atlan

Building a Business Case for a **Data Catalog**

Why do you need to build a Business Case for a Data Catalog?

You know that your data team is struggling to meet expectations because of the inefficiencies in their daily works. This is not a problem that you're facing alone.

Despite the fact that leaders in **more than 65% of enterprises** say that their enterprises risk becoming irrelevant and/or uncompetitive if they do not embrace big data, only **only 27% of data projects** today are successful.

Most common reasons behind data projects failing:

- Failure to make it to production: 85% of big data projects never make it to production.
- Time of coveted data teams spent on menial tasks: 80% of data analyst's time is spent simply discovering and preparing data.
- Increasing governance risk: More than 70% of employees have access to data they should not; creating exposure risks that can amount to fines up to 4% of annual revenues

To tackle these challenges, it is important to invest in foundational activities for your data teams. But….how do you convince the business to invest in foundational activities for your data team?

For any initiative to unlock organic adoption within an organization, it is important for a clear, tangible business outcome to be derived from it.

While earlier data was owned and used primarily by IT teams, today, data is embedded across all functions and used extensively by business users. Today, all business and functional teams depend on data and analytics to support their goals and processes.

To start laying the foundation of success for a data management program, it is essential to conduct an end-to-end inventory of goals, challenges, and priorities of all teams across the organization. This typically requires a top-down approach

where we begin by understanding organizational priorities across teams and map them to value-impacting data systems and challenges to solve.

Steps to Create the Business Case for a Data Catalog

Step 1: Create an inventory of both organizational priorities and challenges and prioritize them

Conduct user interviews with top executives and stakeholders of the organization to understand the potential challenges related to the top 3 - 5 current business priorities.

Value generating initiatives can be mapped to three strategic objectives:

1/ Creating a revenue impact   2/ Improving cost efficiency
3/ Mitigating organizational risk.

Start at one of the three value categories and identify top initiatives that your executives care most about currently.

For instance, let's consider a CPG company. Within each of the three core strategic priorities along with other strategic initiatives for the year, here is the summary of the top organizational priorities for the next fiscal year:

Fig 1: Inventory of strategic initiatives for CPG company in FY2020

© 2020 – Atlan Technologies Pvt. Ltd.

This material may not be copied, modified, distributed and monetized without prior permission.

Next, identify the top 3 initiatives that are mission-critical for organizational success at present.

Fig 2: Top priority strategic initiatives for CPG company in FY2020
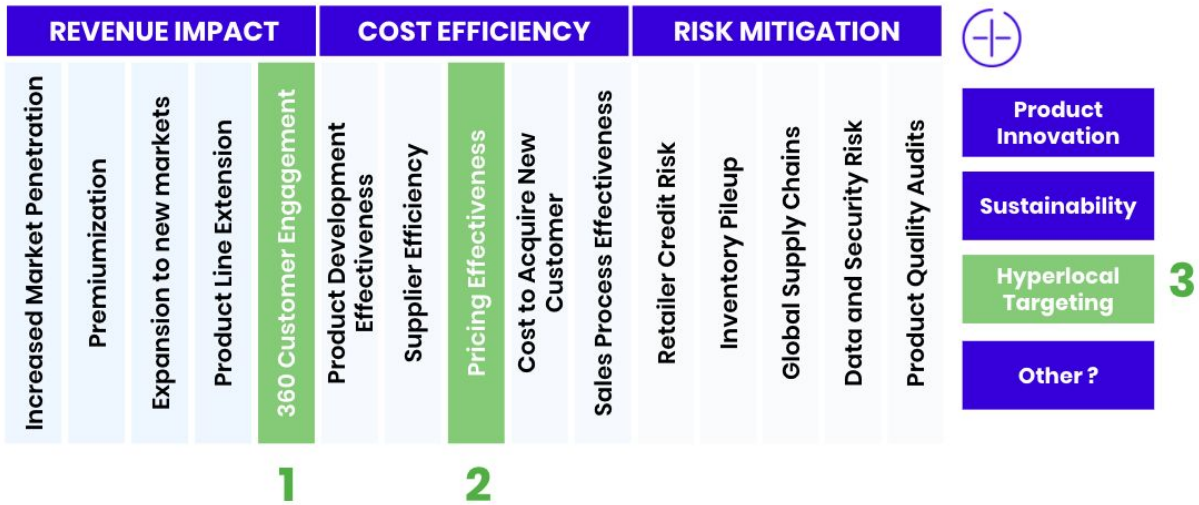
Step 2: Map business value to data-related outcomes and improvements required

One of the recommended tools to carry this out is [Gartner's Value Pyramid](#) that we've found to be effective in linking business value to gaps and priorities in data management.

Fig 3: Gartner's Value Pyramid to connect Business Value to Data
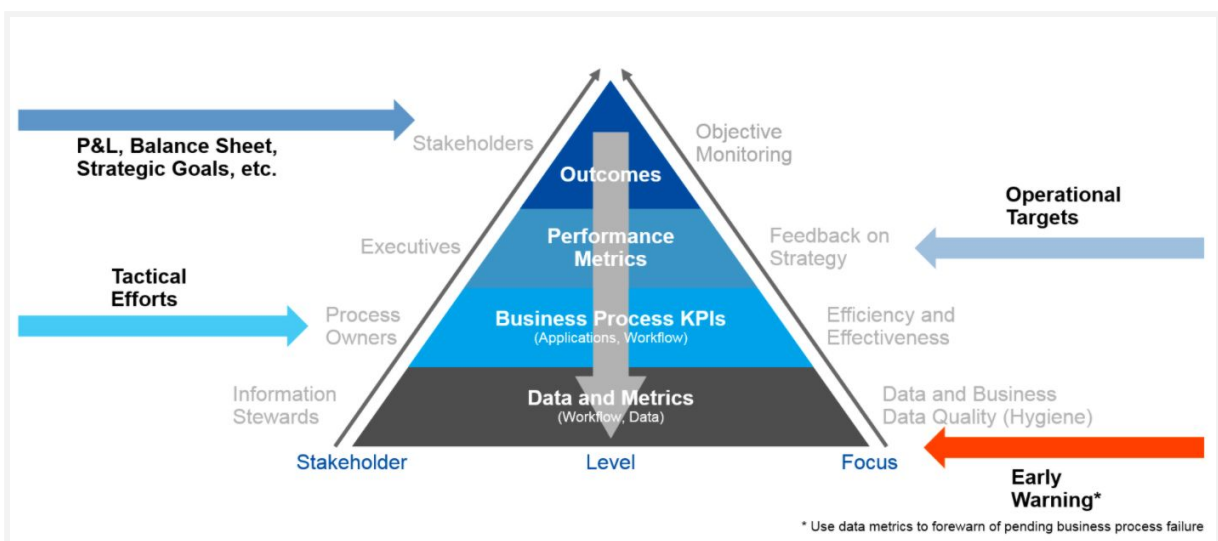
Step 2.1: Select an outcome and identify the core stakeholder

The customer marketing team at Consumer Product company is struggling to improve customer retention and upsell products — due to a problem-ridden 360 Degree Customer Engagement Initiative.

Stakeholder: Customer Marketing team
Outcome: Improve multi-channel customer engagement by enabling sales representatives with predictive insights on which accounts to pursue at a given time to improve win rate

Step 2.2: Determine what change needs to be created to achieve the outcome (along with key performance indicators)

The underlying challenge identified in this case was that of data quality which had 3 key reasons:

- Completeness: Data for only 35% of customers was enriched across touchpoints (in-store, digital, etc)
- Data Quality: Data was ridden with quality challenges (data format issues, missing fields, etc) due to which there are several rounds of feedback and iterations on data output before it's used for decision making
- Data Process Lags: The program was started to provide continuous support to sales representatives to make data-driven decisions on which accounts to pursue at a given time and expected to improve the win rate by 45% as per initial pilot studies. However, due to lags in the data process and productivity challenges, analysts are only able to create 3 - 4 models every quarter.

Underlying Challenge/ Change Required	Key Performance Indicators (KPIs)
Completeness	- Increased % of customers with data enriched across all touchpoints
Data Quality	- Higher data quality score - Reduced number of data quality errors reported at each step of data cleaning and transformation
Data Process Lags	- Create time savings for data analysts so they can service a larger number of projects/ models every month

Step 2.3: Identify the process or application-level changes required to support the desired change (and identified KPIs)

To create the three identified performance metrics, we need to identify the desired application or process level changes that are required:

Underlying Challenge	Performance Metrics	Application or Process-level Change
Completeness	• Increased % of customers with data enriched across all touch points	Change required: Increase visibility around columns available in different tables and facilitate easier data transformations. Help achieve higher completeness % across datasets.
Data Quality	• Higher data quality score • Reduced number of data quality errors reported at each step of data cleaning and transformation	Change required: Improve processes around data quality. Introduce automated data quality checks — and assign data quality scores to each data set to make it easier for analysts to identify and solve data quality issues.
Data Process Lags	• Create time savings for data analysts so they can service a larger number of projects/ models every month	An analyst on the team spends >70% of their time searching for the right data and gathering all the context and information around it. This sustained productivity loss is impacting the efficacy of the entire initiative and sales function of the organization. Change required: Make data easier to discover.

Step 2.4: Identify the new tools or data needed to close the loop on the business change required

The CPG company needs to introduce a data catalog tool in the organization that is focussed on:

1. Improve data discoverability: create time savings for data analysts
2. Increase visibility around quality of data and provide tools to improve effectiveness of data cleaning and transformation processes

Step 3: Go bottom-up by running a survey amongst your data professionals to understand their key challenges

It is critical to also understand the current sentiments and challenges among data professionals in the organization. The most objective way of doing that is to run a survey amongst your data professionals to understand the key challenges and productivity issues that they're facing today.

Fig 4: Sample Survey Questionnaire

CATEGORY	QUESTIONS	OPTIONS (TO BE CUSTOMIZED)
ROLE & USER PERSONA	Which one of the following personas best fit your role at the organization?	- Data Scientist - Data Analyst - Data Engineer - Business Analyst - Business Manager - Data Steward - IT Support (Data Provisioning) - Cloud Team - Other *(If yes, then what persona and how is it different from the ones listed above?)*
	What are your top three priorities this year?	*[Insert text]*
	1. What are your top three challenges that affect productivity? 2. How many hours would you save by solving each of the three challenges every week?	*[Insert text]*
DATA-RELATED ACTIVITIES	1. Which of the following data-related activities do you perform? 2. How much time do you spend on each activity every week?	- Getting access to data - Finding the right dataset - Understanding context associated with data (e.g. understanding of column names) - Running quality checks on datasets - Exploring data & running queries

DATA-RELATED ACTIVITIES	1. Which data-related activities do you face productivity challenges with today? 2. How many hours would you save every week by solving these challenges?	• Getting access to data • Finding the right dataset • Understanding context associated with data (e.g. understanding of column names) • Running quality checks on datasets • Exploring data & running queries
COLLABO-RATION	1. In a typical week, what are the different user personas that you collaborate with on data? 2. Do you face any challenges while collaborating with others on data? If yes, explain. 3. How much time would you save every week if these challenges were solved?	• Data Scientist • Data Analyst • Data Engineer • Business Analyst • Business Manager • Data Steward • IT Support (Data Provisioning) • Cloud Team • Other *(If yes, then what persona and how is it different from the ones listed above?)*

Step 4: Bring it together in a compelling business case

To create a compelling business case, you need to align the top 3 priorities identified from the business perspective and connect them to the key findings from the bottom-up survey i.e. biggest productivity bottlenecks in the organization today.

In this case, by aligning the top priority (i.e. salesforce enablement) with key findings from the bottom-up survey (that reveals the productivity challenge surrounding discoverability of data by analysts) — you will find the overall framework of a compelling business case to align efforts of the entire organization.

Next, it is important to break down the business case into fundamental metrics. There are three key value levers or metrics to measure:
1. Revenue Upside Potential
2. Cost Savings Generated
3. Time Savings Generated (Faster Response Rate and Speed-To-Market)

In this example discussed above, the following could be value levers:

1. REVENUE UPSIDE GENERATED	2. COST SAVINGS GENERATED	3. TIME SAVINGS GENERATED
• 20% jump in win rate attributed to the improved customer engagement models	• Data team is able to service 100% more number of data projects with the same resources (no new hires)	• Each analyst spends 65% less time in looking for the right datasets and related context • 30% fewer data quality issues are arising due to new automated checks

Illustrative Business Case: Mid-Sized Company Implementing a Data Catalog

1/ About the Data Team

DATA PERSONA	#	AVG SALARY
Data Scientists	5	$120,000
Data Analysts	15	$70,000
Business Users (Power Users)	5	*Varied*
IT	2	$80,000
Data Stewards	5	$85,000
Total Team	32	

2/ Key Value Levers Identified:

1. REVENUE UPSIDE GENERATED	• Increased number of Data Science and Analytics Projects: Data Catalog creates time savings which enables team to deliver more successful Data Science and Data Analytics projects with same resources • Upskilling for Business Users: Business users are able to take on 25% of the Data Analyst's task on any Data Analytics Project • Increased Success Rate: By facilitating a more collaborative approach, the data catalog is able to enable the data team to detect roadblocks, and make more projects successful. Success rate goes up from 50% to 65%
2. COST SAVINGS GENERATED	• Cost Savings: To deliver the net new projects without the Data Catalog, the company would be required to hire several data analysts and scientists (on current productivity assumptions) and spend resources for the hiring process
3. TIME SAVINGS GENERATED	• Saves 25% of each Data Scientist's time • Saves 25% of each Data Analyst's time • The IT Team now needs 80% less time to provision data • Each Data Steward now spends 50% less time in managing data exposure risk

3/ Final Business Case

	Type of BV	Atlan ROI	ROI Calculation
1	New Value Generated	81 – 110 x	Revenue of Net New Projects Delivered with the Same Resources / Cost of Adopting Data Catalog
+			
2	Cost Savings	22 – 26 x	(Cost of Acquiring New Resources to deliver Net New Projects *minus* Cost of Adopting Data Catalog) / Cost of Adopting Data Catalog
=			
3	Total Business Value	102 – 136 x	Total $ Business Value Generated / Cost of Adopting Data Catalog

THANK YOU FOR READING THE

Building a Business Case for a **Data Catalog**

Check out our other resources for data teams

The Data Catalog Primer:
Everything you need to know
to setup one for your team

Download Your Copy

The Ultimate Guide on
Implementing **Agile
for Data Teams**

Download Your Copy