Making Data Easier to Find, Access, and Use Should Start With a Data Catalog

Introduction

What if it were true that only half the time, users can easily find, access, and use relevant data and analytic content from a business intelligence (BI) implementation? Would that figure be disconcerting enough to spur a data leader to take action?

The bad news: This is not a what-if scenario. In 2022, 52 percent of respondents indicate analytic consumers and BI use cases find it *impossible, difficult,* or *somewhat difficult to* locate and/or access relevant data and analytic content. And it gets worse: It's been this way since we started tracking this issue and recording data on it in 2020, with percentages rising and remaining above a majority level (50 percent).

The issue of making analytic content easier to find, access, and use is complex and multifaceted; and likely includes technical and non-technical aspects such as (but not limited to) analytic data infrastructure, BI tools, applications, data management, data literacy, data culture, and governance. As such, no single solution or "silver bullet" can address this issue, and any investments in improvements will require time before yielding returns and results.

One of the best places to start is the data catalog, which aims to make analytical content more trustworthy and broadly available by providing the technology necessary to simplify access, enable collaboration, and ensure governance. Organizations that don't yet have a data catalog should get one. Organizations with a data catalog should ensure it is up to date, and that processes exist to ensure regular curation of its content.

If an organization implements a well-maintained data catalog—including a business glossary—that is tightly integrated with self-service BI, data engineering/data preparation and model ops platforms, then positive results will likely ensue. Casual users more likely will find and use the data necessary to make well-informed data-driven decisions. Data engineers and data scientists will find, access, and manipulate the data assets necessary for efficient deployment and management of their analytical models (AI and ML, as well as more traditional rules-based models).

Although these technologies can help address the issue, in the long term, they will prove effective only if the organization implements or revamps an associated change management program and provides appropriate training for each user persona.

Together, these actions will promote adoption and expansion of data literacy and data-driven decision-making practices.

Executive Summary

1. In 2022, 52 percent of users find it impossible, difficult, or somewhat difficult to locate and/or access analytic content and data relevant to their roles.
2. Across all functional areas, a data catalog is critical to supporting and enabling most BI use cases and user personas, and should play a critical role in problem-solving and decision-making processes.
3. Data catalogs continue to show increasing user importance, with 80 percent of organizations in 2022 considering them at least *important*.
4. Perceived importance of data catalogs goes up as organization size increases.
5. Organizations that consider it easy to find and access analytic content are more successful with their BI initiatives.
6. Organizations that regard a data catalog as *critical* or *very important* also are more successful with their BI initiatives.
7. Creating a data-driven culture through change management techniques and training will enable the human and process changes necessary for finding, accessing, and using the organization's data and BI assets effectively.

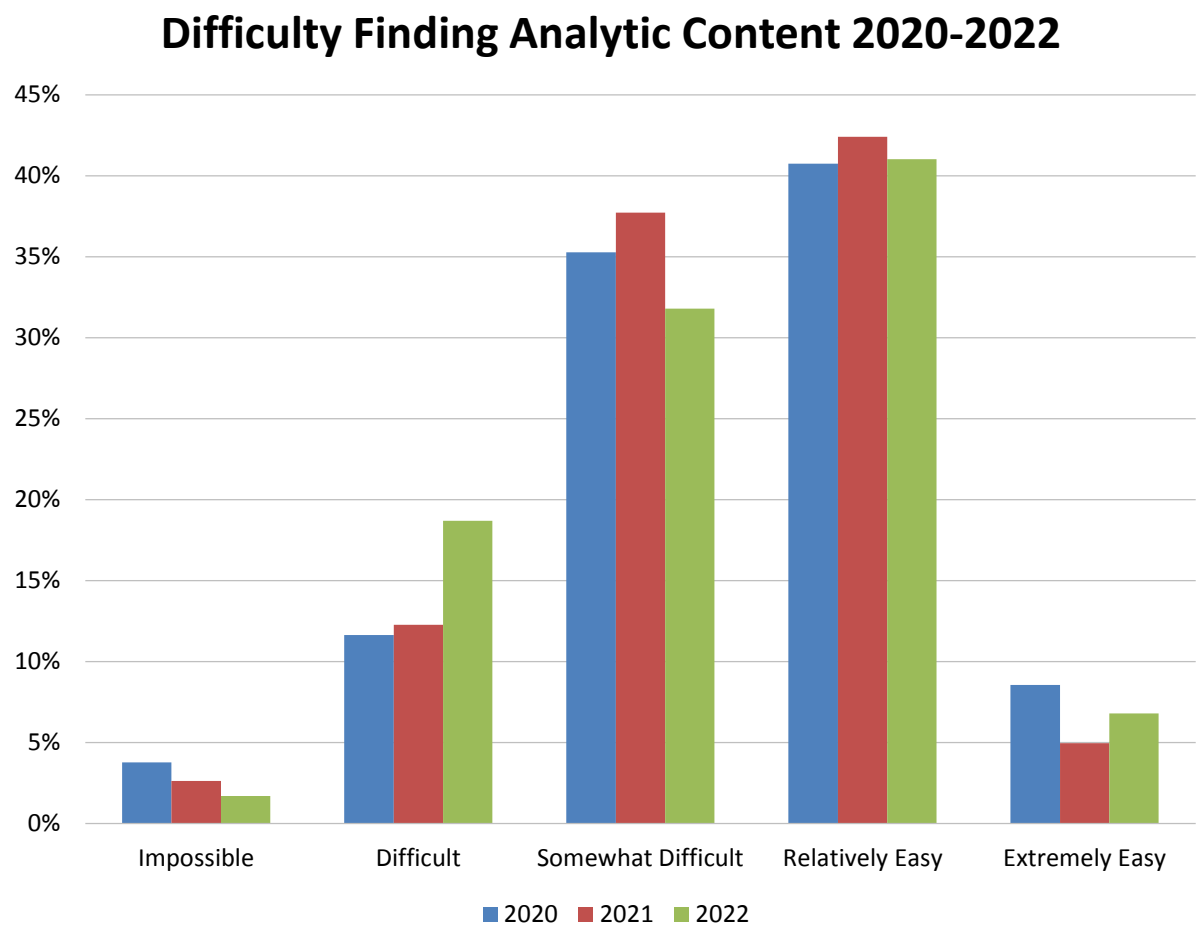
Recommendations

1. To help ensure high levels of usability and trust in a data catalog, organizations should include user participation in their product evaluations and deployments. In addition, organizations should train BI users on product capabilities, best practices, and workflow, with a goal of making analytic content more available to multiple BI use cases and additional users.
2. Carefully scrutinize data catalog vendor offerings to ensure product features and capabilities will meet your needs. This is especially important in larger organizations, which tend to have large numbers of BI users, multiple BI use cases, and significantly more data.
3. Evaluate your data catalog, data engineering / data preparation, and self-service BI tools and platforms to determine their ability to support end-to-end integration and interoperability of systems and processes.
4. Develop end-to-end problem-solving and decision-making process flows. Aim to integrate and automate these as reasonable. Make regular updates to each defined component. Integrate these with the data catalog.

5. If not already present, initiate a change management program that includes a communications plan, stakeholder matrices, and aspects of gamification.

A Majority of Organizations Do Not Consider Finding and Accessing Data and Analytical Content Easy

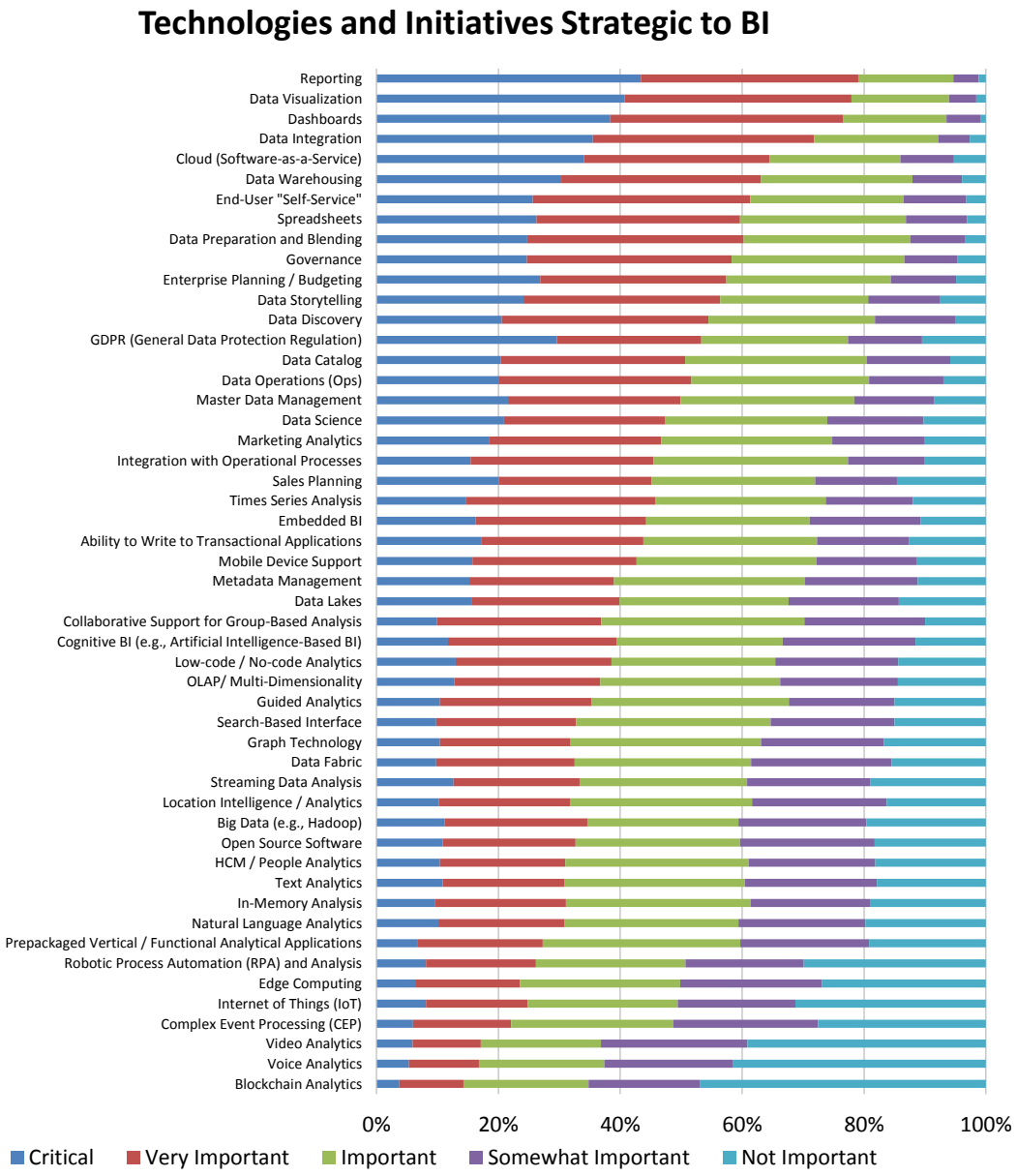
As was the case in 2021, 52 percent of organizations in 2022 consider it *impossible, difficult,* or *somewhat difficult* to locate and/or access relevant analytic content. Year over year, the drop in *somewhat difficult* responses offset the increase in the number of *difficult* responses. Across all years of data, a majority of respondents find it *impossible, difficult,* or *somewhat difficult* to locate and/or access relevant analytic content.



Source: Dresner Advisory Services

Supporting Data-Driven Decision Making

Key components of data-driven decision-making processes—including end-user self-service, data prep, data discovery, a data catalog, and governance—can and should be integrated, and, where practical, automated. Our data show their perceived high importance. In 2022, organizations rate end-user self-service (seventh), data prep (ninth), data discovery (thirteenth), governance (eleventh), and data catalog (fifteenth) in the top one-third of the 51 technologies and initiatives we consider strategic to BI.



Source: Dresner Advisory Services 2022 Wisdom of Crowds® Business Intelligence Market Study

How a Data Catalog Can Help

A data catalog aims to make analytical content more trustworthy and broadly available by providing the technology necessary to simplify access, enable collaboration, and ensure governance.

Collaboration capabilities allow data stewards and/or subject-matter experts to answer questions and make suggestions. Dialogs are preserved and available to everyone who makes similar searches. This helps to make tacit knowledge explicit, providing a mechanism for improving data literacy across the organization.

The business glossary function of the data catalog facilitates data exploration by enabling users to look up relevant business terms, read technical descriptions, and find links to certified sources and related BI content. It also should provide instructions on how to get access to these assets, (preferably) automated workflows for requesting access, and, where appropriate, instructions on how to customize and run data-engineering pipelines.

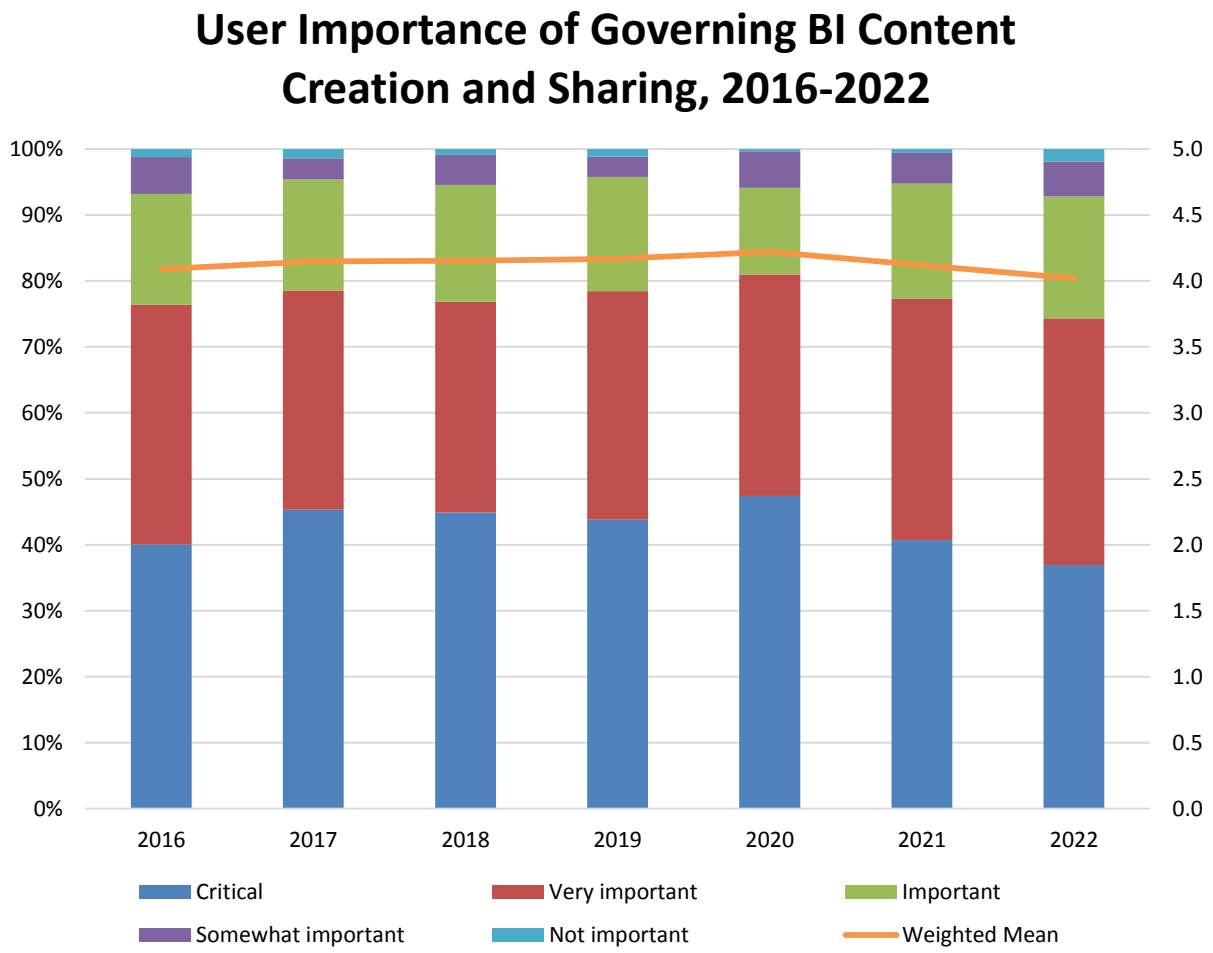
Different user personas will have different requirements. For example, marketing managers may simply want to identify and obtain access to existing BI reports or dashboards that allow them to explore the data. They will likely not need to know the technical details about the data that the reports and dashboards access.

Data scientists or data engineers may need to generate features for AI / ML models and persist these features in an appropriate technology for serving the features to the models. These roles may need to get access to—and be able to customize—existing pipelines in data-engineering and/or data-prep tool(s).

Data governance consists of the processes, policies, roles, and controls that support the creation and use of analytical content as a business asset. Most commonly, this manifests through aspects such as role- and policy-based access controls, defined access levels for analytical content, integration with identity management systems, and an ability to certify analytical content as “official.”

Demand Indicators

In 2022, 74 percent of organizations consider governing content creation and sharing (via policies, controls, and applied technologies)—a key function provided in a data catalog—*critical* or *very important*.

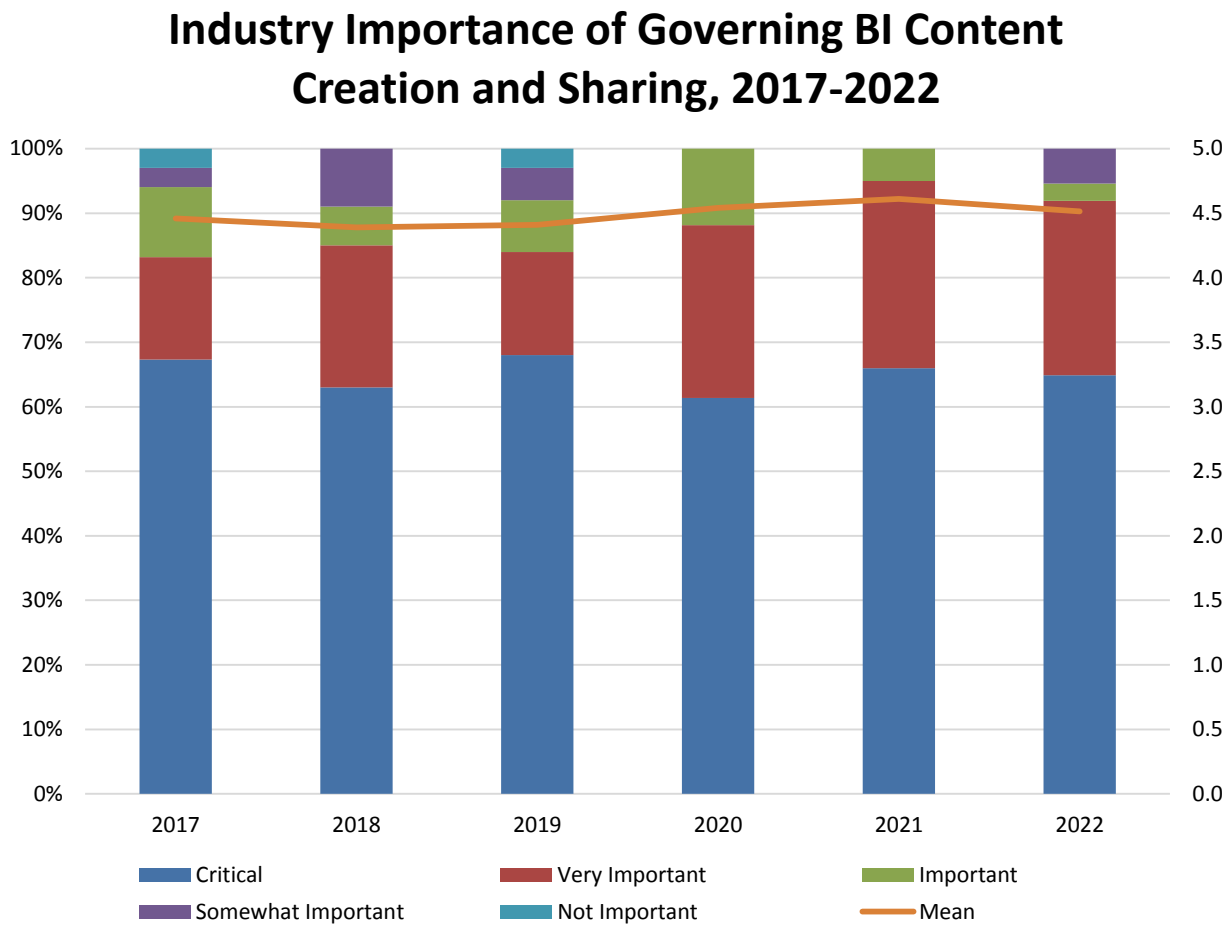


Source: Dresner Advisory Services

Supply-Side Indicators

As our data often show, industry importance ratings often exceed the user-organization importance ratings, as vendors tend to add or enhance features in their offerings (ideally) in advance—or at least in alignment with—user interest and demand. In all years of our data, vendors consider governing BI content creation and sharing as *critical* or *very important* at rates higher than user organizations. In 2022, vendor *critical* or *very important* perceptions (92 percent) exceed the same perceptions by user organizations by 18 percentage points.

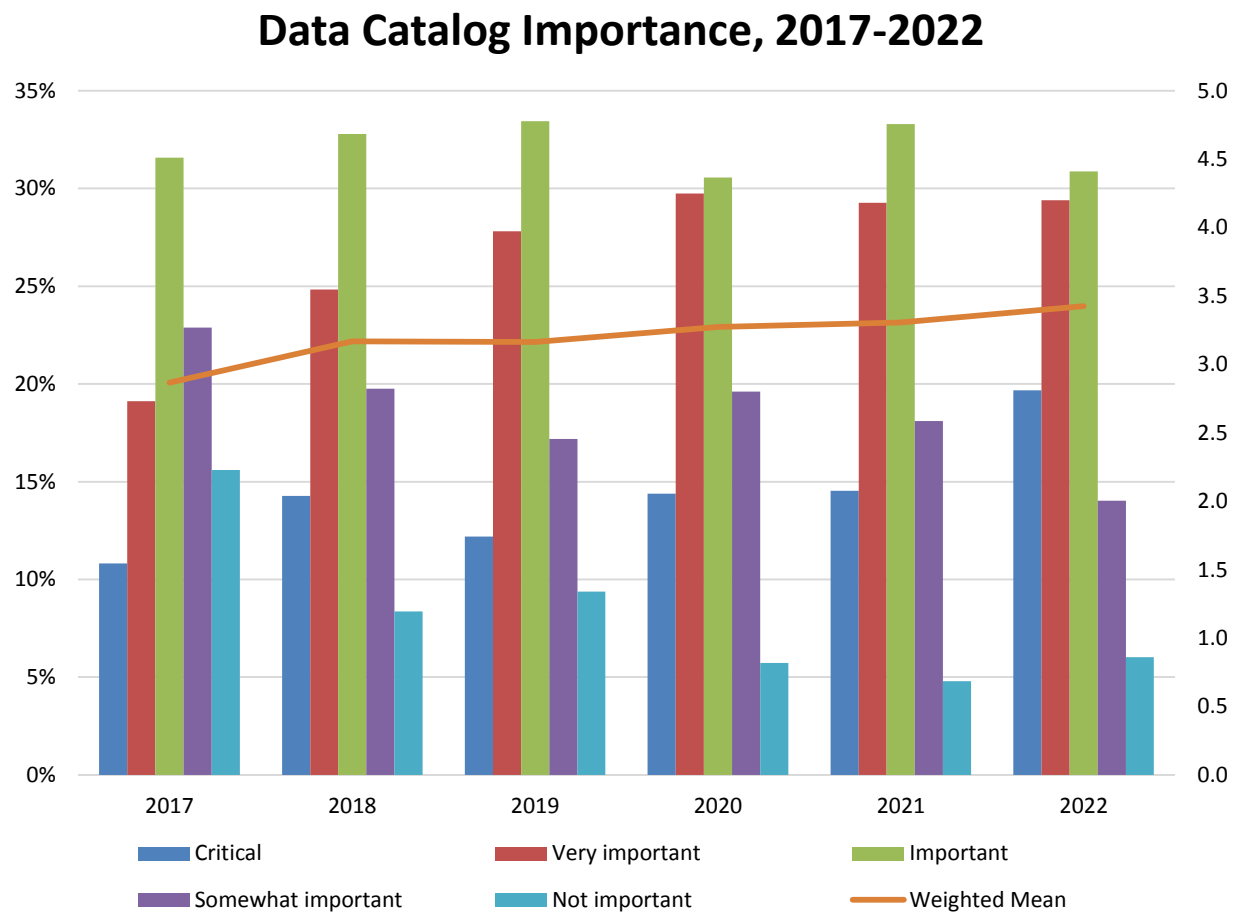
Taken together, these data indicate both strong interest and demand for a key component of a data catalog—features that can help organizations better govern BI content creation and sharing. This governance in turn will help users more easily find, access, and use analytical data.



Source: Dresner Advisory Services

Data Catalog Importance

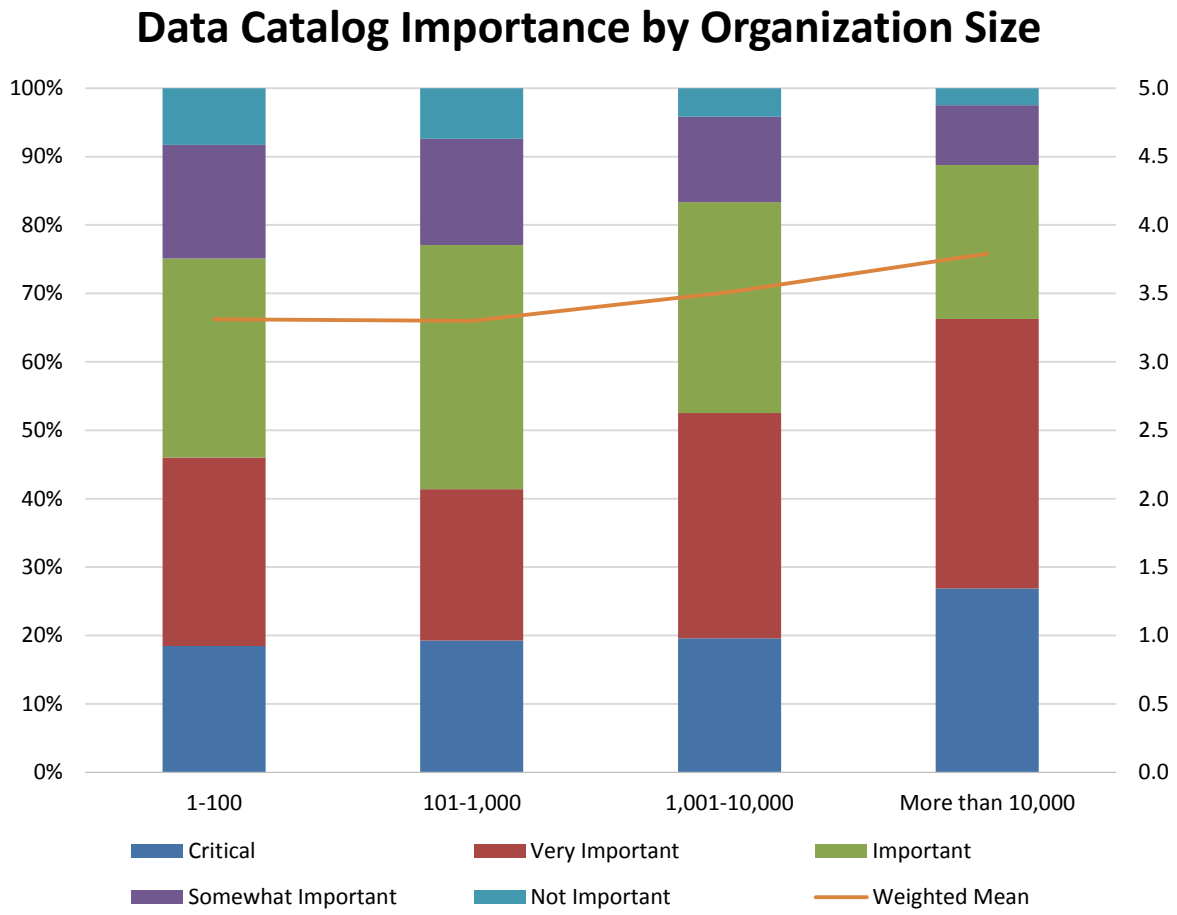
Data catalogs continue to show increasing user importance. In 2022, 20 percent of organizations consider them *critical,* and another 60 percent of organizations view them as *very important* or *important*. In addition, perceptions of *somewhat important* or *not important* stand at or near all-time lows in 2022. This split between higher perceptions and lower perceptions provides one indicator that this market is beginning to mature, as more organizations see the need for—and the potential value associated with using—data catalogs.



Source: Dresner Advisory Services

Data Catalog Importance and the Relationship to Size

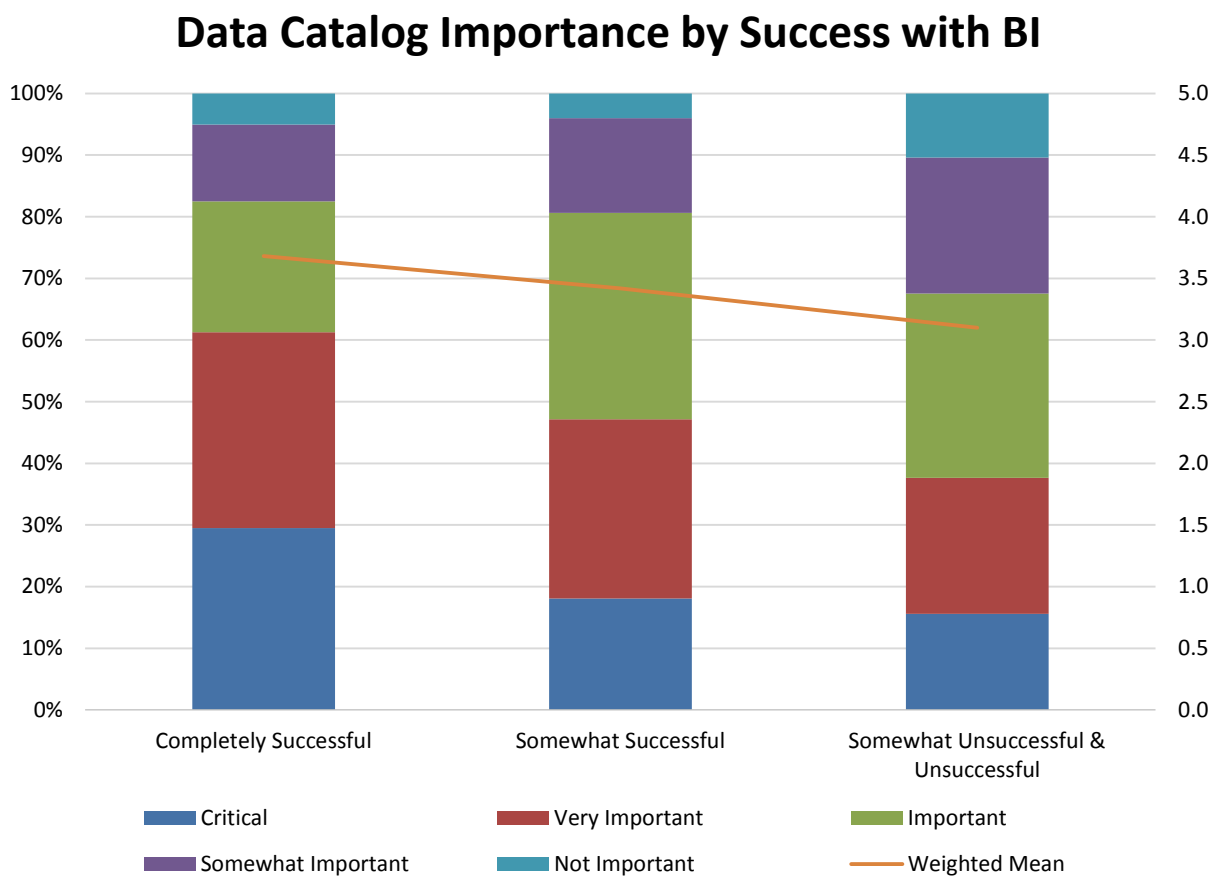
As the weighted-mean average line (in orange) shows, perceived importance of data catalogs goes up as organization size increases. As organizations increase in size, they tend to have multiple data sources and more analytic use cases that require the data governance and collaboration capabilities provided in data catalogs.



Source: Dresner Advisory Services

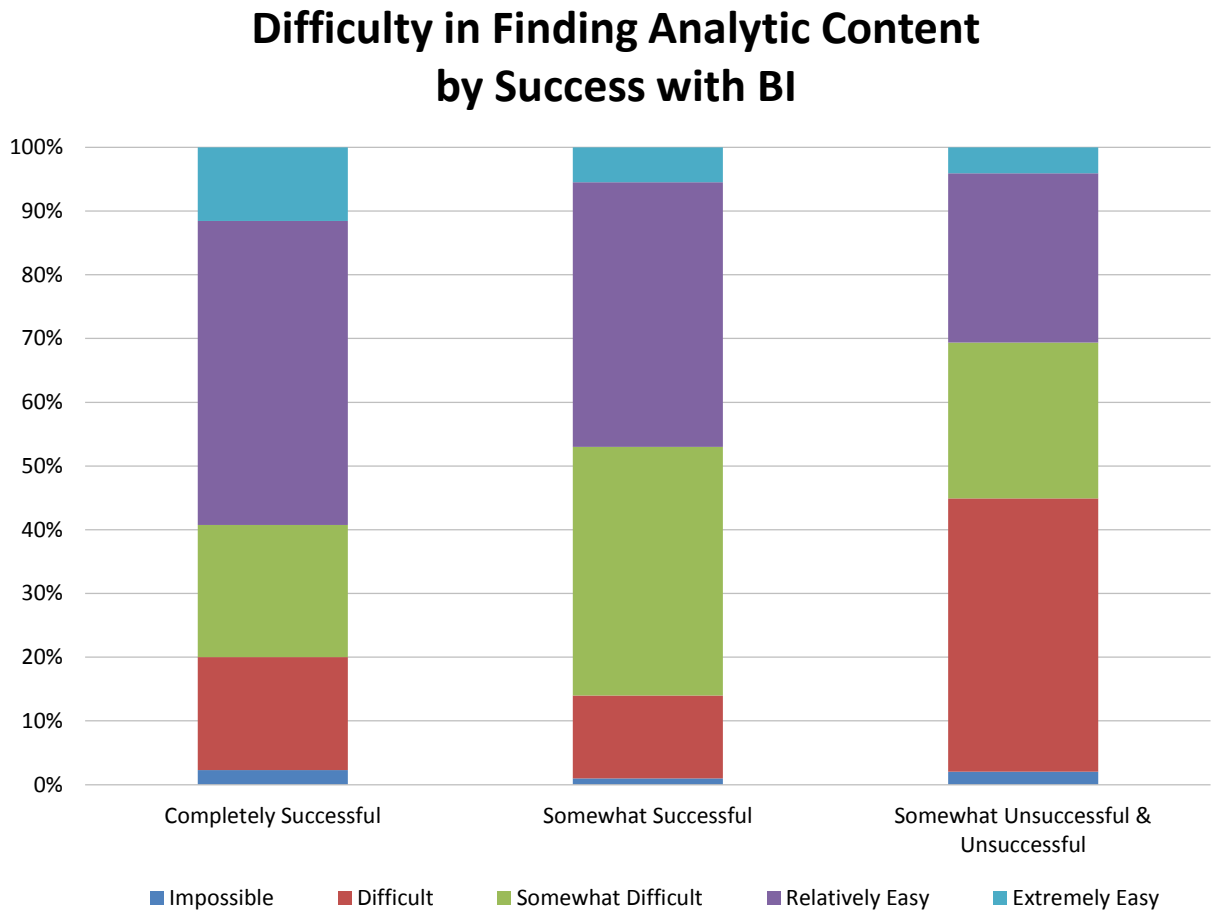
Data Catalog Importance and the Relationship to BI Success

Organizations that perceive a data catalog as *critical* or *very important* also report the most instances of considering their BI initiatives *completely successful* (61 percent). This frequency is 14 percentage points higher than organizations that hold the same perceptions but consider their BI initiatives only *somewhat successful*. The positive correlation between higher perceptions of a data catalog and higher levels of BI success makes sense. A data catalog helps ensure that users can consistently and properly access, collaborate with, explore, and use well-governed analytical content—all factors that, when executed well, contribute to positive achievements from BI and analytics.



Source: Dresner Advisory Services

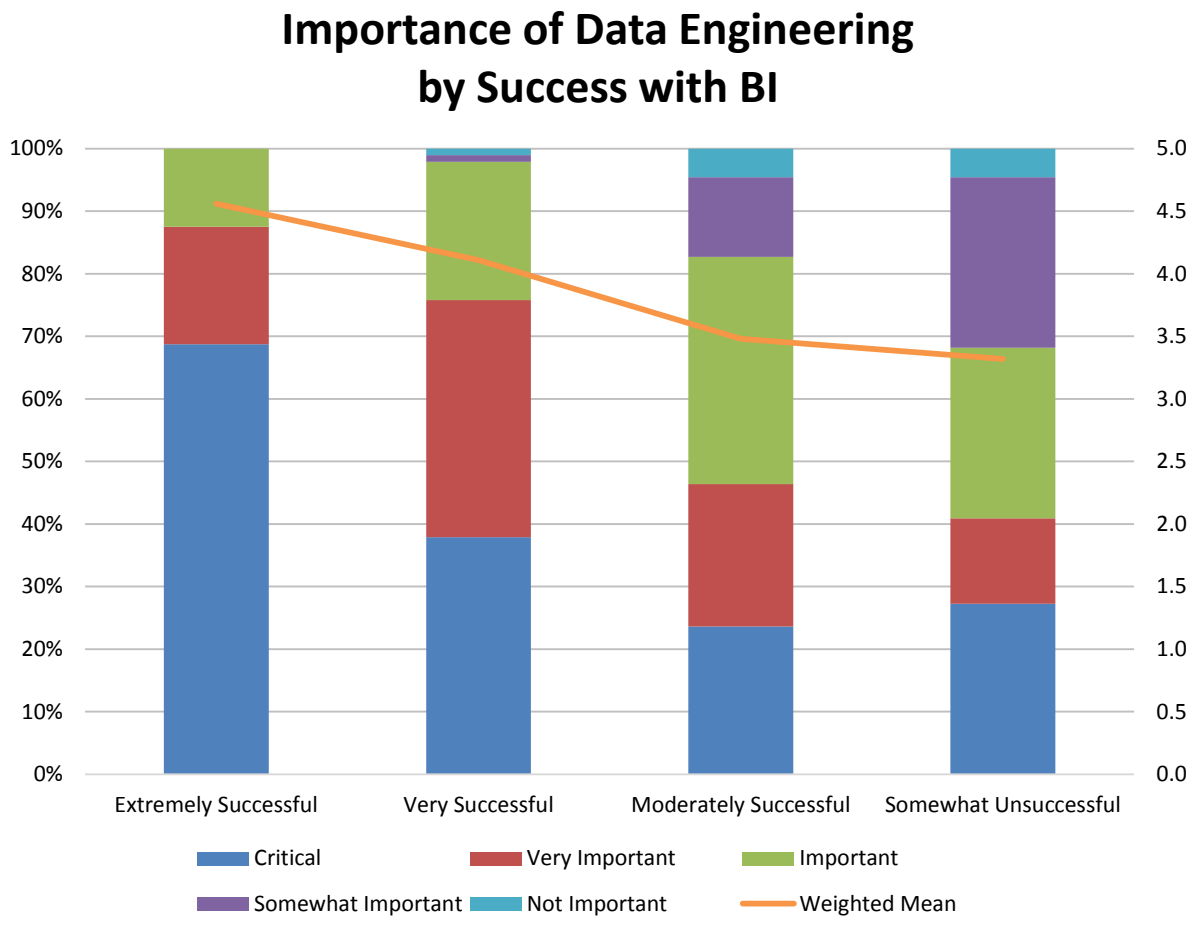
A relationship also exists between the level of difficulty in finding analytic content and success with BI. Organizations that view their BI initiatives *completely successful* also indicate the highest levels of considering it *relatively easy* or *extremely easy* to find and access analytic content within their analytic data infrastructure. Conversely, organizations that deem their BI initiatives *somewhat unsuccessful* or *unsuccessful* report much higher levels of difficulty in finding and accessing analytic content and data. If users perceive finding analytic content as *somewhat difficult*, the BI initiative likewise will likely prove to be only *somewhat successful*.



Source: Dresner Advisory Services

Data engineering represents the best practices and technology capabilities for developing engineered data workflows and pipelines to and between operational and analytic data management infrastructures. Data engineering pipelines and workflows access stored or streaming data, and transform, prepare, catalog, and process that data for the target of an analytic data infrastructure, which provides analytic content for BI users and use cases.

As noted earlier, a data engineer or data scientist can be one of the targeted personas for a data catalog. Our data show another positive correlation between organizations that consider data engineering *critical* also report the most instances of considering their BI initiatives *extremely successful* (69 percent).

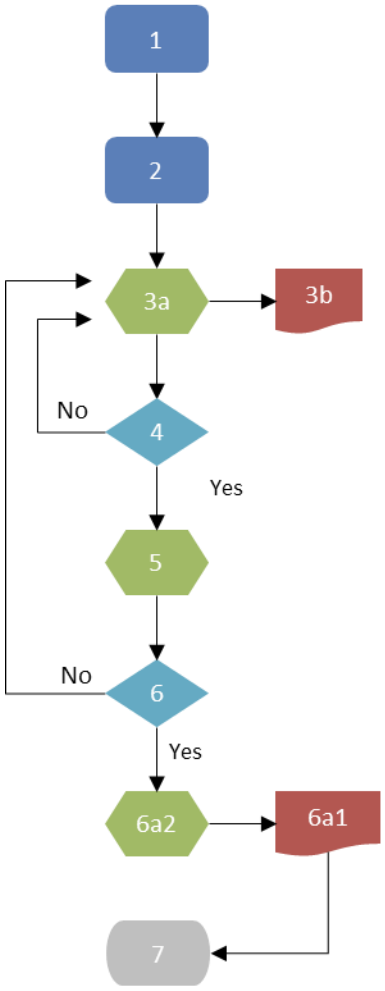


Source: Dresner Advisory Services

Two Example Use Cases and Processes

For two specific use cases and personas, we describe the role the data catalog, data engineering/prep, self-service BI, and model ops tools could play in making it easy to find, access, and use data and BI assets. These examples also show the importance of defining and enforcing appropriate data governance policies and integrating these tools to potentially automate these processes.

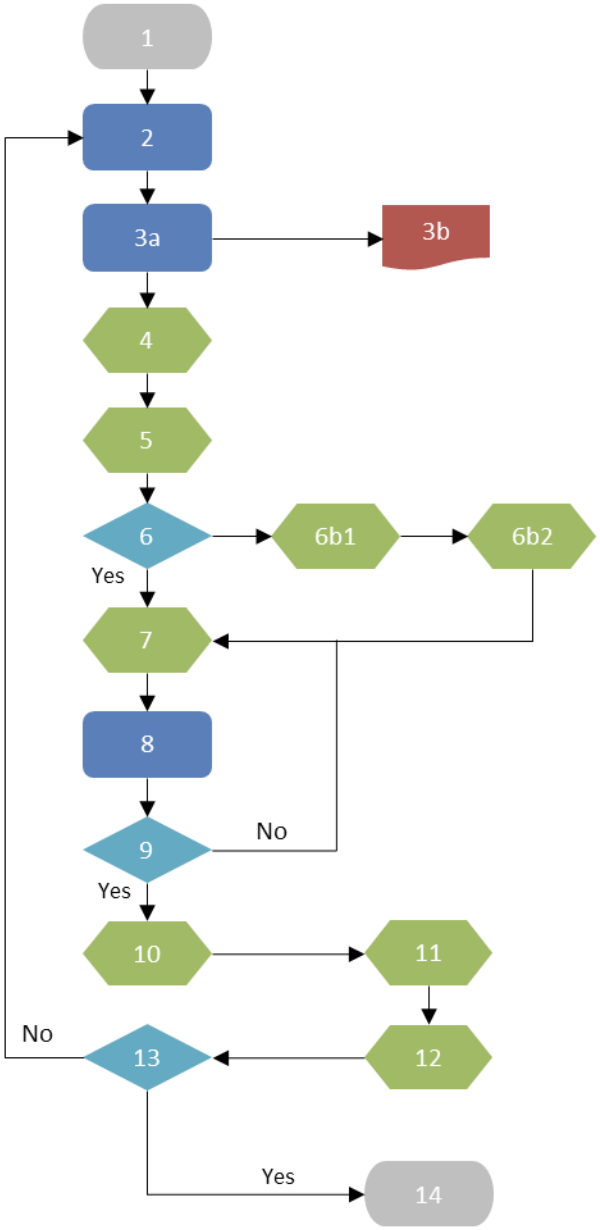
Use Case 1: Marketing Plan



1. Marketing manager is assigned a task to prepare the marketing plan
2. Marketing manager needs to find the current year plan and actual data
3. Marketing manager uses data catalog to look up marketing plan in the business glossary
   a. Follow the links to the existing certified data sources and BI assets (reports and dashboards)
   b. Use the link to the workflow to get access to the selected BI assets which creates an access request
4. Is the access request granted and provisioned?
   a. If no, use link to workflow to revise request as necessary
   b. If yes, next step
5. Explore data in existing BI asset
6. Is data sufficient to complete marketing plan?
   a. If yes,
      1. Use link to initiate workflow to export/link the data to the planning application and go on to next step
      2. Update data catalog with new data set and/or commentary
   b. If no, follow link back to select additional data or BI asset(s)
7. Process complete

Source: Dresner Advisory Services

Use Case 2: Demand Forecast



1. Marketing manager notices a problem with the demand forecast
2. Marketing manager frames the problem including the desired outcome
3. Marketing manager and data scientist discuss the problem (a) and create problem definition document (b)
4. Marketing manager documents problem and acceptance criteria in data catalog collaboration tools
5. Data scientist analyzes existing forecasting data in data catalog and data engineering/prep tool
6. Does forecast include event and new product data?
   a. Yes, next step
   b. No, data scientist finds event and new product data in data catalog
      1. Get access to new data/pipelines
      2. Use link to data pipeline/prep tool and develop proposed features for new model
7. Develop new forecasting model
8. Marketing manager and data scientist review and approve new model
9. New model approved?
   a. Yes, next step
   b. No, go back to develop new forecasting model
10. Data scientist documents new model and features in data catalog
11. Data scientist implements and documents new model in model ops platform
12. Data scientist generates new forecast
13. Marketing manager reviews and approves new forecast
    a. Yes, next step
    b. No, marketing manager goes back to define problem and desired outcome
14. Publish new forecast

Source: Dresner Advisory Services

Not All About the Technology

A well-developed, implemented, and maintained data catalog—integrated with data prep, model ops, and self-service BI tools, automated with processes where appropriate and possible, and overseen by proper governance policies—can help reduce the difficulties associated with finding, accessing, and using analytical content.

However, the fact that a majority of organizations find it difficult to find, access, and use analytical content is not due entirely to technology choices. Lower levels of data literacy and less-frequent data-driven decision making also can exacerbate this challenge.

Change Management

Instituting a formal change management program can help build a data-driven culture and increase the frequency of data-driven decision making. This type of program requires motivating everyone to understand and use data, and teaching them how to find and use available data and BI assets. Such an endeavor generally is harder and takes longer than implementing a technology-first approach—which is likely why few organizations have formal change management programs.

Change management techniques can help ease the transition to a more data-driven culture and customize your program's approach. Stakeholder matrices identify levels of engagement required and support necessary for consumers and producers of data and BI assets. Gamification can increase participation while indirectly providing opportunities for applied learning that don't feel like formal training. For example, a contest with prizes may ask users weekly to provide information about different sets of data and analytical content. Completion of such contest-qualifying activities apply and reinforce data and analytics-oriented training and knowledge, which in turn increases data-literacy levels in the organization.