles from the NeuroLibre project, to explore what is possible when full context to the science is made available (the environment, data, code and narrative).

Another group discussed the importance of working with scientific communities in understanding needs and interests and creating spaces for agile and iterative deployment and feedback with the groups and people who are sharing science. They discussed the importance of communities coming to tool-builders with their needs and ideas, rather than tools being pushed to communities, and the attributes of forming these community relationships. Personas can be valuable in thinking through user needs, and there was interest in collective personas in scientific communication.

2.4. Evening Activity & Day 2

We went to an arcade for dinner and some activities and to continue discussing ideas for the second day. Many other connections were made between the various participants.

The second day was unstructured time in the San Diego Made Factory, where we hacked, learned, built schemas, talked about approaches, and vibe-coded prototypes for the demo-day on Saturday morning.

🗺️ Day 2 - Building with a Plan

The second day was unstructured time in the San Diego Made Factory, where we hacked, learned, built schemas, talked about approaches, and vibe-coded prototypes for the demo-day on Saturday morning.

/day-2

3. Describing the "Bedrock"

3.1. Context and Framing

The discussion centers on **rethinking scientific publishing infrastructure**—moving beyond PDFs and disconnected supplementary materials toward **modular, interoperable, and machine-readable research objects**. Participants refer to this initiative as *Bedrock*, a metaphor for the underlying content layer of science that supports higher-level representations ("soil" as metadata, "flowers" as the presentation layer).

The group's goal is to define or prototype a **schema or standard format** that enables linking, reuse, and interoperability of granular scientific content—figures, data, methods, and even paragraphs—across tools and platforms.

3.2. Problems in the Current System

3.2.1. *Disconnected Elements and Reproducibility Gaps*

- Researchers cannot easily **rebuild or replicate** another scientist's work because key contextual elements (e.g., reagents, SKUs, metadata, data availability) are not structured or linkable.
- Supplementary files are *idiosyncratic, unstructured, and inaccessible*.
- Peer review focuses on narrative, not replication ("peer replication" was emphasized as the real goal).

3.2.2. *Authoring Pain Points*

- Scientists are not incentivized to produce structured metadata; it isn't useful to them during creation.
- Without **author-friendly tools**, metadata curation will not occur.
- Late-stage publication (a year after experimentation) leads to data loss and forgotten details.

3.2.3. *Cultural and Structural Barriers*

- Researchers fear being *scooped* if they share granular results.
- Incentives reward polished narratives over incremental contributions.
- Big-science examples (e.g., LIGO) show structured collaboration and replication mechanisms that small-scale labs lack.

3.3. Desirable Properties of the New Approach

3.3.1. *Modular and Granular Content*

- The **minimum unit of contribution** could be a single experiment or even an observation ("the minimal currency of science").
- Each unit should be **typed and linkable**—a figure panel, paragraph, dataset, or method.
- Enables *peer replication* and building cumulative knowledge graphs.

3.3.2. *Author-Centric Design*

- Metadata entry should be embedded early in the research workflow (lab notebooks, weekly logs).
- Tools should give **immediate value** to authors (e.g., better organization, feedback, or linting-style suggestions).

- Shorter publication cycles (iterative curation) reduce cognitive load and error propagation.

3.3.3. *Linking and Web-Native Infrastructure*

- The web's power lies in **linking**, not redisplaying.
- Embrace *messiness and heterogeneity*—interoperability over standardization.
- Support *bi-directional or resilient links* between research objects while accepting web fragility.

3.3.4. *Layered Metaphor*

- **Bedrock:** raw content and data.
- **Soil:** metadata and schemas enabling discovery.
- **Flowers:** presentation and visualization layers that remix and display content differently.

3.4. Technical Architecture and Standards

3.4.1. *JSON, JSON-LD, and JSON Schema*

- Core consensus: use **JSON as the base data format**.
- JSON Schema defines structure and typing (paragraphs, figures, links).
- JSON-LD adds **linked-data semantics**—shared vocabulary alignment across schemas (via `@context`).
- Schema.org terms and JSON-LD mappings enable translation between domain vocabularies.

3.4.2. *Relationship to Existing Standards*

- **Pandoc AST** provides a precedent but is "hideous" and not ideal for machine processing.
- **JATS** is acknowledged as important but overly rigid and non-interoperable; the group seeks a *successor schema* more like JSON/HDF/Zarr—modular, extensible, and linkable.
- Compatibility with **MyST, Stencila, Quarto, Curvenote**, and other document models is critical; they are viewed as "interfaces" over a shared data model.

3.4.3. *Linking Mechanisms*

- Every element (paragraph, section, figure) should have a **persistent ID or resolvable path**.
- External references (e.g., datasets, code) would use typed links validated against schemas.
- Agreed-upon "carve-outs" for cross-references in ASTs.

3.4.4. *Incremental Adoption*

- Consensus: don't wait for universal agreement.
- Start with **two or three collaborating projects**, publish version 0.1 of a shared schema, and test interoperability.
- Translate existing documents (e.g., JATS arrow.r JSON) to bootstrap adoption.

3.5. Licensing and Attribution Layer

3.5.1. *Machine-Readable Licensing*

- Licenses (e.g., CC-BY, CC-BY-NC, ND) can be **expressed and validated mechanically** within schemas.
- Linking is permissible even between incompatible licenses (not a derivative work), but **reuse and remixing** require compatibility checks.
- Future tooling could include a **license-compatibility matrix** or automated validator.

3.5.2. *Attribution Requirements*

- Schema can embed rules and tooling to enable "if you link to this, include this attribution and license".
- Encourages early sharing by **providing attribution safety**—time-stamping and provenance metadata (a digital "flag-planting").

3.5.3. *Creative Commons Perspective*

- CC representatives emphasized **user awareness** of license implications and limits of automated enforcement.
- Mechanized verification is feasible and desirable but must complement, not replace, human interpretation.

3.6. Broader Philosophical Insights

- **Continuous Science:** Move from static, monolithic papers to continuous, linkable research streams.
- **Trust Circles:** Allow sharing and collaboration at intermediate stages without public release.
- **Open by Design:** Machine readability and openness should be defaults; opacity should require justification.
- **Messy Web as Feature:** Embrace diversity of tools and formats; prioritize translation and interoperability over control.
- **Schema as Conversation:** Rather than one fixed format, define shared *translation points* between evolving community schemas.

3.7. Next Steps and Outcomes

- Draft a **proof-of-concept schema** (version 0.0.1) defining minimal elements:
  - Paragraphs, sections, figures, datasets, and cross-references.
  - JSON Schema typing plus optional JSON-LD context.
- Prototype translation between MyST AST, Pandoc AST, and Stencila Encoda.
- Define a **mechanical validation layer** for licensing compatibility.
- Publish demonstration documents across platforms (Curvenote, Quarto Pub, eLife, etc.) to test interoperability.
- Continue collaboration toward a "Bedrock" format that underlies composable, modular, and interoperable scientific publishing.

3.8. Key Takeaways

1. **The problem is structural:** PDFs and JATS lock science in rigid, narrative-centric silos.
2. **The opportunity is architectural:** A web-native, JSON-based schema can unlock modular reuse.

3. **The transition must be incremental:** Begin with partial interoperability among active communities (MyST, Stencila, Curvenote, JATS, etc.).
4. **Success is cultural as much as technical:** Scientists must see value in the format immediately.
5. **Licensing and attribution are enablers, not afterthoughts:** Machine-readable rights are core to open, reproducible science.

The Bedrock conversation charts a pragmatic path from today's disconnected scholarly formats toward a **composable, interoperable, and link-rich ecosystem for scientific knowledge**—grounded in JSON schemas, aligned through linked data, validated through licensing logic, and grown organically from community-driven collaboration.

4. Day 2

The goals for the second day were to break into ad-hoc groups and work towards one of the six initiatives that we identified in the first day and over dinner.

Open Exchange Architecture (OXA) Define shared schemas and tools for authoring, publishing and archiving scientific content, based from experience in the scientific authoring ecosystem (Quarto, MyST, Curvenote, Stencila) — a new Open Exchange Architecture. We aimed to present on (a) a new schema with examples; (b) how this can export from quarto (as myst/stencila were already relatively compatible); (c) define high-fidelity links to components; (d) discuss the “container” for the research content; and (e) discuss how this could work with other publishing systems and tools (JATS, MECA, Octopus, etc.). Crucially, we wanted to leave with something **concrete** that had a name (which we came up with together on day 2), logo, domain, repository and could be installed from node and pypi.

Creative Commons — Modular Reuse, Licensing and Attribution Work with the open exchange group to ensure the component framework has licensing and attribution built directly into the architecture. This should work for existing narrative containers (articles, papers, preprints, protocols), modular components (panels, figures, tables, datasets, etc.), as well as remixed compositions of research that are a hybrid of the two (an article that references a different figure or panel).

Utilizing and Extending MECA Discuss ways to use existing flexible transfer mechanisms, such as MECA, to bring this format to interoperate with the existing submission and publishing systems. The MECA file format is very flexible and the working group is looking to modernize the NISO recommendation in 2026 to incorporate new cloud-based storage and APIs.

Computational Research Content + AI Produce an experiment of pulling together two computational articles (written in MyST) and explore what can be done with the additional context this provides to research agents.

Spotify for Science Lead an effort to bring many APIs and services together into a new way to browse research that can elevate different trust signals and show research content in new ways.

Model Communities Identify early communities to pilot these approaches with, including from additional computational communities that require more context and different “weird-and-wonderful” research objects to be published (e.g. notebooks, microscopy images, etc.) as well as paths to adoption.

Figure 6. *Working together throughout the day on different ways to communicate scientific content. From left to right: Anton Molina, Nokome Bentley*

Figure 7. *Merging creative commons approach and ideas for how to build attribution and licensing directly into the open exchange architecture. From left to right: Carlos Scheidegger and Taylor Campbell.*

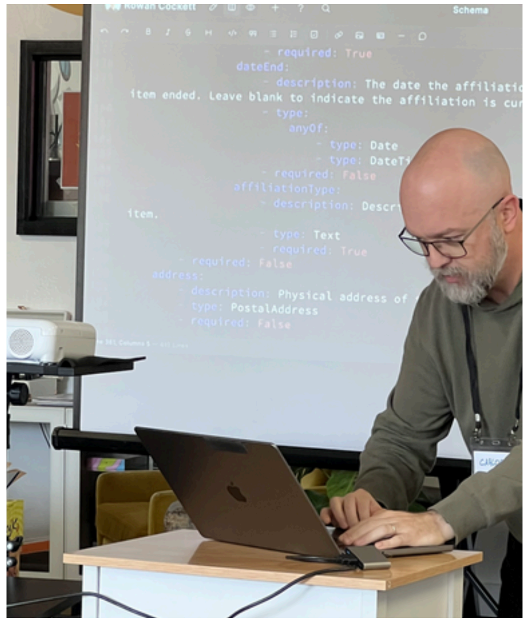

Figure 8. *Carlos presenting on the schemas for the Open Exchange Architecture (OXA).*

5. Day 3

Snippets of what we presented are below. We then talked about next steps, especially in the context of new formats/standards (OXA) and the application development shown by the "Spotify for Science" application.

5.1. Open Exchange Architecture (OXA)

Define shared schemas and tools for authoring, publishing and archiving scientific content, based from experience in the scientific authoring ecosystem (Quarto, MyST, Curvenote, Stencila) — a new Open Exchange Architecture. We aimed to present on (a) a new schema with examples; (b) how this can export from quarto (as myst/stencila were already relatively compatible); (c) define high-fidelity links to components; (d) discuss the "container" for the research content; and (e) discuss how this could work with other publishing systems and tools (JATS, MECA, Octopus, etc.). Crucially, we wanted to leave with something **concrete** that had a name (which we came up with together on day 2), logo, domain, repository and could be installed from node and pypi.

We created a GitHub Repository and website to push the effort forward.

> **Choosing the Name OXA**
>
> A small prototype was pulled together to look at the available three-letter extensions (narrowed from 17,576 to 1756, see available extensions) as well as coming up with back-ronyms using AI. We ultimately didn't take the voting system forward, but the research into extension names helped to seed the conversation.

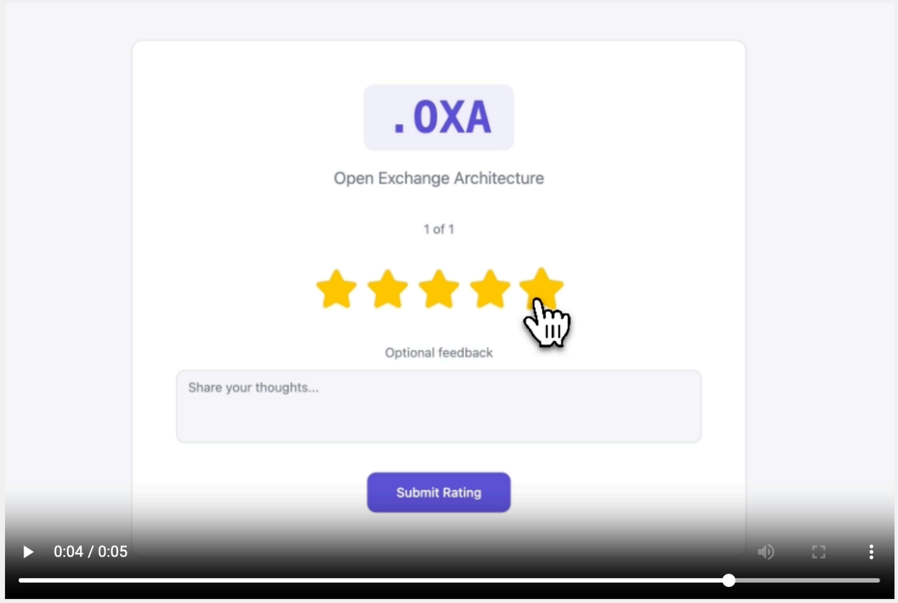

Figure 9. *A prototype for a voting system, that ultimately wasn't used.*

5.2. Creative Commons — Modular Reuse, Licensing and Attribution

Work with the open exchange group to ensure the component framework has licensing and attribution built directly into the architecture. This should work for existing narrative containers (articles, papers, preprints, protocols), modular components (panels, figures, tables, datasets, etc.), as well as remixed compositions of research that are a hybrid of the two (an article that references a different figure or panel).

Researcher produces a "Paper Repo" (made of a collection of objects). Shareable as an object. The paper repo is made of claims (hypothesis headlines), each presented as containers of a panels of evidence. Each panel of evidence (PE) supports its claim with text, figures, and tables. Panels of evidence can exist as an embed or be output as their own PDFs. Each panel of evidence has a corresponding container of its panel dependencies (PD).

Each PD can contain supporting assets:

- Code: Code snippets, source code, linkouts to a repo of code (eg., in Github).
- Research Datasets: Part of the files structure, or in a linked data bucket.
- Design Files: Practicable/manufacturable and/or related to patenting, so functionally distinct. Can be for different types of things that are designed, including formulations.
- Lab Notes: The scratchpad that accompanies the panel.

Researchers can then publish the entire paper repo as a PDF Paper (but not the dependencies, which can just be linked out).

5.3. Utilizing and Extending MECA

Discuss ways to use existing flexible transfer mechanisms, such as MECA, to bring this format to interoperate with the existing submission and publishing systems. The MECA file format is very flexible and the working group is looking to modernize the NISO recommendation in 2026 to incorporate new cloud-based storage and APIs.

Object	License-able?	License Options
PD: Panel Dependencies	YES	License depends on if licensing as a collection or under a single compatible license (machine-determined).
PD: Code	YES	Licenseable with software licenses.
PD: Research Datasets	YES	Normal CC recommendations based on license compatibility.
PD: Design Files	YES	Normal CC recommendations based on license compatibility.
PD: Lab Notes	YES	Normal CC recommendations based on license compatibility.
PDF Paper	YES	License depends on if licensing as a collection or under a single compatible license (machine-determined).

Metadata values for attribution: **Title, Author, Source Link, License Link, Attribution Statement, Rights Statement (TASLAR)**

Figure 10. *The TASLAR framework which stands for Title, Author, Source Link, License Link, Attribution Statement, Rights Statement (Creative Commons).*

5.4. Computational Research Content + AI

Produce an experiment of pulling together two computational articles (written in MyST) and explore what can be done with the additional context this provides to research agents.

5.5. Spotify for Science

Lead an effort to bring many APIs and services together into a new way to browse research that can elevate different trust signals and show research content in new ways.

5.6. Model Communities

Identify early communities to pilot these approaches with, including from additional computational communities that require more context and different "weird-and-wonderful" research objects to be published (e.g. notebooks, microscopy images, etc.) as well as paths to adoption.

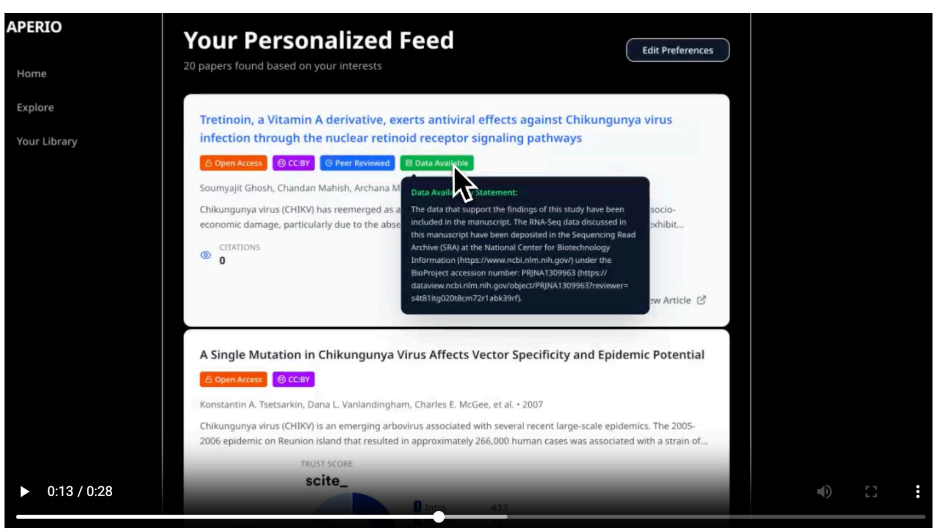

Figure 11. *A walkthrough of the application that was developed in a day that pulled together APIs from a number of different projects in the ecosystem. It is live here: https://v0-aperio-ui-design.vercel.app/*

References

[1] The Navigation Fund, "From Tools to Adoption: A Path to Modular & Interactive Scientific Publishing." [Online]. Available: https://commons.datacite.org/doi.org/10.71707/gn91-ka32

[2] R. Cockett, "From ZIP to Zarr: An Example of Standards Evolution," 2025, [Online]. Available: https://articles.continuousfoundation.org/articles/from-zip-to-zarr-an-example-of-standards-evolution

[3] J. Priem and B. M. Hemminger, "Decoupling the scholarly journal," *Frontiers in Computational Neuroscience*, vol. 6, 2012, doi: 10.3389/fncom.2012.00019.