ZS

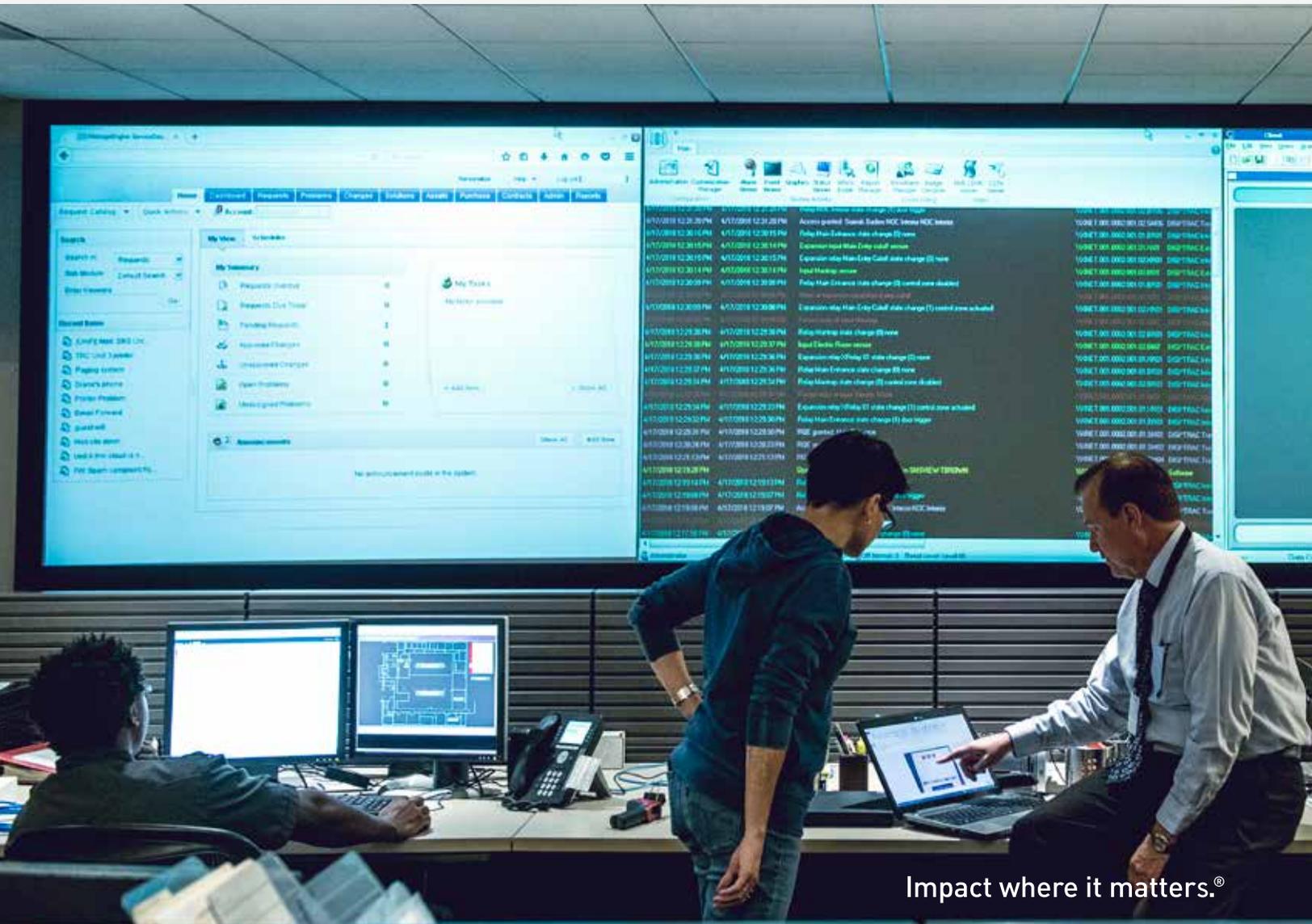# Maximizing your data ROI

How consumer brands can make the most of their data investments

By Srinivas Chilukuri, Russell Evans and Patrick Davis

Impact where it matters.®

The challenge: Brands aren't getting the most from their data

In nearly all industries, heavy investments in analytics have become a "table stakes" cost of doing business. A recent study, for example, found that 70% of senior executives expect the importance of analytics to significantly increase[1] over the next few years. This is particularly true for consumer brands that face pressure from changing shopper behaviors, an evolving retail landscape and a flood of new competitive entrants. A growing list of questions from these businesses—how to acquire more valuable consumers, how to optimize media and trade investments, how to design better products and services, how to navigate supply constraints, etc.—have created a voracious appetite for sophisticated models and the data to power them. Indeed, large consumer organizations can spend millions each year to acquire all manners of data, from broad macroeconomic trends and syndicated product sales to hyper-specific audience targeting metrics.

Yet as data spend continues to rise, our observation is that satisfaction with these investments has generally moved in the opposite direction. One recent study found that only 1 in 4 organizations are fully satisfied with their ability to use their existing data.[2] Business stakeholders still struggle to answer many of the same questions, while data and analytics practitioners spend an inordinate amount of time ingesting and maintaining data. What drives this dichotomy? We believe a primary factor is a bias in many organizations toward acquiring new data. As new business questions arise, traditional analytics processes tend to involve a hunt for the "best" data. Without suitable governance to provide visibility into the current data ecosystem, companies often seek new external sources. Concurrently, an influx of new data providers has emerged with compelling pitches that generate excitement with promises of novel use cases. Over time, the volume of data held by an organization swells, which makes ongoing access and management more painful and dramatically reduces the return on this investment (actual and perceived).

Beyond the (substantial) direct financial cost, there is a real organizational burden associated with unwieldy data expansion. Decision-making cycles are often negatively affected when stakeholders are required to sift through a sea of data with high potential for divergent sources of truth. Holding large volumes of data also introduces legal risk to an organization, as privacy preferences and regulations continue to evolve.

1 Tom Davenport, Jim Guszcza, Tim Smith and Ben Stiller. "Analytics and AI-driven enterprises thrive in the Age of With." Deloitte Insights, July 25, 2019. https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

2 "Data Versus Goliath: Customer Data Strategies To Disrupt The Disruptors," Forbes Insights, June 2018, https://www.treasuredata.com/resources/data-versus-goliath/

A different approach to data ROI

Our recent work with consumer brands suggests there is a different, more purposeful approach that will help maximize return on data investments, improve business decision-making and reduce regulatory risk. Before the wholesale acquisition of entirely new datasets, an organization should instead employ these four tactics (Figure 1):

1. Find new uses for existing data.

2. Procure small, targeted infusions of data to fill gaps rather than bulk datasets.

3. Use artificial intelligence (AI) to enhance the quality of existing data.

4. Dispose of data when it's no longer required.

FIGURE 1

Four tactics to improve data ROI

Here are some real examples of these tactics based on our work with highly successful consumer brands.

1. Find new uses for existing data

Consumer brands are awash in underused data. Most often this arises when data exists in one part of the business but remains unknown to other teams or functions, such as:

- Data that has such an obvious primary application that it is overlooked for additional secondary applications (Example: Syndicated retail sales data)

- Data generated as a byproduct of ongoing operational processes or audience engagement (Example: Email response patterns)

- Data that is incidentally acquired as part of a larger dataset (Example: Timestamp information from Voice of Customer programs)

- Data hosted in previously inaccessible systems (Example: Retail sales associate performance)

Finding new applications for this readily available data is often the easiest way to immediately extract more value from existing investments.

Case study: Trade promotion compliance data

Many manufacturers have co-marketing agreements with retailers where a portion of the retailer's media spend that features the manufacturer's brand will be reimbursed. As part of this process, the manufacturer often receives final marketing collateral (e.g., TV commercials, newspaper inserts, etc.) before and/or after deployment for compliance purposes. This helps the manufacturer ensure that appropriate brand standards were followed prior to reimbursement (e.g., the right logos were used, placement relative to competitive brands was honored, etc.). Arrangements like this represent a significant (and often underused) data opportunity for manufacturers. Because they typically interact with multiple retailers, manufacturers can benefit from this "natural experiment" by applying computer vision and natural language processing to understand the effectiveness of many different marketing and creative approaches, such as:

- What types of offers tend to work best for my brands?

- What messaging resonates most strongly for each of my brands?

- Does a given message perform more strongly in one channel versus another?

- Is it more important to be featured alongside competitive products or as the only product from a given category?

- Against which competitors does each brand perform most favorably?

Answers to these questions can be used directly to improve the manufacturer's creative development as well as indirectly as a source of intelligence for high-value retailer partners.

Case study: Product reviews data

Many consumer brands capture product reviews and report aggregate performance (e.g., average star rating) as a measure of commercial success. But few brands are really leveraging this source of information to its fullest potential to understand things like:

- The underlying needs driving consumer purchasing

- Usage and consumption contexts

- Incremental impact of individual features and functions

- Relative performance versus competitive products

Until recently, acquiring data like this required highly manual, costly efforts. But advancements in natural language processing now enable brands to unlock rich insights from unstructured text at scale with relative ease. For example, ZS's solution, ZS Atlas Intelligence Market, has been specifically trained to extract business-relevant topics (e.g., consumer needs, product features and functions, and usage contexts) from a wide variety of previously underused data sources including product reviews, contact center transcripts and investor reports. These insights can replace other more costly forms of market research and create a deeper, more fundamental understanding of the category to inform efforts such as product or portfolio strategy, brand positioning, product messaging and media targeting.

2. Targeted data infusions

In some situations, acquisition of new data from external sources is necessary. Historically this has resulted in the wholesale purchase of entire datasets, even if a brand ultimately only needed a handful of specific metrics. Recent advancements in the broader consumer data universe now enable a more targeted approach—buying small batches of individual features—that can substantially reduce the required investment by augmenting a brand's existing data. No longer must brands buy the entire farm if all they want is one chicken.

Case study: Feature-level data marketplaces
Data marketplaces that connect brands to data vendors of all varieties have become relatively commonplace. Until recently, however, most marketplaces still required brands to purchase entire datasets. Now players such as Nitrogen.ai and Explorium have emerged that enable feature-level purchases from one or even multiple data sources. Brands can upload a use case or dependent variable of interest and be connected with data providers via automated feature discovery, with a simple checkout process that any Amazon shopper would recognize. The ability to acquire only what's needed helps reduce costs and speed to develop new models. At the same time, many of these marketplaces also enable brands to easily monetize their own first-party data, which can open net-new revenue streams.
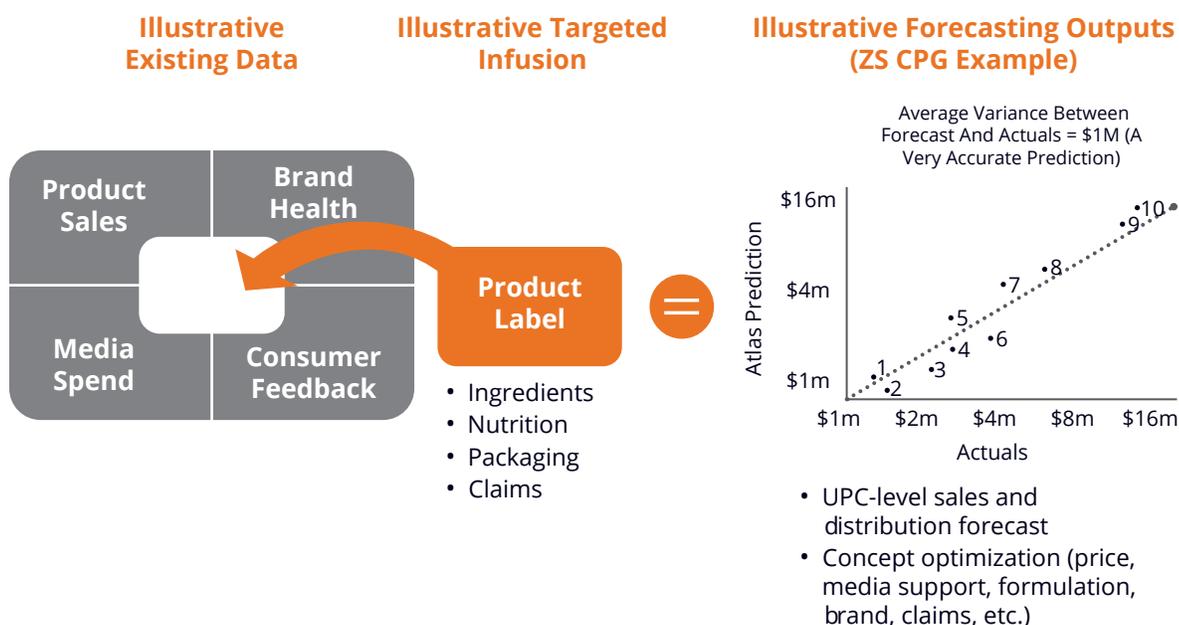
Case study: Product label data
Forecasting the commercial success of new product concepts is old hat for most consumer brands. Current techniques often leverage a combination of category sales data augmented by market research, which can be slow, expensive and not terribly accurate. In fact, many new ideas get little forecasting support before reaching shelves, which results in high launch failure rates. With the targeted infusion of label data, however, consumer brands can apply machine-driven approaches to concept forecasting that are far superior because every idea can be forecast with high accuracy and in real time.

ZS has worked with a variety of consumer packaged goods (CPG) brands to deploy our AI-based new item forecasting solution, ZS Atlas Intelligence Innovation, in just this manner. By infusing a small amount of label data (e.g., nutrition, ingredients, marketing claims, package imagery, etc.) with the considerable data held by most consumer brands (see Figure 2), this solution can determine the incremental value of each product's component parts. These parts can then be dynamically reassembled into new concept ideas with their associated value determined by real consumer behaviors (and not self-reported likelihood to purchase). The resulting forecasts are highly accurate (within 90% of real market launches) and can be used to accelerate innovation pipelines, discover new concept ideas and drive M&A.

FIGURE 2

Illustration of label data infusion

Illustrative Existing Data — **Illustrative Targeted Infusion** — **Illustrative Forecasting Outputs (ZS CPG Example)**

- Product Sales
- Brand Health
- Media Spend
- Consumer Feedback

Product Label
- Ingredients
- Nutrition
- Packaging
- Claims

Average Variance Between Forecast And Actuals = $1M (A Very Accurate Prediction)

- UPC-level sales and distribution forecast
- Concept optimization (price, media support, formulation, brand, claims, etc.)

3. Improving data with AI

In recent years, analytics teams have increased their focus on unstructured datasets such as call center transcripts and customer satisfaction studies—which is understandable, as these sources are generally "messy" and have high potential to produce rich insight. For most organizations, however, the vast majority of data remains in traditional structured forms (e.g., tabular, time series, relational, etc.). While these sources are more familiar, they are often far from perfect and can contain many issues that reduce their value for the purposes of generating actionable insights. Over the years, many organizations have developed hacks or workarounds to address issues like missing values or class imbalances. But artificial intelligence now enables significant improvements in data enrichment that can dramatically increase the business value of structured data.

One such approach relies on a new form of neural networks, called Generative Adversarial Networks (GANs), originally developed to generate images. GANs typically comprise two competing modules—a "generator" that produces new datapoints and a "discriminator" that attempts to determine if that new data is similar to actual data or not. Both modules learn in this dueling fashion over time so that a fully trained GAN can produce high-quality facsimiles of the original source material, either of an image or a structured dataset.

Case study: Synthetic data generation using GANs
Tabular GANs (TGANs) can be used to generate synthetic data that could augment existing datasets and/or enable broader access to data without revealing the actual data. This can be extremely valuable in cases where original source data is insufficient for analysis (e.g., because it's a small dataset), is missing a large number of values or contains sensitive information that can't easily be shared. TGANs are capable of learning the underlying distribution of the training sets so that they can mimic the features and their correlations in the original source data and generate new, artificial data with similar properties.

In work with our clients, ZS has used this approach to:

- Generate synthetic versions of sensitive datasets, like HR data, clinical trial study data and real-world consumer transactional data. These synthetic datasets enable companies to open up sensitive datasets to multiple downstream applications without compromising privacy and security. Some of these downstream applications include:

  a. Enabling ETL systems that can be built using this data and directly deployed to production, reducing the trial and error involved in getting the data structure right

  b. Enabling downstream analytics and reporting for business consumption

  c. Enabling access to external research community to benefit from wider participation

- Augment datasets that have missing values for multiple fields, using GAIN which is an extension of GANs for imputation

Across these efforts, GANs have demonstrated a significant improvement over prior state-of-the-art methods, nearly doubling accuracy in synthetic data generation and boosting rates of successful missing data imputation by more than 20%.

Case study: Oversampling minority class data using GANs
Predictive modelling must take special care of class imbalances when predicting rare events (which by definition appear in datasets less frequently and can skew model performance as a result). GANs can be used to supplement data form minority classes and overcome such challenges. While the architecture is similar to that of the Tabular GANs (TGANs), the generation can be conditioned on the target class and therefore can be used to generate more samples for the minority class. Generated supplementary data can then be widely used to better represent smaller classes (e.g., in better training classification algorithms).

Examples of this approach include:

- Detecting fraud in financial transactions, where the proportion of fraudulent transactions is quite low and within them there can be several different subpatterns

- Predicting the onset of rare conditions in consumer healthcare, where the volume of patients who experience issues and seek treatment can be quite low

Getting started: Mapping your data ecosystem

Maximizing return on data investments ultimately requires coordination across many elements—an organizational data strategy, a comprehensive data management process and an integrated data governance function, to name a few. While these can take years for a larger organization to fully establish, there are ways to make significant progress with more manageable efforts. For example, developing a holistic data ecosystem map (see Figure 3) is ultimately critical to extracting value from existing data in the ways described in this white paper and can be an excellent way to get started.

FIGURE 3

Illustration of a data ecosystem map
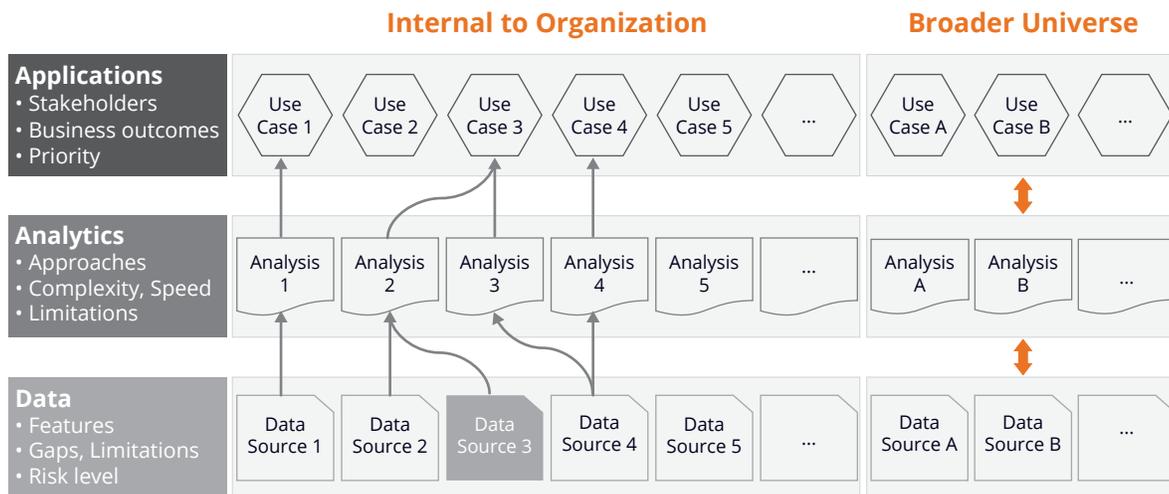
Despite significant investments in data over recent years, many organizations still lack a truly comprehensive taxonomy of their data ecosystem—one that contains both the breadth of information known across all facets of the company and the sufficient depth to maximize usage of this information. Going far beyond a traditional data dictionary, such taxonomies should contain the following elements at a minimum:

1. **Data**: A complete accounting of all raw data held across the organization. For each dataset, this includes the source and cost, specific features held, known strengths and weaknesses, gaps and limitations, and the relative risk of holding this data (e.g., based on external legal regulations but also on internal tolerance and comfort).

2. **Analytics**: An accounting of the analyses being conducted across the organization (by internal and external stakeholders on behalf of the organization). This includes not only the analytical approaches but also the relative complexity, speed and limitations for at least all commonly executed analytics.

3. **Applications**: An accounting of the specific business questions that require answers across the organization (recurring and ad hoc). This includes the requesting stakeholder(s), relative priority and business impact when answered (i.e., what will the answer enable).

To be truly effective, a data ecosystem map must be as comprehensive of the business as possible, which will require active coordination across groups and functions that may not typically share in this manner. While organizations may already be thinking about some of the above elements internally, they also need to maintain awareness of the broader external

data ecosystem in which they operate. In other words, what are the sources of data, the types of analytics and the business applications that other organizations are leveraging that may prove valuable in the future. An infusion of ideas from outside the business can help jump-start your realization of data value.

The benefits to creating a view like this are substantial. With a comprehensive understanding of their data ecosystem, an organization can (see Figure 4):

1. Find new uses for existing data and analyses by identifying a use case that isn't currently being served. Seeking inspiration from outside the business can often be most helpful here.

2. Identify opportunities to enhance existing data with targeted infusions from external sources. This should be done sparingly to avoid introducing data sprawl and complexity.

3. Leverage new analytical approaches to improve the quality of existing data sources. We've highlighted GANs as one example in this article but there are many other techniques available to achieve similar outcomes.

4. Identify opportunities to dispose of data sources that aren't being used. This can reduce immediate costs, while also limiting future liability from data that is inappropriately held or used.

FIGURE 4

Benefits of a comprehensive data ecosystem map
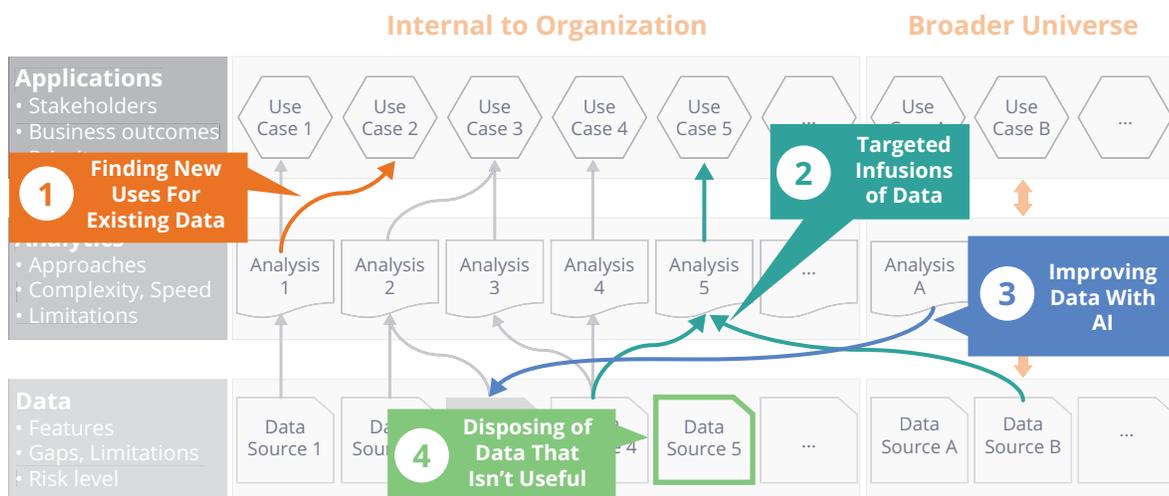
Developing a data ecosystem framework like this is a relatively straightforward effort that can significantly increase the return on data investment for any consumer organization.

About the authors

Srinivas Chilukuri is a principal based out of the Evanston office and is the leader of ZS's North American AI Center of Excellence. This team brings cutting-edge AI solutions to clients across industries. Srinivas's core expertise areas include automated machine learning, natural language processing, customer omnichannel next-best action orchestration and longitudinal customer journey predictive analytics. He has authored several thought leadership articles and speaks regularly at conferences. Srinivas has an MBA from the Indian School of Business in Hyderabad. He holds a bachelor's of technology in mechanical engineering from the National Institute of Technology in Warangal, India.

Russell Evans is a principal in ZS's Chicago office and leads the B2C go-to-market practice area. He has extensive experience leveraging data and analytics as a foundation for developing actionable growth strategies, new products and services, user experiences and cross-channel customer engagement programs. Russell has worked across a number of industries, including packaged goods, technology, retail and hospitality. He helps clients with solutions including customer engagement strategy, personalization, customer journey mapping, customer segmentation, demand-driven innovation, campaign/tactic executional strategy and program measurement. Russell holds a degree with highest honors in ecology and evolutionary biology from Princeton University.

Patrick Davis is based out of the ZS Cambridge office where he leads engagements on strategy and the application of analytics to solve business problems, which has been his focus for more than a decade across multiple industries. He helps clients find new uses for the data they already have, as well as set up their analytics solutions. Patrick graduated with honors from the University of Michigan, where he was awarded a bachelor's degree in mechanical engineering and a master's degree in operations engineering.

About ZS

ZS is a professional services firm that works side by side with companies to help develop and deliver products that drive customer value and company results. We leverage our deep industry expertise, leading-edge analytics, technology and strategy to create solutions that work in the real world. With more than 35 years of experience and 10,000-plus ZSers in more than 25 offices worldwide, we are passionately committed to helping companies and their customers thrive.

Learn more: www.zs.com/industries/consumer-goods