

Driving a Data Strategy Initiative

September 2021

Driving a Data Strategy Initiative

Amit Saini, Sr. Director, Axtria Inc.

1. Introduction
2. Business Unit and Organization Wide Data Strategy
3. Approach for Arriving at An Effective Data Strategy
 - 3.1 Define the Goals and Analysis Requirements
 - 3.2 Gap Analysis - Compare New Requirements Against Existing Process
 - 3.3 Defining the Data Journey
 - 3.4 Governance
 - 3.4.1 Data Ownership
 - 3.4.2 Quality Standards and Quality Checks
 - 3.4.3 Error Reporting and Dashboards
 - 3.4.4 Governing Productionized and User Managed Processes
 - 3.4.5 Quality Improvement Planning
 - 3.5 Data Storage, Retention and Backup
4. Implementing the Strategy
 - 4.1 Identify the Sponsor and Strategy Champions
 - 4.2 Technology Considerations
 - 4.3 An Agile Process for Implementation
5. Conclusions

1. Introduction

At Axtria, while we work with some of the biggest Pharma companies in the world as their partners for delivering actionable insights, we realize that quite frequently one or more of their business units are looking at revamping their data strategy for data analytics.

In almost all large organizations, relooking at data strategy, like many other organization wide initiatives, is a process that is repeated as new needs crop up, technologies evolve, and organization goals and strategies change as well. Such re-examination may be warranted for the whole organization, a set of business units, or even for acquired companies/business units that are being merged.

In such scenarios, it becomes increasingly important to review the existing systems for sourcing and consuming data, both from a business and a technological viewpoint, to ensure that they are adequate for fuelling the business growth.

In this whitepaper, we aim to provide a process to review and evolve the data strategy for a business unit. This methodology can be extended for reviewing the strategy for multiple business units or even at the organization level.

The paper primarily focuses on identifying any shortcomings in the existing processes or gaps in the data requirements and guides the stakeholders toward a process to carry out this review in a methodical and comprehensive manner.

2. Business Unit and Organization Wide Data Strategy

If our initiative focuses on data strategy for a set of business units, we must appreciate that it will invariably be a part of an organization-wide data strategy.

The organisation-wide data strategies, especially for large organizations are rarely established in a single step. They evolve as the organization grows and consolidates its business needs, builds its IT systems, and builds its data and analytics strategies.

The organization-wide strategy generally evolves as data is consolidated and IT systems are integrated across business units. Such consolidations are huge change management exercises that organizations do not conduct very frequently.

At a business unit level, however, the strategy is largely driven by the business requirements for that unit. Depending on the function, these may differ from other business units in terms of the goals, the analysis, data requirements, dependence on external data sources and governance needs.

There are areas where the business unit strategy refers to and depends on the organization wide strategy.

Technological decisions for example, may be guided and owned mainly by the IT teams. Thus, the choices of data management and analysis platforms may be driven largely by the IT teams.

Areas where the business units would likely have more flexibility would be around the data processes, their governance, and integrations with upstream and downstream systems.



3. Approach for Arriving at an Effective Data Strategy

An outcome driven data strategy approach focuses on defining the target state (the outcome) and working backwards through the required analysis and data requirements.

The outcome is the way the final analysis and insights should serve the business needs. These outcomes are established by discussions with the business users who will be the consumers of the system.

If we break down the proposed approach for reviewing the data strategy, it comprises of four main steps:


- 1) Define the goals
- 2) Analyze the gaps in existing strategies (including data and governance) with respect to the goals
- 3) Define the data journey for meeting the goals
- 4) Define the Data Governance strategy (including the data retention, backup and other data handling policies)

3.1 Define the Goals and Analysis Requirements

A strategy must be defined to meet the goals of the business. As businesses evolve over time, their goals change, and the strategy must be re-evaluated and recalibrated to address the revised goals.

Discovering the goals of the business units that the strategy will support is among the first set of actions.





As an example, a business unit may be launching a new drug or enhancing the product line for a certain therapeutic area. In such cases, the business may be interested in measuring the market share of the product over a period. Depending upon the measures, the data may need to be syndicated from one or more data vendors.

In addition, there may be a need to collect and analyse the responses to email campaigns, conference feedback and other digital data inputs.

Such goals should provide the foundation for the strategy that will connect these goals to data and the required analysis.

A Use Case Approach

A use case approach focuses on driving the strategy from the outcome required to the data and process requirements. A use case here refers to the business users' requirements for insights, reports, and various analysis that they require to better govern their business strategies.

Use case definition is one way in which the group working on a data strategy can discover and document the goals of the business. The approach focuses on capturing the business need and the required analysis to meet that need. A use case document would typically capture:

- the comprehensive description of the analysis that the business users require
- the business purpose that it would serve,
- the data required for the analysis, along with its source description
- the validation and data quality checks that should be done before the analysis
- Any specific guidelines about the analysis algorithm, logic, and business rules

A Use Case Template is provided with this paper, as a supplementary aid that can be leveraged for this use case definition exercise.

3.2 Gap Analysis - Compare New Requirements Against Existing Process

Once the business goals are captured along with their data requirements, it is useful to compare the data requirement against the existing data.

If the "as-is" process is well documented, a comparison of the new goals and its analysis can be made easily with the existing data and processes. This will help identify the data gaps and any new data requirements.

If the "as-is" process documentation is not available, it is strongly advised to create that during the goal capturing phase. This will help not only for identifying new data requirements but also in validating that all existing needs are eventually met in the new strategy and approach.

A comparison of the new goals, analysis and data requirements with the existing data, analysis and processes would make it easier to identify the modifications required for each of the areas respectively.

It would also make it easier to summarize and explain the scope of work to adapt to the new strategy and thus also lead to a better overall planning of the initiative.

3.3 Defining the Data Journey

Planning for the journey of the data is as crucial as sourcing the data itself.



Some of the key questions to be asked for planning the data journey strategy are:

- 1) What data sets are required to meet the objectives identified for the data strategy?
- 2) What are the frequencies for the data that is coming in? What are the specific constraints for the upstream data sources in terms of data format and quality?
- 3) What form and shape will the data have at the consumption stage?
- 4) Do we need to prepare data in different formats for consumption – for example, some downstream systems/applications may consume data directly from a data warehouse, while some may need data files to be placed in specific locations?
- 5) Are we planning to review and revise the data needs of the downstream consumers while we plan the data journey?
- 6) Which systems, processes, applications, tools will be required at the end of the journey to consume the data?
- 7) Are we considering changes in the data processing – like combining some processing stages or shifting them from one technology platform to another?



3.4 Governance

Governance is one of the topmost considerations in shaping the data strategy.

Governance around data and data management processes are most important for building trust in data, increasing the value of data for users and meeting the compliance needs for the business units and the organization.

Governance in the data management strategy would cover the data definitions, quality management, ownership of the data and processes and methodologies for planning, oversight, and control over management of data, the use of data and data related sources.

Governance would also include derivation and lineage of the data - where does each consumption-level attribute come from, and how is it transformed as part of the data journey?

This is a key part of understanding, and therefore trusting, data.



3.4.1 Data Ownership

In a robust governance strategy, there must be clear definitions of the various roles that people / groups will play in managing the data and related systems.

There must be clear responsibilities for data owners who are responsible for the updates and quality of data at various stages of processing, and for data originating from different systems and sources (within or outside the organization).

Typically, there are also clear roles for people who will support the usage of data, answer queries about the data and decide on the actions when users have feedback on the data that they consume.

3.4.2 Quality Standards and Quality Checks

Unless the data has the right quality at the time of consumption, it will not be of use.

Quality standards and checks are therefore an integral part of the overall data strategy. Equally important is the way these are implemented and maintained.

Quality management frameworks implemented in the organization must be agile enough to manage the data changes over time and help implement corresponding changes to the data quality checks implementation.

The framework should allow for an easy understanding of the quality of raw data including the problems that it had. It should also allow the governing team to decide how these problems can be addressed and the overall processes improved to ensure that quality issues are managed and prevented eventually.



3.4.3 Error Reporting and Dashboards

The strategy around governance must also allow for constant monitoring of data quality and usage.

Error reporting frameworks are important for highlighting the data related issues in a timely manner and help in identifying, reporting, and controlling data quality issues. They also play an important role in measuring the effectiveness of the governance processes in the overall data strategy.

3.4.4 Governing Productionized and User Managed Processes

Data processing here refers to the end-to-end data ingestion, transformations, provisioning, and final analytics journey of the data in the system. One of the ways to look at this processing journey is to view it as comprising of productionized and user managed processes.

The productionized processes are those that are standardized, regular, fully automated and run typically without any human intervention.

User managed processes are generally the ones that are not automatically triggered but run by users on demand. They may also be the ones that would require more frequent changes to processing as well as business rules.

For example, processes for building data marts and data lakes would largely be standardized and automated while those for running some business unit specific models and insights may be largely user triggered on demand.

The ones that are fully automated end to end would be under stricter control with respect to any changes/modifications.

The ones that will be user triggered would normally need to be modifiable by a larger set of business users by way or business rules. However, to ensure that there is the right level of access, security and data and model re-usability, there must be appropriate processes for governance of these user managed processes as well.


3.4.5 Quality Improvement Planning

The governance of data and related processes is all for meeting a key goal – improve the quality of data. There must be processes for continuously monitoring the governance effectiveness and taking actions for consistently adjusting the processes and improving the data quality and its usage over time. Thus, the governance policies and processes must include appropriate audits, measurements, issue reviews and process reviews.

3.5 Data Storage, Retention and Backup

An important part of the overall strategy is to decide the lifecycle of the data.





This lifecycle defines what data should be stored, how it should be stored and how long it should be retained.

The storage and retention are guided most by the compliance requirements and business needs for an organization.

For Pharma companies, there are several compliance requirements that determine how the organization must handle the data. Regulations also require organizations to keep records of, and be able to show evidence of, how the data has been handled, how it has been accessed and how it is stored.

The business needs too, guide how the data will be stored and how long it should be retained. For example, if the business requires past data from the last three years for analysis and reporting, then that specific data must be made readily available for that period before it can be backed up for long term storage.

Thus, governance of data should comprise of the policies and the processes for handling the data through its complete lifecycle.

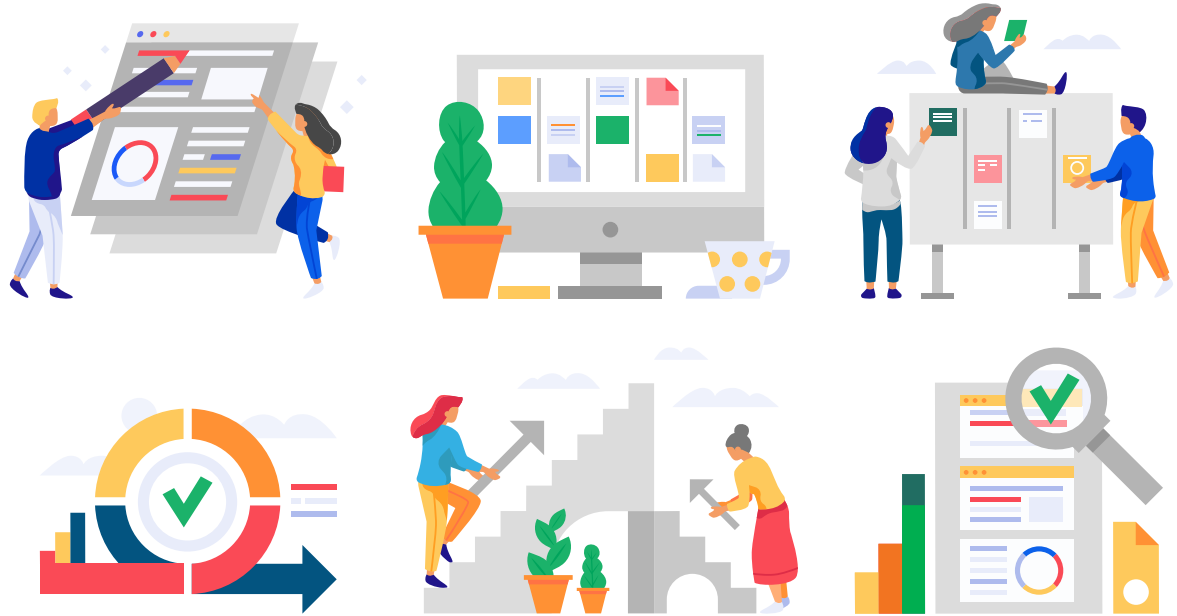
Some points to consider while backing up and retaining data are:

- 1) Classifying data by source – identify whether the data is internal and already stored and backed up in the organization, or if the data is external and must meet some compliance requirements for backup and retention. This classification may also influence the way data will be stored and backed up – in its raw form or processed form
- 2) Identifying and defining the data life cycle – identify which data will be consumed once vs re-used over time
- 3) Identifying if the data will be versioned - for example, sales data coming in every week will be separately stored for each week, but its consolidated monthly version will also need to be stored for analysis and reporting. Consolidated data may also get revised with every weekly update and hence would need to be versioned every time it is updated
- 4) Identifying if inter-related data needs to be stored together – for example, multiple data sets used for a particular process or analysis may need to be stored together to maintain lineage and data processing information for future reference or data diagnosis

4. Implementing the Strategy

We have described the key ingredients for an effective data strategy that will help a business unit, or an organization, define the principles, processes and policies that will shape its overall data strategy.

When it comes to implementing the strategy, an organisation requires a focused effort and a dedicated team to help put the strategy into practice.



4.1 Identify the Sponsor and Strategy Champions

Putting a data strategy into action is a considerable change management exercise, especially if it requires significant changes to the existing systems and processes.

It requires strong communication and follow ups with various stakeholders, especially when there is transition of systems, technologies, processes, or new governance policies involved.

Such initiatives must be managed like any other large change management initiatives for an organization. It is imperative that such initiatives have a senior sponsor in the organization who has the conviction and authority to drive such an initiative.

Like any other change management exercise, it also requires a healthy number of strategy champions who will lead others by example in implementing the strategy and help their peers in resolving related issues.

A data strategy initiative should be led by a team comprised of champions from both business and IT teams. This team would normally be nominated by senior management and should be carefully chosen from among people who are able to collate information from their respective business areas and influence the processes in their respective fields of operations.

There should be a continuation of the team through the implementation phase as well as later for managing the changes and evolution of the data strategy.

4.2 Technology Considerations

In this article we have described a methodology to define an effective data strategy within a business unit or organization.

Implementation of the strategy has several implications for technologies involved and that would typically involve the IT team stakeholders as well.

It is critical that, while taking up a data strategy initiative, an organization looks at its technology implementations and identifies opportunities to revise and optimize them.

With the advent of technologies and new solutions coming up every few months, an organization can look at the cost and benefits for making technology stack changes as part of its data strategy.

4.3 An Agile Process for Implementation

Once the strategy is developed and technology decisions have been made, the implementations take the focus, and the business unit or organization must set up a project team for implementation.

It is best to design the implementation approach in a phased manner compared to a big bang change for all teams involved. This approach not only reduces the complexity of implementation, but also reduces risk of failure by way of early reviews and feedback opportunities.

The business can plan the implementations by teams or products or use cases. This decision would define the overall approach for implementation and would let people understand how and when the strategy implementation will affect their existing practices.

Data reconciliations should be planned for downstream systems to validate that those new implementations bring new value, while servicing the existing systems and their integrations as is.

Based on Axtia's experience with several customers, we suggest following an agile approach.

- 1) In this outcome driven approach, the planning is guided by the outcome – completed and fully functional use cases that business users can start to explore and use.
- 2) A set of related use cases and their required processing may be classified as epics (as per the agile terminology). Each of these epics may be completed in typical two-week sprints with each sprint targeting the completion of one, or more, use cases. These use cases can be validated, demonstrated, and reconciled against existing systems or business requirements.
- 3) Such an approach is very effective for identifying potential issues in the designs, and for securing early feedback from various technical and business stakeholders.
- 4) This approach also addresses a very important need for communication with the stakeholders by involving them in the implementation as reviewers and thus better managing their expectations as well.

5. Conclusions

In this paper we have explored a methodology to guide a business unit's or organization's effort to define or revise its data strategy.

We have examined the key considerations for defining and implementing a data strategy for the business unit or the organization.

The methodology suggested here is that

- 1) We start from identifying the business goals and the required analytics for it
- 2) We work from there to identify data requirements
- 3) We compare the data requirements against the existing system and data
- 4) Based on the gap analysis we list the scope of the implementation
- 5) We can plan for implementation using an agile methodology
 - a) Planning for and delivering by sets of use cases would help in reducing risk for design and approach
 - b) It would also help by involving users and stakeholders early in the process so that the strategy is eventually successful and meets the needs of the business users
- 6) In the whole approach keep the considerations for governance – not only for data but also the governance of the processes
- 7) Aim to build a sustained team that would not only see through the implementation but also maintain and manage the strategy and its evolution with time

Like other organizational change management initiatives, data strategy initiatives also require strong support from senior management in the organization, and champions from within the business unit or organization for taking them forward.

Lastly, communication around the strategy, the need for change, the plan for implementation and early feedback on implementation results, considerably help in successfully driving such initiatives.

Here is a use case documentation template as referred in the use-case based approach for identifying user needs and deriving strategy goals. It can be suitably leveraged as is or with modifications as an aid for identifying the scope of the strategy implementation.

Appendix

6.1 Use Case Documentation Template

Use Case 1	
e.g., Patient Onboarding metrics calculations	
Description	
Owner	Owner of use case – will also be primary for acceptance
Business Use	Who uses this information and for what reason? What is the business goal that this use case will support?
Frequency	What is the frequency of this data use case analysis e.g., daily, every Monday/Thursday etc.
Priority	What is the priority determined by the impact of this use case on business decisions – say High, Medium, Low
User Information	
Function/Subfunction	Which business function/subfunction do the users belong to
Brand(s)	Name the brands that are under consideration here
Delivery of Use Case	
Current Mechanism	How is the outcome delivered? Report or Dashboard or a data file?
Expected Mechanism	How is it expected to be available after the project is implemented? Will it be an email/dashboard etc
Delivery Owner	Who is responsible for delivery – a person, department, role etc.
Use Case Data Sources	
System 1	e.g., Source Data file/table from an upstream system 1
Others sources	Explain the sources, owners/stakeholders of these sources
Use Case Data Check (before running)	
Check 1	e.g., Run use case only if both brands are present
Check 2	Other data quality readiness checks. More checks upstream in the data processing will help in reducing data issues.
Challenges	
Data Lags	What are the current challenges today? e.g., the data is received after 4 days. This should be cut down to 1 day.
Key Issues in Quality	Capturing these issues and their analysis will help in the data quality framework definition as part of the overall strategy
Governance	Any challenges for data collection, ownership, privacy, access, and security
Technology	Any technology challenges regarding storage, processing, delivery
Any other challenges	
Analytics – Use case logic	
Use Case Analytics	A use case logic can be captured if these are to be rebuilt or combined/optimized.
Analytics – Use case Output Validations	
Validation 1	e.g., Metrics should lie between $0 < \text{metric} < 100$
More Validations	Other checks for use case output result validity. These will help in validation of data once the implementation is done

Case Study: Data Challenges for a Large Pharma Company

The pharma company, a client of Axtria, had a huge system setup for data processing, analysis, and reporting for its commercial operations. It was heavily dependent on legacy data platforms and contracts with third parties for its business data needs.

Some of the key challenges that the client's users had identified were:

1. Operational Issues
 - a. Multiple manually triggered processes
 - b. File/data arrival delays and difficulty in analysing impact of those delays
 - c. Long running processing jobs
 - d. Lack of real time log-based monitoring
 - e. Multiple copies of datasets being used by users across business units
2. Data models not flexible enough to handle digital and patient level data
3. Lack of user trust in the data due to frequent quality issues
4. Lack of business user visibility into data processing
 - a. This made it difficult for business users to understand the source and lineage of their data
 - b. Any data issues could take days or weeks to diagnose and sometimes the users would give up on data issue diagnosis

Working with Axtria, the client's chief goals were:

- 1) Organizing data sources so that it was easy to maintain, update and onboard new data sources
- 2) Making it easier for its users to discover data and identify the relevant data for their analysis
- 3) Building trust in the data
- 4) A well organized and scalable processing system
- 5) Complete visibility into the system of data processing with easy diagnosis of data issues

Axtria's team of consultants provided a solution accelerated by Axtria DataMAx offering.

Using DataMAx, the consultant team helped the client achieve significant time savings for data onboarding, setting up new data processes and ease of data provisioning. A typical new data onboarding time was reduced from approximately two weeks to about three days including approvals.

Changes to the data sources were managed significantly better with their time being reduced from more than a week to just one day.

Using the advanced and robust data quality framework of DataMAx, the data processing time and confidence in data quality was very significantly improved. Every business division could therefore use the same data with confidence.

The overall data processing time from ingestion to analytics ready data marts was reduced considerably from a typical five days to less than two days. This is also attributable to a significant reduction in process failures as the overall data and process quality was much better managed and governance was set in place for faster turnaround time of any data issues.

A robust data Catalog and Lineage capability provided business users with much needed understanding of their data. Data scientists could easily search for and discover quality data for their analyses.

This was most appreciated by the data scientists who earlier would take several days to discover, enquire about, and receive appropriate data for the analysis. With the Catalog and its features for community driven reviews and comments, the data scientists could easily discover and share information about relevant data sets all on their own.

Added to that, the governance in the processes and the DataMAx solution, enabled the project sponsors to boost the trust and confidence on the data.



Amit Saini

Sr. Director, Axtria Inc.
Tower B, Building 14
DLF Cybercity, Gurugram – 122002
E: amit.saini@axtria.com

Contact Us

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

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

 info@axtria.com

 [@Axtria](https://twitter.com/Axtria)

 facebook.com/AxtriaInc/

 linkedin.com/company/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™ and Axtria CustomerIQ™ - 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 and LinkedIn

Copyright © Axtria Inc. 2021. All Right Reserved