NeuralMesh™ AIDP in Healthcare and Life Sciences

Continuously current, permission-aware AI knowledge for genomics, drug discovery, pharmaceuticals, and clinical science, all hosted on NeuralMesh AIDP

Challenges

- Stale regulatory indexes create audit and compliance risk
- Consent withdrawal demands true deletion across all indexes
- GxP-validated pipelines multiply per team and per program
- Siloed research knowledge blocks cross-program discovery

Solution

NeuralMesh AIDP continuously vectorizes every document, dataset, and case record the moment it lands or is deleted with POSIX permissions enforced at query time.

Benefits

- Auto-purge on deletion
- Consent withdrawal honored natively
- POSIX ACLs enforce patient and program access at inference
- Superseded documents purged from the index automatically
- One GxP-validated platform replaces per-team RAG silos
- Continuously current indexes reflect every document change on arrival

AI-ready Healthcare and Life Sciences data without manual burden

NeuralMesh™ AIDP is a software and hardware add-on to NeuralMesh that is deployed from the **WEKA App Store**. It transforms how healthcare and life sciences (HCLS) teams operationalize AI factories, by automatically handling ingestion, chunking, embedding, vectorization, and permissions enforced retrieval across research and clinical data, without custom engineering.

Across HCLS AI programs in pharmaceutical regulatory and pharmacovigilance, clinical genomics, secure scientific collaboration, drug discovery, and cryo-EM, workloads stall at the data-preparation layer. **AIDP removes that wall.**

Why choose WEKA's AI Data Platform?

NeuralMesh AIDP deploys from the WEKA App Store without custom pipeline engineering. It resolves the data-preparation bottleneck at the platform layer by offering a high-performance enterprise grade parallel filesystem designed for AI-scale I/O. Every document, dataset, and sequence is vectorized on arrival, updated on change, and purged on deletion.

AIDP is designed for easy deployment in HCLS enterprises and AI factories across pharmaceutical regulatory and pharmacovigilance, clinical genomics, scientific collaboration, drug discovery, and cryo-EM. AIDP extends NeuralMesh from a high-performance storage system into a complete solution that stores research and clinical data at scale, keeps it current, enforces data access controls at query time, and delivers AI outcomes directly.

The problem AIDP solves in HCLS

Most HCLS AI programs fail to reach production because the data infrastructure underneath isn't built for continuous AI operations. For genomics pipelines, regulatory RAG systems, or pharmacovigilance workflows to run cleanly, data must be parsed, chunked, embedded, and indexed.

As clinical documents update, genomic databases refresh, and regulatory filings change, the entire pipeline must stay in sync. **Without AIDP,** this means:

- **Per-project pipeline debt:** Every genomics pipeline, pharmacovigilance workflow, or regulatory RAG system requires a custom pipeline built from scratch, maintained indefinitely
- **Stale research and regulatory context:** When clinical documents, genomic databases, or regulatory filings update, downstream AI systems don't automatically reflect those changes, creating interpretations built on outdated evidence due to stale embedded data
- **Research velocity loss:** Every new program or study rebuilds the same ingestion and vectorization work. Time spent on infrastructure is time not spent on discovery
- **Slow time-to-production:** Selecting, integrating, and validating an AI data stack delays the first AI outcome by months. Clinical and research programs can't wait quarters
- **Compliance and security exposure:** Standard pipelines strip permissions and access lists at ingest. In regulated HCLS environments, this blocks AI deployment until security and compliance teams can verify governance holds at every layer

AIDP offers two fundamental benefits

1. Continuous, automated data preparation simplifies workflows

AIDP watches WEKA NeuralMesh filesystems in real time. As clinical documents arrive or update, are deleted, or regulatory filings change, AIDP automatically handles the full preparation pipeline, without the need for self-created batch jobs or manual intervention. AIDP:

- Detects every file create, update, and delete via continuous monitoring without a batch scheduler, polling, or reindexing cycle
- Parses and chunks clinical documents, genomic datasets, and regulatory filings as they arrive
- Embeds file permissions, UID, and GID metadata at ingest. If a researcher or clinician can't see the source file, they can't see the vector. No middleware, no compliance gap
- Keeps vector indexes synchronized with the live state of your research and clinical data; file deletions, document revisions, and record updates propagate automatically

That continuous data readiness is what makes the second benefit possible: delivering AI outcomes directly to the researchers and clinicians who need them, all without additional engineering overhead.

2. Ready-to-consume data generates better AI outcomes

Your research and clinical teams should be focused on discovery and patient outcomes, not babysitting ETL jobs. AIDP delivers AI outcomes directly to the researchers and clinicians who need them, without requiring engineering teams to deploy or manage AI frameworks:

- **Document Intelligence** surfaces new content across genomic datasets, regulatory submissions, and clinical trial archives the moment it arrives, with results limited to what the requesting user is authorized to see
- **Semantic Search** enables natural language queries over research reports, clinical trial documents, and regulatory filings stored on NeuralMesh and accessed via browser UI or API, backed by NVIDIA NIMs, all managed by AIDP and supported by WEKA

How AIDP works: platform architecture

AIDP is a full-stack system of multiple integrated services that handles the complete journey from raw research and clinical data to AI-ready outcomes. Key outcomes include:

Theme	Without AIDP	With AIDP
Data Freshness	AI projects use stale and outdated data when files change	Create, update, and delete events automatically propagate to the vector store
	Withdrawn SOPs, rescinded labels, and retracted datasets linger in the index	Superseded and deleted documents are purged from the index automatically
Compliance & Governance	AI responses ignore file-level permissions and governance procedures	POSIX ACL enforced from NeuralMesh through to every AI response
	Consent withdrawal requires manual vector purge, creating compliance risk	File deletion propagates automatically to the vector index, honoring HIPAA and GDPR right-to-be-forgotten natively
	Each RAG pipeline requires separate GxP validation, multiplying the compliance burden across programs	One validated platform serves all programs, consolidating the GxP validation surface
Architecture & Maintenance	Custom data pipelines are built and maintained for every project	One platform continuously prepares all NeuralMesh data for AI
	IT teams manage AI frameworks directly	Semantic search and more are delivered out of the box
Visibility & Control	Siloed pipelines offer little or no operational visibility across pipelines	Live queue monitoring, status dashboards, and role-based governance in one UI

AIDP workloads in HCLS

Pharma regulatory, PV & clinical-trial knowledge

RAG and GraphRAG allow regulatory affairs, QA, and PV teams to query massive, constantly changing document collections. These queries surface approved language, detect PV signals, and check compliance against live GxP guidance. However, these queries are only valuable if the underlying data is current and correctly permissioned. And because regulatory documents are continuously versioned and superseded, and batch reindexing lags, citing a withdrawn SOP or rescinded label often results in a compliance event and, ultimately, a direct patient-safety risk.

Traditional data storage architectures force per-team RAG pipelines to undergo separate GxP validation, creating an exponential compliance burden as programs scale. The entire workload is dependent on strict data access control that RA, QA, and PV teams simply can't enforce without the right infrastructure.

NeuralMesh AIDP solves this problem by:

- Purging withdrawn SOPs and rescinded labels from the index **automatically**
- Reflecting PV safety updates **continuously**, eliminating batch reindex gaps
- Replacing per-pipeline GxP revalidation across programs by validating data **one time** at the platform level instead of repeatedly over time

With the unified, constantly updating support from NeuralMesh AIDP, research and regulatory teams are no longer bottlenecked by non-compliance and manual validation of the high-risk regulatory process, allowing them to focus on research and discovery.

Genomics: clinical variant interpretation & rare-disease diagnosis

Using genomic foundation models to match patient variants against candidate genes, clinical genomics teams move from variant calling through annotation to interpretation. This work is grounded in ACMG and AMP guidelines, continuously updated references (like ClinVar, OMIM, and gnomAD), and prior case histories. RAG over that combined corpus is the ideal architecture for rare-disease diagnosis, it is only effective if the index reflects current science and only compliant if it honors patient consent.

Patient genomes are PHI held under consent scoping, so a withdrawal must delete both the files and their vectors to satisfy HIPAA and GDPR right-to-be-forgotten. In legacy approaches that purge is manual and error-prone, and the custom RBAC studies bolt on per consent cohort is exactly the middleware a CISO may refuse to clear for clinical use.

NeuralMesh AIDP solves this problem by:

- Propagating consent withdrawal as automatic vector deletion, honoring HIPAA and GDPR right-to-be-forgotten natively

- Reflecting variant reclassifications the moment they land, preventing interpretation on outdated evidence
- Enforcing file permissions (POSIX ACLs) per study and consent cohort, blocking cross-patient vector leakage

With a genomic knowledge layer that NeuralMesh AIDP keeps current and consent-aware, clinical teams trust every interpretation to the live state of the science and spend their time on diagnosis, not index maintenance.

Drug discovery: protein structure & small-molecule screening

Drug discovery runs two AI pipelines at once. First, HPC pushes millions of tiny files through structure prediction and small-molecule screening, generating millions of candidate files that pile up from one run to the next that are unsearchable. Second, agentic research layer reasons over literature, patents, assay results, ELN entries, and predicted-structure metadata to prioritize targets and generate hypotheses. However, this is only effective if the index stays current and keeps each program's IP isolated.

Because legacy systems cannot easily isolate data on a shared index, prior targets and assays remain siloed. This forces every new program to rebuild its own search and governance stack from scratch, delaying cross-program discovery and forcing teams to recreate existing work. This leads to expensive, repetitive GPU recompute cycles and longer discovery timelines.

NeuralMesh AIDP solves this problem by:

- Isolating therapeutic programs in a shared index with program-level File level Permissions, protecting IP across teams
- Reflecting new assay results and ELN entries in the agentic assistant the moment they land
- Enabling long-context queries across screening corpora without recompute

NeuralMesh AIDP governs the knowledge layer, allowing discovery teams to search everything they have folded, screened, and measured across every campaign and keep their attention on the next candidate.

Cryo-EM: single-particle analysis & structure determination

Cryo-EM instruments generate thousands of micrographs daily, feeding deep-learning particle pipelines that resolve 3D atomic models from raw density maps. Because the core pipeline is a throughput and GPU problem, the AI knowledge opportunity sits alongside it. By making every dataset, structure, and result searchable across the archive, researchers can instantly detect structural overlaps across different runs, match novel targets to successful historical sample preparation conditions, and track dynamic protein conformations.

In legacy environments, manual cataloging cannot keep pace with the massive flood of petabyte-scale instrument data. To enable cross-instrument comparisons, organizations need a shared space to query all active imaging data, a requirement that makes purging rejected runs even more critical. Without this clean-

up, lingering junk datasets clutter the unified storage system and leave the shared archive progressively harder to search as it scales.

NeuralMesh AIDP solves this problem by:

- Tracking every micrograph and dataset on landing, replacing manual cataloging
- Purging discarded datasets from the index automatically, keeping archive searches clean
- Enabling semantic and similarity search across the full structural archive, in one namespace spanning every instrument

With NeuralMesh carrying the throughput and NeuralMesh AIDP governing the archive, structural biology teams search every dataset they have collected and every structure they have solved, without the need for hand-cataloging.

Bioinformatics HPC: sequence alignment & simulation

High-throughput bioinformatics pipelines generate massive, complex datasets of sequence alignments, variant calls, and molecular dynamics simulations. Applying agentic RAG over this complex corpus allows researchers to search and compare past runs, instantly linking newly detected gene variants to historical mutation patterns or physical molecular behaviors.

In legacy approaches, outputs accumulate faster than teams can index them, locking prior runs in isolated project directories where they are functionally invisible, while outdated or retracted datasets linger to quietly pollute downstream queries with bad evidence. This leaves teams with fragmented, untrustworthy search results and forces researchers to repeat expensive alignment and simulation runs that have already been conducted.

NeuralMesh AIDP solves this problem by:

- Indexing simulation outputs and alignment results on landing, making them immediately searchable
- Purging retracted runs from search results automatically, keeping downstream queries clean
- Governing all bioinformatics outputs across teams and programs from one platform, with permissions enforced at query time

With NeuralMesh running the pipelines and NeuralMesh AIDP governing what they produce, research teams find and reuse results from any run instead of rebuilding the same search stack for all projects.

Secure scientific collaboration & knowledge sharing

Research teams spread across therapeutic areas, studies, and partner sites need to be able to discover and reuse prior scientific work without exposing sensitive data. Applying permission-aware RAG over diverse repositories such as experimental results, ELN entries, and historical datasets creates a unified, searchable knowledge layer that dynamically enforces IP, PHI, and consent boundaries at the moment of query. However, this collaborative layer is only effective if data is uniformly cataloged across all programs and only viable if compliance controls are natively integrated at the storage layer.

Traditional legacy approaches force a choice between sharing data and protecting it, allowing strict compliance boundaries to block broad scientific discovery. By trapping teams in isolated project silos and preventing them from safely collaborating across adjacent programs, traditional architectures forced researchers to recreate expensive analyses, bury critical institutional insights, and drive up research costs because teams cannot safely learn from discoveries made in adjacent programs.

As a result, teams rebuild analyses that already exist in adjacent programs, institutional knowledge sits inaccessible, and scientists cannot benefit from discoveries made outside their own project. This leads to both inefficient and more expensive research outcomes. Traditional architectures push access control into separate middleware, which IT security teams may never approve for sensitive research.

NeuralMesh AIDP solves this problem by:

- Enforcing PHI restrictions, IP boundaries, and consent scoping at query time, with no separate middleware
- Surfacing prior findings across therapeutic areas and sites, eliminating redundant analysis
- Tying every retrieval to POSIX-enforced access, so audit lineage is native rather than bolted on

With one governed knowledge layer from NeuralMesh AIDP, scientists build on each other's work across programs instead of rediscovering it, keeping their focus on new science.

Ready to get started?

Filesystem level permissions (POSIX-ACL)

HCLS AI programs stall at the access-control layer. AIDP resolves this natively, no custom middleware required.

- POSIX permissions are already the standard in regulated Linux environments
- AIDP enforces them at inference time, not as a post-retrieval filter
- Structurally, not by policy, PHI-scoped queries can't surface consent-restricted vectors
- Audit lineage is native: every retrieval event is tied to the enforced permission set
- Result: the same access control model the CISO already manages governs the AI

AIDP turns "the CISO blocked AI" into "the CISO approved AI."

0 hours	1	Auto
Deletion-to-searchable latency vs. batch purge window	Validated platform replaces per-program RAG pipeline silos	Consent withdrawal honored natively; deletion propagates to vector index

Ready to bring WEKA AIDP to your HCLS AI program?

Visit weka.io/contact to schedule an architecture briefing for your research or clinical AI program.

→ **See NeuralMesh AIDP in Action**