FROM DATA TO DISCOVERY:
ACCELERATING LIFE SCIENCES BREAKTHROUGHS WITH DATA STRATEGY SOLUTIONS FROM DELOITTE AND DATABRICKS

Leveraging data strategically can help health care, life sciences, and biopharmaceutical organizations gain a competitive edge, drive innovation, and achieve business objectives.

Together, Deloitte and Databricks offer powerful solutions that empower organizations to build and execute effective data strategies, transform data into actionable insights, and help drive breakthroughs in drug discovery, personalized medicine, and beyond.

LIFE SCIENCES: AN INDUSTRY IN FLUX

Life sciences organizations face rapidly evolving challenges, especially with:

Regulatory complexity
The number of new regulations and guidelines is growing by approximately 15% annually[1].

Technology adoption
70% of biopharma companies are investing in artificial intelligence (AI) and machine learning (ML), though only 25% have successfully integrated AI and ML into their operations[2].

Supply chain pressures
78% of life sciences executives are prioritizing supply chain resilience and agility to mitigate risks and disruptions[2].

FOUR PILLARS OF A SOLID DATA STRATEGY

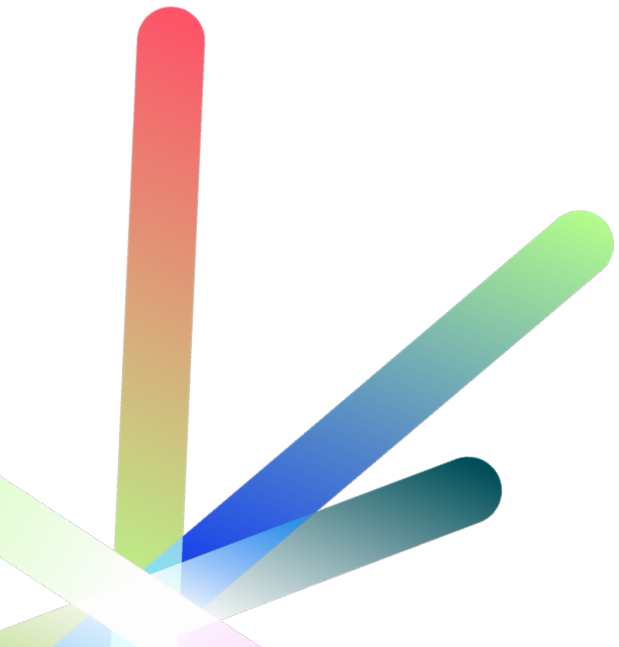
To address these challenges and drive biopharmaceutical transformation, organizations should have plans, collaborators, and solutions that can deliver on the four key pillars of a solid data strategy:

01 **Data governance and management**
Ensuring data quality, security, and compliance through best practices like data cataloging, lineage and stewardship

02 **Data integration and accessibility**
Providing seamless data integration and convenient accessibility through user-friendly, secure platforms and tools

03 **Advanced analytics and AI**
Turning data into actionable insights using predictive analytics, machine learning, and AI

04 **Scalability and flexibility**
Building a scalable and flexible data architecture to support growth and adapt to evolving business needs



IDENTIFYING NEW OPPORTUNITIES TO INNOVATE

Emerging digital trends are accelerating the need for cross-domain analytics, data, and platforms in life sciences, driving new opportunities in:

- **Improved customer experience**
  Leverage agentic AI, backed by high-quality data and insights, to improve the customer and patient experience
- **Personalized medicine**
  Individualized manufacturing and supply chain solutions for cell and gene therapies
- **Regulatory approval pathways**
  Expedited pathways and reliance on real-world evidence (RWE) and post-market studies can help ensure safety data efficacy
- **Clinical development**
  Increased focus on patient enrollment and engagement can help meet clinical development timelines
- **Manufacturing processes and supply chain**
  Enhancing complex ecosystems that rely on multiple teams and collaborators can increase process quality for biologics
- **Competitive therapies/indications**
  Focusing on patent expirations, biosimilars (medicines close in structure to biologics), and crowded therapeutic areas
- **Outcomes-based reimbursement**
  Payment for drugs based on real-world safety and efficacy data, often times on per-patient basis
- **Digitally connected patients**
  Supporting disease management and access to online medical information
- **Internet of Things (IoT) adoption**
  Improved clinical data capture and disease management through IoT-enabled solutions

RESHAPING THE LIFE SCIENCES VALUE CHAIN THROUGH MODERNIZED DATA STRATEGY

Leading biopharmaceutical organizations are redefining their data strategies by taking a holistic view of their enterprise value chain. This includes research and development (R&D), manufacturing and supply chain operations, and commercial operations, supported by enabling functions like human resources and finance—as seen in Figure 1.

Research

- Accelerate drug discovery
- Facilitate data sharing among researchers
- Uncover novel insights & predict trends

Development

- Enhance drug development efficiencies
- Accelerate market readiness
- Optimize clinical trial design

Connected R&D

Manufacturing & Quality

- Achieve operational excellence
- Reduce waste
- Enhance product quality

Supply Chain

- Optimize inventory management
- Mitigate supply chain risks
- Ensure operational resilience

Connected Manufacturing & Supply Chain

Sales & Marketing

- Targeted marketing
- Dynamic pricing
- Sales channel optimization

Patient

- Patient services
- Health equity
- Integrated access

Market Access & Med Affairs

- Reimbursement scenario
- Planning evidence generation

Connected Commercial

Finance

- Empower strategic decision-making
- Increase forecast accuracy

HR & Enabling Functions

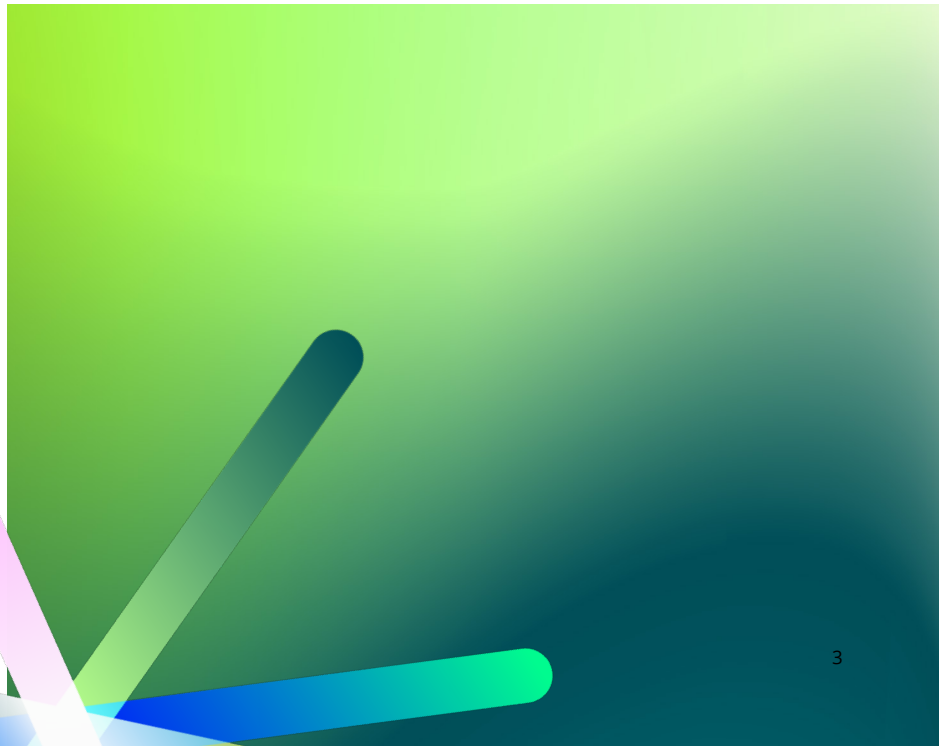
- Streamline operational processes
- Improve talent experience

Connected Enabling Areas

Figure 1. The life sciences enterprise value chain.

Across R&D, manufacturing and supply chain, and commercial operations, life sciences organizations can use a "data-as-a-product" approach—where the data itself is treated as a valuable asset that can be managed and utilized to help enhance processes and deepen insights, leading to numerous business benefits.

We refer to this kind of valued data as a "data product."



DATA PRODUCTS, ACROSS THE LIFE SCIENCES VALUE CHAIN

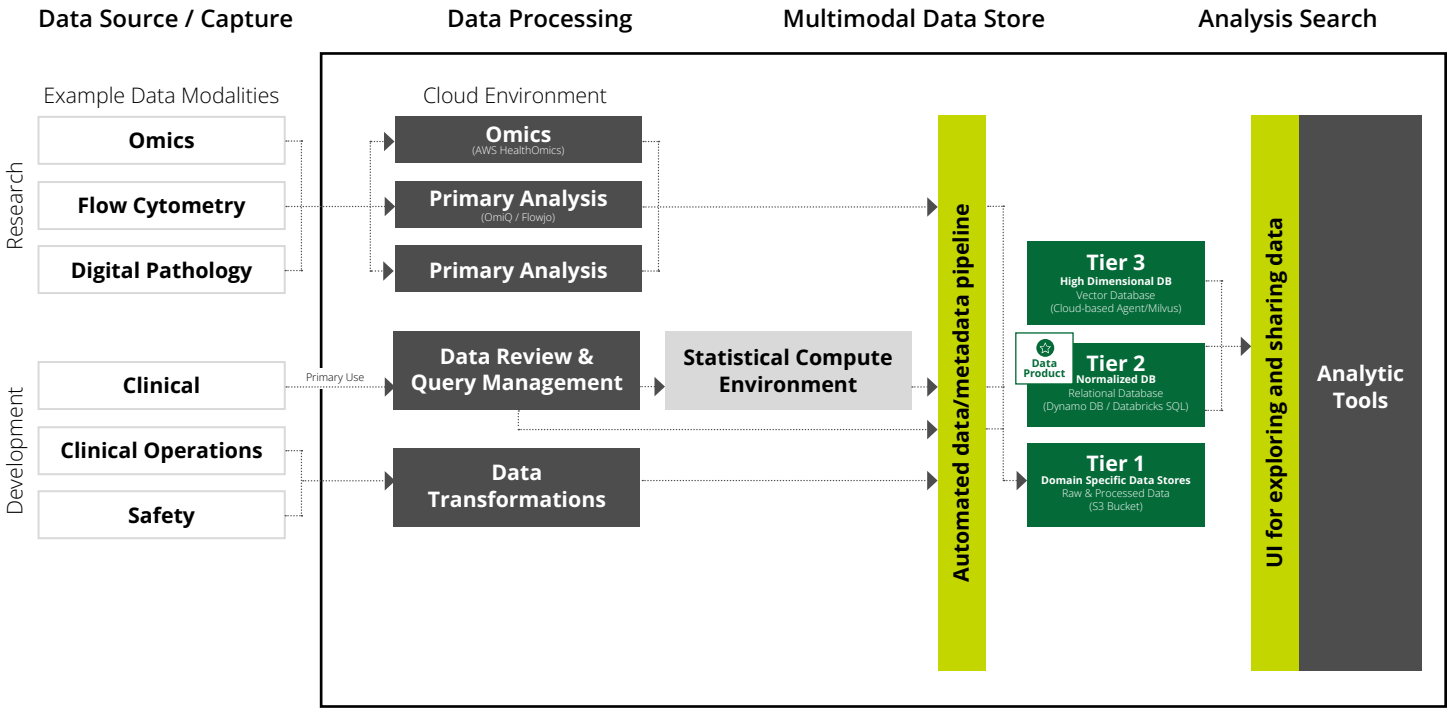
RESEARCH AND DEVELOPMENT

In R&D, leveraging data can drive improved science and bring more effective drugs to market faster through:

Real-world evidence and insights
Utilizing real-world data to inform clinical and commercial decisions.

Next-generation therapies
Innovating in personalized medicine and advanced therapies.

To put this into motion, life sciences organizations can combine a multi-modal data pipeline with storage, analytics, and search capabilities into a single platform. Such a platform can drive enhanced discoverability, analysis, collaboration, and scalability during the R&D phase. When each modality has its own data product implementation, R&D teams can then leverage the unified framework seen in Figure 2:



Automated Data Ingestion

Workflows for data intake/ data consumption from internal and external sources (CROs/vendors) to land data in a common landing zone with automated metadata extraction.

Standardized Data Processing

Automated extraction with a central framework that leverages consistent tools and systems for each modality to process data from raw to "analysis ready" data in a repeatable and standardized manner.

Multimodal Data Store

3-tiered, modular data mesh approach to support storage and retrieval scenarios optimized for modality specific recall and analysis.

Multimodal Data Integration

Leverage LLMs + Knowledge graphs to start driving rapid insight from multimodal datasets through textual anaylsis, semantic relationship, and data visualization.

Analysis Ready Multimodal Data

Easier access to multimodal data based on FAIR principles through exploration UI to search, find and access multimodal data for custom secondary analysis to enable the use cases.

Figure 2. An example of an early-stage R&D data product environment.

MANUFACTURING AND SUPPLY CHAIN

Once discoveries have been made and new therapies ready for production, manufacturing and supply chain become crucial next steps to utilizing data products.

With diverse inputs for biopharmaceuticals manufacturing and supply chain, and unpredictable global economic factors like tariffs, seeing the entire value chain end-to-end is critical.

The data-as-a-product approach can enhance production processes and quality control through:

- **Integrated evidence planning:** Coordinating evidence generation across the product lifecycle.
- **Dynamic business planning:** Adapting to market changes with agile data strategies.
- **Sustainability practices:** Ensuring compliance and operational resiliency through data transparency.

Figure 3 highlights an example of this manufacturing and supply chain data environment.

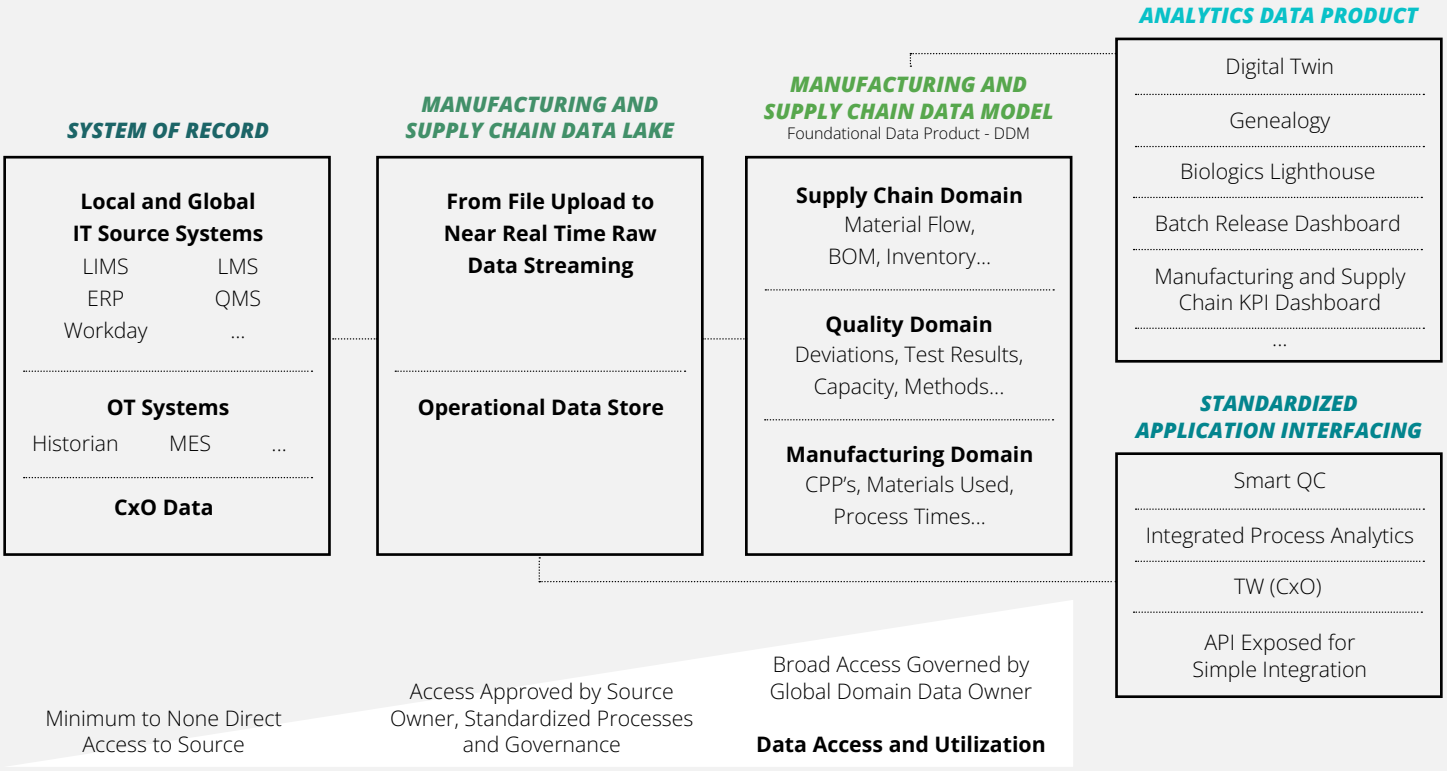


Figure 3. An example of a manufacturing and supply chain data environment.

The example outlines a data environment with a central repository and key systems like laboratory information management, enterprise resource planning, and quality management.

Data is categorized into domains: supply chain (material flow, bill of materials, inventory) and manufacturing (process parameters, materials, process times).

This structured approach enables advanced analytics across manufacturing and supply chain use cases like genealogy tracking, biologics analytics, and dashboards for batch release and KPI monitoring. Further, we're seeing these use cases across some of the largest biopharmaceutical companies today.

MANUFACTURING USE CASES

- **Digital twins**
  Apply digital twins to develop new products and therapies or improve existing ones, as well as modify or redesign equipment that products those goods.
- **Quality control**
  Monitor external factors (temperature, humidity, and other environmental conditions) that can impact drug production and storage, helping to ensure that the highest quality standards are met. resiliency through data transparency.
- **Overall equipment effectiveness and predictive maintenance**
  Measure manufacturing equipment productivity based on real-time sensor and equipment data. Use AI for predictive maintenance to prevent failures.

SUPPLY CHAIN USE CASES

Inventory optimization
Automate inventory replenishment systems triggering pre-approved requisitions when stock falls below safety levels.

Demand forecasting
Analyze historical data and market trends to accurately predict future product demand.

Delivery time prediction
Estimate when goods will arrive at their destination, streamlining production and distribution activities.

Supplier risk monitoring
Develop N-tier supplier maps using internal and external data to assess and mitigate risks in the supply chain.

Order processing automation
Automate order processing with virtual agents to reduce administrative costs and improve efficiency.

COMMERCIAL

Enhancing commercial operations through data-driven insights can help life sciences organizations boost customer engagement, product positioning, and sales performance, as outlined in the use case in Figure 4.

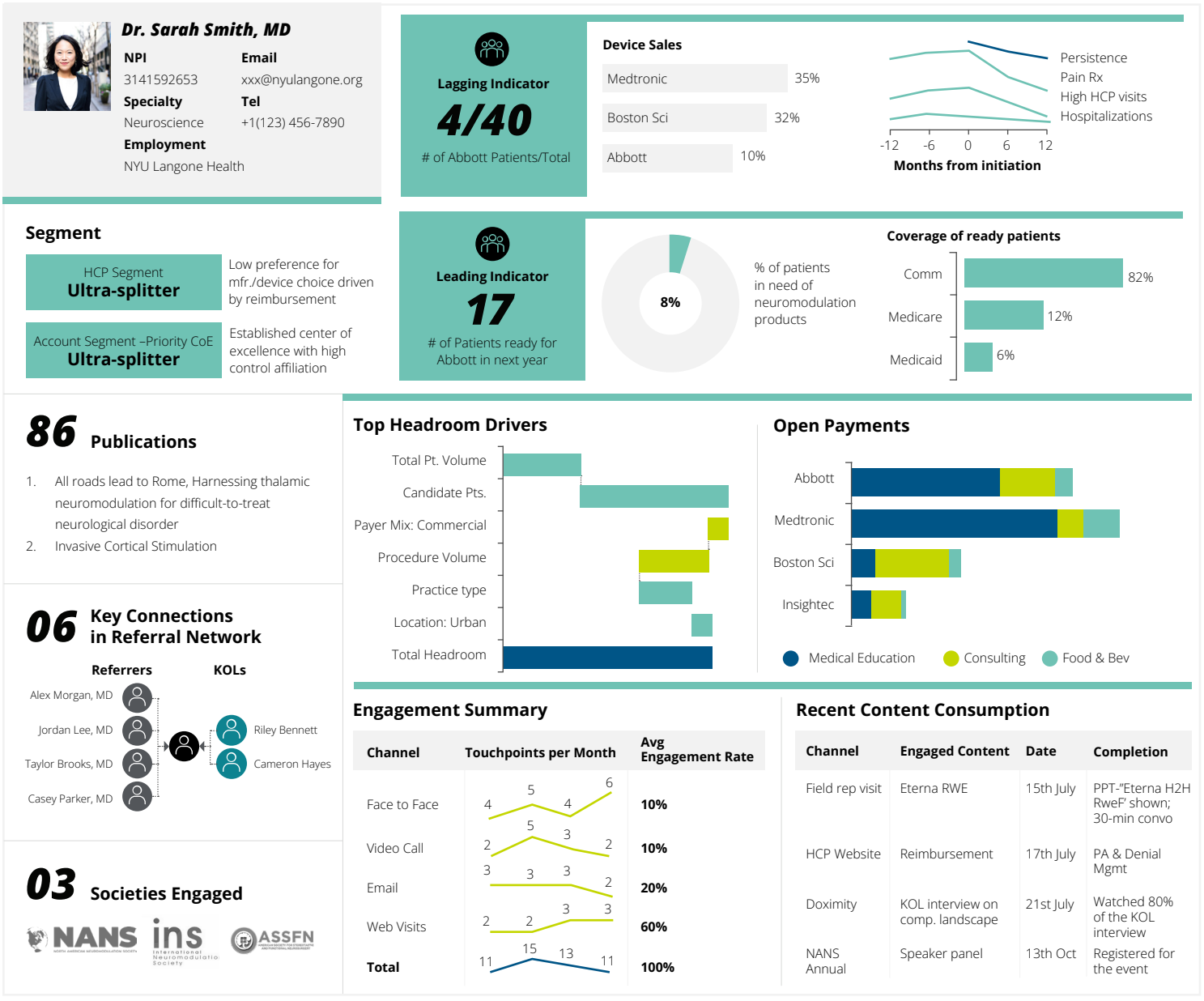


Figure 4. An example of a "customer 360" data product for enhanced customer engagement. This example is for illustrative purposes only.

In Figure 4, a life sciences organizations created commercial data layers to generate unified insights. These insights span beyond traditional data management capabilities to drive better predictive customer insights, personalization, and improved business outcomes.

Customer 360 data products can also provide integrated intelligence of target customers through AI-driven algorithms for demographics, behavioral profile and practice profiles.

They can be customized to include:

- **AI analytics models**
  of health care professional (HCP) and patient behaviors
- **Headroom analysis**
  to predict future script growth opportunities for a brand
- **Patient profiles**
  summarized by socioeconomics and ethnicity
- **Treatment patterns**
  categorized by diagnosis, lab test, treatment, adherence
- **Visualization**
  of all information encompassing a 360 data model
- **Promotion responsiveness**
  to dynamically adjust channel frequency
- **HCP brand share vs. account**
  for account selling, relevant HCP messaging
- **Patient demo trends vs. new diagnosis rates**
  to message HCPs on brand-approved gender and age indicators

AI-powered customer data products and decision systems provide dynamic services, creating hyper-personalized experiences for HCPs, patients, field operations, care coordinators and caregivers. These data products enable several high-value use cases, including:

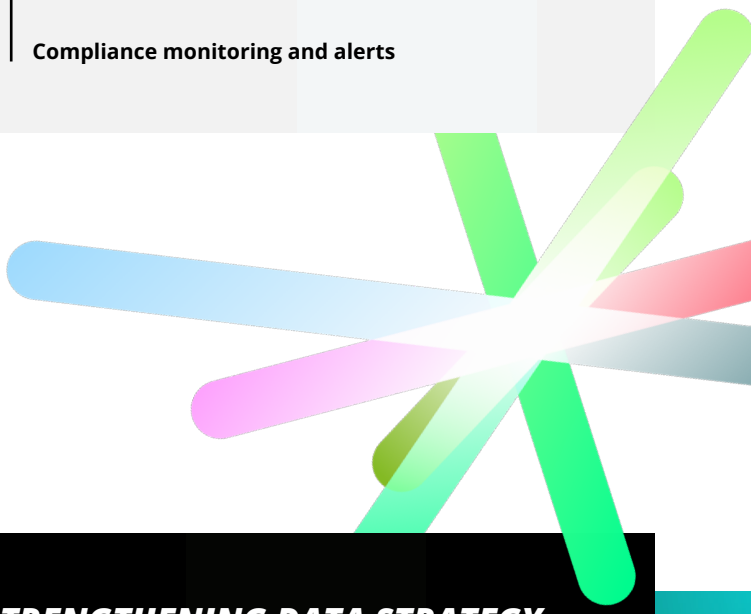
Omni-channel marketing and reporting

AI/ML-driven recommendations for sales reps

Dynamic targeting for HCPs and patients

Predictive models for demand forecasting

Compliance monitoring and alerts



THE KEY ELEMENTS OF A SUCCESSFUL DATA STRATEGY

As the examples across the enterprise value chain illustrate, a successful data strategy helps life sciences organizations with:

- Connecting business strategy to data strategy
- Emphasizing shared investments, capabilities, and governance models
- Building agility in technology architectures and investments
- Monitoring return on investment (ROI)

STRENGTHENING DATA STRATEGY THROUGH DATA MESH AND GENERATIVE AI (GenAI)

Data mesh is a new approach to data architecture that decentralizes data management, emphasizing domain-oriented data ownership and treating data as a product. This paradigm shift aims to improve scalability and agility in data handling. Four principles anchor a data mesh approach:

- Domain-oriented decentralized data ownership
- Quality data products
- Self-serve data infrastructure
- Federated computational governance

These principles create a scalable and efficient data management framework, enhanced by GenAI, which automates data processes to generate valuable insights from decentralized sources. Figure 5 illustrates this framework in detail.

LLMs and other core components of the stack are being commoditized rapidly – More and more, value will come from proprietary data powering tailored solutions.

PURPOSE SPECIFIC USE CASES

Domain-specific models trained on proprietary data

Examples:

- Automated Medical Review (MLR) of promotional materials for pharma company
- Evaluating donation impact for global healthcare organization

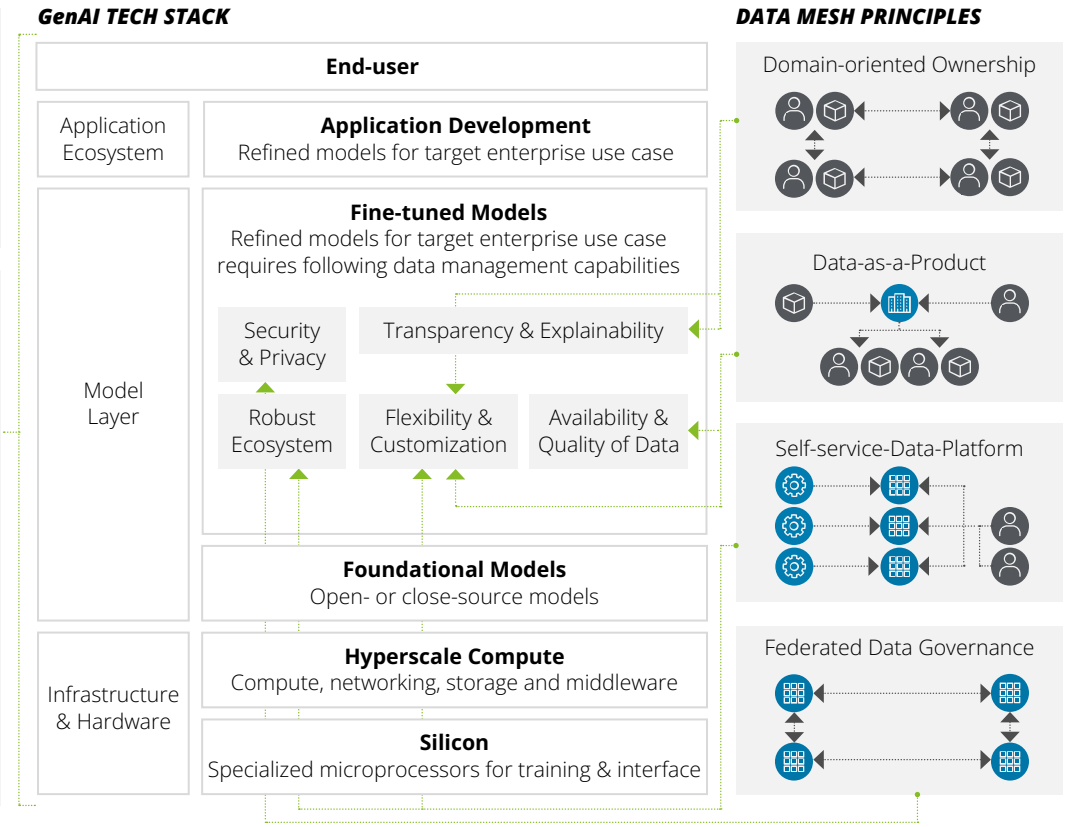


Figure 5. A data mesh-powered GenAI framework.

The data mesh/GenAI integration enhances AI capabilities with domain-specific data, enabling life sciences teams to leverage decentralized data for more accurate and actionable insights. It is already making significant impacts by:

Transforming clinical trials through integrated data analytics

Advancing personalized medicine by synthesizing genomic and clinical data

Optimizing supply chains by connecting diverse data sources to enhance efficiency and visibility

THE ROLE OF STRUCTURED AND UNSTRUCTURED DATA IN ENHANCING DATA STRATEGY

Understanding the distinctions between structured and unstructured data is crucial for leveraging new data approaches to drive innovation and improve patient outcomes.

Structured data
is highly organized, easily searchable, and typically stored in relational databases like spreadsheets and SQL databases. It offers ease of analysis and efficient retrieval but is limited by its rigidity and need for a defined schema.

Unstructured data
includes formats such as text files, emails, social media posts, videos, and images, like research documents and multimedia files related to patient feedback and clinical trials. It offers greater flexibility and richer insights but requires advanced tools like natural language processing (NLP) and machine learning for deeper analysis.

Leveraging unstructured data helps organizations understand patient needs, improve clinical trial efficiency, and enhance research and development. This dual approach to data management enables life sciences companies to innovate and respond effectively to the evolving healthcare landscape.

PUTTING DATA MESH AND DATA-AS-A-PRODUCT APPROACHES INTO PRACTICE

Transitioning to a data mesh or data-as-a-product architecture involves defining data standards, enabling Master Data Management (MDM) and Reference Data Management (RDM) capabilities, and ensuring continuous compliance and efficiency, as seen in Figure 6.

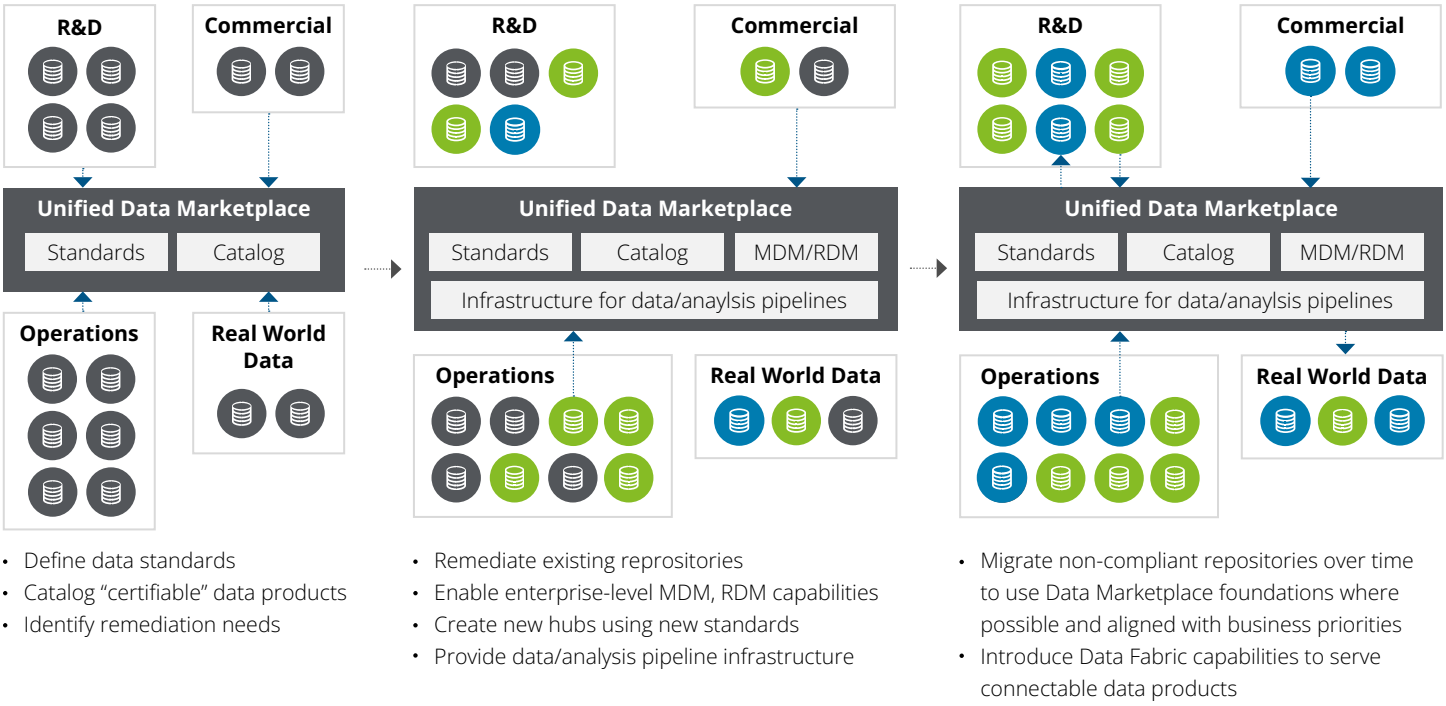


Figure 5. A data mesh-powered GenAI framework.

ACTIVATING A CONNECTED DATA STRATEGY

We've identified three operating models for activating a connected data strategy:

Decentralized
Grants autonomy to individual departments, allowing them to tailor data initiatives to specific research and patient care needs.

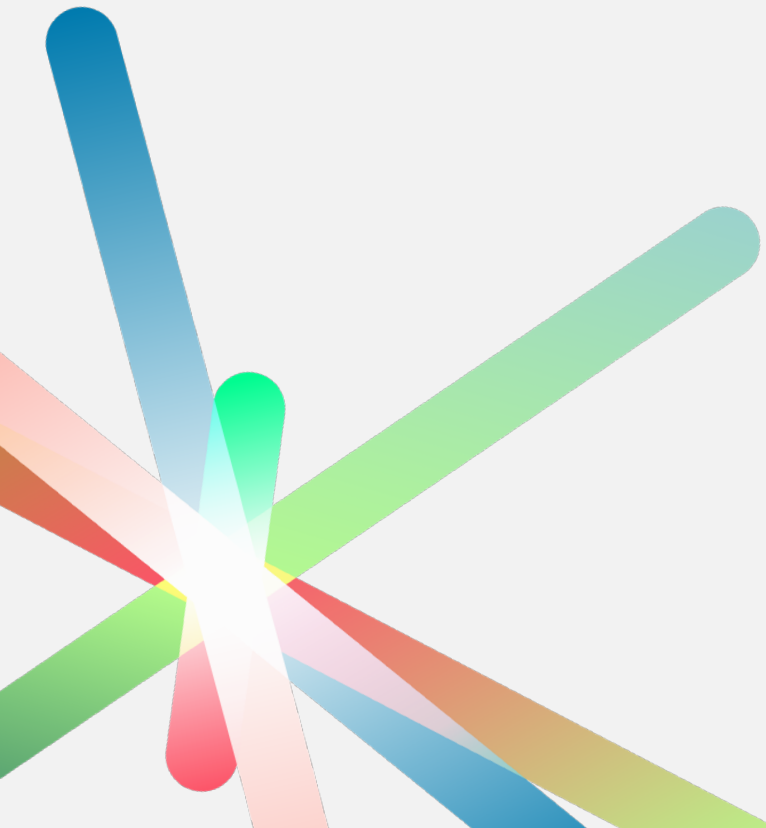
Informally federated
Balances autonomy with centralized guidelines, enabling diverse teams to operate independently while adhering to shared standards.

Formally federated
Centralizes data governance under a Chief Data Officer, enforcing enterprise-wide standards to help ensure strategic alignment.

By adopting the right operating model, life sciences organizations can integrate their data strategy with enterprise IT, enhancing innovation, regulatory compliance, and patient outcomes, as outlined in Figure 7.

Business Context	Operating Model	Highlights	Illustrative Examples
Product line organized, highly customer-focused	**Decentralized** 	• Increased autonomy over domain specific priorities • Increased focus on domain • Increased focus on value realization for business	Selection of a platform will be identified as a need by individual business function data leads / CDOs and will be brought up to Enterprise IT for platform vendor selection
Diverse businesses with an enterpreneurial culture and focus	**Informally Federated** 	• Mix of autonomy & federated guidelines • for funding Mix of a domain and cross- • functional approach Mix of Business and IT view for value realization	Each business function data leads / CDOs manage their own data assets and platforms however, there is a council convened regularly to discuess shared guidelines that can be leveraged by all functions
Highly centralized organization structure that cost-focused with a top-down culture	**Formally Federated** 	• Balanced Autonomy and Control over funding • Increased focus on the cross functional data usage • Increased focus on the • Enterprise IT view for value realization	The central data lead / CDO defines and enforces the standard, platforms, and other decisions at a federated level while individual business function data leads / CDOs control the data-to-day operations

Figure 7. Operating models to activate a connected data strategy.



HOW THE DATABRICKS PLATFORM CAN ENHANCE LIFE SCIENCES DATA STRATEGIES AND ACTIONS

Databricks offers unique technological advantages beneficial to life sciences organizations:

- **All data types**
  Databricks lakehouse architecture processes all structured, unstructured, and semi-structured data.
- **Unified data platform**
  For data engineering, data science, and machine learning.
- **Scalable cloud infrastructure**
  Handles large volumes of data.
- **Advanced analytics and AI**
  Advanced analytics and AI capabilities allow organizations to derive actionable insights from data.
- **Real-time data processing**
  Real-time patient monitoring and predictive analytics.

Deloitte. | databricks

WHERE TO NEXT?

A robust data strategy is essential for overcoming challenges and seizing opportunities in health care and life sciences.

The Deloitte and Databricks alliance offers a strategic advantage by combining deep industry experience with cutting-edge technology to transform data into actionable insights.

By focusing on the four pillars of data governance, integration, advanced analytics, and scalability, organizations can break down data silos, improve data quality, and accelerate AL adoption. These strategies can not only enhance operational efficiency but can also drive innovation in areas like personalized medicine and clinical trials.

Let's get started.

GET IN TOUCH

Shri Chary
AI & Data Managing Director
Life Sciences & Healthcare
Deloitte Consulting LLP
shrchary@deloitte.com

Roy Anjan
AI & Data Managing Director
Life Sciences & Healthcare
Deloitte Consulting LLP
anjroy@deloitte.com

Nirav A. Mehta
AI & Data Principal
Life Sciences & Healthcare
Deloitte Consulting LLP
niravmehta@deloitte.com

Steven Zhang
AI & Data Managing Director
Life Sciences & Healthcare
Deloitte Consulting LLP
shenzhang@deloitte.com

Raveen Sharma
Core Business Operations
Life Sciences & Healthcare
Deloitte Consulting LLP
ravesharma@deloitte.com

SOURCES

1. https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/global-trends-in-r-and-d-2024-activity-productivity-and-enablers
2. https://www2.deloitte.com/us/en/insights/industry/health-care/life-sciences-and-health-care-industry-outlooks/2025-life-sciences-executive-outlook.html