The Business Case for a Data Catalog

A Unifi Perspective

By Dave Wells

May 2018

Research sponsored by:

About the Author

Dave Wells is an advisory consultant, educator, and research analyst dedicated to building meaningful connections throughout the path from data to business impact. As an educator he has written dozens of courses and taught hundreds of classes about data warehousing, data modeling, data architecture, and business intelligence for professional organizations such as Dataversity, eLearningCurve, and TDWI.

About Eckerson Group

Eckerson Group is a research and consulting firm of veteran practitioners who help business analytics leaders use data and technology to drive better insights and actions. Its researchers and consultants each have more than 20 years of experience in the field and are uniquely qualified to help business and technical leaders optimize their investments in business intelligence, analytics, big data management, and the internet of things.

To see our research and learn about our consulting services, go to www.eckerson.com.

About This Report

The research for this report is made possible by unifi

Abstract

The value and benefits of a data catalog are often described as the ability for analysts to find the data they need quickly and efficiently. Data cataloging accelerates analysis by minimizing the time and effort that analysts spend finding and preparing data. Anecdotally it is said that 80% of self-service analysis without a data catalog is spent getting the data ready for analysis. Using the data catalog can cut that percentage from 80 to 20. Although there is a high degree of truth in this anecdotal view, it is insufficient to build a business case for technology investment.

This paper quantifies the value of a data catalog to help estimate return on investment for a data catalog. Starting from a single analytics use case, it illustrates the revenue creating and cost saving potential of data catalogs. Revenue and cost are important—making and saving money—and are the primary focus. Additional benefits of time savings, quality improvement, trust and confidence, frustration reduction, and risk avoidance are also described.

This Unifi product profile is a companion to the Eckerson Group report, "The Ultimate Guide to Data Catalogs" (May 2018).

The Value of a Data Catalog

A data catalog is an investment with real costs. Beyond the cost of software, a data catalog requires investment to hire and train people, and populate the catalog with metadata. Companies often struggle to describe the business case for a data catalog. Anecdotal cost justifications focus on a frequently described situation where analysts spend 80% of their time finding and preparing data and only 20% analyzing it. Higher quality of analysis is listed as a benefit when analysts can find and use the best datasets instead of those that are easy to find. Centralizing and sharing tribal knowledge is seen as an advantage that prevents loss of knowledge and increases collaboration, communication, and sharing.

A difficult but important question: How is ROI for a data catalog determined?

These are all good examples of how data cataloging delivers value, but they don't really answer a common and important question: How is data catalog ROI determined? The examples are soft, anecdotal, and non-quantitative. They provide no guidance when the goal is to estimate or calculate return on investment (ROI) for a data catalog.

A Data Analysis Example

We'll begin with a data analysis scenario to illustrate the process working with and without a data catalog. A single analysis project is viewed through three lenses: (1) Analysis through an IT project, (2) Self-service analysis without a data catalog, and (3) Self-Service analysis with a data catalog. As each option unfolds we'll note the time and cost of analysis—quantitative variables that provide the foundation for ROI calculations.

The Scenario

Kate manages the online sales line of business for a midsize retailer with $216 million in e-retail sales annually. She is evaluating product bundling as a way to maximize the value of every customer sale, and she needs to make decisions about product bundling and pricing. Bundled items sell at a lower price than if the same items are sold separately. Kate has questions about the optimum number of items in a bundle, the right mix of items, and the cost savings for bundle purchases. She wants to find the sweet spot where price and sales volume combine to drive the most revenue.

Option 1 – An IT Project

Kate's first option is to request that IT build a bundle sales simulator that she can use to explore the sales volume and revenue impacts of different bundling options. The timeline for this approach is:

Week 1	Develop and submit project request
Weeks 2–3	Wait in project queue
Week 4	Requirements gathering and definition
Weeks 5–9	Design and development
Week 10	Testing and debugging
Week 11	Bundle scenario simulations
Week 12	Bundle and pricing decisions made with moderate uncertainty

Three months after the analysis need is identified Kate can finally implement bundling. Then, after operating bundled sales for a few months she will want to evaluate the performance. Continuing to work through IT, she will request a report of bundle sales performance showing trends in sales volume and revenue generation. Then she will experience:

- Some delay while waiting in the project queues
- A few days of requirements definition
- Two to three weeks of design and development
- A few days of testing

Finally she will be able to view and evaluate bundle sales performance and compare actual results to simulated results. She may then choose to revise and refine bundling and pricing. She may also want to request modifications to the bundle simulator.

Option 2 – Self-Service without a Data Catalog

Kate's second option is to perform the analysis using a self-service data visualization and analysis tool. Although comfortable with Excel she recognizes that the analysis she needs is too complex to work well with Excel. She decides it is time to step up to Tableau, which already has some presence in the company. The timeline for this approach is:

Day 1	Acquire Tableau and learn the basics
Day 2	Frame the problem and determine data needs
Days 3–5	Search for data • Ask IT if they can provide any of the data needed • Ask colleagues and work the tribal knowledge network • Find some "close enough" data through tribal knowledge • Use some spreadsheet data from a past project (although not quite the right fit and somewhat dated, it is easy to acquire) • Wait for IT to deliver a data extract
Days 6–9	Data preparation • Load the data into Excel • Manually blend the data from different sources in a single workbook • Add formulas to calculate new variables for analysis • Determine needs for additional data
Days 10–11	Find and acquire additional data
Days 12–13	Continue data preparation • Manually blend in newly acquired data • Format for Tableau compatibility
Day 14–15	Data visualization and analysis
Day 15	Bundle and pricing decisions made with some uncertainty

Three weeks after the analysis need is identified Kate can make decisions and begin to implement bundling. After implementing she will want to monitor her bundle sales performance. To continue to work with Tableau she will need to:

- Find and acquire data for sales performance monitoring
- Build the data preparation workflow for her sales performance dashboard
- Build the sales performance dashboard
- Operationalize the sales performance dashboard

Finally Kate will be able to continuously monitor bundle sales performance and compare actual results to goals. She may choose to revise and refine her bundling and pricing strategy. She may also want to refine the dashboard as new questions arise.

Option 3 – Self-Service with a Data Catalog

The third and preferred option is to perform the analysis using a self-service data visualization and analysis tool together with a data catalog. This option is not entirely in Kate's control and is only available to her if the company has acquired and implemented a data catalog. Using a combination of Unifi for data catalog, discovery and data preparation, along with Tableau she can quickly find and prepare data, perform the needed analysis, and get on with her decision making, bundling implementation, and real business impact. The timeline for this approach is:

Day 1	Acquire Tableau and learn the basics Frame the problem and speculate about data needs
Day 2	• An hour or so to find the data o Search using the Unifi data catalog o Find similar analyses conducted by others o Find trusted data • A few hours of data preparation o Move seamlessly through the Unifi catalog from data selection to data preparation operations, workflows, and tips o Build project specific data preparation workflows • Visualize and analyze
Day 3	o Deliver prepared data in the right format for Tableau analysis o Visualize and analyze iteratively Bundle and pricing decisions are made with confidence in the quality of data and analysis

Three days after the analysis need is identified Kate has made decisions and started to implement bundling. As the implementation proceeds she will want to monitor the bundling initiative daily. To do this she can:

- Revisit Unifi to find and prepare data for monitoring
- Use Tableau to build her monitoring dashboard

After quick analysis and informed decision making, Kate can actively monitor and manage bundles and pricing with confidence. Data cataloging, data preparation, and self-service analysis tools will enable her to quickly find answers as new questions arise and the bundling program evolves.

Comparing Time and Cost of Analysis

Looking at the results of the three options it is clear that Kate gets what she needs much faster with self-service analysis than through an IT project. She saves even more time with self-service analysis in conjunction with Unifi for its data catalog, data discovery, and data preparation capabilities. The time gains are significant, but cost savings are equally important. Kate's annual salary plus overhead is $120,000. The average cost of an IT analyst in her company is $100,000. Working from those numbers the calculated direct cost of analysis for each option is shown in the table below.

	Option 1	Option 2	Option 3
Time to analyze	3 months	3 weeks	A few days
Cost to analyze	$17,3001[1]	$3,4602[2]	$6903[3]

[1] 9 weeks of IT analyst @ $100,000 annual salary
[2] 50% of Kate's time for 3 weeks @ $120,000 annual salary
[3] 50% of Kate's time for 3 days @ $120,000 annual salary

Analysis cost using Unifi for data cataloging and data preparation is only 4% of IT project cost!

The real difference among these options emerges when viewed as percent of time and cost saved. With Unifi, the cost of analysis is less than 20% of self-service analysis without Unifi, and approximately 4% IT project cost. Time saved is similarly impressive at 23% and 3.3% respectively.

Business Impact

The effect on the analysis process is only the beginning of the story. The greatest business impact of a data catalog often appears in revenue generation. Let's assume that Kate's product bundles generate a modest 0.5% increase in revenue, and recall that her line of business generates $216 million annually prior to bundling. Bundles will grow revenue $1.08 million annually. The three-month delay to implement bundling in option 1 represents $270,000 in lost revenue opportunity. The three-week delay of option 2 represents $124,600 in lost revenue opportunity. With the rapid analysis and implementation of option 3, the company realizes maximum revenue opportunity.

Comparing option 3 to option 1, the combined cost savings and revenue realization total $286,610 ($270,000 + $17,300 - $690). The combined cost savings and revenue realization for option 3 vs. option 2 total $127,370 ($124,600 + $3,460 - $690).

A Data Analysis Example

Direct Financial Impact

At this point we've seen an example of the direct financial impacts of a data catalog for one analysis use case—a single analysis need in a single line of business. Assuming this example is a representative and typical use case we can project annual return on a data catalog investment based on the anticipated number of use cases per year across the entire enterprise.

Number of use cases per year	Shift from IT analysis projects to self-service with Unifi	Shift from self-service without a data catalog to self-service with Unifi
100	$28,661,000	$12,737,000
200	$57,322,000	$25,474,000
500	$143,335,000	$63,685,000
1000	$286,610,000	$127,370,000

Your numbers will depend on the size of your enterprise, annual revenue, and frequency of analytics use cases. Using the pattern and calculations illustrated here you can plug in your own data and assumptions to calculate a reasonable estimate of direct data catalog value for your organization. ROI calculation, of course, must consider the costs of catalog implementation, operation, and administration including software licensing, staffing, training, and data curation.

Additional Savings

Though difficult to quantify, the waste and inefficiency from redundant and overlapping analysis in large corporations is clearly substantial. Business and data analysts with self-service tools frequently reinvent analysis previously performed by others simply because they are unaware that it exists. The discovery capabilities in Unifi bring visibility to existing analysis and dashboards. When analysis work done by colleagues becomes visible, redundant effort shrinks, time and costs are saved, and shared knowledge of business dynamics and insights result.

Value from Governance

Unifi's integrated data governance and security features add value in two areas—quality improvement and risk avoidance. Cataloging key metrics ensures that they are calculated consistently, which improves the quality of analysis and promotes trust in data and analytics. Protecting privacy and securing sensitive data helps ensure compliance with data protection regulations. Active security

measures reduce the risk of data corruption and loss. It is difficult to quantify the value of increased trust and confidence, compliance violations avoided, and data breaches prevented, but the benefits are undeniable.

Workplace and Cultural Value

A data catalog delivers human and cultural value as well as financial value. The reduction of workforce friction is visible and tangible. In many organizations the interactions described in options 1 and 2 cause friction as delays in business initiatives have visible impact on business results. When business urgency conflicts with IT overload, tensions rise and working relationships suffer. The problem is compounded when time-sensitive information and analysis needs amplify stress levels. For example, if Kate's objective was product bundle promotions targeting Mother's Day shoppers and she experienced weeks or months of delay, the benefit of a timed promotion might be lost completely, Timely analysis contributes directly to positive working relationships, collaborative culture, and timely business actions. Delays have the opposite effect. A common complaint from marketers relates to social media sentiment that may have positively or negatively driven sales. If it takes weeks or months to find the anomalies, it is too late to address the issue head on and respond to or exploit the sentiment.

It's All About Business Value

The business value of Unifi data cataloging, discovery, and data preparation technology is real. Quantifiable benefits such as revenue realization and cost savings help to calculate ROI, but that is only part of the story. Less readily quantified benefits of trust, confidence, and risks avoided are equally important. Whether your current analytics processes rely on IT projects, on self-service analysis without cataloging, or a combination of the two, you are sure to see tangible benefits with Unifi's technology.

Need help with your business analytics or data management and governance strategy?
Want to learn about the latest business analytics and big data tools and trends?
Check out **Eckerson Group** research and consulting services.

About Unifi

The Unifi Data Platform breaks down the barriers of operational data silos and makes the information that matters more accessible across the enterprise. At the heart of the platform is a comprehensive suite of self-service data discovery and preparation tools to empower business users. Employing machine learning and artificial intelligence technologies, and optimized for the cloud, Unifi predicts what the business user wants to visualize and then connects the resulting data natively to the BI tool for fast, accurate results.