

AI-Ready by Design: How Adimab's 15+ Years of Data Integration Makes Every Antibody Program Smarter Than the Last

ADIMAB

Bingjie Ouyang, Kirt Johnson, Chris Sumner, Cody Williams, Jason Aze, Drew Prescott, Max Vasquez · Adimab, LLC
 PEGS Boston 2026 · Session C6A: ML and Digital Integration in Biopharmaceuticals

INTRODUCTION

Antibody discovery generates interconnected data across library construction, selection, clonal characterization, and therapeutic lead delivery. In many organizations, this information is distributed across separate LIMS, ELN, and spreadsheet systems, limiting cross-program analysis.

Retrospective queries such as "Get all thermal stability data for Fab samples derived from Adimab 5.0 library selections" or "How do selection protocols compare across target classes?" require substantial manual data reconciliation when systems are not integrated.

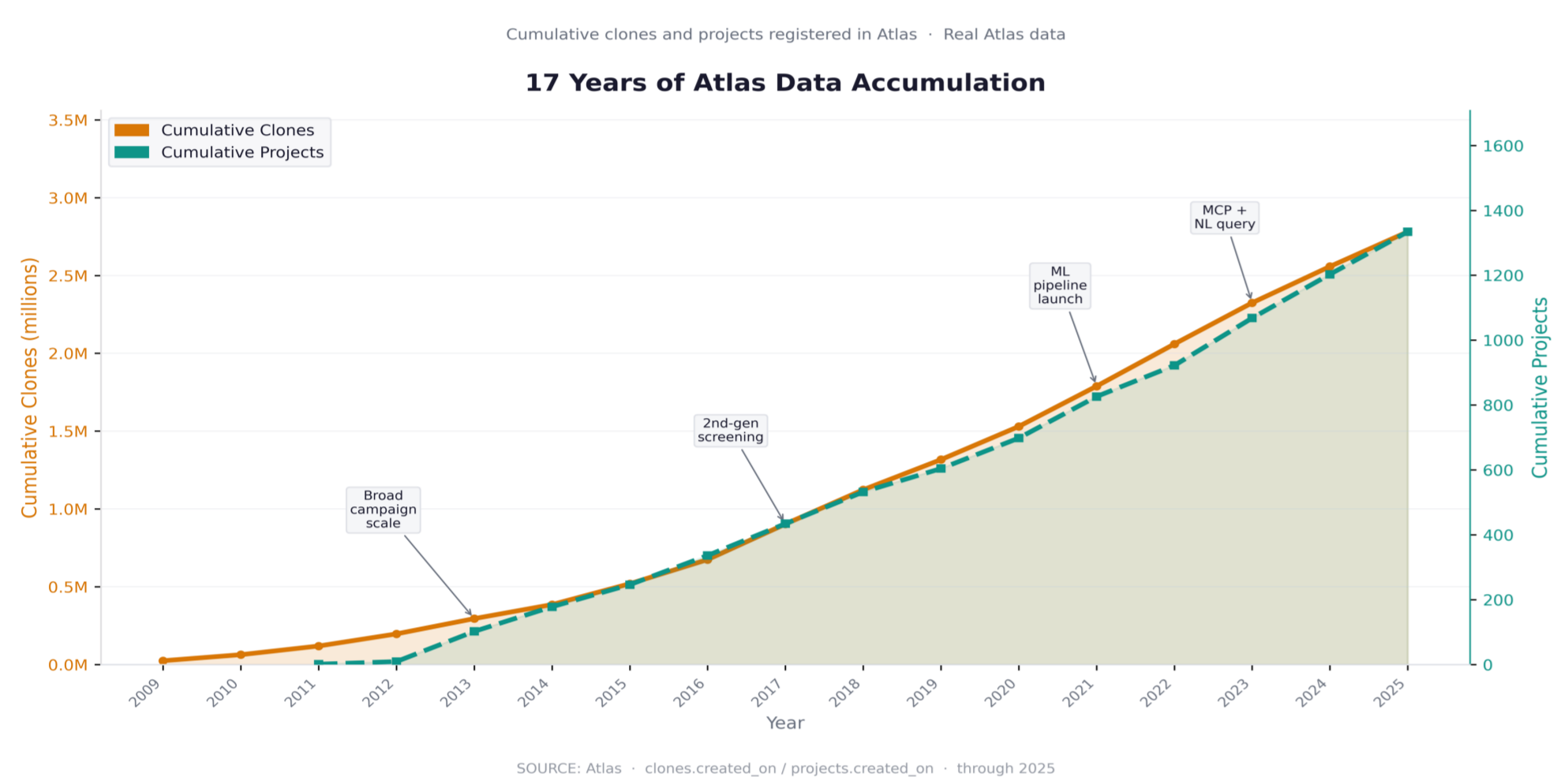
Effective application of machine learning to antibody development depends on structured, linked experimental data rather than algorithmic sophistication alone. Here we describe Atlas, Adimab's integrated data platform, and how 15+ years of consistent data capture provide a foundation for machine learning models that depend on large-scale, linked experimental data, a complement to structure-based computational methods in antibody discovery.

OBJECTIVES

- Illustrate how structured, linked operational data supports downstream machine learning applications
- Demonstrate end-to-end clone traceability across 1,431 programs and 2.85M clones
- Present natural language database querying via the Model Context Protocol (MCP)
- Quantify the relationship between accumulated experimental measurements and ML prediction throughput

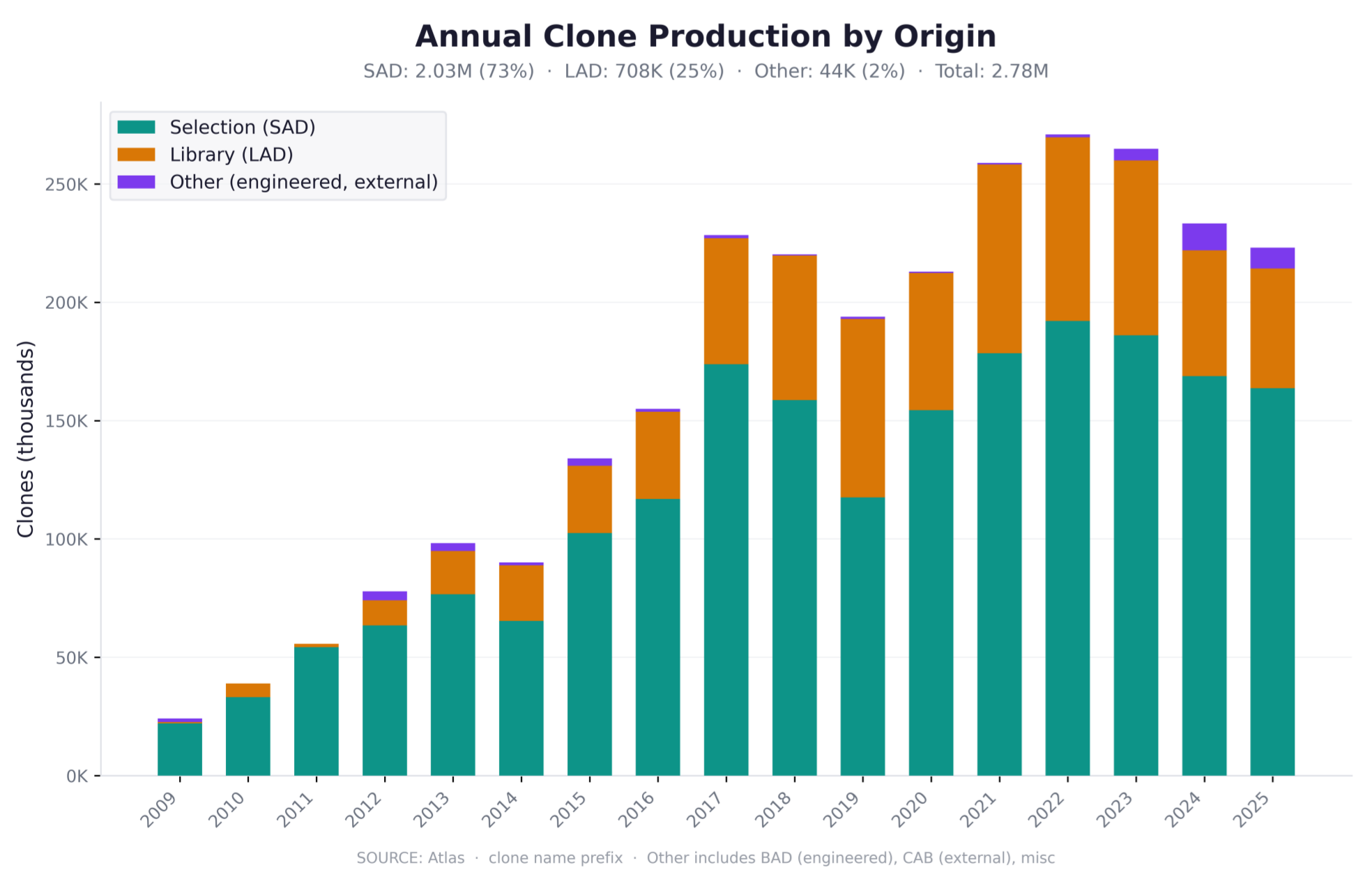
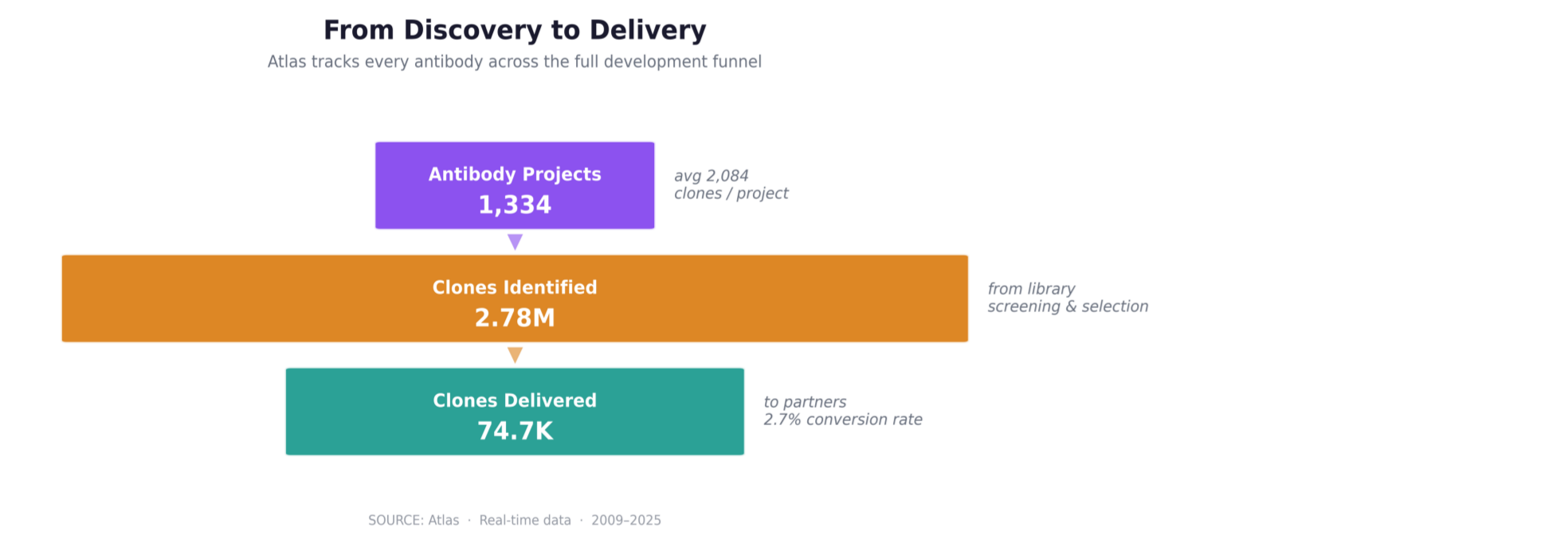
DATA SCALE: 15+ YEARS OF GROWTH

Atlas has captured antibody discovery data continuously since 2009. The cumulative growth of projects and clones reflects the expanding scope of Adimab's discovery operations, with each new program contributing to a shared, queryable dataset.



FROM DISCOVERY TO DELIVERY

Atlas tracks clone progression through the full development pipeline: from project inception through clone identification and delivery to partners. Conversion rates at each stage are quantifiable from the linked dataset. Clone origins include selection-acquired (SAD, 73%), library-acquired (LAD, 25%), and other sources including engineered variants and external clones (2%).

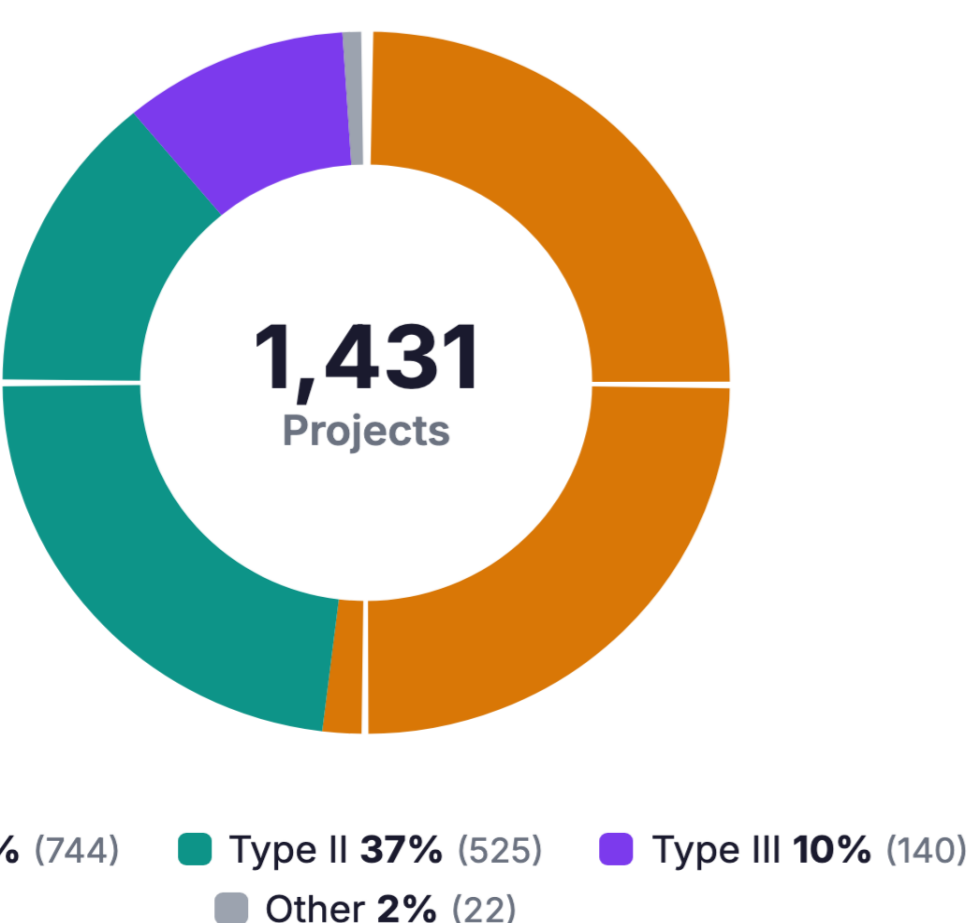


PROJECT DIVERSITY

Atlas captures data across the full range of Adimab's discovery programs: 1,431 projects spanning Type I (52%), Type II (37%), Type III (10%), and other formats.

This breadth of the data spans multiple project types, therapeutic targets, and antibody formats. This contributes to the diversity of the training dataset. Models trained on data from a single project type or target class may not generalize well; Atlas's cross-program dataset mitigates this limitation.

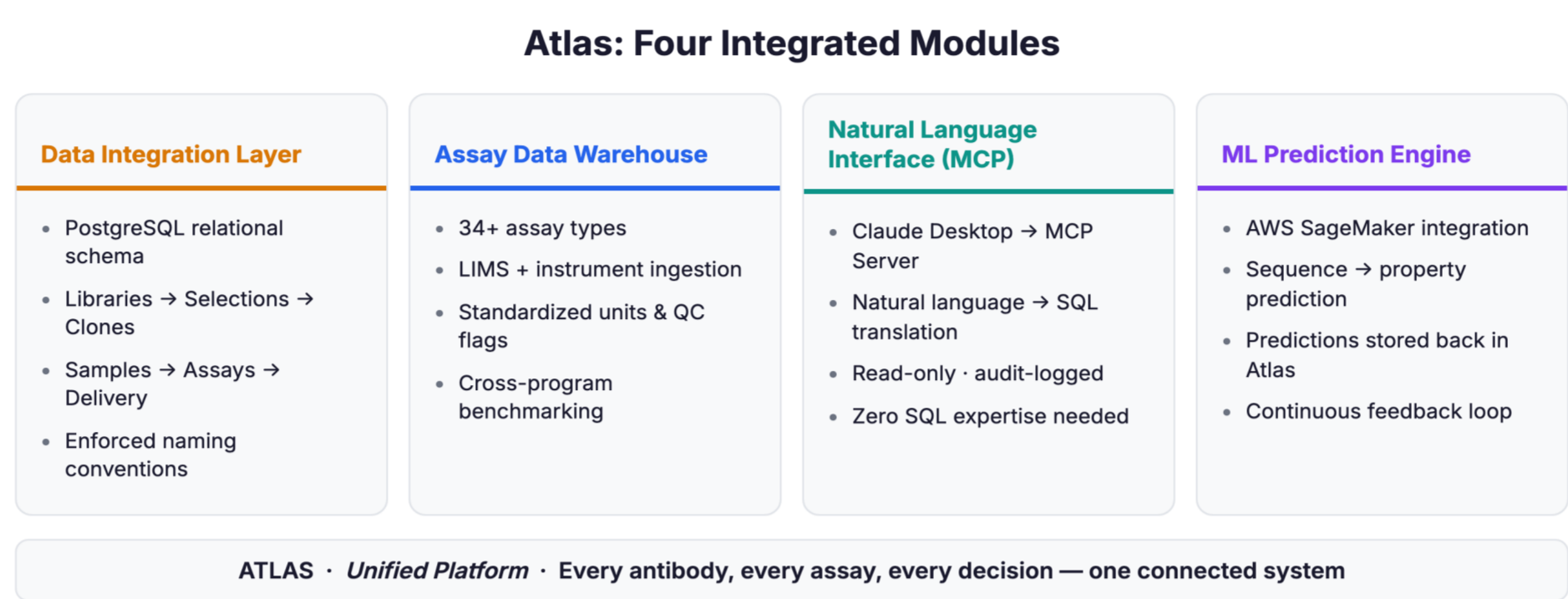
Project Type Distribution



ATLAS: FOUR INTEGRATED MODULES

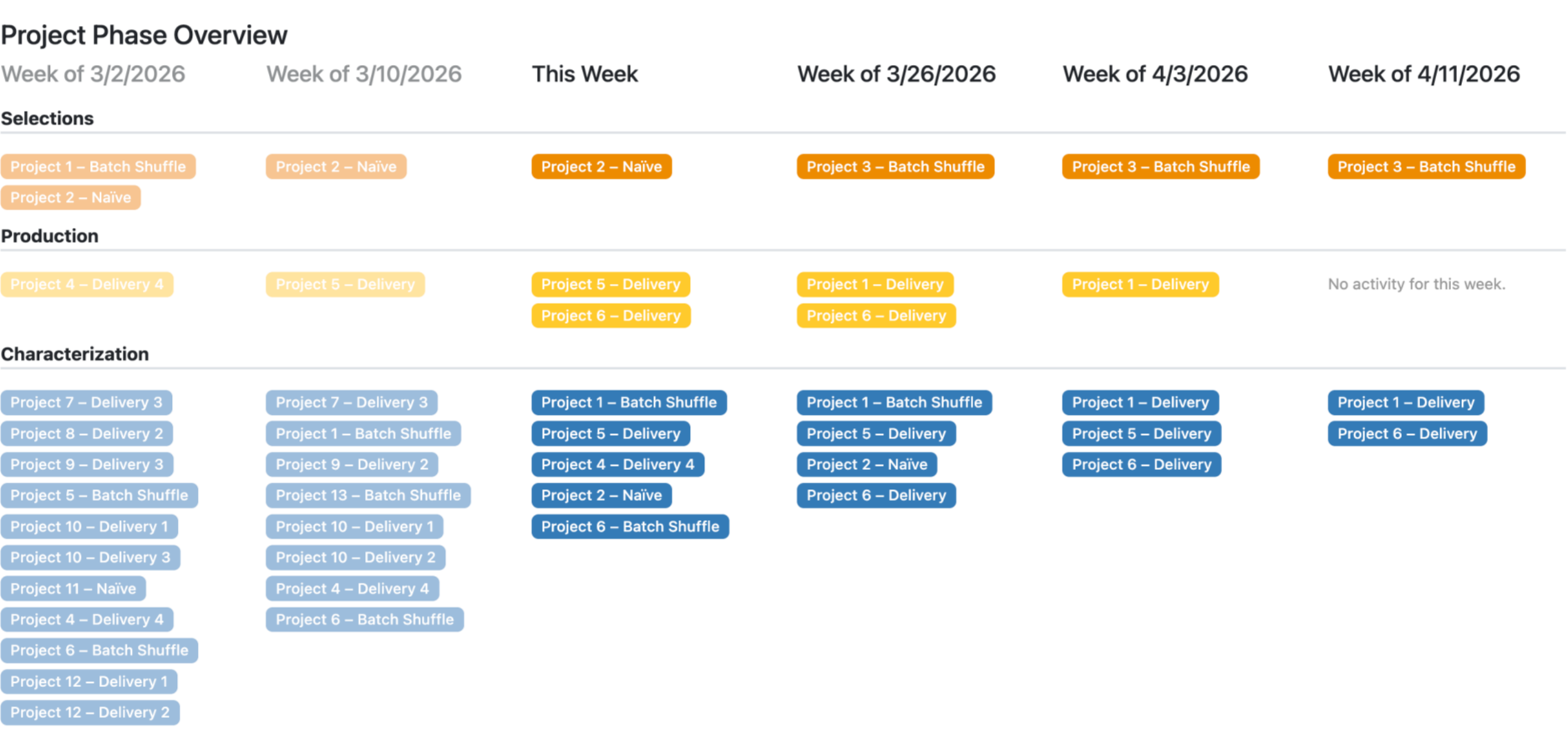
Atlas is not a conventional LIMS. It is a purpose-built relational data system designed to support scientific decision-making by linking experimental data across the full antibody discovery workflow.

- Data Integration Layer — PostgreSQL relational schema enforcing naming conventions and entity relationships at creation time
- Assay Data Warehouse — 34+ assay types with standardized units, QC flags, and cross-program benchmarking
- Natural Language Interface — MCP-based querying enabling plain-English access to the full database
- ML Prediction Engine — AWS SageMaker integration for sequence-to-property prediction



ATLAS USER INTERFACE

The Atlas web interface provides each researcher with a personalized view of active projects, resource allocation, recent productions, and selection campaigns. Because scientists use Atlas as their primary operational tool, data is captured at the point of origin. They are structured, linked, and complete without retroactive curation.



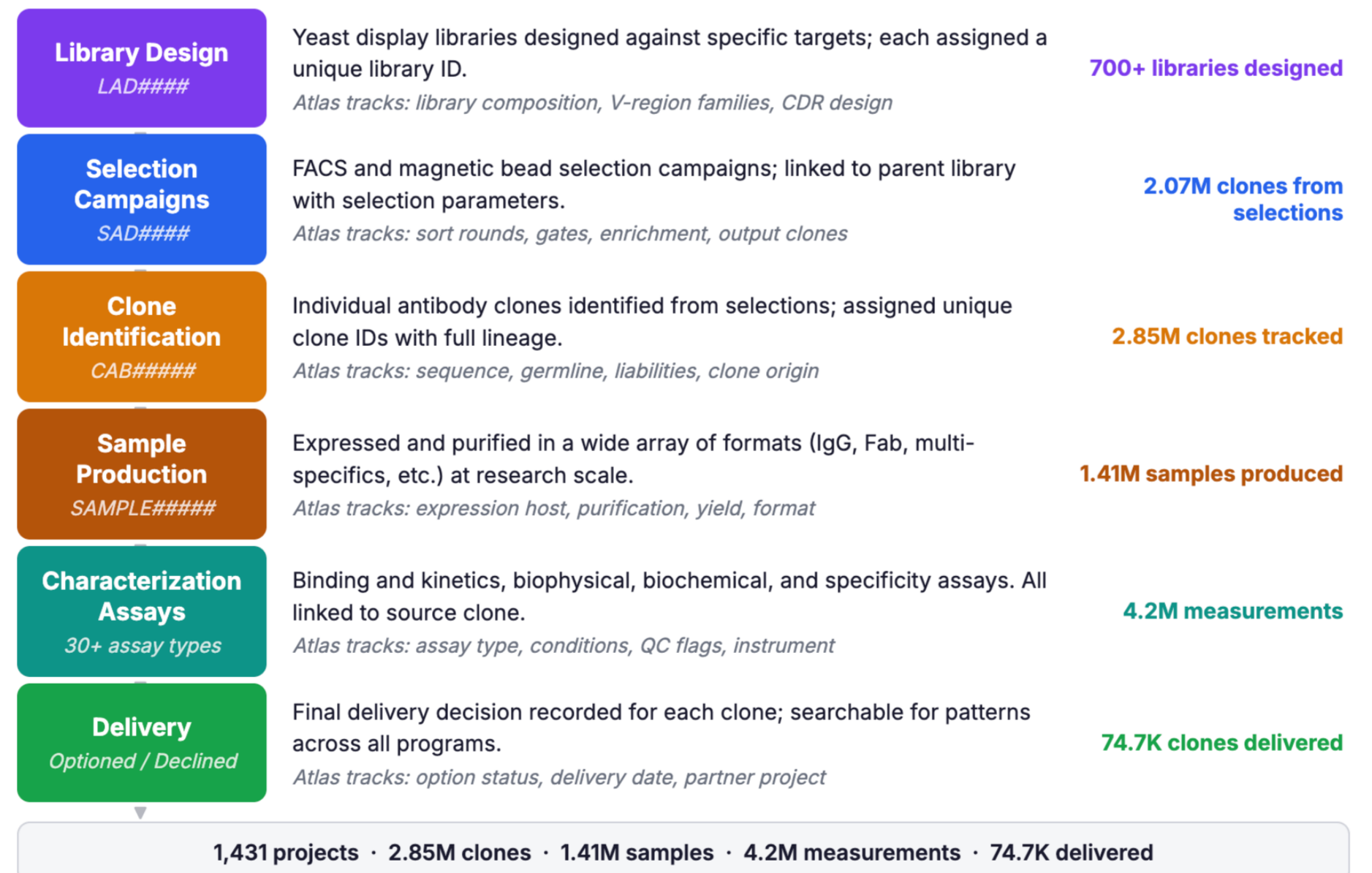
THE CONNECTED CLONE JOURNEY

A central feature of Atlas is end-to-end lineage tracking. Each clone record is automatically linked to its source library, selection history, characterization data, and delivery status at the time of creation.

This traceability is bidirectional: users can trace forward from a library to all derived clones, or backward from a delivered clone to its original library composition. Six stages are tracked: Library Design, Selection Campaigns, Clone Identification, Sample Production, Characterization Assays, and Delivery.

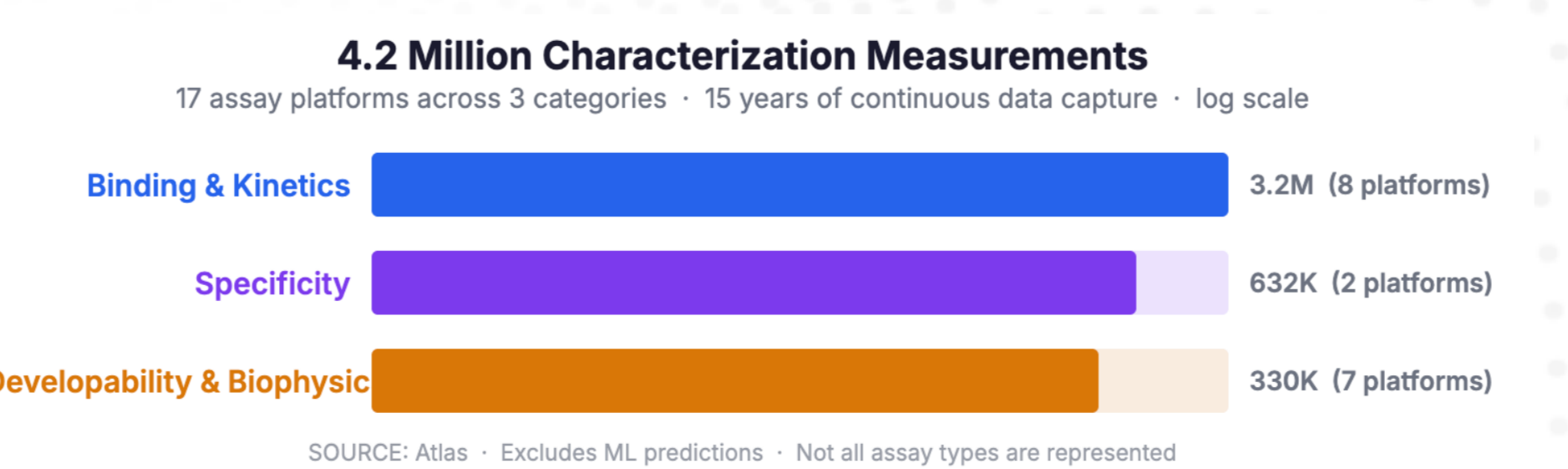
The Connected Clone Journey

Six stages. One connected record.



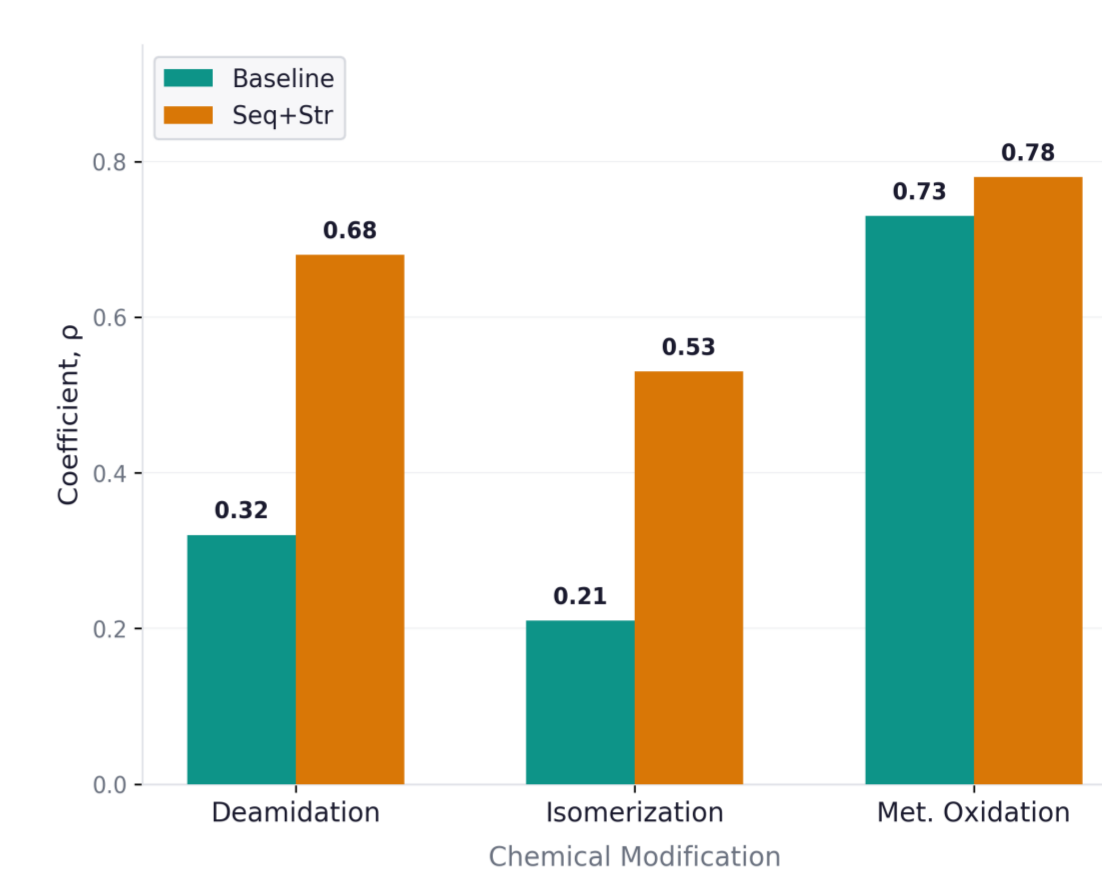
THE CHARACTERIZATION DATA WAREHOUSE

Atlas contains 4.1 million characterization measurements across 16 assay platforms, each linked to its source clone and project. This structured data warehouse supports cross-program analysis: HIC retention time predictions, developability assessments, and other ML applications draw on the full history of measurements accumulated over 15 years of routine operational use.



ML ENABLED DEVELOPABILITY PREDICTION

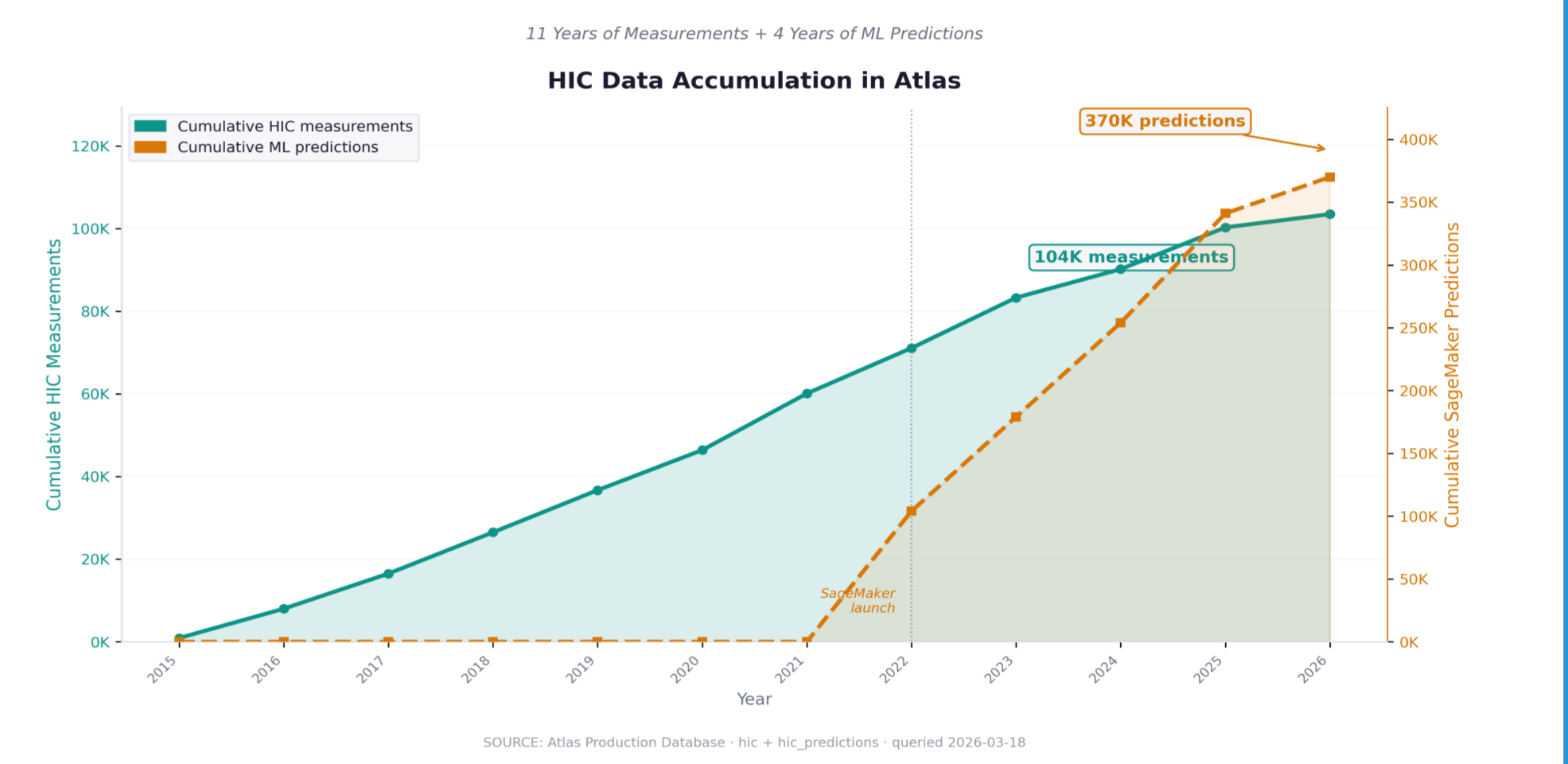
Sequence and Structure Improve Prediction Score



Atlas warehouses 11,000+ site-specific accelerated stress measurements across 700+ antibodies, enabling ML models that combine sequence motifs with structural context. Prediction scores (p) improve 2–2.5x over baseline for deamidation and isomerization liabilities.

HIC retention time is a key developability indicator reflecting antibody hydrophobicity. Atlas's linked sequence and characterization data, with 103K experimental measurements accumulated over 11 years, provided the training foundation for a predictive model deployed on AWS SageMaker in 2022.

Predicted values are stored in Atlas alongside experimental results, enabling direct comparison and iterative model refinement. The integration of predictions back into the same system that generated the training data creates a closed feedback loop: each experimentally validated prediction improves future model iterations.

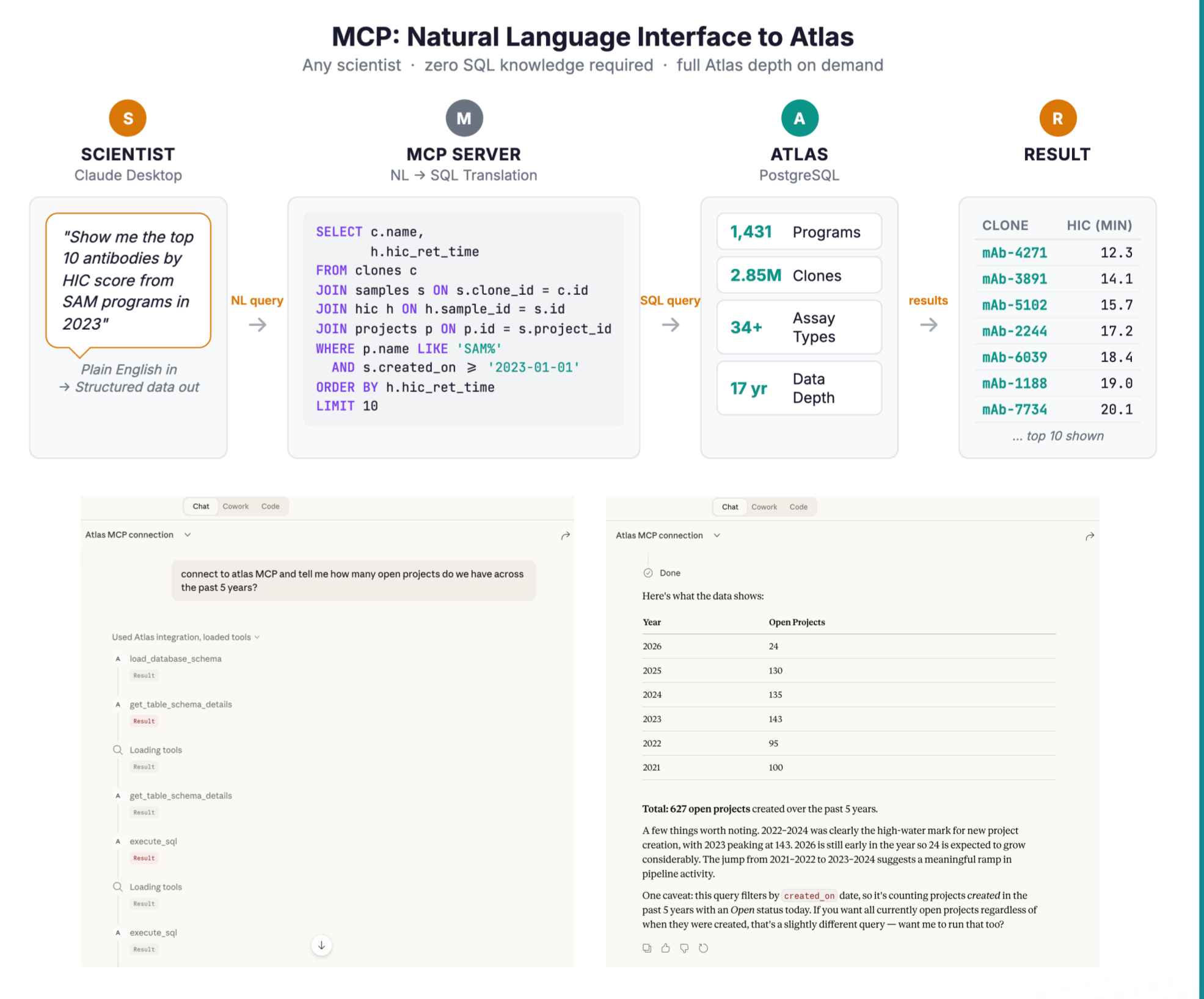


NATURAL LANGUAGE ACCESS VIA MCP → MODEL PERFORMANCE

Atlas is accessible via the Model Context Protocol (MCP), which translates natural language queries into validated SQL and returns structured results. This enables researchers without database expertise to query 17 years of linked experimental data directly.

The query pathway is: Claude Desktop → Atlas MCP Server → SQL generation and validation → structured results. All queries are read-only and audit-logged. For example, a researcher can ask "What are the HIC retention times for clones from project X?" and receive tabulated results without writing SQL.

The utility of this interface depends directly on the consistency and completeness of the underlying data model. These qualities result from Atlas's long-term use as an operational system.



CONCLUSIONS

- 1 Data architecture decisions precede and constrain ML capability
- 2 Schema and workflow design choices made in 2010 enabled machine learning applications not deployed until 2022. Retrospective integration of fragmented data systems is substantially more difficult than prospective design.
- 3 Domain-specific relational models support cumulative analysis
- 4 Atlas's schema is organized by scientific workflow rather than sample type, and links 1,334 programs across 17 years. Each new project is immediately queryable in the context of all prior data.
- 5 Natural language interfaces lower the barrier to institutional data access
- 6 MCP enables researchers to query the full Atlas database without SQL expertise, making 17 years of linked experimental data accessible to all team members.
- 7 ML prediction throughput scales with accumulated experimental data
- 8 SageMaker-based predictions exceed experimental measurement throughput by approximately 10x (104K predictions vs. 11K measurements per year). Predicted values are stored alongside experimental data, supporting iterative model evaluation.

Contact Adimab

If you are interested in partnering with Adimab, please reach out to our Business Development department at bd@adimab.com. The QR code on the right links to Adimab posters and other resources.

