

Electronic Laboratory Notebook (ELN) Data Archiving Enabled by ASM Format

Sreeni Yetukuri, Christian Del Valle, Vinny Antonucci
Merck & Co., Inc., Rahway, NJ, USA

Fall 2024 Allotrope Connect Meeting

Electronic Laboratory Notebook (ELN)

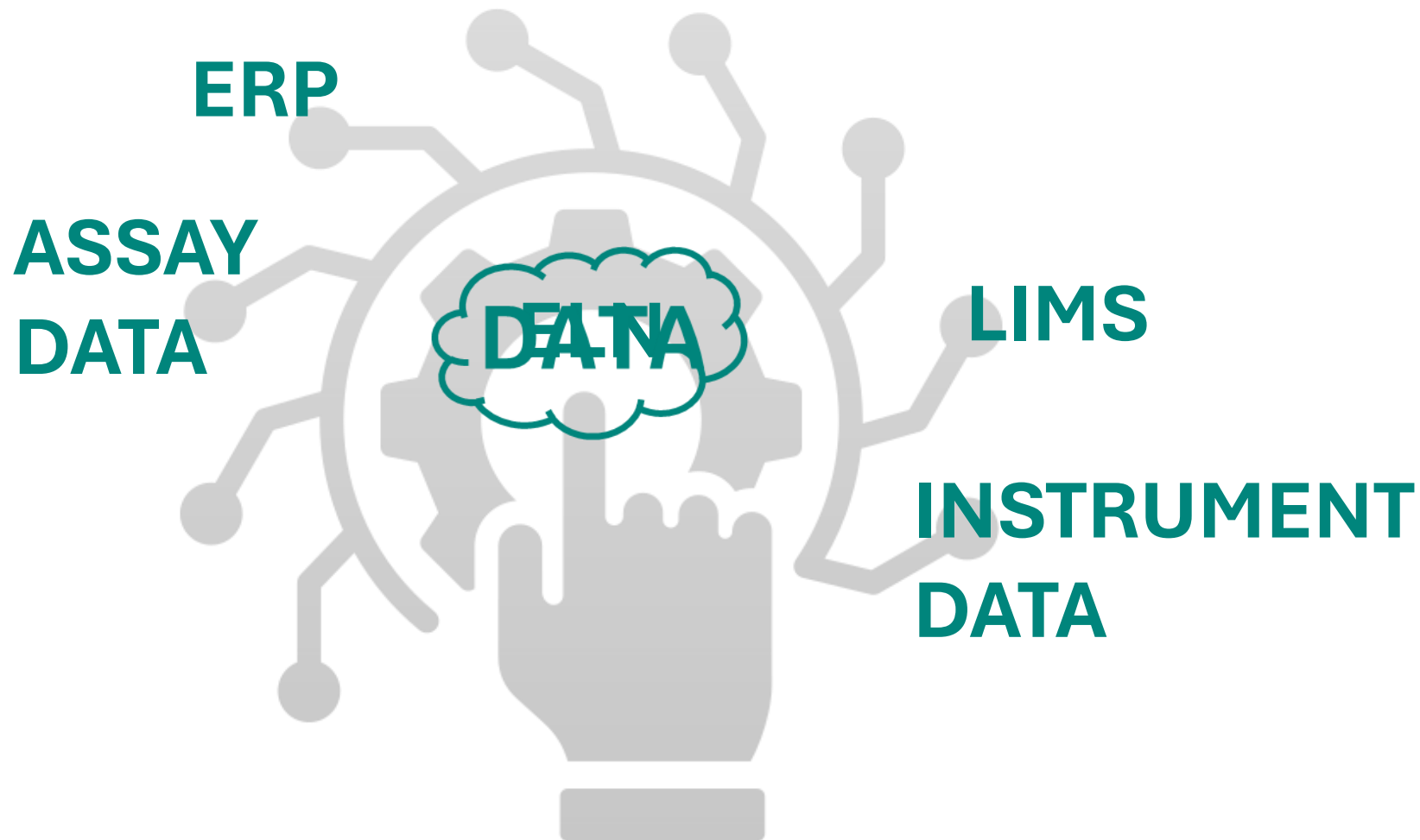
Digital platform for recording, managing, and sharing research data and experimental procedures. ELN applications have served as the “hub” of R&D data for decades, but the role of ELN is evolving...

Benefits of ELNs

- Data organization
- Data access
- Data security
- Facilitate collaborations
- Version control
- IP protection
- Facilitate compliance

Key Features of ELNs

- Electronic data capture
- Data structure & context
- Search & data mining
- Integration & automation
- Configurable workflows



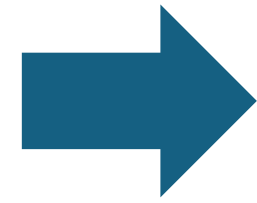
Electronic Notebook Modernization

From



CambridgeSoft[®]
Life Science Enterprise Solutions

- >20 years use
- >30TB data
- 4 Key configurations
- Aged technology



To

 **signalsNotebook**

- Modern cloud-native app
- Highly configurable to meet workflows
- Help achieve a data-centric ecosystem
- The app is a tool, not a strategy, so patterns for standardizing template building blocks are critical to success



Application-Neutral Digital Archive in the Cloud

Why Archive Electronic Notebook Data ?



Compliance

Help meet quality & regulatory expectations



Intellectual Property

Protect innovations & experimental results



Knowledge Preservation

Provide a historical record of research activities



Data Reliability

Ensure provenance & reproducibility of research



Data reusability

FAIRify data for use in analytics, modeling, etc

Data Migration Principles Used

- ✓ Adherence to data lifecycle & retention and data integrity policies
- ✓ Consistent with evolving reference architecture for R&D ecosystem
- ✓ Pursue vendor agnostic solutions for data storage and access
- ✓ Deliver expected user experience for consumption of migrated data
- ✓ Manage total cost of ownership of solution



Design principle

No data silos please ! Target a dynamic, source-independent "data product" for experiments

High Level Requirements Summary

Universal Requirements

- Ability to maintain appropriate data authorization controls
- Ability to easily search for any data type using a single search operation across multiple sources.
- Data must be able to be presented in a human readable format.
- Durable references must be maintained (images, instrument data, other experiments, etc.).
- Data should be migrated incrementally over time vs. all-at-once.
- Data must be stored in a vendor-neutral format.

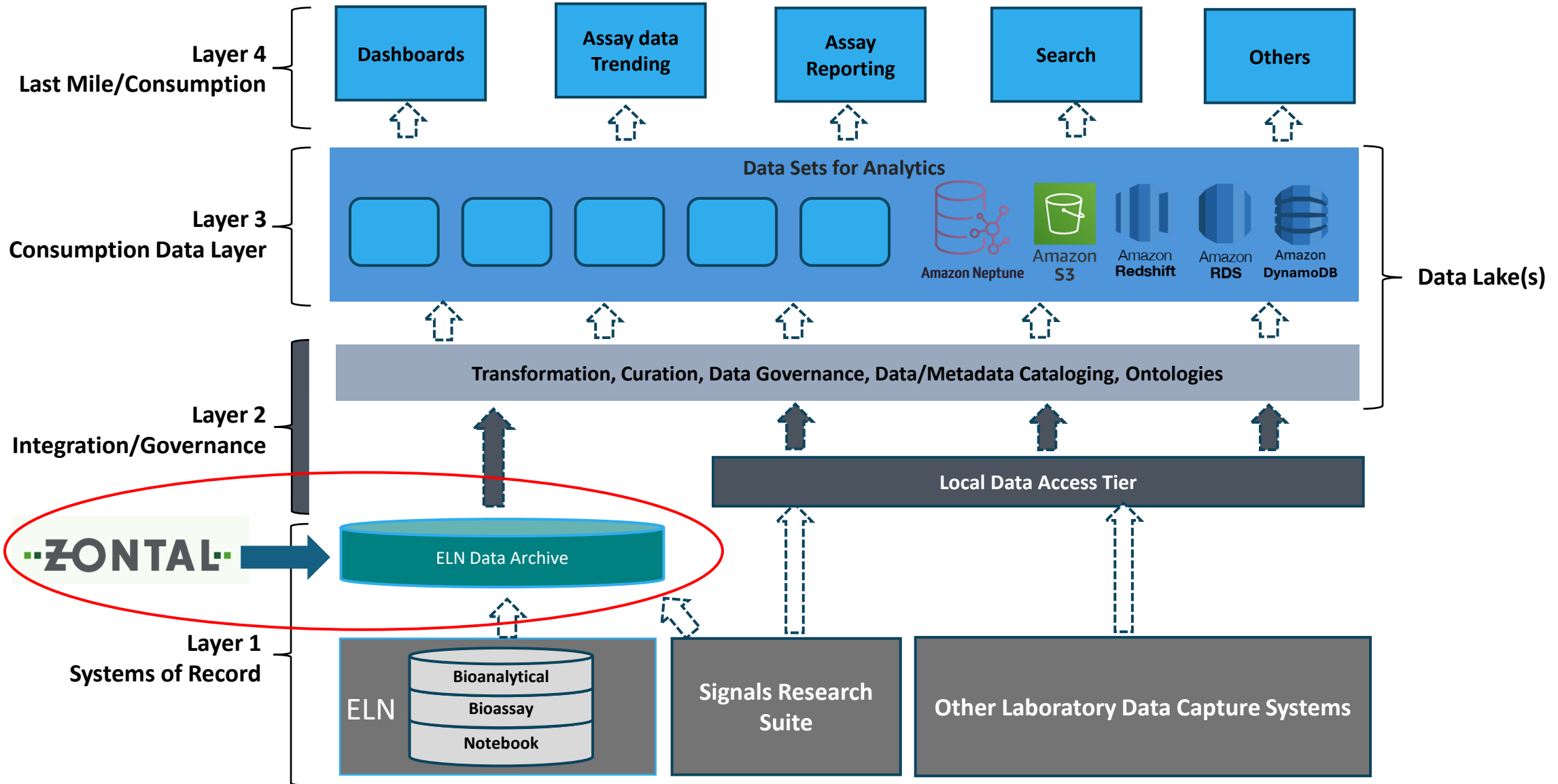
Notebook Archive

- Ability to access experiment in its original form.
- Ability to create/access PDF of experiment quickly.
- Ability to export legacy notebook data including audit trails & signatures.
- Ability to meet data lifecycle & legal requirements.
- Solution must meet GxP 21CFR part 11 & 58 compliance.

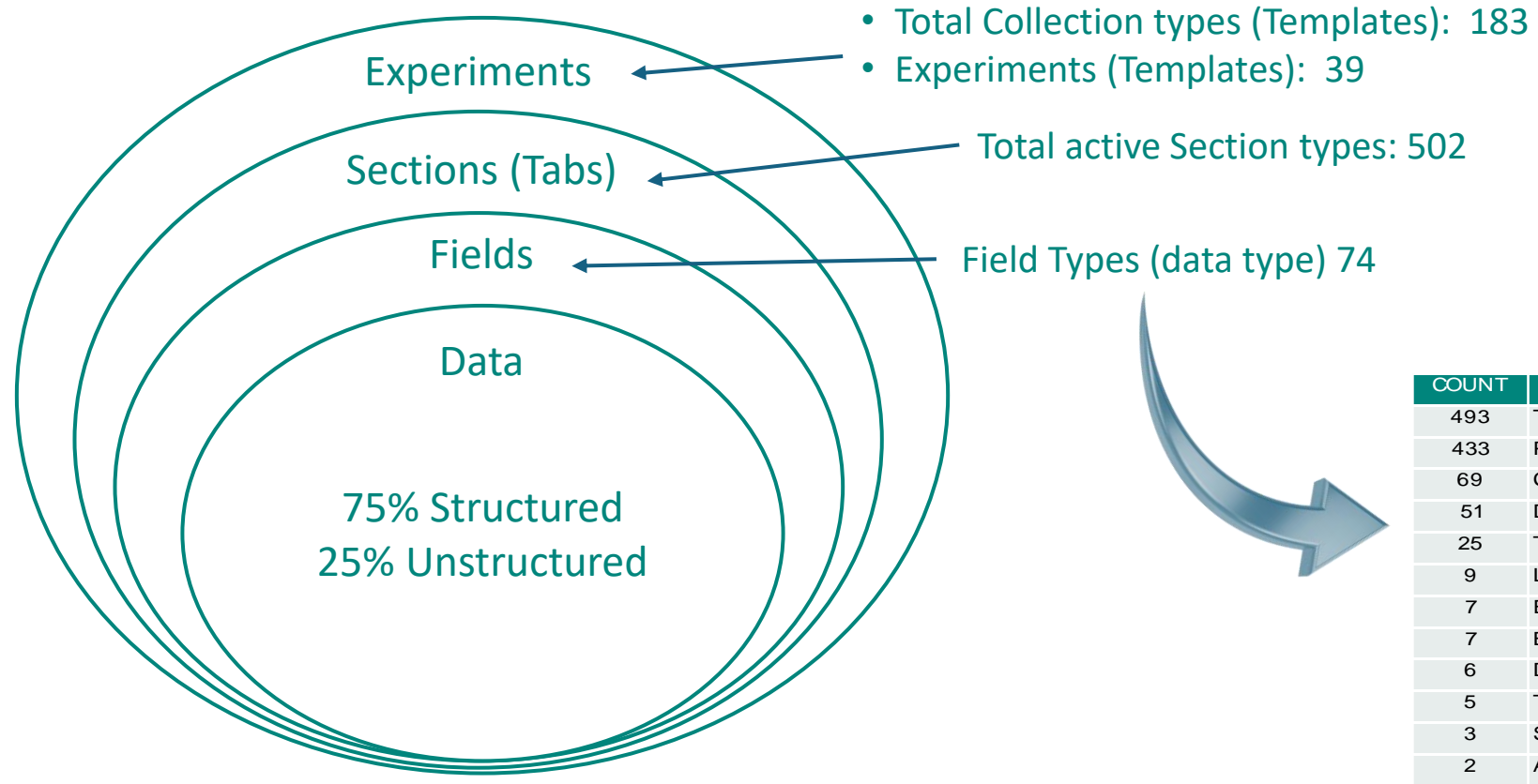
Analytics

- Maintain notebook data in an accessible form for future analytics use cases (FAIR)
- Capability to combine structured and unstructured data for mining & insights.
- Ability to maintain existing suite of reports (Spotfire, others.)
- Ability to segment & extract data using API's so it's consumable for other apps.

Target Conceptual Architecture



Diversity of Legacy ELN Data to be Archived



- Total Collection types (Templates): 183
- Experiments (Templates): 39
- Total active Section types: 502
- Field Types (data type) 74

ELN Configuration Types

Chemistry
Paper on Glass

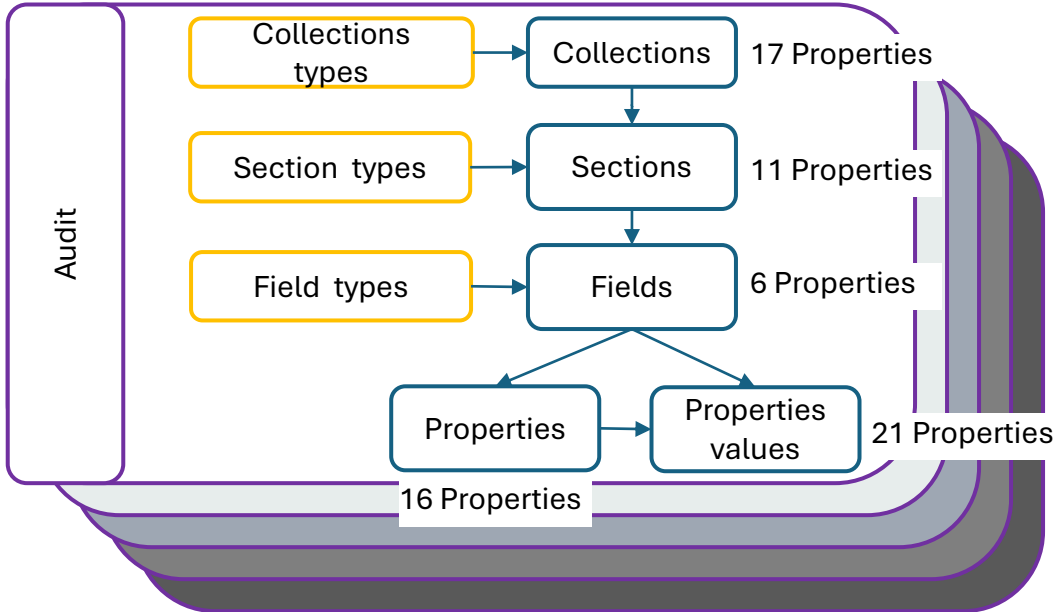
Bioassay
Bioanalytical (preclinical)



COUNT	Structured Field Type
493	Table
433	Property List
69	Chemical Structure
51	Database Table
25	TableEx
9	Label
7	Excel Spreadsheet
7	BioAssay Primary Data
6	Database Value
5	Table of Contents
3	Sort Results
2	Audit Trail Viewer
2	Child Table of Contents
1	HTML DB Table
1	NuGenesis Data List
1	BioAssay Summary Data
1	Workflow Action List
1	Batch Explorer
1	BioAssay Field Dictionary

COUNT	Unstructured Field Type
155	Styled Text
86	Auto Text
62	Subsection
42	Active Document
23	Reaction Toolbar
23	Favorites
14	PDF Viewer
10	MasterDetailTable
6	Spectrum
6	Image
4	Join Type
4	Experiment Info
4	Text Summary
4	PowerPoint SlideShow
3	Summary VCR
3	Ancillary Data
2	Unannotated Version Query
1	Stored Document
1	Hosted Document
1	E-Mail
1	ENCC Reactant
1	Advanced Query Text
1	Context Sensitive Help
1	Drag and Drop Field
1	E-Mail Viewer
1	ENCC Library
1	URL Display

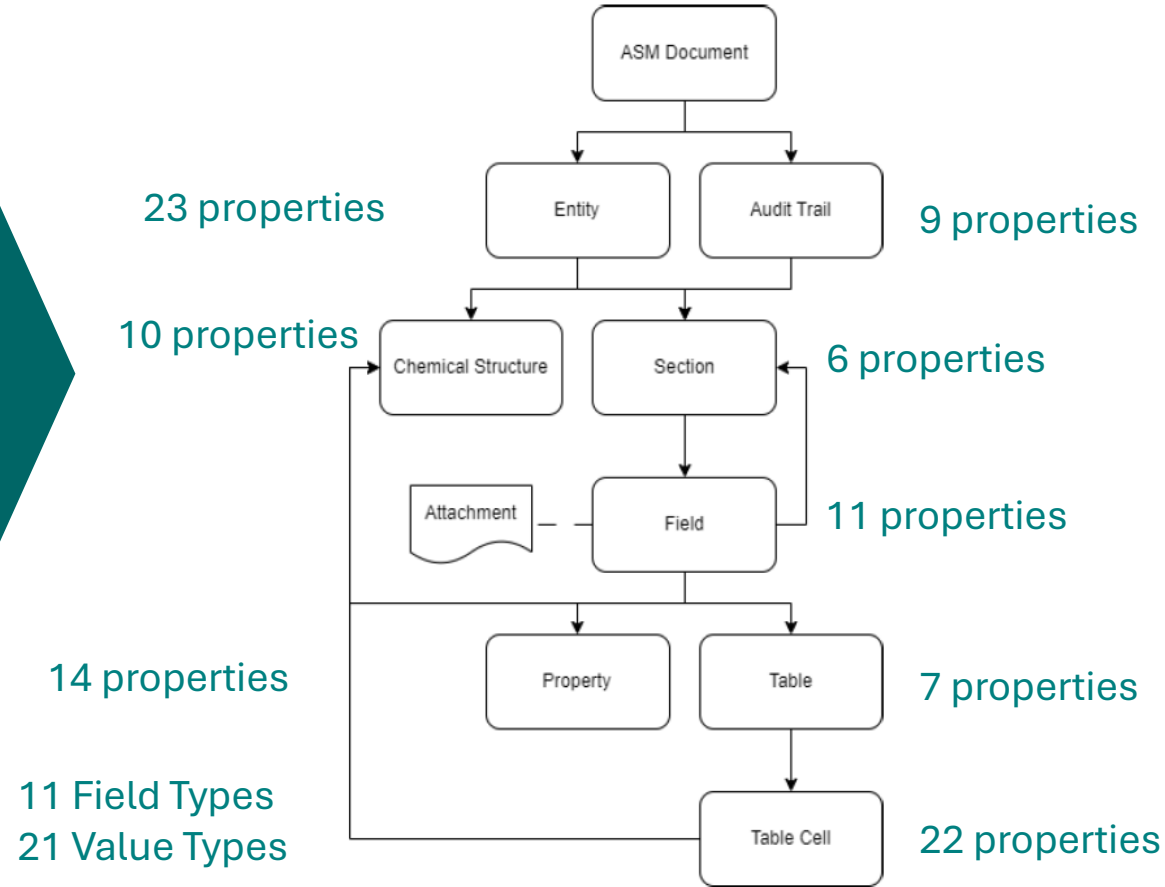
Draft ELN Archive Data Model – Unified Representation



254 types of tables (entities)
 Currently used in legacy ELN



Entities



~25x consolidation to ~10 reusable entities

Path Forward with new Allotrope ELN Working Group

- ✓ First step of Allotrope expansion into experiment domain !
- ✓ Model from Merck – ZONTAL collaboration was donated to kickstart Allotrope ELN WG effort
- ✓ Model includes 10 reusable entities and 134 total properties in first draft
- ✓ Governance underway with community SMEs to improve / extend model as required (e.g., suggestion to include Chemistry Reaction Table and Materials Table components into ELN data model based on initial feedback)
- ✓ Goal is to publish first candidate model in Q4 2024 Allotrope Product release and finalize in Q1 2025 release for archiving use case, but potential for model expansions if supported by the community

ELN Data Archiving Solution Demonstration

