ARIEL: AGENTIC RETRIEVAL INTERFACE FOR ELECTRONIC LOGBOOKS

T. Hellert*, K. Iliev, A. Sulc, G. Martino, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
C. Tennant, Thomas Jefferson National Accelerator Facility, Newport News, VA, USA
H. Hoschouer, Fermi National Accelerator Laboratory, Batavia, IL, USA

Abstract

Operational logbooks are essential for documenting accelerator performance, interventions, and operator experience, but their content is often inconsistent, unstructured, and difficult to search. This limits both human retrieval and the use of AI systems that rely on high-quality historical data. Project ARIEL (Agentic Retrieval Interface for Electronic Logbooks) introduces a modular, facility-agnostic framework that standardizes how logbook information is ingested, enriched, and searched across accelerator laboratories. Each participating site hosts its own ARIEL database while adopting a shared schema, data-enhancement modules, and interoperable search components. Enhancement modules provide semantic metadata, text and figure embeddings, and optional machine-state snapshots at ingestion time. On the retrieval side, ARIEL supports keyword, embedding-based, multimodal, and machine-state search, forming a unified foundation for an agentic retrieval layer capable of orchestrating multiple search strategies. This contribution presents the architecture, schema design, and early cross-facility prototypes, and describes how ARIEL fits into the broader DOE Genesis mission to establish shared, interoperable AI infrastructure for accelerator facilities.

INTRODUCTION

Operational logbooks are among the richest sources of institutional memory at accelerator facilities, yet their practical utility is severely constrained by the way they are written and stored. For example, a meta-analysis of the Jefferson Lab CEBAF (Continuous Electron Beam Accelerator Facility) electronic logbook spanning more than 1.25 million entries over fourteen years illustrates the scale and character of this challenge: only about 15% of entries are human-authored operational notes, with the remainder dominated by auto-generated records such as shift sign-ins and alarm handler outputs [1]. At the other extreme, Fermilab's Accelerator Division logbook (ADEL) is almost entirely human-authored, illustrating how widely the auto-generated-to-human ratio varies across DOE facilities. The human-authored fraction carries the genuine operational knowledge. These entries are free-form notes written under time pressure, laden with facility-specific shorthand, bare acronyms, beamline region codes, and device identifiers whose meaning is opaque to anyone outside a narrow expert community. Equivalent concepts may appear in a dozen different surface forms across sites and time periods, so literal keyword search is brittle, and operators must often already know the precise phrase an entry used in order to find it. Critical troubleshooting knowledge is effectively locked away, inaccessible for training new operators or for AI systems that depend on high-quality historical records. These limitations are not merely an inconvenience for human operators; they directly constrain the downstream AI systems that the accelerator community is now actively building. Prior work has explored AI-assisted logbook analysis on a per-team basis [2], and a recent survey has systematically catalogued the resulting gaps [3]. A central challenge identified across that body of work is how to share enhancements and standards without centralizing sensitive operational data. Each facility has legitimate reasons to keep its raw logbook records local. Without a coordinated approach, individual sites will continue to solve the same infrastructure problems independently, and retrieval tools built at one laboratory will not benefit its neighbors. Controlled-vocabulary expansion systematically maps shorthand such as "BPM" to its canonical form "Beam Position Monitor". This technique can partially bridge the vocabulary gap, but only when applied consistently to both the indexed entries and the user queries at retrieval time. This requirement is non-trivial: expanding queries without also expanding the indexed documents degrades dense retrieval by shifting the query vector away from the region of embedding space where the un-expanded document vectors reside [4]. Getting this right requires a shared infrastructure layer, not ad-hoc per-site scripts. ARIEL addresses this challenge by providing a modular, facility-agnostic framework that defines a shared schema, a library of enhancement and search modules, and an agentic retrieval layer, while leaving each site in full control of its own data. The work is complementary to sister efforts on EPICS event-log anomaly detection [5] and machine-state embeddings as an operational coordinate system [6], both of which interact with or build upon the enriched logbook substrate that ARIEL produces.

ARIEL ARCHITECTURE

The ARIEL architecture is organized around two principles: federated deployment and modular enrichment. Each participating facility hosts its own ARIEL database instance, so raw logbook data never leave the site. A shared layer ensures that independently deployed instances remain interoperable. It comprises a common schema, a module API, and a growing library of enhancement and search components. Figure 1 summarizes the resulting ingestion-to-enhancement-to-search architecture. Facility-specific *adapter* code is the entry point for every site. Adapters

* thellert@lbl.gov

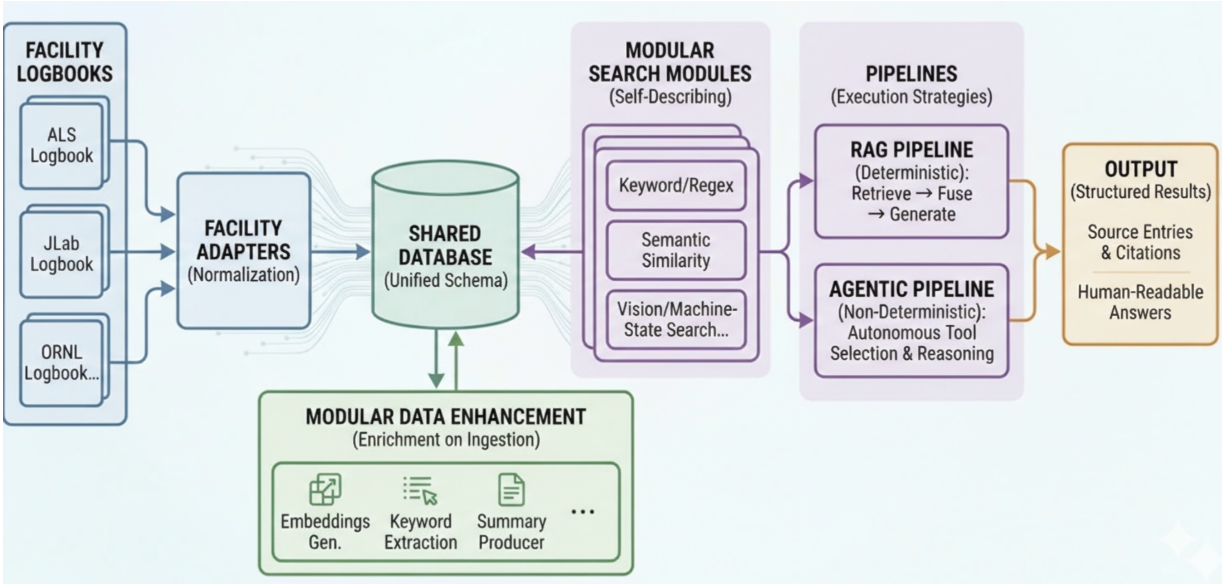

Figure 1: ARIEL system architecture. Facility-specific adapters convert raw logbook entries into a shared schema while preserving local data sovereignty. Enhancement Modules then analyze each entry, extracting text and generating semantic embeddings, to produce enriched "Enhanced Entries" suitable for unified search. Query modules, including semantic retrieval and RAG, can be used directly by users or composed by the agentic retrieval layer to resolve complex queries. (Vision search is planned for future releases)

translate native logbook formats (which vary considerably across laboratories in structure, field naming, and attachment conventions) into the shared ARIEL schema, normalizing fields such as timestamps, author identifiers, and attachment references without discarding any facility-specific metadata. Once an entry has been normalized, it passes through a configurable pipeline of *Enhancement Modules* that operate at ingestion time to produce an "Enhanced Entry." Shared modules available to all sites include LLM-based summarization and keyword extraction, controlled-vocabulary expansion that maps local acronyms and device identifiers to canonical forms, and the generation of one or more text embeddings using selectable encoder models. Figure embeddings for attached images are generated in parallel. Sites that expose machine-state data via EPICS may also attach optional state snapshots to each entry, linking the textual logbook record to the contemporaneous machine configuration. The schema is designed to be simultaneously flexible, consistent, and modular. Modules may add fields to the Enhanced Entry at ingestion time, and site-specific extensions are explicitly supported, but all entries share a common set of required fields and naming conventions that guarantee cross-facility interoperability. Multi-embedding support is built in: a single entry can carry embeddings from several encoder models simultaneously, along with acronym-expansion tables, machine-state metadata, and ontology links. This design is motivated by empirical findings that embedding-model performance varies substantially across models and retrieval tasks, making per-site and per-task model selection preferable to a single shared default. Across six models evaluated on the CEBAF logbook, top-10 recall on title-to-body retrieval ranged from roughly 0.15 to 0.52, a spread large enough to make model choice a first-order engineering decision rather than a minor implementation detail [7]. Notably, a sentence-embedding model fine-tuned on accelerator-physics literature [8] sat at the lower end of that band on logbook retrieval. The model was trained on long-form journal articles and proceedings, but operator notes are short and full of facility shorthand, and a model tuned for one style does not necessarily work well on the other. Enhancement modules that are domain-generic, such as auto-tagging and summarization, can be shared across sites, while modules that depend on facility-specific infrastructure, such as EPICS process-variable attachment, remain site-local.

SEARCH MODES AND AGENTIC RETRIEVAL

ARIEL exposes a tiered spectrum of search modules that range from classical keyword and regular-expression matching through embedding-based semantic retrieval, hybrid ranking, and retrieval-augmented generation (RAG), up to machine-state search and, in future releases, vision-based retrieval over attached figures. Because all modules operate on the shared ARIEL schema, the majority work cross-facility without modification; site-specific modules coexist in the same registry. The interaction between controlled vocabulary and retrieval mode has direct practical consequences. CEBAF evaluations constructed a purpose-built acronym query set of 276 controlled-vocabulary triggers, each paired with ground-truth logbook entries, and measured Precision@10, Hit Rate@10, and MRR@10 under four permutations: no enhancement, entry-only, query-only, and both sides [4]. Applying expansion to both entries and queries improved precision by 29%, hit rate by 13%, and

MRR by 19% over the un-enhanced baseline. Expanding queries without also expanding the indexed documents, by contrast, degraded all three metrics by 17–22%: a counterintuitive result that arises from the way dense embedding models compress entire inputs into single fixed-dimensional vectors. Adding canonical long forms to a query shifts the whole vector into a different region of embedding space, away from the un-expanded document vectors that encode only the bare acronym. This asymmetry trap does not exist in keyword search and is the primary reason symmetric expansion is a hard requirement for any production deployment of CV-enhanced dense retrieval [4]. Dense retrieval also remains weak for rare device identifiers, highly ambiguous shorthand, and non-discriminative expansions; these are natural candidates for hybrid ranking, which combines embedding similarity with keyword and regular-expression matching and explicit metadata filters to recover precision in exactly the cases where pure semantic methods fail. Machine-state search operates over the snapshot embeddings attached at ingestion time, enabling nearest-neighbor retrieval in the operational coordinate system [6]: given a current or target machine configuration, the system can surface historically similar operating conditions and the logbook entries associated with them, directly connecting the textual record to the quantitative machine history. Above the individual search modules sits the ARIEL agentic retrieval layer. Rather than presenting a static search bar, the agent can decompose a complex natural-language query (such as "summarize all injector faults over the last month") into a sequence of tool calls, selecting the appropriate module or combination of modules for each sub-task: consulting the facility ontology to resolve component identifiers, querying the logbook for related fault entries, and synthesizing the results into a human-readable summary or report. The layer is explicitly modular: reasoning, planning, and tool-orchestration components are independently swappable, so new orchestration strategies, including emerging chain-of-agents or deep-research frameworks, can be integrated as the field advances.

CROSS-FACILITY PROTOTYPES

Two prototype deployments currently demonstrate the architecture in concrete settings. The first is a self-contained mock-up logbook that ships with the Osprey reference implementation [9, 10]. Because it carries its own synthetic data and requires no connection to any real facility system, it serves as a self-consistent validation environment for the full agentic stack and as a tutorial and onboarding vehicle for new sites or developers integrating ARIEL for the first time. The second prototype connects ARIEL to a production system. The `als-profiles` configuration wires ARIEL directly to the Advanced Light Source logbook database, which holds approximately 120,000 real operational entries. This deployment is a concrete instance of the federated model: ALS data remain on-site and under facility control, while the ARIEL enhancement and search stack operates locally against them. Figure 1 shows the ingestion-to-enhancement-to-search flow as deployed in these prototypes. Together, the two deployments validate the federated approach and establish a working baseline for the cross-facility benchmarking planned in the next phase of the project.

CONCLUSION AND OUTLOOK

ARIEL operationalizes the DOE Genesis mission's call for shared, interoperable AI infrastructure by standardizing the logbook substrate on which downstream AI systems across DOE accelerator facilities will depend. The federated architecture ensures that each facility retains full sovereignty over its operational data while benefiting from a shared library of enhancement and search components and a common agentic retrieval layer. Early prototypes at the ALS demonstrate the end-to-end pipeline on a production-scale logbook, and the empirical results from CEBAF on embedding-model selection and controlled-vocabulary enhancement already provide actionable guidance for deployment decisions at any participating site. Immediate next steps include finalizing the Enhanced Entry schema and database technology, developing the initial suite of enhancement modules covering text embeddings and basic semantic enrichment, and deploying baseline keyword and semantic search across additional participating facilities. Longer-term work will focus on cross-facility benchmarking and quantitative retrieval metrics, systematic controlled-vocabulary evaluation across sites, hybrid retrieval studies combining dense and sparse signals, deeper agentic integration capable of multi-step reasoning over the logbook corpus, and a unified user-interface design developed in coordination with shared working groups.

ACKNOWLEDGMENTS

This material is based on work supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics under Contract No. DE-AC05-06OR23177, and by the Director of the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This document was prepared in part using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, Office of High Energy Physics HEP User Facility. Fermilab is managed by Fermi Forward Discovery Group, LLC, acting under Contract No. 89243024CSC000002.

REFERENCES

[1] C. Tennant, "A meta-analysis of the electronic logbook: 2012–2025", Jefferson Lab, Newport News, VA, USA, Tech. Note JLAB-TN-25-068, 2025.

[2] A. Sulc *et al.*, "Towards unlocking insights from logbooks using AI", in *Proc. IPAC'24*, Nashville, TN, USA, May 2024, pp. 3579–3582. doi:10.18429/JACoW-IPAC2024-THPR37

[3] A. Sulc *et al.*, "eLog analysis for accelerators: Status and future outlook", arXiv:2506.12949, 2025. doi:10.48550/arXiv.2506.12949

[4] C. Tennant, “CEBAF logbook controlled vocabulary enhancement evaluation: Summary and results”, Jefferson Lab, Newport News, VA, USA, Tech. Note JLAB-TN-26-024, 2026.

[5] A. Sulc, T. Hellert, and S. Hunt, “Unsupervised anomaly detection in ALS EPICS event logs”, arXiv:2509.13621, 2025. doi:10.48550/arXiv.2509.13621

[6] C. Tennant, J. Li, and S. Wang, “Machine-state embeddings as an operational coordinate system for accelerator operation”, arXiv:2604.05914, 2026. doi:10.48550/arXiv.2604.05914

[7] C. Tennant, “CEBAF logbook embedding model evaluation: Summary and results”, Jefferson Lab, Newport News, VA, USA, Tech. Note JLAB-TN-26-021, 2026.

[8] T. Hellert, J. Montenegro, M. Venturini, and A. Pollastro, “Domain-specific text embedding model for accelerator physics”, *Phys. Rev. Accel. Beams*, vol. 28, p. 044601, 2025. doi:10.1103/PhysRevAccelBeams.28.044601

[9] T. Hellert, J. Montenegro, and A. Sulc, “Osprey: Production-ready agentic AI for safety-critical control systems”, *APL Mach. Learn.*, vol. 4, p. 016103, 2026. doi:10.1063/5.0306302

[10] T. Hellert *et al.*, “Status of Osprey: A production framework for agentic AI in control systems”, presented at IPAC’26, Deauville, France, May 2026, paper MOP6322, this conference.