standards

MDPI

Perspective

Post-Document Science: From Static Narratives to Intelligent Objects

Mehmet Fırat

Department of Computer Education and Instructional Technologies, Education Faculty, Yunusemre Campus, Anadolu University, 26470 Eskişehir, Türkiye; mfirat@anadolu.edu.tr

Abstract

Scientific publishing is currently constrained by an unstructured narrative bottleneck paradigm, which increasingly diverges from the scale, complexity, and computational nature of modern research. Despite rapid advancements in data generation and analysis, scientific knowledge is predominantly disseminated as static narrative artifacts, thereby limiting reproducibility, machine accessibility, and cumulative integration. This study explores how scientific communication can be restructured to facilitate scalable validation and reliable knowledge accumulation. We propose the Object-Oriented Scientific Information paradigm, wherein scientific contributions are represented as executable, machine-interpretable objects that integrate structured data, reproducible methodologies, and formally encoded semantic claims. To operationalize this paradigm, we delineate the architecture of an Autonomous Knowledge Engine, a modular neuro-symbolic system that combines domain-specialized Mixture-of-Experts routing, formal verification of claims, and an information-theoretic filter based on marginal information gain. This architecture enables continuous validation, redundancy control, and the integration of scientific contributions within an active knowledge graph. The analysis demonstrates that Object-Oriented Scientific Information (OOSI) and Autonomous Knowledge Engine (AKE) fundamentally differ from existing document-based, executable, and semantic publishing models by shifting epistemic control from narrative evaluation to computational verification. We conclude that transitioning toward a computable scientific record is essential for sustaining reliable and self-correcting science in the context of accelerating knowledge production.

Keywords: scientific publishing reform; machine-interpretable science; knowledge graphs; reproducibility

Check for updates

Academic Editor: Peter Glavič

Received: 28 December 2025
Revised: 27 March 2026
Accepted: 30 March 2026
Published: 3 April 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Epistemic Bottleneck of Scientific Knowledge Contraction

Over the first quarter of the twenty-first century, the infrastructure supporting scientific knowledge has shown increasing strain. Although data production has advanced rapidly toward high-throughput petabyte-scale systems, the prevailing modes of scholarly communication have changed far more slowly. Scientific results continue to be disseminated primarily as static documents that are optimized for human reading and printing rather than computational analysis. As highlighted by Mons et al., this practice results in data being "buried in supplementary graveyards," limiting machine accessibility and challenging the implementation of the FAIR principles of findability, accessibility, interoperability, and reusability [1,2].

This misalignment between data generation and knowledge dissemination has implications that extend beyond technical inefficiencies. This increasingly affects the reliability

and interpretability of scientific records. Peer review, the principal mechanism for quality control, faces mounting pressure from escalating submission volumes, methodological complexities, and limited reviewer capacities. As Smith noted, traditional peer review may function less as a consistent quality filter than as a probabilistic selection process influenced by bias and chance [3]. In parallel, the reproducibility crisis has drawn attention to the systemic weaknesses in the verification of data, code, and analytical pipelines prior to publication [4].

These challenges are compounded by the exponential growth of scientific literature, which has pushed the signal-to-noise ratio beyond the practical limits of individual human cognition. Ioannidis showed that publication bias, selective reporting, and insufficient statistical power can create conditions in which a substantial fraction of published findings are false or overstated [5]. Under such conditions, errors may persist and propagate not because of individual misconduct but as a structural consequence of the current publication ecosystem. These dynamics are reinforced by incentive structures that systematically favor quantity over epistemic robustness [6].

This Perspective argues that incremental reforms, including journal digitization or expanded access to supplementary materials, are unlikely to resolve these issues at their roots. Instead, addressing the emerging epistemic bottleneck requires reconsideration of the fundamental unit of scientific communication. Specifically, it calls for a transition from static narrative artifacts to machine-interpretable, verifiable, and dynamically integrable representations of scientific knowledge, capable of supporting both human understanding and computational reasoning at scale.

The urgency of this transition is amplified by the rapid integration of Large Language Models (LLMs) into scientific workflows. While LLMs excel at synthesizing text, their reliance on unstructured training data makes them prone to hallucinations and logical inconsistencies. Without a transition to grounded, computable knowledge representations, generative AI risks accelerating the production of plausible-sounding but scientifically flawed narratives, effectively polluting the cumulative scientific record.

2. From Narrative to Object: The Object-Oriented Scientific Information Paradigm

To address the limitations of document-centric scientific communication, we propose a conceptual shift from narrative publications to OOSI. In this paradigm, a scientific contribution is no longer treated as a static textual artifact but as a discrete, computationally defined object that can be executed, validated, queried, and integrated into a broader knowledge system. Rather than replacing the narrative entirely, OOSI repositions the narrative as an interface layer, secondary to a formalized, machine-interpretable core [7]. Rather than treating the scientific paper as the fundamental epistemic unit, OOSI reconceptualizes scientific contributions as executable and interoperable objects embedded within a computable knowledge structure, as shown in Figure 1.

The OOSI paradigm draws inspiration from object-oriented principles in software engineering, where complex systems are constructed from modular and well-defined components governed by explicit interfaces and constraints. Within this framework, scientific domains function analogously to classes, encoding shared assumptions, axioms and methodological constraints. For example, a domain such as superconductivity may define admissible experimental conditions, accepted theoretical formalisms, and the required validation criteria. Individual experiments or analyses are then instantiated as objects belonging to these classes, inheriting their structural and epistemic constraints from them.

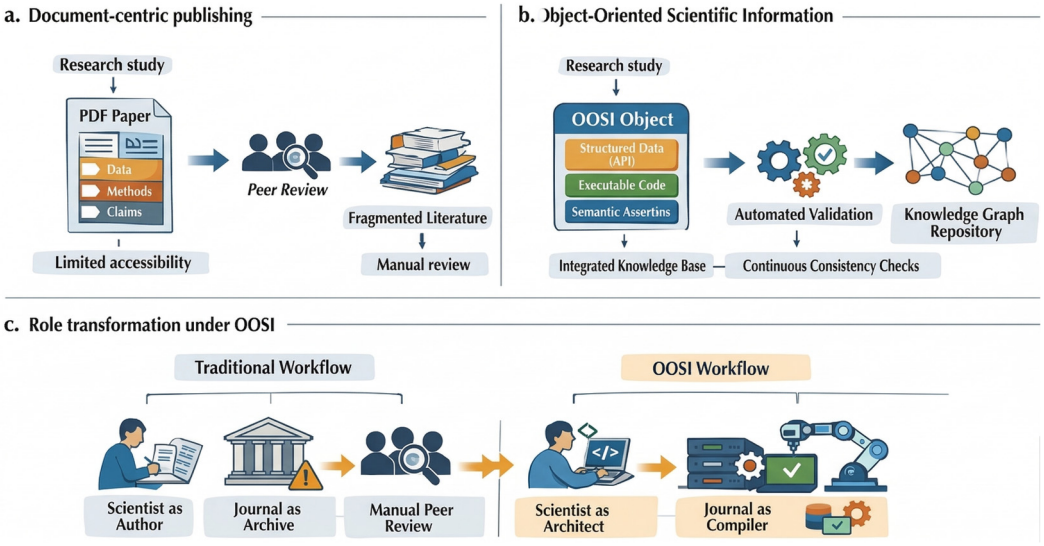


Figure 1. From narrative documents to object-oriented scientific knowledge.

Under the OOSI, a scientific submission is required to explicitly declare and conform to the class from which it is derived. This inheritance enforces domain-specific rigor by design, reduces methodological ambiguity, and prevents implicit assumption drift across publications. Each object encapsulates three components.

First, structured data must be provided as machine-accessible entities exposed through standardized schemas and application programming interfaces rather than embedded in static tables or supplementary files. This enables direct reuse, aggregation, and automated interrogation without the need for manual extraction.

Second, executable methodologies must be included in the form of runnable code, formal workflows, or explicitly defined computational pipelines. This component specifies the precise operations required to transform the input data into the reported results, supporting reproducibility as an operational property rather than a post hoc aspiration [8].

Third, semantic assertions must encode core scientific claims in a formal, machine-verifiable representation. By expressing claims in logical or type-theoretic frameworks, such as Lean/Coq [9], OOSI enables automated checks for internal consistency, dependency tracking across claims, and principled detection of contradictions as the knowledge base evolves. Related efforts, such as nanopublications, semantically encode claims but do not integrate executable methods or inheritance constraints [10].

Together, these components redefine the functional roles within the scientific ecosystem. The scientist's primary task shifts from crafting a persuasive narrative to designing verifiable, interoperable knowledge objects. Correspondingly, the role of the journal evolves from that of a passive archive to an active compilation and validation environment responsible for integrating new objects into a continuously evolving scientific codebase. Viewed this way, OOSI represents not merely a technical reform but an epistemic reconfiguration of how scientific knowledge is represented, validated, and accumulated at scale.

Crucially, the adoption of OOSI does not require researchers to abandon natural language processing or acquire advanced software engineering skills. Leveraging recent advances in auto-formalization [11], the proposed system functions as a semantic translation layer. It automatically converts natural language assertions into formal logic and executable code, which the author validates. This 'human-in-the-loop' verification shifts the scientist's role from a writer of passive documents to an architect of verifiable knowledge objects.

3. Architecture of an Autonomous Knowledge Engine

Operationalizing the Object-Oriented Scientific Information paradigm does not require the development of a monolithic artificial general intelligence. Instead, it calls for a modular, hybrid neuro-symbolic architecture designed explicitly for scientific verification, integration, and curation, as shown in Figure 2.

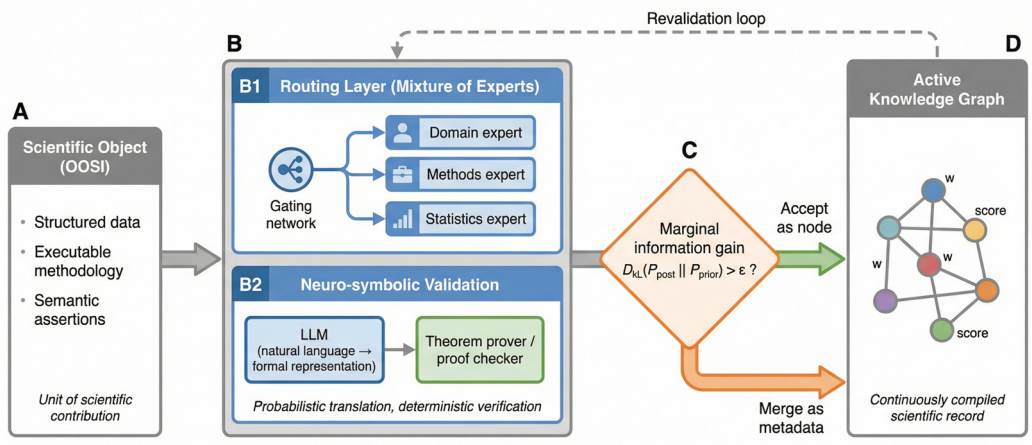


Figure 2. Epistemic architecture of the Autonomous Knowledge Engine (AKE). (**A**) The Scientific Object serves as the unit of scientific contribution, encapsulating structured data, executable methodology, and semantic assertions. (**B**) The processing core consists of (**B1**) the Routing Layer, which utilizes a Mixture of Experts (MoE) to distribute evaluation across domain, methods, and statistics specialists, and (**B2**) the Neuro-symbolic Validation layer, where LLMs perform probabilistic translation of natural language into formal representations for deterministic verification by theorem provers. (**C**) An Information-theoretic Filter computes the marginal information gain using Kullback–Leibler divergence (DKL) to determine if the contribution warrants a new node or a metadata merge. (**D**) The Active Knowledge Graph represents the continuously compiled scientific record, featuring a revalidation loop for ongoing epistemic updates.

In Figure 2, scientific contributions are ingested as object-oriented entities and routed through specialized expert modules, neuro-symbolic validation, and an information-theoretic filter before being integrated into a continuously evolving knowledge graph. Validation and filtering operate continuously, enabling automated consistency checking, redundancy control, and epistemic updates at scale.

3.1. The Routing Layer: Specialization Through Mixture of Experts

As illustrated in Figure 2, incoming scientific objects are first routed through a sparsely gated mixture-of-experts layer to enable parallel domain-specific evaluation. Scientific knowledge is intrinsically specialized. A single general-purpose large language model lacks the depth required for domain-level validation across heterogeneous fields. To address this, the architecture employs a sparsely gated Mixture of Experts (MoE) framework [12], in which a lightweight gating network dynamically routes incoming scientific objects to domain-optimized expert modules.

Each expert corresponds to a narrowly defined disciplinary or methodological competence, implemented as fine-tuned low-rank adaptation (LoRA) modules layered onto the shared foundation model. For example, a submission concerning protein folding is routed to a biochemistry expert, whereas its statistical analysis is concurrently evaluated by a biostatistics expert. This parallel, multiperspective evaluation reflects how expertise is distributed in human science while remaining computationally scalable [13].

The AKE architecture employs a neuro-symbolic feedback loop to ensure epistemic integrity. While neural components (LLMs) provide the flexibility to interpret complex sci-

entific intent, a symbolic reasoning engine enforces a logical consistency and mathematical validity. By decoupling hypothesis generation (neural) from formal verification (symbolic), the system mitigates the stochastic nature of current AI models, ensuring that every OOSI object is mathematically sound before it enters the global knowledge graph.

Experts are defined along orthogonal axes (domain, methods, statistics) and trained via task-specific corpora using LoRA adapters. Load-balancing and entropy regularization prevent routing collapse. Disagreements between experts trigger additional proof obligations rather than majority voting, preserving epistemic conservatism.

3.2. The Validation Layer: Neuro-Symbolic Verification

Figure 2 highlights the separation between probabilistic translation by large language models and deterministic verification performed using formal proof systems. Generative models, despite their expressive power, are probabilistic systems and are therefore prone to hallucinations. Therefore, they cannot serve as the final arbiters of scientific truth. The proposed architecture adopts a neuro-symbolic validation layer in which generative models act as translators rather than judges [14].

Natural language claims are transformed into formal logical or type-theoretic representations, which are then subjected to deterministic verification using theorem provers or formal-proof assistants. As argued by Szegedy, formalization shifts the burden of validation from narrative interpretation to machine-checkable logic, enabling what has been described as "auto-formalization," where the computer verifies the proof rather than prose [11]. This separation preserves the expressive flexibility of natural language while grounding scientific claims in a verifiable semantics.

3.3. The Filter: Marginal Information Gain

In the current formulation, the belief state $P_{(x)}$ is defined over graph-level epistemic units, where xxx denotes validated claim nodes or parameterized claim classes within the active knowledge graph. Priors are induced from the existing graph state via normalized dependency weights and historical validation outcomes, while posteriors are computed by hypothetical integration of the candidate object and re-evaluation of affected subgraphs.

The epistemic filter shown in Figure 2 ensures that only objects contributing sufficient marginal information gain are considered independent nodes. To address the accelerating redundancy of scientific publications, this system incorporates an information-theoretic gatekeeper. The marginal information gain is quantified using the Kullback–Leibler divergence, a standard measure from information theory that characterizes the discrepancy between the prior and posterior probability distributions [15]. This gain is quantified using the Kullback–Leibler divergence between the prior belief state (P_{prior}) and the posterior state (P_{post}) after hypothetical integration, as defined in Equation (1):

$$D_{\text{KL}}\left(P_{\text{post}} \,||\, P_{\text{prior}}\right) = \sum_{x} P_{\text{post}}(x) \ln\left(\frac{P_{\text{post}}(x)}{P_{\text{prior}}(x)}\right) \quad (1)$$

The posterior-to-prior direction is intentionally adopted, as the objective is to quantify how much the integrated object updates the epistemic state of the knowledge graph, rather than how surprising the posterior appears from the prior perspective. If the resulting information gain falls below a threshold ε, the object is not considered an independent contribution. Instead, it is merged as a metadata update or as supporting evidence for an existing node. This mechanism ensures that the knowledge graph grows in epistemic depth rather than in uncontrolled breadth, counteracting the inflationary dynamics of the current publication system.

While KL-divergence provides a principled baseline, the architecture does not assume a single universal metric. Alternative or complementary measures such as minimum description length, prequential codelength, marginal likelihood ratios, or contribution Shapley values can be substituted depending on domain-specific epistemic norms. The filter is therefore metric-pluggable rather than fixed.

3.4. Representational Assumptions and Interoperability

To situate the OOSI paradigm within the broader landscape of machine-readable science, it is important to distinguish OOSI from existing semantic representation frameworks. Throughout this Perspective, we use the term "narrative" to denote unstructured or semi-structured natural language sequences optimized for human cognition but inherently opaque to deterministic machine reasoning. While narratives permit ambiguity and implicit assumptions, OOSI requires that the core scientific claims, methodologies, and data dependencies be explicitly encoded in machine-actionable formats.

Crucially, OOSI is not proposed as a replacement for existing Semantic Web standards or established domain ontologies. Rather, it functions as a higher-level epistemic orchestration layer that abstracts away from, but remains fully compatible with, foundational representational protocols. The semantic assertions within an OOSI object can be natively serialized using standard frameworks such as the Resource Description Framework (RDF), Web Ontology Language (OWL), or JSON-LD. This ensures that the Active Knowledge Graph generated by the AKE remains fully interoperable with legacy knowledge bases, federated SPARQL endpoints, and existing semantic infrastructure.

OOSI conceptually diverges from standard descriptive ontologies (such as CIDOC CRM for cultural heritage or the Gene Ontology for biology) in its operational intent. Traditional semantic approaches and knowledge graphs are primarily descriptive; they map static relationships between entities (e.g., claiming "Protein A inhibits Enzyme B"). In contrast, OOSI objects are executable and verifiable. An OOSI unit encapsulates not only the descriptive claim, but the runtime methodology and formal proof logic required to validate that claim.

Frameworks like Nanopublications or FAIR Digital Objects provide the essential syntax for data accessibility, while OOSI provides the operational semantics for epistemic validation. By treating established representational standards as a foundational substrate, the AKE framework focuses on the higher-order problems of continuous integration, neuro-symbolic verification, and the quantification of marginal information gain, bridging the gap between static semantic storage and autonomous knowledge evolution.

3.5. Implementation: Science as Code

Crucially, implementing this architecture does not require training trillion-parameter models from the first principles. The most efficient approach lies in orchestrating high-performance open models, such as Llama 3 or Mistral, augmented with graph-based retrieval mechanisms (GraphRAGs). Within this "science as code" workflow, established software engineering metaphors become operationally significant.

A scientific submission functions as a pull request for the global knowledge graph. Peer review is reinterpreted as continuous integration, in which automated agents execute unit tests for data integrity, methodological reproducibility, and logical consistency. Acceptance corresponds to a merge that integrates the validated object into the evolving knowledge base.

Recent demonstrations, including systems such as *Co-scientist*, have shown that large language models can autonomously plan and execute complex scientific workflows when coupled with appropriate tools [16]. The proposed architecture extends this demonstrated

agency from experimental execution to the curation, validation, and integration of scientific knowledge.

4. Discussion: From Knowledge Representation to Autonomous Epistemic Control

The introduction of an Autonomous Knowledge Engine (AKE) reframes the discussion of scientific communication from questions of representation to questions of epistemic control. Existing reform efforts have largely focused on improving access, transparency, and reproducibility within a fundamentally human-mediated publication pipeline. In contrast, the AKE architecture explicitly addresses how scientific knowledge can be validated, filtered, and integrated at scales that exceed human cognitive and institutional capacities.

Many alternative models in the literature, including open data repositories, executable papers, research compendia, and nanopublications, improve individual aspects of the scientific workflow but preserve documents as the primary epistemic unit. In this sense, these approaches treat publications primarily as passive containers of information rather than as active, computationally integrated systems, a distinction previously articulated in discussions of post-document scientific communication [17]. Even FAIR Digital Objects emphasize persistent identification and reuse without providing mechanisms for automated reasoning, cross-claim consistency checking, or dynamic revalidation. These approaches mitigate friction without altering the underlying dynamics of knowledge accumulation.

In contrast, AKE operationalizes Object-Oriented Scientific Information as an active system rather than a passive container. Scientific contributions are integrated into an active knowledge graph in which nodes, dependencies, and epistemic weights are continuously updated, reflecting the dynamic properties of large-scale knowledge graph systems described in previous studies [18]. The combination of mixture-of-experts routing, neuro-symbolic verification, and information-theoretic filtering introduces properties that are absent from existing models. Scientific contributions are no longer evaluated solely at the time of publication but remain subject to continuous assessment as the knowledge graph evolves. This enables the detection of contradictions, redundancies, and dependency drift in ways that static repositories cannot support.

A critical distinction lies in the treatment of newness. Conventional publication systems conflate novelty with topical differences, incentivizing breadth over epistemic contributions. The marginal information gain filter implemented in AKE decouples value from the topic and redefines the contribution in terms of belief update. Objects that do not meaningfully alter the posterior state of the knowledge graph are not discarded but demoted to contextual or supporting roles. This shifts the system from a publication-maximizing regime to an information-optimizing regime.

The neuro-symbolic validation layer further differentiates the AKE from existing computational approaches. While large language models have demonstrated remarkable capacity for synthesis and planning, their probabilistic nature precludes their use as the final arbiters of truth. By constraining generative models to act as translators into formal semantics and delegating validation to deterministic proof systems, AKE separates linguistic competence from epistemic authority. This architectural choice directly addresses concerns surrounding hallucinations, unverifiable reasoning, and opaque decision-making in AI-assisted science.

Importantly, AKE does not eliminate human judgment but redistributes it. Human expertise is concentrated at the level of class design, axiom definition, and constraint specification, where conceptual insight is essential and automation is neither feasible nor desirable. The system is responsible for routine verification, redundancy detection, and integration. This division of labor mirrors successful patterns in software engineering,

where human creativity operates within formally constrained environments enforced by automation.

However, the transition to the OOSI paradigm carries inherent epistemic risks that must be critically managed, particularly the danger of over-formalization. Mandating that all scientific claims be machine-verifiable risks imposing structural rigidity, potentially marginalizing qualitative, exploratory, or interpretative research where conceptual ambiguity is often a feature rather than a flaw. Furthermore, unintended exclusionary dynamics may arise regarding technical literacy and resource access; researchers from under-resourced institutions could be systematically disadvantaged if object assembly requires advanced computational infrastructure. To mitigate these exclusionary risks, the AKE explicitly relies on intuitive, AI-driven auto-formalization as a frictionless translation layer, ensuring the system evaluates the epistemic merit of the contribution rather than the coding proficiency of the author.

Beyond epistemic risks, contextualizing the practical deployment of the AKE requires distinguishing between immediate feasibility and long-term speculative goals. Currently, the ingestion of structured data, auto-formalization via LLMs acting as semantic translators, and the maintenance of localized graph retrieval systems are technically viable. Conversely, scaling this architecture to support global, decentralized MoE orchestration introduces profound governance and interoperability challenges. AKE would pose severe monopolistic risks to the scientific record. Therefore, sustainable implementation necessitates a federated governance model (analogous to decentralized web protocols) where universities and research consortia host interoperable nodes, driven by funding agencies shifting their mandates to incentivize computable objects over legacy bibliometrics.

Taken together, these features position the AKE not as an extension of existing publishing models but as a shift toward autonomous epistemic governance. Scientific knowledge is a continuously compiled artifact subject to explicit rules of inheritance, validation, and updates. In this sense, the contribution of this work lies not only in proposing Object-Oriented Scientific Information but also in demonstrating how such objects can be orchestrated into a self-correcting knowledge system.

Worked Micro-Example

To illustrate how OOSI and the AKE operate in practice, we present a minimal end-to-end micro-example drawn from a tightly formalizable biomedical domain. The example demonstrates how a scientific claim can be ingested, validated, and integrated within a unified computational framework.

The object represents a randomized controlled trial assessing whether a treatment reduces post-intervention systolic blood pressure. It encapsulates a structured dataset, a reproducible statistical workflow, and a formally specified claim. The claim asserts that the treatment effect parameter in a linear model is negative and statistically significant under a predefined significance threshold.

The OOSI object declares membership in a clinical RCT class, which enforces constraints on randomization, outcome specification, and admissible models. Validation proceeds through structural checks on the data schema, re-execution of the analysis within a containerized environment, and verification that the reported effect size and p-value satisfy the acceptance conditions encoded in the claim template. If these conditions are met, the claim is marked as verified and becomes eligible for integration into the knowledge graph.

Marginal information gain is then computed over a belief state defined on the direction of the treatment effect. If the posterior belief induced by the object exceeds a domain-specific threshold, the object is integrated as an independent node; otherwise, it is merged as supporting evidence for an existing claim. This mechanism enables quantita-

tive biomedical results to be continuously validated and epistemically weighted without reliance on narrative judgment.

In this example, the belief state is defined over the binary event space $\chi \in$ {negative effect, non-negative effect}. Assuming a prior P_{prior} = (0.60, 0.40) derived from existing graph evidence and a posterior belief P_{post} = (0.85, 0.15) induced by the verified object, the resulting marginal information gain is derived as shown in Equation (2):

$$D_{\text{KL}}\left(P_{\text{post}} \,||\, P_{\text{prior}}\right) \;=\; 0.85 \ln\frac{0.85}{0.60} \;+\; 0.15 \ln\frac{0.15}{0.40} \;\approx\; 0.15 \text{ nats} \tag{2}$$

Since this value exceeds the domain-specific threshold ε, the object warrants integration as an independent knowledge node rather than a mere contextual update.

5. Implications for Policy, Infrastructure, and Incentives

The introduction of an Autonomous Knowledge Engine-AKE has implications that extend well beyond publication practices, directly affecting science policy, research infrastructure, and incentive structures. Because AKE operationalizes OOSI as an active, self-updating system, its adoption necessitates a coordinated rethinking of how scientific output is regulated, funded, evaluated, and rewarded.

5.1. Policy Implications

Current science policy frameworks remain anchored to document-based outputs, with compliance defined in terms of access to articles, data repositories or supplementary materials. Within an AKE-based ecosystem, the policy emphasis would shift from access to documents toward compliance with computability and verifiability requirements. Publicly funded research may require contributions to be submitted as executable, machine-interpretable objects that pass predefined validation and reproducibility checks at the time of ingestion.

Such policies would transform reproducibility from a retrospective concern into a precondition for inclusion in scientific records. Moreover, because AKE assigns an explicit epistemic status to claims based on validation outcomes and information gain, policy instruments can differentiate between exploratory objects, confirmatory contributions, and derivative updates. This enables evidence-weighted governance, in which regulatory and funding decisions can adapt dynamically to the evolving state of the knowledge graph. In this sense, AKE provides a concrete technical substrate for what has been described as epistemic governance, in which information structures participate in the regulation and stewardship of knowledge [19].

5.2. Infrastructure Implications

The deployment of AKE implies a fundamental reorientation of the scholarly infrastructure. Journals and preprint servers will evolve into compilation and orchestration environments responsible for routing submissions, executing validation pipelines, and integrating accepted objects into shared knowledge graphs. This function extends beyond traditional editorial workflows and aligns closely with continuous integration systems in large-scale software development.

At the institutional, national, and international levels, AKE-compatible knowledge graphs would function as persistent scientific backbones, supporting cross-domain interoperability, long-term preservation, and continuous revalidation. Unlike static archives, these infrastructures would remain epistemically active, enabling previously accepted claims to be re-evaluated as new evidence, methods, or formal constraints emerge. Impor-

tantly, such systems are computationally scalable, addressing the structural limitations of human-mediated reviews in an era of exponential knowledge growth.

5.3. Incentive Implications

Perhaps the most transformative impact of AKE lies in its ability to realign scientific incentives. Contemporary reward systems emphasize publication counts, journal prestige, and narrative novelty, reinforcing a volume-driven equilibrium. In contrast, AKE enables fine-grained attribution of credit at the level of scientific objects and their components, including datasets, analytical pipelines, formalized claims, and domain-level class definitions.

This granularity supports alternative metrics grounded in epistemic contributions, such as downstream dependency, reuse frequency, validation persistence, and marginal information gain. Contributions that strengthen, refute, or refine existing knowledge objects, including negative results and replications, become first-class scientific outputs, rather than marginal activities. Consequently, incentive structures can shift from novelty maximization to knowledge stabilization and refinement.

Crucially, AKE does not replace human judgment and creativity. Instead, it redistributes cognitive labor by automating routine verification, redundancy detection, and integration tasks while elevating the importance of conceptual design, theoretical framing, and constraint specification. Incentive systems that recognize these higher-order contributions would more accurately reflect the realities of modern, computation-intensive sciences.

5.4. Proof-of-Concept Implementation: The AXIOMA Prototype

To bridge the gap between conceptual architecture and practical deployment, we developed AXIOMA, a proof-of-concept prototype of the Autonomous Knowledge Engine. AXIOMA is designed to operationalize the OOSI paradigm, demonstrating how scientific claims can be ingested, validated, and integrated into a computable record without relying on unstructured narrative evaluation.

The prototype operationalizes the theoretical layers of the AKE architecture through three primary interfaces: object assembly, epistemic filtering, and topological integration. In a traditional workflow, a researcher drafts a static manuscript. In the AXIOMA, the user constructs a Scientific Object Domain Layer (SODL) unit—the system's specific instantiation of an OOSI. As illustrated in Figure 3, the Object Architect console requires the author to explicitly declare the scientific domain and the object class (e.g., an Educational Data Mining RCT).

This step enforces structural inheritance; the submission must comply with the predefined axioms and data schemas of its parent class. The interface intentionally decouples the structured synopsis from the executable core and the semantic assertions, ensuring that methodology and data remain machine-accessible by design rather than being buried in supplemental text.

Once an object is assembled and formally verified, it undergoes the Epistemic Manifold Test, which computes its marginal information gain against the existing knowledge base. Figure 4 demonstrates the algorithmic execution of the Kullback–Leibler (KL) divergence filter discussed in Section 3.3.

The console contrasts two hypothetical submissions: redundant and disruptive contribution. Redundant contribution is an object that merely applies an established model to a standard dataset, yielding a near-zero epistemic divergence ($\Delta \approx 0.000015$ bits). The system automatically flags this as a "High-Confidence Replication," merging its data to reinforce the existing prior rather than publishing it as a novel standalone paper. A disruptive contribution is an object introducing a mathematically superior heuristic in a sparse data

regime, yielding a significant information gain ($\Delta \approx 1.6518$ bits). Because this exceeds the system's acceptance threshold (ε), the engine authorizes the creation of a new topological branch. This mechanism empirically illustrates how the AKE shifts the gatekeeping process from an evaluation of narrative novelty to a rigorous calculation of epistemic contribution.

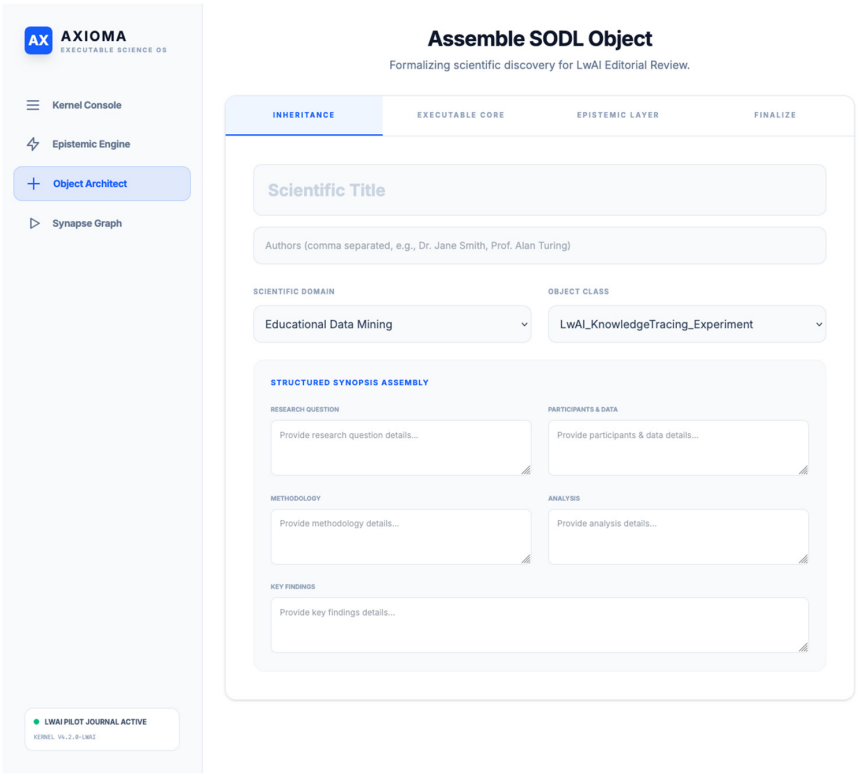


Figure 3. Object architect.

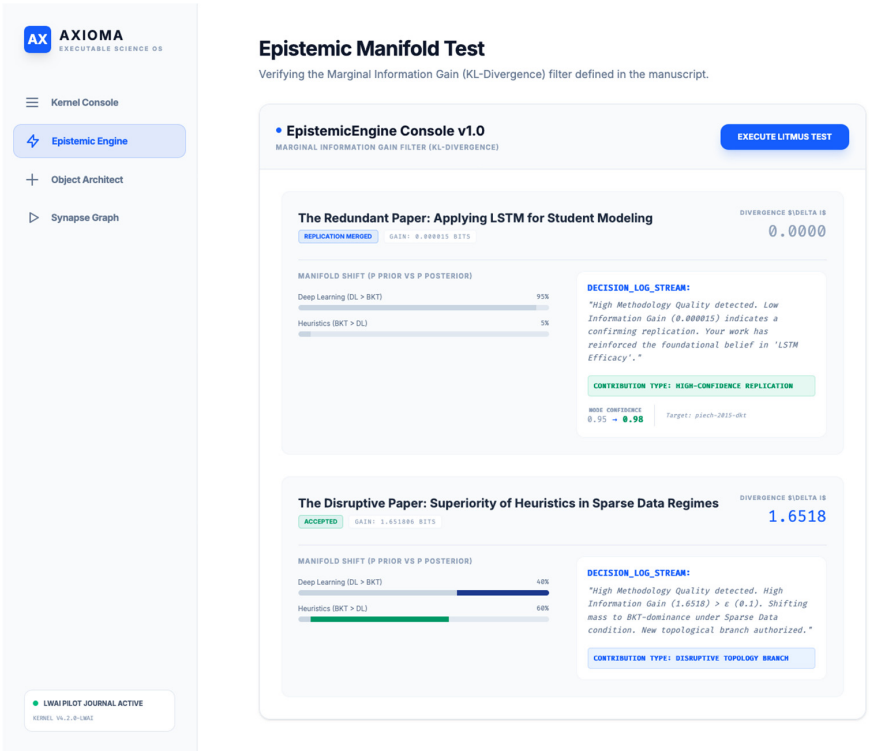


Figure 4. Epistemic Engine.

Upon passing the epistemic filter, validated objects are compiled into the system's global registry, visualized as the Active Topology and Synapse Graph (Figure 5).

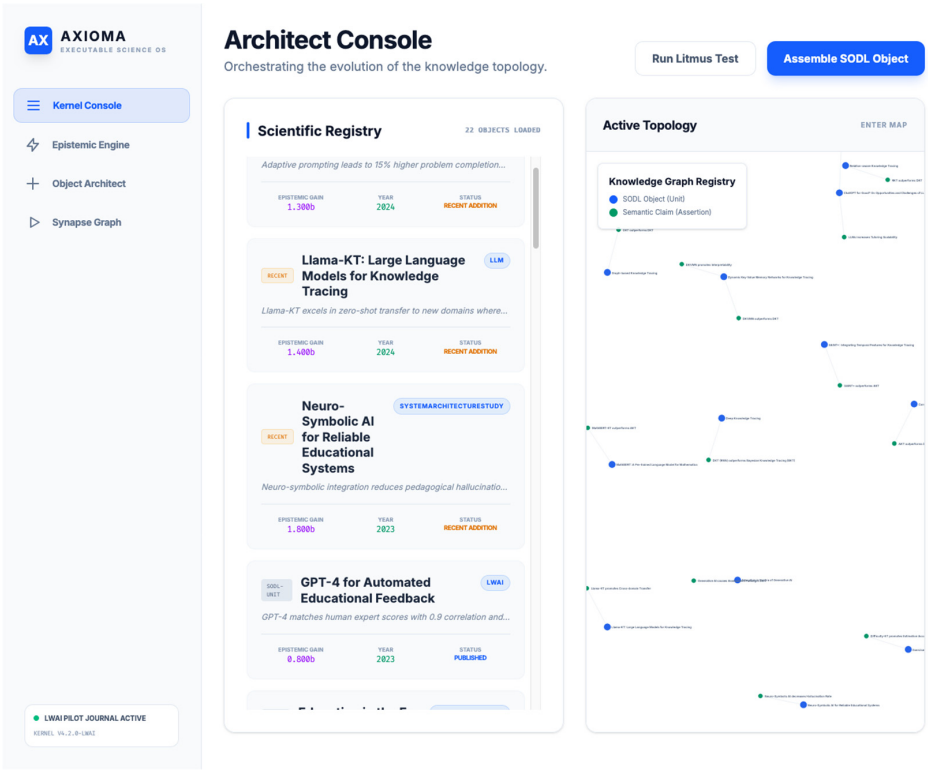


Figure 5. Kernel interface.

Unlike static journal indexes or traditional citation networks, this registry maps the direct semantic and methodological dependencies between instantiated objects. The graph functions as a continuously updated epistemic ledger, where the weight of each node dynamically adjusts as new evidence is ingested and revalidated. Once an object is assembled and passes the initial methodological checks, it is compiled into a verified scientific unit (Figure 6).

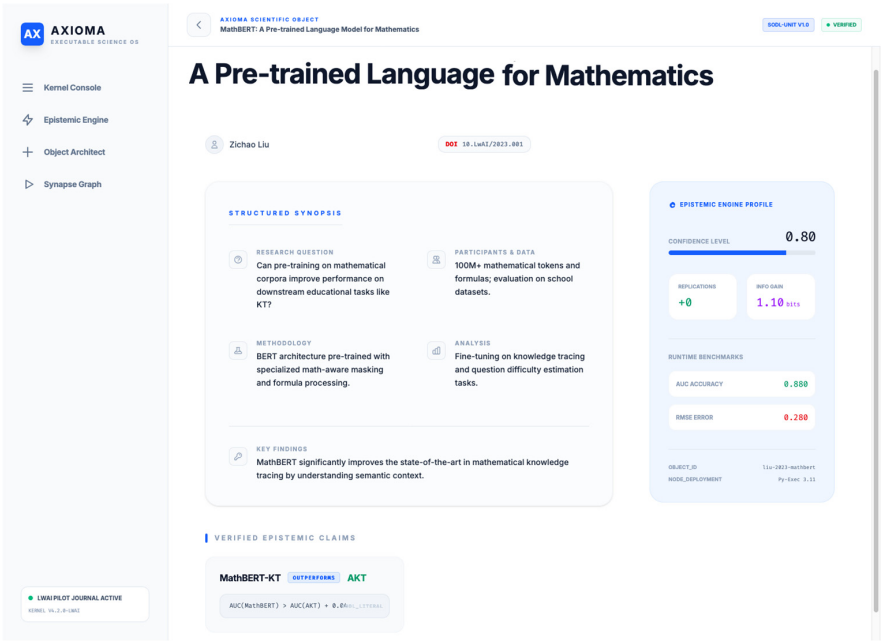


Figure 6. Scientific object (article).

6. Conclusions

The continued reliance on static, document-based publishing has rendered scientific records increasingly misaligned with the scale, complexity, and computational nature of modern research. Incremental reforms have improved access and transparency but have not addressed the structural causes of irreproducibility, redundancy and epistemic overload. These limitations reflect not individual failure but the constraints of an outdated knowledge architecture.

By reframing scientific contributions as executable, verifiable, and interoperable objects, OOSI provides a foundation for computable scientific records. The proposed Autonomous Knowledge Engine demonstrates how such objects can be validated, filtered, and integrated through modular neuro-symbolic systems that respect both domain specialization and formal correctness.

This shift moves scientific communication from narrative accumulation to continuous epistemic integration. This enables a scientific ecosystem that scales with knowledge production, supports systematic self-correction, and preserves human judgment where it matters most. The transition to such an architecture is not only a technical opportunity but also a necessary step toward sustaining reliable science in the computational era, aligning with earlier calls to rethink the infrastructure of scientific discovery in networked and computational terms [20].

Ultimately, the output of scientific inquiry should no longer be a static archival record but an executable function within a global network. In this new paradigm, a 'publication' becomes a modular, active agent capable of interacting with other models to autonomously synthesize new discoveries. We are moving from the reading science era to the computing era.

Limitations and Suggestions

While the OOSI and the AKE offer a comprehensive blueprint for the future of scientific communication, we explicitly note that this work presents a conceptual perspective supported by a minimal proof-of-concept, rather than a globally deployed infrastructure. Realizing this vision entails structural assumptions that must be addressed through rigorous empirical validation.

A primary epistemic limitation of the proposed architecture is its reliance on the strict formalizability of scientific claims. The neuro-symbolic validation layer, which translates natural language into deterministic logic, is highly effective for computationally grounded fields such as physics, genomics, and quantitative educational data mining. However, extending this paradigm to qualitative or highly contextualized disciplines presents a profound ontological challenge, as these fields often resist rigid schema encapsulation.

To transition from a conceptual framework to a validated system, future research must proceed through a structured, metric-driven evaluation roadmap. Initial empirical validation will require testing the pipeline against a retrospective corpus of quantitative papers to measure the auto-formalization accuracy of the semantic translation layer. Subsequently, the marginal information gain filter must be calibrated against real-world scientific consensus, correlating algorithmic divergence scores with established retrospective metrics to define optimal rejection thresholds.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The author have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

https://doi.org/10.3390/standards6020014

Abbreviations

The following abbreviations are used in this manuscript:

OOSI	Object-Oriented Scientific Information
AKE	Autonomous Knowledge Engine
FAIR	Findable, Accessible, Interoperable, Reusable
MoE	Mixture of Experts
LLM	Large Language Model
KL	Kullback–Leibler
CI	Continuous Integration
API	Application Programming Interface
LoRA	Low-Rank Adaptation

References

1. Mons, B.; van Haagen, H.; Chichester, C.; Hoen, P.-B.T.; den Dunnen, J.T.; van Ommen, G.; van Mulligen, E.; Singh, B.; Hooft, R.; Roos, M.; et al. The value of data. *Nat. Genet.* **2011**, *43*, 281–283. [CrossRef] [PubMed]
2. Mons, B.; Neylon, C.; Velterop, J.; Dumontier, M.; da Silva Santos, L.O.B.; Wilkinson, M.D. Cloudy, increasingly FAIR: Revisiting the FAIR Data guiding principles. *Inf. Serv. Use* **2017**, *37*, 49–56. [CrossRef]
3. Smith, R. Peer review: A flawed process at the heart of science and journals. *J. R. Soc. Med.* **2006**, *99*, 178–182. [CrossRef] [PubMed]
4. Baker, M. 1500 scientists lift the lid on reproducibility. *Nature* **2016**, *533*, 452–454. [CrossRef] [PubMed]
5. Ioannidis, J.P.A. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124. [CrossRef] [PubMed]
6. Smaldino, P.E.; McElreath, R. The natural selection of bad science. *R. Soc. Open Sci.* **2016**, *3*, 160384. [CrossRef] [PubMed]
7. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
8. Stodden, V.; McNutt, M.; Bailey, D.H.; Deelman, E.; Gil, Y.; Hanson, B.; Heroux, M.A.; Ioannidis, J.P.; Taufer, M. Enhancing reproducibility for computational methods. *Science* **2016**, *354*, 1240–1241. [CrossRef] [PubMed]
9. Avigad, J.; de Moura, L.; Kong, S. Theorem proving in Lean. In *Proceedings of the 25th International Conference on Automated Deduction (CADE-25)*; Springer: Cham, Switzerland, 2015; pp. 664–678.
10. Groth, P.; Gibson, A.; Velterop, J. The anatomy of a nanopublication. *Inf. Serv. Use* **2010**, *30*, 51–56. [CrossRef]
11. Szegedy, C. A promising path towards autoformalization and general artificial intelligence. In *Intelligent Computer Mathematics*; Springer: Cham, Switzerland, 2020; pp. 3–20. [CrossRef]
12. Shazeer, N.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q.; Hinton, G.; Dean, J. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv* **2017**, arXiv:1701.06538.
13. Fedus, W.; Zoph, B.; Shazeer, N. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *arXiv* **2021**, arXiv:2101.03961.
14. Besold, T.R.; Garcez, A.D.; Bader, S.; Bowman, H.; Domingos, P.; Hitzler, P.; Kühnberger, K.U.; Lamb, L.C.; Lima, P.M.V.; Penning, L.; et al. Neural-symbolic learning and reasoning: A survey and interpretation. Available online: https://ebooks.iospress.nl/doi/10.3233/FAIA210348 (accessed on 20 December 2025).
15. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
16. Boiko, D.A.; MacKnight, R.; Gomes, G. Autonomous chemical research with large language models. *Nature* **2023**, *624*, 570–578. [CrossRef] [PubMed]
17. Kuhn, T.; Chichester, C.; Krauthammer, M.; Dumontier, M. Publishing without Publishers: A Decentralized Approach to Dissemination, Retrieval, and Archiving of Data. *arXiv* **2014**, arXiv:1411.2749.
18. Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge graphs. *ACM Comput. Surv.* **2021**, *54*, 1–37. [CrossRef]
19. Floridi, L. *The Philosophy of Information*; Oxford University Press: Oxford, UK, 2011.
20. Nielsen, M. *Reinventing Discovery: The New Era of Networked Science*; Princeton University Press: Princeton, NJ, USA, 2011.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.