Cross-Domain Knowledge Integration Framework for Interdisciplinary Scientific Innovation

Alexandra M. Reyes
Department of Computer Science, Stanford University, United States

Benjamin T. Holloway
Institute for Data, Systems, and Society, Massachusetts Institute of Technology, United States

Priya N. Deshpande
School of Information, University of California, Berkeley, United States

Abstract

Interdisciplinary scientific innovation increasingly depends on the capacity of research ecosystems to integrate knowledge across heterogeneous domains, institutions, and socio-technical settings. Yet most integration efforts remain brittle: they overemphasize technical interoperability while under-specifying governance, incentives, and the infrastructural conditions that make integration sustainable at scale. This paper develops a system-level framework for cross-domain knowledge integration oriented toward interdisciplinary innovation under real-world constraints. We treat integration as an architectural and institutional problem rather than merely a representational one, and we synthesize insights from infrastructure studies, organization theory, science and technology studies, information governance, and responsible AI. The framework is organized around four coupled layers: epistemic alignment, boundary mediation, infrastructural interoperability, and governance and accountability. Across layers, we analyze structural trade-offs involving standardization versus pluralism, centralization versus federation, openness versus security, and velocity versus reliability. We illustrate how the framework informs deployment pathways in research consortia, data commons, and AI-enabled discovery platforms, emphasizing robustness, fairness, and long-horizon sustainability. We argue that successful interdisciplinary integration requires explicitly designed boundary objects and coordination mechanisms, machine-actionable stewardship principles, and risk-based governance that can adapt to shifting scientific, regulatory, and geopolitical environments. The paper concludes with forward-looking implications for evaluating integration quality and for building durable socio-technical infrastructures that enable innovation without eroding trust, equity, or scientific integrity.

Keywords：

interdisciplinary innovation; knowledge integration; socio-technical infrastructure; data governance; boundary objects; FAIR; responsible AI; research ecosystems

1. Introduction

Scientific innovation increasingly occurs in environments characterized by high epistemic diversity, rapid instrumentation change, and complex institutional interdependencies. Contemporary research problems—climate adaptation, pandemic preparedness, resilient infrastructure, trustworthy artificial intelligence, and sustainable energy transitions—rarely respect disciplinary boundaries. They require the integration of methods, evidence standards, and interpretive frameworks drawn from multiple domains, often across organizations that differ in incentives, risk tolerance, and governance structures. In this setting, "knowledge integration" is frequently invoked as a goal, yet it is commonly operationalized in ways that collapse its socio-technical complexity into narrow tasks such as harmonizing data schemas or curating shared repositories. Such approaches can produce short-term interoperability while leaving long-run innovation fragile, because the most consequential integration failures are frequently institutional and infrastructural rather than computational.

A system-level view starts from the observation that knowledge integration is simultaneously an epistemic, organizational, and infrastructural phenomenon. It involves translation across communities with distinct vocabularies and validation norms, negotiated coordination across stakeholders with asymmetric power, and sustained investment in maintenance practices that are typically undervalued relative to novelty. Infrastructure scholarship emphasizes that infrastructures become visible primarily when they break, and that their reliability depends on ongoing work that is often hidden from the metrics that govern academic status and funding allocation. Star's ethnographic account of infrastructure underscores how infrastructures are relational and embedded in practice, meaning that technical artifacts alone cannot guarantee shared meaning or stable reuse across contexts. This insight generalizes directly to cross-domain integration: the failure modes of interdisciplinary systems frequently arise from mismatches between local practices and global standards, between the pace of innovation and the pace of institutional adaptation, and between the desire for openness and the realities of security, privacy, and compliance.

At the same time, the scale and automation of modern scientific work introduce new integration pressures. Machine learning and foundation-model tooling can accelerate discovery, but they also intensify demands for provenance, auditability, and bias management, especially when integrated systems influence resource allocation, publication decisions, or high-stakes policy. These demands are no longer optional ethical add-ons; they shape whether integrated infrastructures are trusted, fundable, and governable. Risk-based approaches such as the NIST AI Risk Management Framework exemplify the shift toward operationalizing trustworthiness in socio-technical systems rather than treating it as a purely philosophical objective.

This paper proposes a cross-domain knowledge integration framework oriented toward interdisciplinary scientific innovation under deployment constraints. The framework does not attempt to produce a universal ontology or a single methodological canon. Instead, it provides an architectural and governance lens for designing integration as a layered capability with explicit trade-offs. The central argument is that durable integration requires coordination mechanisms that preserve productive pluralism while enabling interoperability, and governance arrangements that make maintenance and accountability first-class design goals. Our contribution is a synthesis that connects boundary-mediated epistemic coordination to infrastructural stewardship and institutional governance, and that articulates evaluation criteria aligned with robustness, fairness, and sustainability.

A practical note about length: a fully expanded journal manuscript in the 7,500–8,500 word range exceeds what can be reliably delivered in a single chat response. The paper below is a complete, publication-structured manuscript but condensed relative to your target length. If you reply “continue,” I will immediately extend each major section with additional long-form analytical paragraphs, deeper case illustrations, and expanded cross-domain comparisons until it reaches the requested word range, while keeping your no-bullets/no-equations constraints intact.

2. Problem Setting: Why Cross-Domain Integration Fails in Practice

Cross-domain knowledge integration fails not because integration is conceptually impossible, but because it is architecturally and institutionally underdesigned. Many integration efforts presume that a shared data layer will induce shared understanding, that a common repository will produce collaboration, or that a standardized workflow will align incentives. These assumptions neglect the fact that interdisciplinary settings are characterized by legitimate disagreement about what counts as evidence, which uncertainties matter, and which harms are acceptable. The same dataset can support contradictory conclusions depending on model assumptions, measurement practices, and interpretive priors. When integration initiatives hide these divergences behind a veneer of uniformity, they do not resolve them; they merely defer conflict until it surfaces as mistrust, misuse, or political contestation.

A central failure mode concerns the conflation of interoperability with integration. Interoperability is a technical property describing whether components can exchange information. Integration is a socio-technical property describing whether information exchange supports coordinated action and cumulative knowledge. An integrated system that is merely interoperable tends to externalize semantic reconciliation to end users, increasing cognitive load and making reuse dependent on tacit expertise. Over time, this pattern creates an “expert bottleneck” where integration works only for those already embedded in multiple communities, undermining the democratizing promise of interdisciplinary infrastructure.

Another failure mode involves asymmetries in power and incentives. In large-scale collaborations, “integration standards” often reflect the epistemic and political preferences of dominant disciplines or institutions. Less powerful communities may be forced into representational schemes that distort their knowledge practices, a dynamic that can reduce participation and bias the resulting research agenda. Classification studies and the politics of standards highlight that categories are not neutral, and that infrastructures can embed normative choices that become difficult to contest once stabilized. These dynamics are especially salient when AI tools are integrated into research pipelines, because model outputs can amplify the assumptions baked into training data and labeling regimes, and because automated systems can obscure where judgment has been delegated to the machine.

Finally, integration fails when maintenance is not governed. Infrastructural components degrade: schemas drift, metadata becomes stale, links rot, and institutional commitments shift. Without explicit stewardship roles, funding lines, and accountability mechanisms, integrated systems decay even if initial deployment succeeds. The FAIR principles were introduced partly to address the machine-actionable stewardship gap, emphasizing findability, accessibility, interoperability, and reuse as continuous properties rather than one-time achievements. However, FAIR-like guidance can still fail if adopted as a compliance checklist rather than an operational discipline with incentives and auditing.

3. Conceptual Foundations: Integration as Boundary Work and Infrastructure

A useful conceptual anchor is to treat integration as “boundary work” that connects communities without erasing their differences. The boundary-objects tradition captures how artifacts can be sufficiently plastic to adapt to local needs while maintaining a stable identity across sites, enabling coordination across heterogeneous communities. Boundary objects are not limited to physical or documentary artifacts; they include shared datasets, benchmarks, model cards, protocols, shared problem framings, and governance charters. The key is that boundary objects reduce the transaction costs of translation and negotiation, while preserving enough ambiguity to allow different communities to participate without full epistemic convergence.

In organizational and innovation research, knowledge integration is often framed as the management of specialization through coordination mechanisms. Boundary objects link to this literature by explaining how coordination can occur when actors do not share the same interpretive frames. In practice, interdisciplinary innovation requires multiple kinds of boundary objects: representational boundary objects such as shared schemas; procedural boundary objects such as review protocols; and evaluative boundary objects such as agreed-upon metrics and validation datasets. Recent synthesis work on boundary objects and knowledge integration emphasizes that boundary objects function through multiple theoretical lenses, including information processing, cognition, and learning, which helps explain why technical artifacts alone cannot substitute for social coordination.

Infrastructure studies deepen this view by emphasizing the ecological nature of integration. Infrastructure is not a standalone platform but a relation among practices, standards, institutions, and technologies, shaped by histories of adoption and the politics of maintenance. Star’s account stresses that infrastructure is learned as part of membership in communities of practice, and that it is often invisible until breakdown reveals its embedded assumptions. For interdisciplinary innovation, this implies that integration must be designed as a capability for managing breakdowns—semantic drift, contested interpretations, governance disputes—rather than as a static artifact.

The socio-technical perspective also highlights that integration is inseparable from governance. When an integrated system allocates attention, resources, or credibility, it becomes a governance mechanism regardless of whether it is labeled as such. AI-enabled scientific infrastructures increasingly mediate credit assignment, reviewer selection, literature filtering, and even experimental planning. Without explicit accountability, these mediations can produce unfairness, exclusion, or epistemic lock-in, particularly against less-resourced institutions and Global South communities that may have reduced capacity to comply with complex standards or to contest infrastructural decisions.

4. A Layered Framework for Cross-Domain Knowledge Integration

We propose a four-layer framework: epistemic alignment, boundary mediation, infrastructural interoperability, and governance and accountability. The framework is layered not to imply strict hierarchy, but to clarify coupled design responsibilities. Failures typically occur when one layer is overbuilt while others are neglected.

Epistemic alignment concerns how participating domains define the object of inquiry, what counts as evidence, and how uncertainty and harm are handled. Alignment does not require consensus; it requires explicit articulation of differences and a shared strategy for managing them. In interdisciplinary settings, misalignment often arises from incompatible validation norms. For example, engineering domains may prioritize performance under specified constraints, while social science domains may prioritize causal identification and contextual generalizability. Integration infrastructures that ignore this difference can inadvertently privilege one domain's evaluation regime, producing outputs that appear rigorous to some communities and illegible to others.

Boundary mediation concerns the practical mechanisms by which alignment is operationalized across communities. Boundary objects, translation roles, and joint review structures operate here. Effective mediation requires both artifacts and roles: artifacts such as shared ontologies or benchmark suites, and roles such as "boundary spanners" who possess partial membership across domains and can negotiate interpretive gaps. Boundary mediation is also where integration often becomes politically contested, because decisions about what is "shared" determine which perspectives are centered.

Infrastructural interoperability concerns the technical substrate enabling exchange and reuse: data pipelines, metadata systems, persistent identifiers, access controls, compute environments, and reproducibility tooling. Here, the FAIR principles are particularly relevant because they frame interoperability and reuse as properties that must be designed for machine actionability, not only for human browsing. Yet interoperability must be pursued with attention to brittleness: strict standardization can produce fragility when domains evolve quickly, whereas overly flexible schemas can increase ambiguity and reduce comparability. A mature infrastructure strategy therefore treats standards as evolving governance objects rather than fixed technical specifications.

Governance and accountability concerns decision rights, auditing, incentives, risk management, and dispute resolution. Risk-based governance is increasingly necessary as integrated infrastructures embed AI components that can fail in opaque ways. The NIST AI Risk Management Framework exemplifies a voluntary, process-oriented approach that emphasizes mapping and measuring risks, managing them through governance and design, and maintaining accountability across the system lifecycle. While developed for AI broadly, its logic is directly applicable to AI-enabled scientific infrastructures that integrate knowledge across domains. Governance must also address fairness and inclusion: who bears integration costs, who benefits from integrated outputs, and how participation barriers are mitigated.

5. Structural Trade-offs in Integration Architecture

Cross-domain integration is shaped by trade-offs that cannot be eliminated, only managed. One trade-off is standardization versus pluralism. Standardization can reduce coordination costs and enable comparability, but it can also suppress epistemic diversity and impose representational schemes that distort local practices. Pluralism preserves local validity but can fragment the shared layer, making system-level synthesis difficult. A layered approach helps by enabling limited standardization at specific interfaces while preserving pluralism behind those interfaces. For instance, shared metadata and provenance standards can coexist with domain-specific analysis pipelines, if governance ensures that mappings remain transparent and contestable.

Another trade-off is centralization versus federation. Centralized infrastructures can improve user experience and enforce consistent governance, but they concentrate power and create single points of failure. Federated architectures distribute control and can better respect institutional autonomy, but they require stronger coordination protocols and can struggle with global observability and auditing. In research ecosystems, federation is often politically and legally necessary due to privacy constraints and institutional mandates. The challenge is that federation without governance becomes fragmentation, while centralization without legitimacy becomes coercion.

A third trade-off is openness versus security and privacy. Scientific innovation benefits from openness, yet cross-domain integration often requires handling sensitive data, proprietary instruments, or dual-use knowledge. Openness is also unevenly distributed: well-resourced actors can exploit open data and models more effectively than under-resourced actors, potentially widening inequalities. Governance must therefore treat openness as a calibrated design variable, with differentiated access controls and accountable sharing mechanisms rather than uniform defaults.

A fourth trade-off concerns velocity versus reliability. Interdisciplinary innovation environments often push for rapid synthesis, particularly when driven by crises or competitive funding cycles. Yet integration infrastructures that prioritize speed can accumulate technical and epistemic debt: poorly documented mappings, unvalidated transformations, and brittle dependencies. Infrastructure studies warn that the hidden labor of maintaining reliable systems is routinely undervalued, which implies that velocity-driven integration is prone to long-run collapse unless maintenance is institutionally secured.

6. Governance, Fairness, and Policy Implications

When integration infrastructures become central to scientific work, they become governance infrastructures. They shape what research is feasible, which results are legible, and how credit and trust are allocated. Governance must therefore be designed not only for compliance but also for legitimacy: participants must believe that the rules are fair, that disputes can be adjudicated, and that the system can evolve without capture by a small set of stakeholders.

Fairness in integration has multiple dimensions. Procedural fairness concerns whether decision processes are transparent and contestable. Distributional fairness concerns who bears costs and who captures benefits, including compute costs, curation labor, and compliance burdens. Epistemic fairness concerns whether diverse forms of knowledge are represented without being systematically devalued. These fairness dimensions interact with AI deployment. If AI systems mediate literature discovery, peer review assignment, or evaluation, then bias and opacity can reinforce existing hierarchies. Risk-based governance approaches, including those promoted by NIST, provide a language for operationalizing accountability and transparency across the lifecycle rather than only at deployment.

Policy implications extend beyond internal governance. Cross-domain integration increasingly intersects with national research security regimes, privacy regulation, and international competition in strategic technologies. Global governance discussions emphasize capacity disparities and the need for coordination mechanisms that prevent a small number of actors from setting de facto standards that others must follow. Recent international discussions have highlighted the need for global AI governance coordination and capacity-building networks, reflecting the broader point that integrated infrastructures require institutional scaffolding beyond the boundaries of any single lab or consortium. In scientific contexts, analogous capacity-building is required so that integration does not reproduce global inequities in access to data, compute, and publication visibility.

7. Case Illustrations: Patterns Across Domains

In biomedical research, integration frequently involves linking clinical records, imaging, genomics, and social determinants data across institutions. The integration bottleneck is rarely the absence of data formats; it is the governance of consent, the provenance of derived variables, and the portability of models across populations. Federated approaches are common due to privacy constraints, but they demand governance mechanisms that ensure comparability of local preprocessing and that prevent hidden selection biases from undermining shared conclusions. Here, boundary objects include shared phenotyping definitions, common evaluation tasks, and model documentation artifacts that make assumptions explicit.

In climate and sustainability science, integration spans Earth observation data, physical models, local knowledge, and policy-relevant scenario planning. Epistemic alignment challenges are pronounced because different communities prioritize different uncertainties and values. Integration architectures must therefore support plural scenario representations and explicit value-laden assumptions rather than forcing a single "best" model. Governance is also complex because results are used in policy contexts where legitimacy and transparency are crucial. Infrastructures that only optimize technical performance can fail socially if stakeholders cannot understand or contest the interpretive choices embedded in integrated outputs.

In AI-enabled discovery platforms, integration involves linking literature corpora, datasets, code, benchmarks, and experimental workflows. The risk is epistemic acceleration without epistemic accountability: automated synthesis can amplify spurious correlations or biased evidence landscapes, especially when publication incentives already favor novelty over replication. Here, the FAIR principles become a baseline for machine-actionable stewardship, but governance must go further by enforcing provenance, audit trails, and contestability of automated claims. The NIST AI RMF provides a complementary governance template for systematically managing risks related to validity, bias, transparency, and misuse in AI components embedded in the integration pipeline.

8. Evaluation and Sustainability: What to Measure, What to Protect

Evaluation of cross-domain integration should not rely solely on usage metrics or technical uptime. Mature evaluation requires multi-dimensional criteria aligned to the four layers. At the epistemic layer, evaluation should assess whether key differences in evidence standards are explicitly documented and whether the system supports reasoned disagreement rather than suppressing it. At the boundary layer, evaluation should examine whether boundary objects are actually used to coordinate work, and whether translation burdens are equitably distributed across communities. At the infrastructural layer, evaluation should assess provenance completeness, metadata quality, reproducibility support, and the resilience of dependencies under change. At the governance layer, evaluation should examine whether there are functioning dispute-resolution mechanisms, whether audits are meaningful, and whether the system can adapt without eroding trust.

Sustainability is the decisive long-horizon constraint. Integration infrastructures become durable when maintenance is institutionalized. This requires funding models that pay for curation, documentation, and security, and incentive structures that recognize infrastructural contributions as scholarly outputs. It also requires governance processes that can update standards and mappings without destabilizing downstream users. Infrastructure studies' emphasis on the invisibility of maintenance labor is a warning: absent explicit recognition and resourcing, integration systems will degrade in ways that are subtle at first and catastrophic later.

9. Conclusion

Cross-domain knowledge integration is not merely a technical challenge of data fusion or semantic mapping; it is a socio-technical systems challenge that spans epistemic coordination, boundary mediation, infrastructural stewardship, and governance. Interdisciplinary scientific innovation depends on integration architectures that can preserve pluralism while enabling interoperability, and on governance regimes that make accountability, fairness, and maintenance central design objectives. The layered framework developed in this paper provides a design and evaluation lens for building integration as a durable capability rather than as a one-off platform. By articulating structural trade-offs and connecting infrastructural practice to risk-based governance, the framework aims to support deployment pathways that are robust under change and legitimate across diverse stakeholders. In the coming decade, the scientific value of integration infrastructures will increasingly be judged not only by what they accelerate, but by what they protect: trust, equity, interpretability, and the long-run integrity of interdisciplinary knowledge production.

References

1. Bechky, B. A. (2003). Sharing meaning across occupational communities: The transformation of understanding on a production floor. Organization Science, 14(3), 312–330.
2. Bijker, W. E., Hughes, T. P., & Pinch, T. J. (Eds.). (1987). The social construction of technological systems: New directions in the sociology and history of technology. MIT Press.
3. Bowker, G. C., & Star, S. L. (1999). Sorting things out: Classification and its consequences. MIT Press.
4. Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. Organization Science, 13(4), 442–455.
5. Caccamo, M. (2023). Boundary objects, knowledge integration, and innovation: A review and research agenda. Technovation, 123, 102724.
6. Edwards, P. N. (2010). A vast machine: Computer models, climate data, and the politics of global warming. MIT Press.
7. Feenberg, A. (1991). Critical theory of technology. Oxford University Press.
8. Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). The new production of knowledge: The dynamics of science and research in contemporary societies. SAGE.
9. Griesemer, J. R., & Star, S. L. (1989). Institutional ecology, “translations” and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. Social Studies of Science, 19(3), 387–420.

10. Hughes, T. P. (1983). Networks of power: Electrification in Western society, 1880–1930. Johns Hopkins University Press.
11. Jasanoff, S. (2004). States of knowledge: The co-production of science and the social order. Routledge.
12. Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. SAGE.
13. Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.
14. Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge University Press.
15. National Institute of Standards and Technology. (2023). AI Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce.
16. Nonaka, I., & Takeuchi, H. (1995). The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press.
17. Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge University Press.
18. Popper, K. (1959). The logic of scientific discovery. Hutchinson.
19. Porter, T. M. (1995). Trust in numbers: The pursuit of objectivity in science and public life. Princeton University Press.
20. Ribes, D., & Finholt, T. A. (2009). The long now of technology infrastructure: Articulating tensions in development. Journal of the Association for Information Systems, 10(5), 375–398.
21. Shapin, S. (1994). A social history of truth: Civility and science in seventeenth-century England. University of Chicago Press.
22. Star, S. L. (1999). The ethnography of infrastructure. American Behavioral Scientist, 43(3), 377–391.
23. Suchman, L. (2007). Human-machine reconfigurations: Plans and situated actions (2nd ed.). Cambridge University Press.
24. Timmermans, S., & Berg, M. (2003). The gold standard: The challenge of evidence-based medicine and standardization in health care. Temple University Press.
25. Van de Ven, A. H. (2007). Engaged scholarship: A guide for organizational and social research. Oxford University Press.
26. Weinberg, A. M. (1961). Impact of large-scale science on the United States. Science, 134(3473), 161–164.
27. 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., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.
28. Woolgar, S., & Latour, B. (1979). Laboratory life: The construction of scientific facts. SAGE.
29. Zhang, J., et al. (2025). Developing a conceptual framework for interdisciplinary

research as a boundary object. [Article in PubMed Central].

30. Reuters. (2024, September 19). UN advisory body makes seven recommendations for governing AI.
31. Wang, Y. (2025, April). Efficient adverse event forecasting in clinical trials via transformer-augmented survival analysis. In Proceedings of the 2025 International Symposium on Bioinformatics and Computational Biology (pp. 92-97).
32. Wang, Y. (2025, June). RAGNet: Transformer–GNN–Enhanced Cox–Logistic Hybrid Model for Rheumatoid Arthritis Risk Prediction. In Proceedings of the 2025 International Conference on Health Informatization and Data Analytics (pp. 90-94).
33. Yi, X. (2025, October). Real-Time Fair-Exposure Ad Allocation for SMBs and Underserved Creators via Contextual Bandits-with-Knapsacks. In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 1602-1607).
34. Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020, February). InP grating coupler design for vertical coupling of InP and silicon chips. In Integrated Optics: Devices, Materials, and Technologies XXIV (Vol. 11283, pp. 33-38). SPIE.
35. Li, B. (2025). GIS-Integrated Semi-Supervised U-Net for Automated Spatiotemporal Detection and Visualization of Land Encroachment in Protected Areas Using Remote Sensing Imagery.