Co-build to Accelerate Pipeline Decisions: TMI Framework for Comprehensive Therapeutic Evaluation

Authored by:

Krutika Gaonkar - Senior Manager, Partnerships, Elucidata

Chhavi Dawar, PhD - Senior Scientific Engineer, Elucidata

WHITEPAPER

elucidata

Contents

The Authors

Kritika Gaonkar, Senior Manager (Partnerships) at Elucidata brings over 10 years of experience in significantly improving data quality and reducing operational costs for bioinformatics and data science teams. This experience enables her to design and implement customer-focused solutions for academia, hospitals and pharmaceutical companies. She believes that the most successful solutions are born from collaboration and a relentless focus on the end-user. She is constantly exploring emerging technologies to deliver cutting-edge solutions that provide a lasting competitive advantage.

Chhavi Dawar, PhD, Senior Scientific Manager at Elucidata leads the implementation of AI-driven multi-omics solutions to accelerate drug discovery. She spearheads the development of advanced computational pipelines, AI-enabled data harmonization & insights generation and rigorous QA/QC workflows for ML-ready biomedical datasets. Her work integrates genomics, transcriptomics, proteomics, and real-world data for biomarker discovery, patient stratification, target identification, and validation. Collaborating with global pharma and biotech partners across oncology, cardio-metabolic, infectious diseases, and rare disorders, she facilitates cutting-edge AI innovation with practical applications in translational research.

Executive Summary

The rise of **novel therapeutic modalities** like antibody drug conjugates (ADC), bispecific antibodies, tumor-infiltrating lymphocyte (TILs), miRNA therapeutics, RNA interference (RNAi), gene editing and cell therapies have transformed the biopharma landscape in the last decade. These innovations offer the potential to **address previously undruggable targets** and deliver more precise treatments.

As these modalities mature, they occupy distinct positions along the innovation curve each with different levels of scientific validation, translational readiness, and capacity to address unmet clinical needs.

Some are on the brink of commercialization, while others remain in exploratory phases. Biopharma companies **need to adopt a tailored, data-driven approach for modality assessment**, that aligns with internal strengths, uses public data wisely and balances investment based on risk, timelines, and strategic fit.

Elucidata's TMI (Target-Modality-Indication) solution transforms this high-risk modality assessment problem into timely, actionable insights through a strategic target-modality-indication assessment framework. This framework leverages multi-omics, clinical trials and real world data to assess target-modality-indication fit through harmonized integration of internal and public data. This will enable early identification of promising modality-indication pairs and indication expansion opportunities. Enabling de-risked investment decisions and guide resource allocation based on biological and clinical evidence.

TMI (Target-Modality-Indication) framework redefines how strategic decisions are made in drug discovery by moving from siloed, intuition-driven choices to a unified, data-driven framework.

This framework delivers three key advantages:

(1) Strategic clarity in choosing the right modality with minimized risk,

(2) Accelerated identification of new indication opportunities, and

(3) Streamlined access to high-quality, multi-modal evidence for better cross-functional collaboration.

The framework doesn't just improve existing processes but provides a data-driven framework, enabling biopharma teams to navigate complexity with precision and seize first-mover advantage in new therapeutic spaces.

Strategic & Business Impact

Elucidata's AI-powered framework delivers impact across critical value levels for biopharma investors by accelerating asset evaluation and enhancing capital efficiency.

The platform enables rapid, evidence-based triage of target-modality-indication combinations, resulting in faster go/no-go decisions and optimized R&D resource allocation. This translates to improved pipeline productivity and higher return on early-stage investments, while minimizing sunk costs on non-viable programs.

Value for R&D Teams and Stakeholders

Through automated and scalable integration of internal and external datasets, Elucidata de-risks portfolio decisions and identifies high-potential expansion opportunities with clinical and commercial relevance. **The framework will future-proof data infrastructure for R&D teams** by creating reusable, modular data products that support:

Multi-program assessments

Reduce duplication

Lower long-term operational costs

Business Impact

- The platform improves **strategic decision-making** across discovery, translational, and business functions by aligning scientific evidence with commercial viability.
- It also fosters **cross-functional and partner collaborations, streamlines regulatory preparedness, and shortens development timelines.**
- For investors, this means **accelerated time-to-market, reduced risk exposure, and sustainable competitive advantage** in an increasingly crowded and high-stakes therapeutic landscape.

Introduction

Despite substantial investments in pharma R&D, a substantial proportion of clinical trials continue to fail often due to weak biological rationale at single cell level or incomplete evidence to back the modality and indication of interest.

> *A recent analysis using genetic evidence from the Open Targets Platform revealed that clinical trials lacking strong genetic support for their target-indication pair were* ***significantly more likely to be stopped*** *(OR = 0.73, $P = 3.4 \times 10^{-69}$). This pattern held true across withdrawn, terminated, and suspended studies (Razuvayevskaya, 2024).*

These findings underscore a critical gap in early-stage decision-making: the absence of integrated, multi-dimensional evidence at single cell level focussed on modality and indications of interest to enable deeper understanding of target interactions and modalities mechanism for a given patient population.

This whitepaper presents a comprehensive **Target–Modality–Indication (TMI) assessment framework,** designed to mitigate such risks. By creating comprehensive data products including genetic, clinical, and real world data, the approach enables timely decisions to drive competitive differentiation and higher clinical success rates and more efficient resource allocation.

Broader Context

Driven by deeper understanding of disease mechanisms and cutting edge technology, biotech has been developing innovative drug modalities at an astonishing pace.

MORE THAN

17

new drug modalities have been developed in the past 20 years.

These include technologies like gene therapies, cell therapies, RNA-based drugs (e.g., siRNA, mRNA), antibody-drug conjugates, protein degraders, and CRISPR-based genome editing.

For example, in oncology currently **70-75% of molecules in the clinic are ADC, biologics and cell therapies (Mullard, 2025)** and this trend is well known to increase in the future. What makes these modalities transformative is their ability to offer curative potential for genetic and rare diseases, and provide more precise control over biological systems.

In some cases, new modalities may not replace but complement existing therapies for greater efficacy. These synergistic effects allow pharmaceutical companies to achieve better and safer clinical outcomes.

Traditional Approaches and Limitations

Conventional drug discovery frameworks focus heavily on identifying the right target, tissue, patient population and commercial opportunity. While these steps are essential, they are often conducted in fragments across the company, leading to delay in collaboration. This incomplete view of the target may lead to lower successful candidates.

Key insights of the target at single cell level, understanding the genetic profile of patients that respond to the modality and those who don't, could be the critical layer that turns potential failure into first-in-class success. Modality selection, whether based on antibody, RNA, cell or gene therapy, directly impacts bioavailability, dosing frequency, safety profile, and mechanism of action.

Without incorporating modality fit into early-stage assessments, teams risk:

- **Misjudging feasibility**
- **Exposing patients to needless safety concerns**
- **Failing to differentiate in crowded markets**

It's not just about the right target in the right tissue, it's about delivering it with the right modality, to the right patient, in the safest and most effective way to target the specific diseased cell types.

Objectives

Pharmaceutical companies are increasingly building multi-modal data platforms to deliver high-quality data to their organizations. We propose a Strategic Data-Driven Assessment Framework called TMI framework to unlock the use case for comprehensive insights into target, modality and indication.

The objectives of this study are to:

01 **Introduce TMI (Target, Modality and Indication) framework** based on multi-modal evidence from internal and public data.

02 **Showcase how it accelerates insights** to de-risk modality choices based on efficacy, safety, and competitive landscape.

03 **Establish its capability to enhance access** to and interpretation of unstructured scientific content, omics data, and imaging data.

Scope

Systematically identifying the most promising therapeutic modalities and indication for a given target will require AI-driven processes to find, ingest, and evaluate quality of relevant preclinical and partner data. This requires expertise to build ETL pipelines to stream public and internal data, purpose-built tools to extract information from unstructured texts ,omics and imaging data and develop and maintain data products that are specific to the company's goals.

Aim of the Study

To highlight the data infrastructure, organizational capability, and analytical maturity required to move from fragmented experimentation to strategic execution.

This approach will enable pharma companies to maintain relevance and competitiveness in a rapidly evolving therapeutic landscape.

Problem Statement

How can biopharma companies improve capital efficiency when prolonged evaluation cycles and disconnected data increase time-to-decision and burn rate? Are existing tools truly enabling timely, data-backed choices around which targets, modalities, and indications can be prioritized? What is the investment risk of delayed go/no-go calls, missed market opportunities, and non-strategic asset allocation?

This paper explores how the TMI Evaluation Framework addresses the challenge of **long, costly evaluation cycles for targets, modalities, and indications**. By streamlining early-stage assessments, it enables faster, data-driven decisions and reduces R&D expenditure.

Specific Challenges and Underlying Causes

Core Challenge

Large biopharma organizations struggle to systematically evaluate and prioritize modality-indication pairs that align with both R&D strategy and commercial objectives, while accounting for internal capabilities, timelines, and risk appetite.

This challenge is compounded by fragmented data ecosystems, siloed internal knowledge, inconsistent evaluation frameworks, and limited integration of public and proprietary data sources.

The consequences are significant across multiple dimensions. **Suboptimal modality-indication alignment can be:**

Scientific

Delay translational progress and hinder biomarker-informed patient stratification

Operational

Inefficient resource allocation, extended development timelines, and limited scalability of insights across programs.

Regulatory

Uncertainty in trial design, endpoints, and evidence generation, risking approval delays and increased costs.

Impact on R&D Efficiency

Existing tools are often retrospective, static, or narrowly focused, and fall short of supporting dynamic, cross-modality decision-making at scale.

Critical Need

Data-driven, modular, and scalable framework that integrates biological, clinical, and strategic data to inform early decision-making.

Such a solution will empower R&D and investment teams to confidently de-risk portfolios, accelerate development, and focus investments on high-impact, evidence-backed opportunities in an increasingly competitive and complex innovation landscape.

Elucidata's Solution

Rethinking Modality Evaluation

Elucidata's solution is a disruptive, first-of-its-kind, AI-powered framework designed to help large biopharma organizations evaluate and prioritize emerging therapeutic modalities with unprecedented speed, precision, and conviction.

Elucidata addresses this by

- Consistently processed and annotated single cell datasets
- Structured Clinical trial data for intel into current therapy performance and safety profiles
- Unstructured literature for latest insight into the indication biology
- Harmonized inhouse data for insights on modality-target-indication fit

This novel approach de-risks early-stage R&D decisions and supports smarter, more strategic investments in high-potential programs across disease areas.

Unique Strengths

What makes this solution truly innovative is its tiered data product architecture, which delivers targeted insights at three strategic levels:

01 ***Target-centric layer*** provides deep biological and mechanistic insights at single cell resolution for prioritizing therapeutic targets using internal and public data.

02 ***Modality-centric layer*** incorporates insight into responder and non-responders, patient heterogeneity, and unmet clinical needs to evaluate differentiation of their asset.

03 ***Indication-centric layer*** contextualizes internal and external R&D in understanding the indication biology, comprehensive single cell atlas for indication of interest and latest

These layers are powered by **Polly**, **Elucidata's industry-leading data platform** that ensures FAIR data harmonization, multi-modal integration, and robust AI-cloud scalability. The result is a reusable, secure, and enterprise-ready solution that aligns scientific feasibility with business objectives and commercial viability.

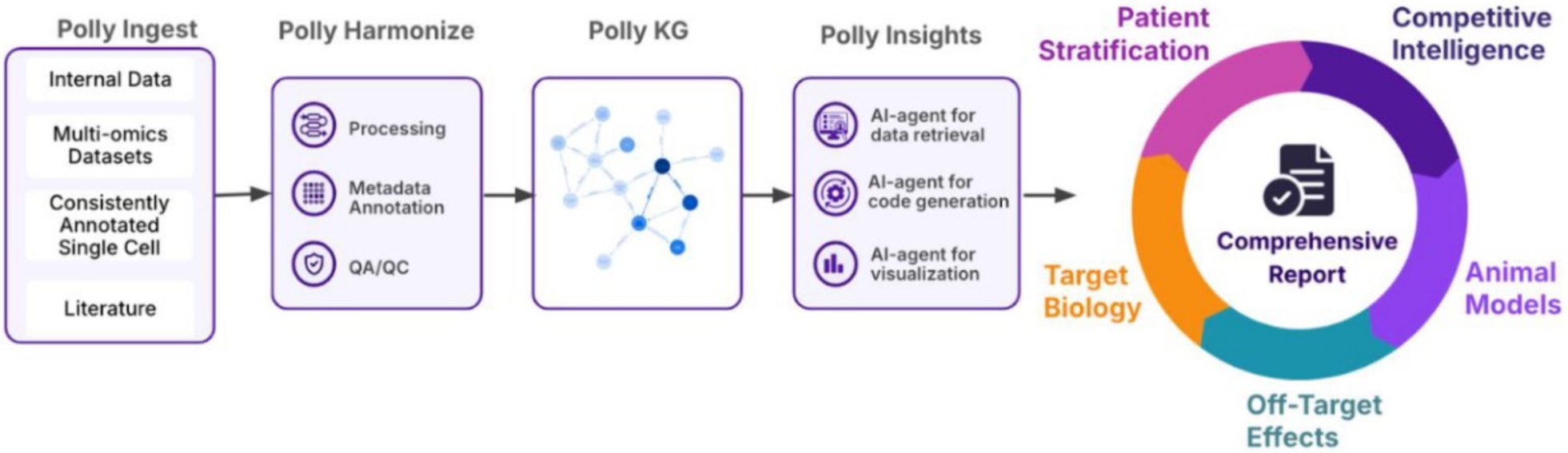


Figure 1: Overview of the Polly platform that supports the target-modality-indication assessment

Alignment with Elucidata's Goals

Elucidata is uniquely positioned to deliver this solution.

Recognized as the **Most Innovative Biotech Company of 2024** by Fast Company (FastCompany, 2024) and **winner of the NCI CRDC AI Data Readiness Challenge** (NCI, 2024), we bring deep expertise across **25+** data modalities.

Our platform is **trusted by over 35 organizations**, including but not limited to, leading pharmaceutical companies, academic research institutions, clinical and healthcare organizations.

This solution exemplifies Elucidata's strategic mission, to transform complex biomedical data into AI-ready, fit-for-purpose data products that accelerate drug discovery and development.

By enabling R&D and investment teams to build once and apply broadly, we empower the life sciences industry to make faster, evidence-based decisions and unlock new therapeutic possibilities at scale.

Technical Architecture

Elucidata employs a **three-tiered analytical framework** that integrates target-centric, modality-centric, and indication-centric data. This structured approach ensures a comprehensive evaluation of biological relevance, clinical impact, and strategic fit.

Tiered Approach to build Data Products

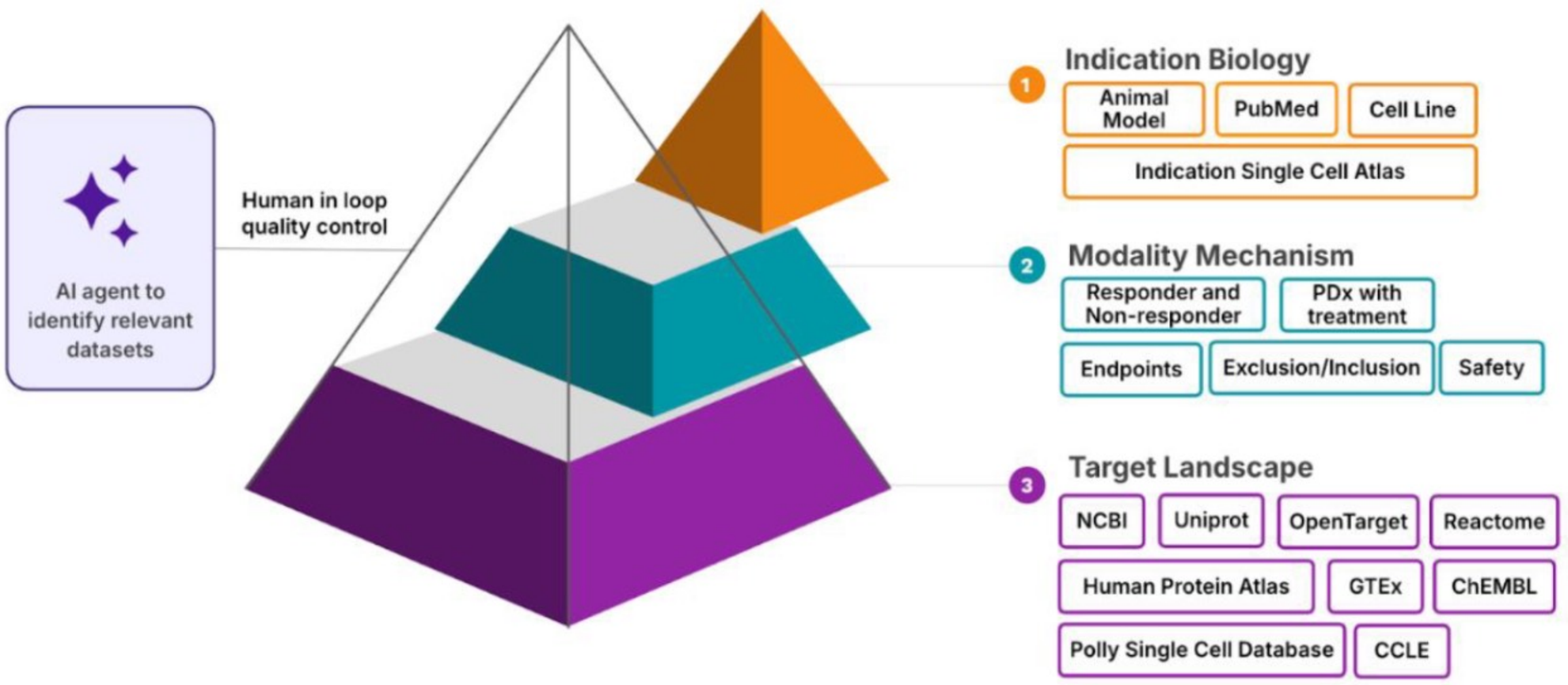


Figure 2. Tiered Approach to Build data products for target, modality and indication of interest

Indication Assessment: Strategic Portfolio Alignment for Indication of Interest

An indication-specific data product delivers an **integrated view of disease biology by -**

Combining disease vs. normal datasets, multi-omics insights from GEO, PRIDE and other public sources, and literature evidence to uncover key molecular pathways.

Including animal models for preclinical validation, a robust single-cell atlas for high-resolution analysis of cell-type specific expression, cell–cell interactions, and spatial organization when available.

Incorporating in-house or public CRISPR knockout data, enabling assessment of gene essentiality in diseased cell survival crucial for identifying and validating high-confidence therapeutic targets.

Modality Assessment : Bridging Biology to Therapeutic Impact

Modality-centric data products are invaluable for **competitive intelligence and translational insight**.

- It enables precise comparison of how competitors target the same molecule or pathway using different modalities, revealing strategic gaps or overlaps.
- By integrating responder vs. non-responder expression profiles, it helps identify biomarkers for patient stratification. Coupling this with mouse PDX and cell line models annotated with treatment responses supports preclinical validation.
- When layered with clinical trial data, including inclusion/exclusion criteria and reported adverse events, it allows teams to assess competitor trial design, anticipate safety risks, and refine their own development strategy for differentiation and de-risking.

Target Assessment : Unlocking Biological Potential

Target-centric data focuses on **comprehensively assessing a target's suitability across drug modalities**.

Databases	Functionality
NCBI, UniProt	Foundational genomic and protein-level annotations, offering clarity on sequence, structure, and known functions.
Human Protein Atlas	Contextualizes tissue-specific and subcellular localization, aiding modality selection
ChEMBL	Bioactivity data linking targets to small molecules and biologics
GTEx	Adds insights into tissue-wide gene expression in healthy states
Reactome	Maps target pathways and interactions, enabling a systems-level understanding crucial for evaluating downstream effects and combinatorial strategies
OpenTargets	Integrates disease associations and tractability scores, guiding therapeutic relevance.
A consistently annotated single-cell database	Reveals target expression and cell-type specificity, critical for precision medicine

Tools and Technology

Elucidata has built a scalable data ingestion and enrichment framework that integrates public and proprietary sources into a unified platform, to enable real time, and evidence driven insights across biomedical R&D.

At the heart of the solution, is a **knowledge graph** powered by structured ontologies that captures complex, multi-scale relationships across genes, diseases, patients, and clinical endpoints.

Together, these components form a semantic foundation that supports dynamic data integration, hypothesis generation, and strategic decision-making in drug discovery and translational research.

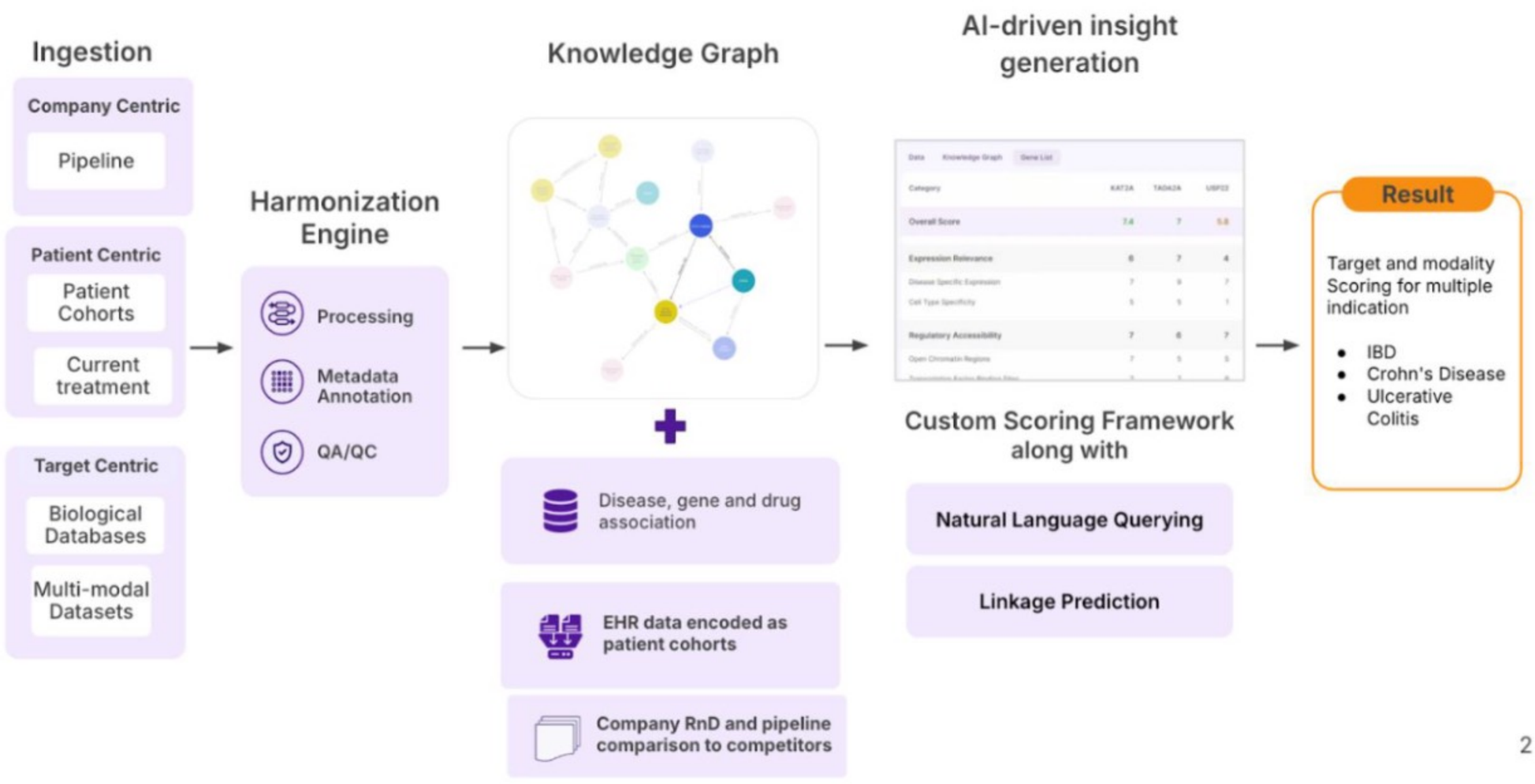


Figure 3. Ingest, harmonize and build knowledge graphs for Causal Insight.

The following components enable building a flexible data infrastructure to enable accelerated insights into target, modality and assessment:

01 **Modular connectors** ingest high-volume data streams from sources like Open Targets, ClinicalTrials.gov, and ChEMBL, normalizing diverse formats into a common schema.

02 **Large language models (LLMs)** based AI systems to intelligently identify and extract high-value datasets from repositories such as GEO and PRIDE, minimizing manual curation and accelerating discovery.

03 **Processing and aggregation** of multimodal in-house and public data at scale to create high quality consistently processed data.

04 The data quality data products are then stored as a **knowledge graph to enable a semantic layer.** This enables reasoning and hypothesis generation with scoring framework custom built for use case of interest.

Modular Connectors for Multimodal Data Ingestion

Scalable and modular adapters are capable of ingesting data in real time from a diverse set of biomedical and clinical databases, including Open Targets, UNIPROT, and ChEMBL.

These adapters are designed to handle varying data formats, update frequencies, and access protocols, ensuring seamless integration of up-to-date information.

The architecture supports high-throughput processing, normalization, and mapping of incoming data to a common data model, **enabling continuous enrichment of downstream applications such as:**

Drug discovery

Target validation

Clinical Trial planning

LLMs for Dataset Identification and Curation at Scale

Large language models (LLMs) that can i**ntelligently search, filter, and extract metadata** from large public repositories like GEO and PRIDE have been deployed by Elucidata to help accelerate data discovery (Mondal, 2025).

These models are fine-tuned to recognize relevant experimental designs, sample types, diseases, and omics modalities, allowing automated identification of datasets that match predefined research criteria.

This approach significantly **reduces manual curation time and enables a dynamic pipeline** where high-value datasets are continuously identified and queued for ingestion and analysis.

Scalable Processing and Support for Multi-Modal Data

Valuable multimodal datasets are generated or acquired **from different sources** to enable biological insights into the biology or patient cohorts of interest. Consistent processing of data with productionized pipelines allows for scale processing while also being cost-effective.

With a clear goal of aggregating the data into target, modality or indication data products enable a **transparent and standardized** processing and representation of data to enable FAIR at source.

Knowledge Graph to Enable Custom Scoring Framework

At the core of the solution is a robust knowledge graph built on a structured data model and enriched with domain-specific ontologies. This graph is designed to capture granular and meaningful relationships, such as:

- **Gene-to-gene interactions**
- **Gene-pathway-disease associations**
- **Gene expression across tissues**
- **Links between indications, phenotypes and patient characteristics**

*It also **models relationships** between patients and their comorbidities, biobank sample availability, and multi-modal data including omics and imaging.*

The graph serves as a semantic integration layer that supports powerful querying, hypothesis generation, and cross-domain reasoning delivered as a scoring framework for each use case.

AI-assisted Data Ingestion, Harmonisation and Processing

Elucidata's modular harmonization pipeline ensures semantic consistency, comparability, and context across datasets. By standardizing key identifiers (e.g., HGNC, ICD, ChEMBL), normalizing omics data for cross-cohort analyses, and structuring unstructured text via NLP, the platform enables seamless integration and high quality analysis.

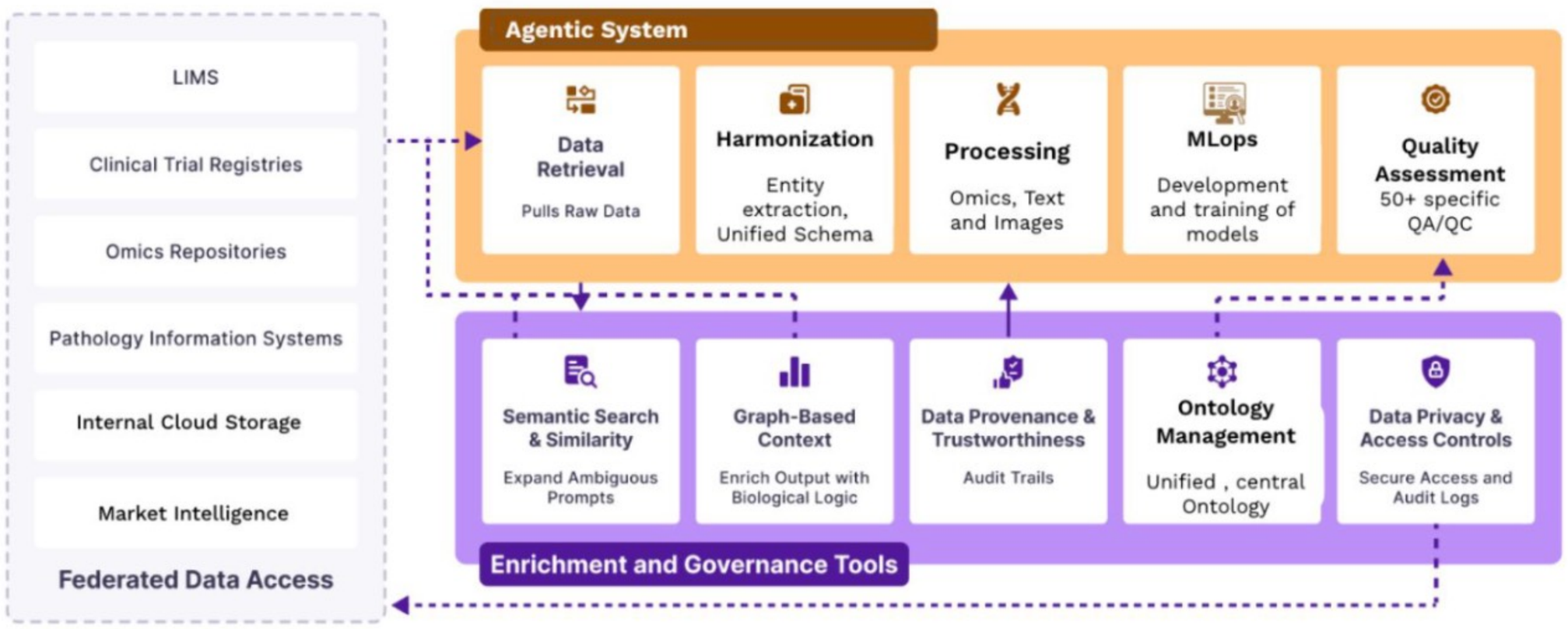


Figure 3. Ingest, harmonize and build knowledge graphs for Causal Insight.

Data Retrieval

Data retrieval spans both public and internal sources, with a focus on capturing high-value datasets relevant to specific modalities and indications.

Polly's Agentic search capabilities enable automated discovery of public datasets using filters such as:

- Disease indication,
- Therapeutic modality,
- Patient metadata (e.g., responder/non-responder status), and
- Contextual evidence from literature and ClinicalTrials.gov to inform competitor trial design and outcomes.

Data Processing

Newer modalities target lesser-known mechanisms and require custom references, tools, and analysis to derive meaningful insights. This is performed using the following steps:

01 Raw data is analyzed at scale to extract high-quality biological insights, focusing on pathways of interest, especially **distinguishing responders vs. non-responders and disease vs. normal conditions** across human, mouse, and cell-line models. This comprehensive approach reveals how diseases and modalities affect biology across human, animal, and cell-line systems.

02 **Variant and GWAS data is annotated** with predicted variant effects and population frequency to contextualize variants within disease biology.

03 Internally processed data, including comparative genomics and validation data, is **reintegrated into the knowledge graph** with appropriate processing to support knowledge aggregation.

04 Single-cell data, internal and public, is uniformly processed using **Cellranger v8** to produce cell-type-specific expression matrices and detailed cell-type annotations for both disease and normal conditions.

05 Cell types are consistently annotated across **330M cells** using aggregated cell marker **databases from cellmarkerDb, pangloDB, Hubmap**, and manually curated publications; differential expression is captured across cell types to allow deeper exploration at the cellular level.

MLOps

Elucidata enables development and training of models for in-silico predictions and causal inference from literature and molecular data provides a powerful framework for target, modality, and indication assessment by uncovering mechanistic relationships rather than mere correlations.

Algorithms such as bnlearn and GRNBoost are particularly valuable in this context. These tools enable the construction of **Bayesian networks** to model probabilistic dependencies between genes (Carapito, 2022), pathways, and phenotypes and infer gene regulatory networks from single-cell or bulk transcriptomic data.

When integrated with curated scientific literature, these approaches can highlight causal drivers of disease, predict modality-specific mechanisms, and uncover shared molecular signatures across indications.

Quality Assessment

Elucidata's pipeline incorporates a multi-layered data quality assessment and validation framework which includes the following:

01 ***Schema validation:***
Ensures data conforms to predefined structures and types during ingestion.

02 ***Ontology mapping verification:***
Confirms accurate alignment of terms to standard vocabularies (e.g., HGNC, ICD, MedDRA)

03 ***Outlier and missing value checks:***
Identifies anomalies or gaps in omics and clinical datasets.

04 ***Source provenance tracking:***
Captures and maintains detailed metadata about data origin, version, and transformations.

05 ***Reproducibility audits:***
Validates that harmonized data and outputs are consistent across reruns

06 ***Manual curation checkpoints:***
Applies expert review to high-impact entities like targets, endpoints, and modality labels

06 ***Data Quality Assessment Report:***
Visualizes metadata completeness, data source credibility, and integrity of data through plots and quality metrics for each data type.

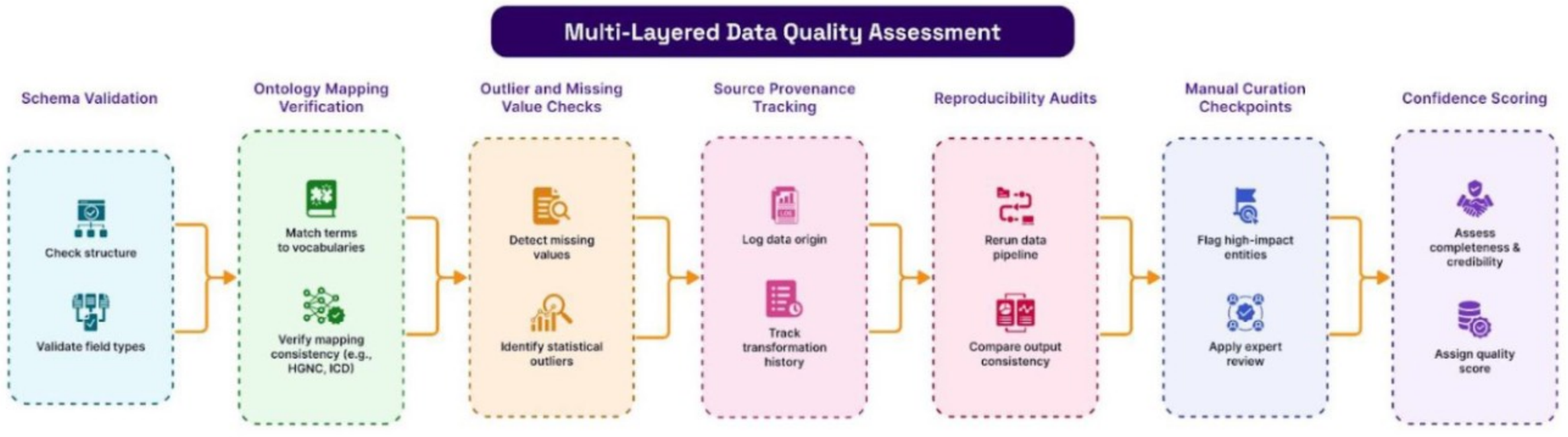


Figure 5: Multi-layered data quality assessment and validation framework

Governance and Enrichment Tools

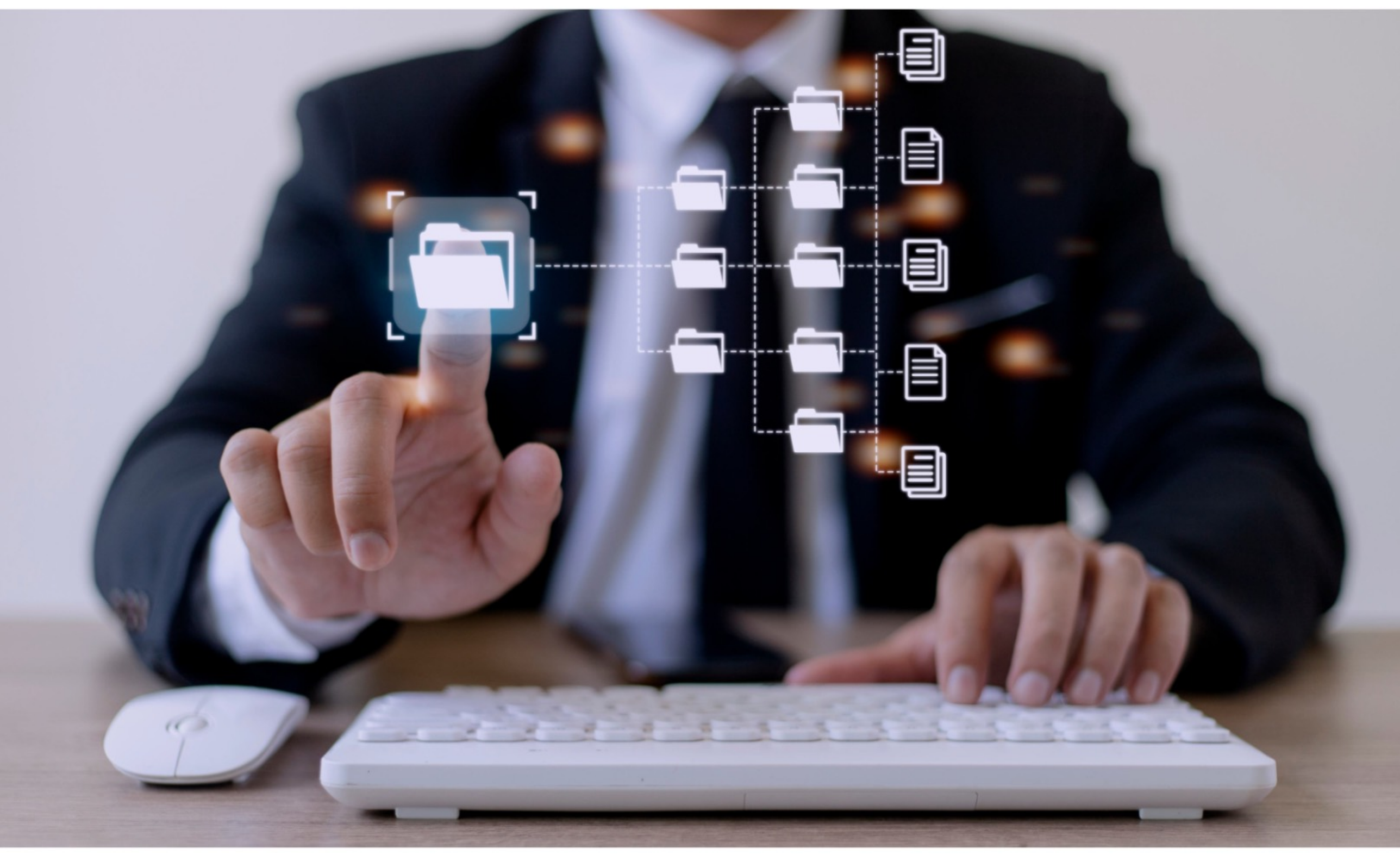
The Polly platform empowers organizations with robust data governance and seamless data consumption tools, ensuring that biomedical data is both accessible and compliant.

Key capabilities

- It supports **semantic search and graph-based context building**, enabling researchers to intuitively navigate complex relationships across datasets.

- Polly also integrates **ontology management**, allowing consistent annotation and harmonization of internal and external biomedical vocabularies.

- Coupled with its **built-in data privacy controls and access management**, Polly ensures secure, auditable, and policy-compliant use of sensitive biomedical data accelerating insights while maintaining trust and regulatory alignment.



Future Roadmap and Enhancement

Elucidata's AI-powered TMI framework is not only designed to address today's challenges in therapeutic modality triage but is also built with a future-facing architecture to continuously evolve with the needs of R&D teams in the pharma space.

The future roadmap prioritizes expanding intuitive, scalable ways to consume insights most notably through user-friendly interfaces for knowledge graphs that surface modality-target-indication connections in an interactive and decision-ready format. Enhancements also include periodic updates of public, proprietary datasets and seamless integration of partner data to ensure assessments are grounded in the most current evidence.

Peripheral Challenges

This data strategy offers a powerful foundation to address a broad array of challenges across the drug discovery and development continuum. It can be extended to:

- **Accelerate target discovery**
- **Enable biomarker identification**
- **Optimize clinical trial design**
- **Inform mechanism of action studies**
- **Support label expansion and post-market surveillance**

The same multi-modal, AI-ready infrastructure can be applied across functional domains, offering ***a unified and reusable approach*** *to complex decision-making.*

As organizations expand into new modalities or therapeutic areas, this solution provides a consistent and scalable analytical backbone.

Scalability

The **TMI framework is inherently modular** and designed for horizontal and vertical scaling.

Horizontally, it can be deployed across multiple therapeutic areas, modality types, and target classes without reengineering. Vertically, it can integrate deeper layers of proprietary and partner data, including imaging, proteomics, lab automation outputs, and manufacturing data to ensure continued relevance as internal datasets mature.

The platform's cloud-native design ensures seamless integration into existing enterprise data ecosystems.

Built on Elucidata's Polly platform, it supports:

- Automated data ingestion,
- FAIR harmonization, and
- Continuous learning models, enabling reuse of pipelines and insights across programs and teams.

This reusability drastically reduces marginal cost per insight and accelerates time-to-decision across portfolios. In the long term, the solution's scalability ensures that as scientific questions evolve and business needs grow, biopharma organizations are equipped with **a future-ready, AI-augmented infrastructure** to lead the next wave of innovation.

Conclusion

As therapeutic modalities, data types, and data sources continue to diversify and evolve, biopharma organizations face **increasing complexity** in extracting meaningful intelligence for selecting and prioritizing the right target-modality-indication combinations.

Elucidata's AI-driven, multi-modal data integration framework addresses this challenge by unifying fragmented internal and external datasets into a single scalable platform.

This solution empowers faster, evidence-based triage and strategic decision-making that de-risks investments, accelerates time-to-market, and enhances both scientific rigor and commercial outcomes across the drug development pipeline.

The timing to adopt such an integrative framework has never been more critical. The biopharma industry's strategic shift towards precision medicine and the expansion of novel therapeutic modalities demand comprehensive, data-driven approaches to align innovation with the most promising clinical indications. Additionally, the current market gap for an end-to-end solution that integrates public and proprietary data presents a unique opportunity for organizations to gain a competitive edge by leveraging scalable, reusable data products and AI-powered insights.

We strongly encourage biopharma investors, R&D leaders and pharma organizations at large to critically evaluate whether their current frameworks meet the evolving demands of therapeutic innovation and data complexity.

This **first-of-its-kind framework** offers a transformative approach to modality assessment and indication expansion by integrating your internal pipelines and partner data with high-quality public datasets; it opens doors to deeper, more actionable insights.

Imagine the impact of piloting this system on your current prioritization efforts for targets, modalities, and indications. It could illuminate unseen opportunities, reduce uncertainty, and give your therapeutic strategies a decisive edge in a highly competitive space.

References

[1] Razuvayevskaya, O., Lopez, I., Dunham, I., & Ochoa, D. (2024, July 29). Genetic factors associated with reasons for clinical trial stoppage. Nature Genetics, 56(9), 1862–1867. https://doi.org/10.1038/s41588-024-01854-z

[2] Mullard, A. (2025). Back to therapeutic modality basics. Nature Reviews Drug Discovery, 24, 327–328. https://doi.org/10.1038/d41573-025-00072-4

[3] FastCompany (2024, March 19). The most innovative companies in biotech in 2024. https://www.fastcompany.com/91034883/biotech-most-innovative-companies-2024

[4] NCI (2024, May 21). Winners' Presentation for the NCI CRDC AI Data Readiness Challenge.https://datascience.cancer.gov/news-events/events/winners-presentation-nci-crdc-ai-data-readiness-challenge

[5] Carapito, R., Li, R., Helms, J., Carapito, C., Gujja, S., Rolli, V., Guimaraes, R., Malagon-Lopez, J., Spinnhirny, P., Lederle, A., Mohseninia, R., Hirschler, A., Muller, L., Bastard, P., Gervais, A., Zhang, Q., Danion, F., Ruch, Y., Schenck, M., Collange, O., ... Bahram, S. (2022). Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. Science Translational Medicine, 14(628), eabj7521. https://doi.org/10.1126/scitranslmed.abj7521

[6] Mondal, R., Sen, M., Sengupta, S., Maity, W., Palapetta, S., Dhruw, N. K., & Jha, A. (2025, June 10). Multi-agent AI system for high quality metadata curation at scale. bioRxiv. https://doi.org/10.1101/2025.06.10.658658v1

About Elucidata

Elucidata is a San Francisco and Boston-based deep tech company which enables organizations to become AI-ready. Powered by a 110+ strong multidisciplinary team, the company's technology delivers AI-ready data 10X faster with 60% reduction in data processing costs through automated workflows. Keeping data quality at its foundation, Elucidata's proprietary data quality checks ensure 99% accuracy in multi-modal data through a human-in-the-loop approach. Rooted around people, technology and process, Elucidata's mission is to keep AI 'Real' for precision medicine.

Have questions or want to discuss the ideas shared in this whitepaper?

We'd be happy to hear from you.

For any queries, clarifications, or to explore how these approaches could apply to your work, please reach out to the experts at krutika.gaonkar@elucidata.io and chhavi.dawar@elucidata.io.

Book a Demo