



Unified Data Platform Build vs. Buy: Key Imperatives for Life Sciences IT Organizations

How Commercial Life Sciences Maximize Data
Investments with a Data Fabric

August 2023

Unified Data Platform Build vs. Buy: Key Imperatives for Life Sciences IT Organizations

How Commercial Life Sciences Maximize Data Investments with a Data Fabric

Puneet Bhatia, Senior Director, Axtria Inc.

Should I Build or Buy a Data Management Solution?

Life sciences companies continually face the question of building or buying their critical data management solutions. A challenge IT organizations face is managing the expanding talent required to build an in-house data ecosystem. The market continually pressures commercial organizations to invest in the latest tools and technologies to gain a competitive advantage. Maximizing how far your dollars go and weighing the pros and cons of a build versus buy approach needs constant balance.

Custom-built solutions offer the promise of more control. However, the proliferation of these investments created data silos, leading to ever-increasing costs to operate, maintain, and extend the solutions. And an organization's spiraling investment to accommodate even further market changes creates a risk to business operations.

An excellent example of this is the process surrounding operational data quality. As life sciences companies go through cycles of commercial products, they may scale back resources or contractors. This inevitably leads to "technical debt" and a data quality operation that hasn't been modernized. Demands become more complex and require further expense to keep up with new compliance, regulations, and technical standards. Therefore, purchasing a pre-built platform managed within a company's research and development function results in an efficient investment.

Control and customizability do not have to be sacrificed when buying a life sciences commercial data platform. This is especially true if the data platform capabilities are integrated with the principles of a data fabric. Leading industry analyst Gartner defines data fabric as the design framework "for

attaining flexible, reusable, and augmented data integration pipelines, services and semantics."¹

A data fabric approach enables life sciences organizations to leverage capabilities that are available off-the-shelf while customizing additional elements. This approach mitigates the organization's triple investment dilemma of first building, then integrating, then managing an ongoing investment to keep each component up to date.

This white paper aims to guide life sciences companies on maximizing the value of their current data investments in a build versus buy environment. It highlights the opportunity to stitch together their data ecosystems using a standardized, commercial, off-the-shelf data fabric approach.

Technology Changes Impacting the Shift from Custom to Commercial Solutions

Technological advancements in data and analytics are constantly increasing in pace and complexity. Life sciences firms cannot continually invest to keep up with the changes. As an enterprise technology stakeholder, your responsibility is to keep the costs in line with the business operations and value achieved.

Choosing an off-the-shelf solution places the burden of improving capabilities on the provider rather than the buyer. These solutions also mitigate some of the challenges of custom-built approaches by spreading the cost of constant ecosystem advancements among a broad customer base.

There is a long history of companies migrating from custom-built to commercial solutions. In the late 1990s, customer relationship management (CRM) systems were almost all custom-built, but very few are today. The late 2000s saw

the successful migration of solutions from custom legacy technology to cloud-based software-as-a-service (SaaS) technology. In these cases, companies shifted to buying commercial solutions because of the complexities of managing technological change.

The same trend is happening now across data technology implementations, and data warehouses are a good example. If companies still choose to go with custom-built solutions, they often compromise on features. Unrelenting implementation costs and integration hurdles make building a complete custom solution difficult. The approach to tackling these issues merits an introduction to the data fabric concept, which is being widely adopted for unified data platform implementations. In the next section, we will delve into the key elements and reasons IT leaders find value in implementing a data fabric across their data landscape.

The Data Fabric Approach and its Advantages

What makes the data fabric approach so advantageous is that each layer is not a siloed capability. Each layer must be decoupled and interoperable with the next. In the same way microservices broke down monolithic applications and created isolated components that were maintained and scaled separately, the data fabric insists each component adhere to the same principles but with a specific reference to data. Figure 1 below shows the core data fabric layers and how critical commercial data platform capabilities are called out within each layer.

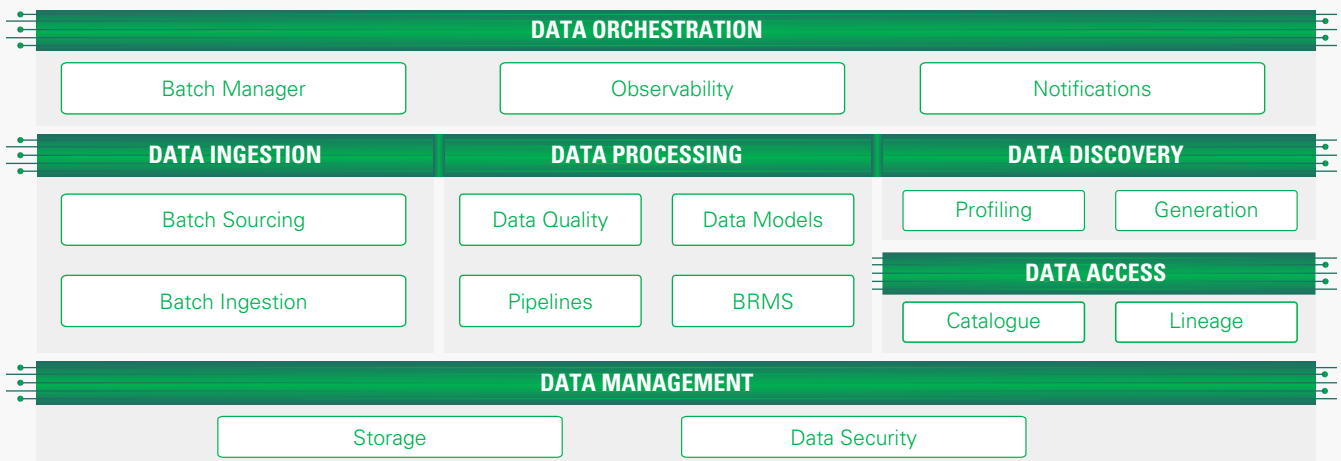
The advantages of data fabric can be summarized as follows:

1. *Data democratization and interoperability:* It enables organizations to use existing investments and integrate them more widely across multiple data products. For example, when the organization wants to swap out a solution for data quality, the effort to make that change is efficient since integrating data quality into the data pipelines is only an application programming interface (API) call.
2. *Improved data governance with integrated metadata:* When a data platform has fully integrated metadata, the data consumer can easily explore the data object's structure, data quality rules, when the last time data was loaded, and the business rules version applied to the process. This integration removes roadblocks to governing data and the associated rules that process it.
3. *Streamlined data product maintenance to increase time to insights:* By having a streamlined data engineer and consumer experience, the time to create and manage data products reduces. In 2021, Gartner stated that adopting a data fabric approach "reduces time for integration design by 30%, deployment by 30% and maintenance by 70%."²

Platform Buy Versus Custom Build: The Case for a Data Fabric

A modern data management stack incorporates several integrated processes that must be managed across end-to-end data workflow layers. While a custom-built stack covers some crucial components, deciding how to handle future optimization has proved difficult. Two possible alternatives include:

Figure 1. The data fabric layers



Source: Atria Inc.

Option one: Extend investment in a custom build (every two to four years)

With this, companies continue to invest in the present ecosystem by adding more custom technology, which necessitates tremendous work and has a high probability of breaking current system integrations with each new update.

Motivation for option one

- A custom build, tailored to your company's specific needs, is most useful when tech stacks and feature sets do not change much each year.
- A bespoke system allows for more control over the existing elements as well as elements the company believes would best fit into the legacy environment in the future.



OPTION ONE

Concerns with option one

- Technology acceleration forces frequent updates or changes across the entire build stack every two to four years, triggering a trade-off between value and capability. In addition, sourcing knowledge on newer technologies and tools will impact ongoing implementations.
- There is a high probability of breaking current integrations with each new update.



Many companies still build and operate data stacks. However, the life sciences industry is gradually adopting a unified data fabric, moving to a more stable environment.

Source: Axtria Inc.

Option two: Buy a commercial data fabric capability (safeguard custom investments)

With this option, a company can buy a modern commercial data platform optimized and pre-configured for different pipelines and procedures for use cases specific to life sciences.

Motivation for option two

- Deploying a uniform data fabric will help stitch existing technology investments, providing enhanced agility, lower operational costs, scalability, and self-documentation advantages.
- Investing in a commercial data platform will mitigate the organization's ongoing two- to four-year upgrade costs.
- Reducing your organization's R&D costs will be easier by leveraging a vendor's larger R&D investments. The vendor will keep up with changes through continuous updates while maintaining data pipeline integrity, end-to-end data quality, and functional workflows.



OPTION TWO

Concerns with option two

- Care must be taken when choosing a vendor partner who knows your organization's commercial life sciences needs.
- Evolving your organization's data strategy into a data fabric for the first time will take effort and support. But once completed, incremental changes would be intuitive and far less time-consuming.
- Resources would need to be trained on the platform, but once hands-on, clients will be able to utilize the self-service capabilities of the platform.



Source: Axtria Inc.

From a long-term perspective, option two proves to be more effective since a data fabric will preserve current investments and optimize capital expenditures and operating expenses prudently for the future.

A buy versus build analysis demands a detailed comparison across key parameters to determine operational and deployment value based on prior implementations (as seen in the tables below). Based on Axtria’s observations, operational value or day-to-day efficiency is becoming far more critical than deployment value and build investments.

Table 1: Operational Value Impacting the Build vs. Buy Decision

	BUILD (Custom solution)	BUY (Platform)
Agility	<ul style="list-style-type: none"> • Six weeks: the amount of time reports from new data sources can take • Three months: the average time required for building projects for brand analytics use cases with legacy infrastructure • Delays to onboard specialty hires, design, and execute advanced artificial intelligence/ machine learning (AI/ML) use cases 	<ul style="list-style-type: none"> • Four weeks: reduction in time from query and report generation • Three-fold increase in analyst productivity with self-serve modules • Design and implement new datamarts in less than one day
Operational Cost	<ul style="list-style-type: none"> • Change request time and opportunity cost due to legacy systems • Cost of adding new regions or business units • Limited automation of the current ecosystem resulting in higher turnaround time use cases 	<ul style="list-style-type: none"> • 30% reduction in costs through increased productivity and lower cost of hiring • 35% faster change fulfillment with automated processes and self-service • 20% saved from custom implementation
Data Quality	<ul style="list-style-type: none"> • Observational data quality pipelines identify but do not fix issues 	<ul style="list-style-type: none"> • Operational pipeline-ready data quality support identifies errors and addresses them using prescriptive rules, saving 35% in processing time
Product Upgrade	<ul style="list-style-type: none"> • Custom data warehouse (DWH) needs rebuilding every four to five years at a cost of \$5 million to \$12 million • Product upgrades not available; requires the generation of a new project 	<ul style="list-style-type: none"> • Cost of Product upgrades covered in the SaaS license; no additional charge • Innovation in the form of continuous new releases; no need to worry about technology obsolescence
Scalability	<ul style="list-style-type: none"> • Requires employee training across multiple platforms 	<ul style="list-style-type: none"> • Single platform simplifies organizational rollout of training • Ability to scale the acquisition and processing of large quantities of data
Data Efficiency	<ul style="list-style-type: none"> • Three months: significant effort needed for data onboarding and processing across the ecosystem with limited visibility on business rules applied 	<ul style="list-style-type: none"> • 40%+ effort reduction on data onboarding, collection, and preparation activities • Four weeks: average time needed to build projects using recipes

Table 2: Deployment Value Impacting Build vs. Buy Decision

	BUILD (Custom solution)	BUY (Platform)
Speed to Deployment	<ul style="list-style-type: none"> • 14 months to stitch a custom data fabric 	<ul style="list-style-type: none"> • Create multi-tenant environments in less than 10 minutes • Save up to 50% in deployment time • Seamless integration of new services
Cost of Deployment	<ul style="list-style-type: none"> • Service development costs, SME time (duration of development) • Large, expensive resource pool required for custom build 	<ul style="list-style-type: none"> • Save over 50% in deployment cost from custom build
Self Documentation	<ul style="list-style-type: none"> • None 	<ul style="list-style-type: none"> • Self-documentation: DWH business rule configuration, data marts, business processes, workflows, extract, transform, and load (ETL)
Pre-Tested Data Domains	<ul style="list-style-type: none"> • All from scratch 	<ul style="list-style-type: none"> • Multiple domains added per release

Key Benefits of Data Fabric-led Buy Implementation

As evident in the analysis from the tables above, a unified data platform has distinct advantages across operational and deployment parameters. Data fabric-based platform deployments have the following key benefits:

1. Innovative, unified architecture.

Axtria DataMAX™ is a unified data fabric, the product of years of experience helping life sciences customers with agile and integrated data management. Axtria DataMAX™ abstracts the tribal knowledge into code to manage pipelines and processes. The out-of-the-box solution provided by Axtria DataMAX™ is less expensive than hiring a large team of programmers attempting to recreate the code.

2. No “rip-and-replace” of the existing investment.

The Axtria DataMAX™ intelligent data management fabric seamlessly fits into an existing ecosystem, helping to preserve the investments already made while providing trusted analytic and report-ready data sets for critical business decisions.

3. Avoids unnecessary business costs.

Seamless integration into the existing ecosystem minimizes unnecessary capital expenditure.

4. Confidence in data security at rest, in motion, and in use.

Deploying in the company’s virtual private cloud (VPC) environment is achievable because of an adaptable fabric. No environment or data security is compromised as a result.

5. A self-documentation platform.

Axtria DataMAX™ “self-documents” as data discovery happens in parallel with product installation, making data profiling and analysis of existing data sources much quicker and automated. This approach also helps understand data attributes and identify quality issues with built-in checks. Compared to traditional methods of building a DWH, this self-documenting capability allows companies to glean information directly from the network by simply attaching the fabric and letting it self-document.

6. Flexibility: the inflexible truth in data management.

The Axtria DataMAX™ platform enforces best practices across any life sciences data stack to speed up deployment, simplify maintenance and enable seamless changes to both the tech and business layers. Additionally, it provides a wide range of options used to orchestrate and manage data, including DWH, data quality, ETL, and cataloging.

7. A self-service approach.

The Axtria DataMAX™ no-code data fabric empowers users to create, experiment with, and apply business rules to analytic processes with minimal assistance.

8. Significant cost reduction.

Quicker turnaround times and lower licensing costs result in increased financial control and the possibility to scale up or down within companies’ budgets.

9. Rapid time to value.

A pre-built library and marketplace of data connectors, reports, and key performance indicators enable rapid insight generation across business functions.

10. A lower total cost of operations (TCO).

Technology and design choices minimize technical debt, lower TCO, and facilitate future improvements without the need to rip and replace the tech stack.

Bringing a Data Fabric into Reality for Life Sciences

A unified data fabric strategy provides life sciences companies with a more promising path to effectively managing increased market agility and predictable data-driven decision-making. Your data fabric needs the following components to be effective:

Operational Data Quality: Data quality is at the heart of any data management program and is essential for successfully adopting a data fabric architecture in your organization. Most enterprise data quality solutions are great for observability. However, Axtria DataMAx™ provides operational data quality. It integrates directly with the customer's data repository and stores all error records in the customer database, allowing workflows to branch execution or filter out troublesome data efficiently. In addition, your data remains secure as it never leaves the confines of your repository.

Operational Data Catalog: For a data fabric solution to be most cost-effective, it should embed a catalog into the working processes, where management can become part of operations, not merely the recipients of a final data presentation.

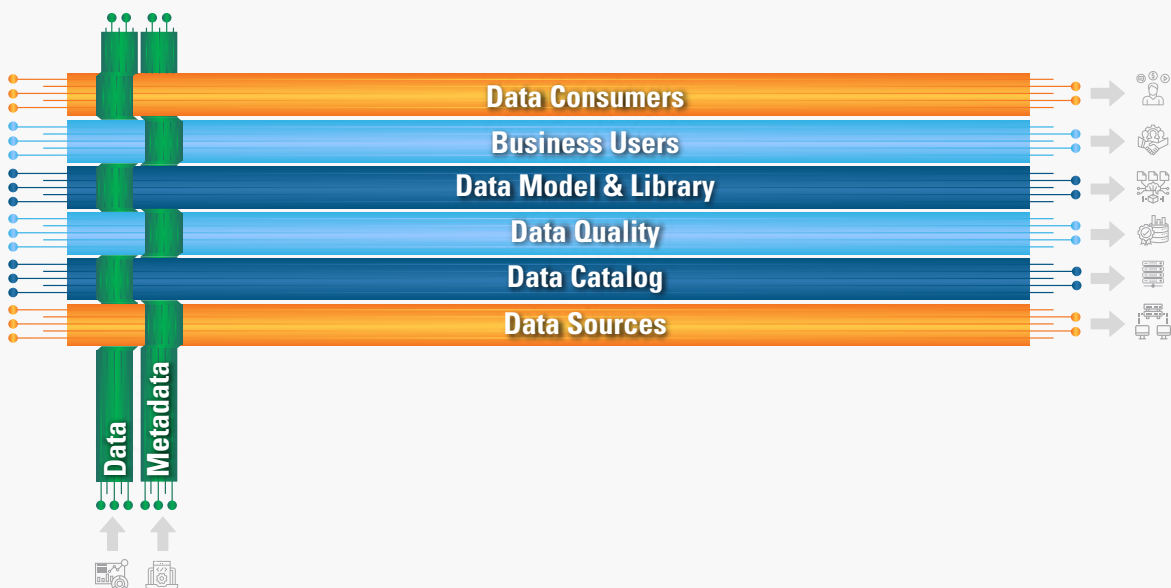
The catalog component is at the heart of the Axtria DataMAx™ data management fabric. It is an operational catalog that includes all interim states. It allows businesses to see the data quality in each temporary state, view the lineage connecting those interim states, and verify attribute definitions at each point.

Don't Rip & Replace, Integrate with a Data Fabric

Axtria DataMAx™ incorporates integral data components, such as data integration, operational data quality, active metadata, AI/ML, data cataloging, and more. It also employs traditional methods while allowing the organization to incorporate new technologies to avoid a "rip and replace" experience.

Pre-built Life Sciences Data Models: The key to any data implementation's return on investment is modeling "once for all" instead of repeating data modeling for each new application. It is less expensive to deploy and uses a pre-built common data model with the advantages of the collective knowledge of data scientists and engineers. Furthermore,

Figure 2. Data Fabric



Source: Axtria Inc.

Axtria DataMAx™ expands pre-built support for more data sources and analytical use cases compared to a custom-built environment.

Platform Agnostic: The data fabric approach bridges and strengthens connectivity between data lakes, warehouses, and other applications. There is no “lift and shift” as applications and data remain where they are. Your present investments are protected because it works with various data warehouses, lakes, and more. The data fabric simply uses them in its operational pipeline.

Solid Data Integration Foundation: The data solution should support a variety of data delivery types, including ETL, streaming, and data virtualization, among others. Axtria DataMAx’s™ data fabric approach enables the orchestration of pre-existing ETL tools. It includes pre-built structured query language (SQL)-based templates for transferring data across all levels of the typical life sciences industry data model. Additionally, there is a custom transformation option for this SQL-based architecture.

Operational Lineage: Visual end-to-end data lineage is essential for ensuring a data fabric’s reliability. Data lineage in Axtria DataMAx™ is “design-time” rather than just “run-time,” as in most data lineage systems. Because the workflows are developed with a visual, drag-and-drop interface that gathers documentation on the fly, operational lineage is captured.

No-code Business Rules Engine: Axtria DataMAx™ helps incorporate and enforce business rules in the processing and reporting of data. It accommodates data challenges by managing a few integrated design patterns, providing flexible support, and maintaining business rules.

Assessing the Current State of Your Data Ecosystem

Licensed commercial data management platforms are gaining traction over custom-built environments. They enable rapid integration across all life sciences data sources while maintaining security, accuracy, and compliance. Data quality and business management standards can be implemented smoothly to guarantee that the data being processed is fit for purpose and trustworthy. Users can specify the

establishment of data marts for later consumption and reporting by analytics systems, models, or individual data stewards.

To help analyze where their company’s data ecosystem is headed, all life sciences companies must consider four questions. Having the right strategies for these questions will help improve the performance of data stacks and stitch actual operational value into their data fabric.

1. Are you able to manage data quality across the entire fabric?

Data quality is often related to the ability to incorporate point data checks for each stack component. In reality, data quality should manage quality across the orchestrated pipeline. Is it necessary to check whether your custom-built stack has stitched operational quality where the platform can monitor end-to-end data from when it is sourced and processed until provisioning? In Axtria DataMAx™, data quality is centralized, and the framework allows end-to-end quality management across the pipeline to identify, report, and plug gaps as they emerge with the most robust data linkage.

2. Is your custom-built data ecosystem self-sustaining?

While your company has already built a data stack, you might not have spent enough effort to overlay a state-of-the-art data management fabric layer where any changes to present custom code (due to ongoing technology upgrades) are constantly monitored and updated in the platform. With such a fabric, no single change will break the overall solution and you won’t need to spend massive amounts of effort and capital updating the stack every four to five years.

3. How much do you focus on operational value over deployment?

Traditionally, companies focus on the deployment of the data stack and expect to see operational value once completed. This notion couldn’t be more misplaced. Forward-thinking life sciences companies have realized that while deployment is important, robust operational value is more critical on an ongoing basis. That can be achieved only with structured, effective data fabric stitching right from the start.

4. Does your data ecosystem empower self-service operations?

The most significant leverage large life sciences companies are missing in their data stacks is a high degree of self-service, where each persona can tweak data pipeline components such as data sources, data quality rules, catalog classification, and more.



Unified Data Platform: The Clear Way Forward

The verdict between build versus buy is clear. Deploying a unified data platform with a data management fabric at its core will be critical to IT teams' decisions across the life sciences industry. Axtria DataMAx™ is the only available out-of-the-box platform that integrates and unifies existing legacy systems using a revolutionary approach to managing data. This innovative data fabric is highly scalable and effectively manages data from a multitude of sources, thereby reducing

costs and enabling IT and business leaders to collaborate effectively—resulting in increased deployment and operational efficiencies.

Axtria offers a structured engagement approach to help life sciences companies dive deeper. We share a true assessment of where your company stands and how it can benefit, even from the initial steps of implementing a robust commercial data management platform like Axtria DataMAx™.

Learn more at

<https://www.axtria.com/cloud-products/axtria-datamax-cloud-data-management/>



References

1. Gartner. Information technology: Gartner glossary. Data fabric. Accessed August 7, 2023. <https://www.gartner.com/en/information-technology/glossary/data-fabric>
2. Panetta K. Gartner top 10 data and analytics trends for 2021. March 15, 2021. Accessed August 7, 2023. <https://www.gartner.com/smarterwithgartner/gartner-top-10-data-and-analytics-trends-for-2021>



Puneet Bhatia

Senior Director, Axtria Inc.

E: puneet.bhatia@axtria.com

Contact Us


+1-877-9AXTRIA
connect@axtria.com

300 Connell Drive,
Suite 5000
Berkeley Heights,
NJ 07922

Disclaimer


Axtria® understands the compliance requirements behind personalization and we do not work with any personally identifiable data that can identify an end-customer of a business.

We have the strictest data security guidelines in place as we work with businesses to improve the experience for their customers.

 www.axtria.com

 facebook.com/Axtria

 connect@axtria.com

 Axtria – Ingenious Insights

 @Axtria

Founded in 2010, Axtria is a global provider of cloud software and data analytics to the life sciences industry. We help life sciences companies transform the product commercialization journey to drive sales growth and improve healthcare outcomes for patients. We continue to leapfrog competition with platforms that deploy artificial intelligence and machine learning. Our cloud-based platforms - Axtria DataMAX™, Axtria SalesIQ™, Axtria InsightsMAX™, Axtria CustomerIQ™, and Axtria MarketingIQ™ - enable customers to efficiently manage data, leverage data science to deliver insights for sales and marketing planning, and manage end-to-end commercial operations. We help customers in the complete journey from data to insights to operations.

For more information, visit www.axtria.com.

Follow Axtria on Twitter, Facebook, Instagram, and LinkedIn.

Copyright © Axtria Inc. 2023. All Rights Reserved.