MAKE INSIGHTS ACTIONABLE WITH AI & BI

Collective advice from 20 + Data Leaders

ATSCALE

Contents

ATSCALE

From the Editor’s Desk

Data and Analytics (D&A) is considered the next frontier for innovation and productivity in business. A Mckinsey report says data-driven organizations provide EBITDA increases of up to 25 percent. According to Boston Consulting, the first 9 of the top 10 innovative companies are data firms. But achieving a sustainable competitive advantage from D&A is challenging. Many D&A projects are unsuccessful. Gartner says, only 20 percent of the Data and Analytic solutions deliver business outcomes. So, how can organizations get value from Data and Analytics? Research by DBP Institute found there are three main areas to get improved business results from D&A and they are (a) Data culture (b) Quality data and (c) Data literacy.

Against this backdrop, after the roaring success of the first edition of the book - “Make AI & BI Work At Scale,” where over 10,000 eBooks were downloaded, I worked with numerous Data and Analytics experts and put together the second edition of the book, “Make Insights Actionable with AI & BI” or rather of a Body of Knowledge (BoK) on how Data and Analytics including the Semantic Layer can help organizations achieve AI and BI at scale. Specifically, this eBook is centered on how can enterprises leverage the Semantic Layer to achieve AI (Artificial Intelligence) and BI (Business Intelligence) at scale, given that the Semantic Layer helps business users access data using common business terms? The Semantic Layer, if managed well, can play a pivotal role in managing this change and improving business performance.

While some of the positive elements that worked during the preparation of the first edition of the book were carried forward, the book was further expanded to bring more thought leaders to share their experience. Specifically, the second edition of the book gives a holistic perspective to the Data and Analytics community covering data management, data engineering, data science, and decision science to improve the odds of delivering Data and Analytics solutions successfully. The authors in the 20 chapters who have contributed to this book include industry practitioners, subject matter experts (SMEs), and thought leaders who have a stellar track record in leveraging Data and Analytics solutions for improved business performance.

Today, every company intends to leverage data and analytics to fuel the business performance and build a sustainable competitive advantage. The benefits of data and analytics solutions is not just in creating new revenue streams but also in achieving quicker time-to-market, reducing the cost of business operations, and minimizing business risk. I sincerely hope this eBook will help you and your organization get improved business results from data and analytics. All the best!

Your Sincerely

Prashanth Southekal, PhD, MBA, MS
Editor of Make AI & BI Work At Scale (2nd Edition)
Managing Principal, DBP Institute & Professor at IE Business School, Spain

Foreword

Kamayini Kaul

Global Head Information Insights and Analytics, CSL Behring

“Data is the new oil.” — Clive Humby, British mathematician, Chief Data Scientist of the consumer insights company, Starcount

“Errors using inadequate data are much less than those using no data at all.” - Charles Babbage, mathematician, philosopher, inventor, and mechanical engineer

“Information is the oil of the 21st century, and analytics is the combustion engine.” - Peter Sondergaard, Senior Vice President and Global Head of Research at Gartner, Inc.

“In God we trust, all others bring data.” — W. Edwards Deming, American engineer, statistician, professor, author, lecturer, and management consultant

“With data collection, ‘the sooner the better’ is always the best answer.” — Marissa Mayer, American Investor and former CEO Yahoo

The digital news and media today are replete with idioms and metaphors like the ones above. A myriad more are added daily to the ever-growing quotable quotes' annals. Modern enterprises, small and large, are trying to hire, train and upskill data leaders who need to navigate the everevolving and competing challenge of value delivery while designing, building and futureproofing an enterprises investment in data and analytics

In a world where speed to outcome is the surest path to a durable competitive edge, data leaders are expected to walk with a magic bullet of executing with accuracy and precision at scale. Balancing accuracy with speed, it turns out, requires knowing how to leverage collective learnings and communal wisdom while deftly side-stepping known pitfalls and dead ends. From a data strategy that delivers a digital moat to the corporate strategy of your enterprise, to analytics roadmaps and plans guided by that data strategy; and from technology engineering and orchestration to the last mile of front-line operations optimization charged with the insights generated – it's a tall order of asks in a hyper dynamic and volatile business and economic context. A leader's burden lies in knowing what tools to wield when, what solutions buy/build/rent and what skills and workforce capabilities to insource/outsource/partner cocreate

The authors in the chapters ahead are colleagues and thought leaders in data with well over 450 years of cumulative corporate and start-up experience. The lessons and teachings that follow are a masterclass from some of the best minds in data and analytics. Some, I've had the privilege of being an avid pupil of and others a colleague and partner in arms to. Successful scaling is what I like to call the art of surrounding yourself with innumerable coaches, mentors, advocates, teachers, allies, and partners, all at the same time.

To the readers who join us in the pages and hours ahead –

May our collective hindsight become your personal leadership foresight!

Kamayini Kaul
VP, Global Head Information Insights and Analytics
CSL Behring

Data and Analytics Value drivers

Andy Keller

Vice President, Enterprise Data,
Cardinal Health

The innovation enabled by data and analytics is all around us and new capabilities seem to emerge daily. A couple recent examples are sensors that recognize smells and visually identify changes in appearance that could alert patients and providers to potential health issues. The speed of innovation will continue to increase as AI and new technologies unleash the power of data to augment people, continuously learn and accelerate how we work and live.

While few would argue that data is foundational to accelerating and realizing the potential of this exciting new digital world, many leaders struggle to gain quick access to the core metrics and reporting critical to improving their business performance. When they do have access, there are often gaps in trust or the predictive capabilities necessary to target proactive actions for improving performance that analytics can provide.

Enabling digital capabilities and maximizing the power of data and analytics is increasingly recognized as a strategic priority. Many of the principles core to accelerating this transformation are covered in this book and require a disciplined approach and operating system to mature to world class. Generating visible value while building these capabilities is recognized as a key to success in this journey to transform and change.

Because of this, focusing on the presentation of this data and the analytics to produce insights and enable action is often a priority. Improving the speed to find, access, understand and trust the data results in faster decision making and project delivery. As the foundation of trusted data improves, the time spent finding, understanding, and curating the data shifts to increased time generating insights, automating, and innovating.

This focus on improving the business intelligence and analytics at the point of consumption with the emergence of powerful tools like digital twins of the supply chain and 360 views of customers will continue to increase the impact and visible creation of value. While enabling and advancing these capabilities is important, ensuring a focus on managing and enhancing foundational master and other operational data is critical to performance across the value stream. Core data assets like product and customer data are used continually as customers search for products, order products, and track their shipments. The quality and fidelity of this data enables a world class ordering experience across channels powered by agile, resilient, and optimized supply chains. Improving data management capabilities and the quality of the data reduces “data friction” across supply chain partners and leveraging data from sensors and/or unstructured external data can help predict changes before organizations see the changes in their supply chains.

This book covers foundational principles and examples of the world class capabilities that enable and accelerate the value of data and analytics. The value and impact of the digital world will continue to unfold as automation, the metaverse other technologies change how we work and live – the speed and impact of this change all have the common thread of data.

Andy Keller
Vice President, Enterprise Data, Cardinal Health

Chapter 1

Analytics for The Win through the Data Lens

Kirk Borne

Chief Science Officer, DataPrime Inc

About 20 years ago, when I was still working at NASA, NASA issued a general announcement requesting concept papers from the community describing potential new technologies to support future human missions to the Moon, Mars, and beyond. By that point in my career, I saw everything through the "data lens". I was already (at the time of that announcement) 15+ years into my career as an astrophysicist, during which I spent many long hours on data analysis. But there was more to data than scientific analysis.

As I worked on massive data systems for NASA space astronomy missions for most of those 15+ years, I discovered a "new" set of mathematical algorithms and methods for insights discovery from big data: data mining and machine learning (now collectively called analytics and data science). As a lover of mathematics since childhood, I was drawn to these new methods, like a bee to honey.

To acknowledge my growing fascination with insights discovery from massive data, one of my NASA colleagues gave me a data-related gift when I left NASA in 2003 to join George Mason University – he gave me a children's toy hammer with a handwritten label that read, "Kirk's data mining hammer." This declaration referred to the saying, "to a child with a hammer, all the world is a nail." My colleague would be amused at how I could find an application of data mining (machine learning and data science) to about every problem I encountered. One might say, then and especially now, that I could drone on about data for hours. Data mining was my hammer, and the world of data was my nail. Well, we now know that last part was right – the big data world is becoming everyone's "nail"! We also know analytics and data science are becoming a universally required "hammer" in every domain.

Getting back to the story about that NASA announcement: from the perspective of my datacentric universe, I envisioned a predictive analytics concept for NASA's future Moon (and ultimately Mars) missions. Specifically, it was a supply chain analytics concept. I imagined there would be a time when the distant orb would have multiple bases with multiple teams of humans and robots doing scientific, engineering, mining, and other activities across that distant landscape. To support these distributed activities for extended periods (months or years), an efficient steady supply chain would be needed – preferably (for cost reasons) there would be a single delivery node servicing all receiver nodes.

As I developed my lunar supply chain concept, I noted that one couldn't simply order a needed item and expect quick delivery from Earth. So, I conceived an in-orbit warehouse orbiting the Moon (or Mars) with most of the needed supplies. The warehouse would be replenished autonomously via insights learned through mining the database logs of available supplies, past usage rates, predicted future usage needs, and so forth. The predictive analytics system would trigger timely shipments from Earth for just-in-time delivery to the distant celestial surface. In addition, bespoke "space-to-surface" supply replenishment deliveries from Moon orbit to a specific lunar base could also be automatically invoked using those same database logs of usage statistics and supply chain analytics workflows for each lunar base. Hence, exploring distant worlds would be supported by essentially the same just-in-time global inventory replenishment systems that major retailers already use worldwide!

That was then – and this is now: into the data-driven world of supply chain analytics, enter the drone! Several large retailers (both brick-and-mortar and e-commerce companies) have proposed to use (or are already using) drones to deliver product shipments to your home: "airto-surface" supply replenishment deliveries! It takes just one more step of analytics logic to imagine that this can be done through autonomous just-in-time replenishment and proactively sending items to your doorstep as determined through predictive supply chain analytics.

Not only that, but also consider this: nearly all e-commerce organizations use recommender engines for both upsell and cross-sell opportunities, using predictive data analytics models of consumer buying patterns and behaviors. A drone-powered delivery e-commerce business can use their customer data to make offers and recommend other products to consumers before the delivery of their initial requested items, such as: “special discount on your new HDTV if you order it in time for your next cheese and popcorn shipment.” The potential for increased sales and revenues from such drone-powered data-driven analytics-informed supply chain delivery systems could be astronomical, to the moon even!

A Question of Strategy

Before we continue, we need to address business strategy around data and analytics. We hear a lot of organizations say they are “Data-first”, or “AI-first, or “Data-driven”, or “Technologydriven”. A better prescription for business success is for an organization to be analytics-driven and thus analytics-first, while being data-informed and technology-empowered. Analytics are the products, the outcomes, and the ROI of our Big Data, Data Science, AI, and Machine Learning investments!

AI strategies and data strategies should therefore focus on outcomes first – your organization’s mission, its “north star”, itsraison d'être. Such a focus explicitly inducesthe corporate messaging, strategy, and culture to be better aligned with what matters the most: business outcomes! I call this Analytics by Design. Why? When considering the importance of an analytics-first strategy in business, we should think about our outcomes first, then our data. Analytics by Design is analogous to similar principles in education: Understanding by Design, where the focus is on outcomes first, then secondarily on designing programs and activities that will achieve those outcomes, plus defining metrics to measure progress toward achieving those outcomes. Thus, as analytics are the outcomes of the organization’s data activities, including data science, machine learning and AI, an Analytics by Design organizational strategy is beneficial. The outcomes of programs and activities should meet business goals and objectives, and business metrics (KPIs = Key Performance Indicators) are derived from those. The analytics metrics, therefore, measure performance against desired analytics outcomes.

A Plethora of Analytics Use Cases

The above lunar supply chain story illustrates just one example of a broader category of business analytics applications: predictive analytics on real-time (perhaps streaming) business data. Here are a few other examples of such applications:

1. Real-time credit risk prediction.
2. Real-time fraud risk prediction.
3. Real-time personalized customer interactions.
4. Real-time context-specific and location-sensitive product and/or content marketing to consumers.
5. "Do Not Pay" classification on fraudulent insurance claims before payment.
6. Real-time determination of benefits eligibility to mitigate underwriting fraud.
7. Detection of insurance rate evasion tactics within the policy quote process.
8. Optimal actuarial price determination at the point of policy-quote decision-making.
9. Health risk prediction at the point of healthcare decision-making.
10. Real-time detection of anomalous and adversarial cyber network behaviors.
11. Data breach prediction and intervention before it happens.
12. Illegal funds transfer prediction and intervention before they happen.
13. Non-compliant business transactions intervention and risk reduction before they happen.
14. Optimized supply chain and warehouse product flows: positioning the right products in the right quantities at just-in-time locations.
15. Predictive product demand and pricing by finer levels of product subcategories.

For essentially each of these business applications, there is potentially a use case across the full spectrum of analytics dimensions: detection, prediction, and intervention (Descriptive, Predictive, and Prescriptive Analytics).

Signals (data) from ubiquitous digital sources within our enterprise systems carry transactional information (what happened to what?), contextual metadata (where and when is it happening?), and analytics information (what insights do the patterns in the data encode?). The analytics can be descriptive/diagnostic (describing what has happened or what is happening within a transaction), predictive (providing behavioral insights into the interests, intentions, and preferences of the actors within that transaction), and prescriptive (providing actionable insights into the right decisions and actions to take).

Those different analytics applications alone can expand into a vast number of unique use cases across hundreds of different disciplines and industries: agriculture and forestry, arts, charities, commerce, construction, cyber, defense, education, energy, engineering, entertainment, fashion, finance and insurance, food services, gaming, government, health, information services, innovation, investments, manufacturing, media, mining, music, oil and gas production, policy, public service, publishing, real estate and property management, recreation, retail, science, social services, technology, textiles, tourism, transportation, travel, utilities, waste management, wholesale, and others.

I compiled a big list of some of those analytics use case possibilities (in the attached table). Nearly all of these have applications in multiple diverse industries. If each one of those industry-specific applications (e.g., the 15 listed above) were enumerated for each of the 70+ listed analytics use cases (tabulated in the attached table), the expanded list would easily consist of over 1000 ways to use analytics for the win, through the lens of data (informative insights-carrying digital signals from enterprise systems).

Common Analytics Challenges and Myths

As analytics practitioners, leaders, and stakeholders, who are we? We are digital (data) professionals entrusted with our organization's massive digital (data) assets. Our mission is to discover insights and to deliver value from these data. We are tasked with doing that efficiently (measured by time-to-solution) and effectively (measured by the completeness and accuracy of the solution). We can categorize the challenges we encounter with this assignment into three categories:

1. Finding competitive advantage for the organization in data analytics, including data science, machine learning, artificial intelligence (AI), and automation.
2. Acquiring, nurturing, benefiting from, and retaining key data science and analytics talent.
3. Avoiding the hype, shiny object distraction, and FOMO (Fear Of Missing Out) can drive misguided motivations and "busy work" around those big data assets.

Unfortunately, each challenge is often associated with one of these three myths and our corresponding natural reactions to them, respectively:

Myth #1: **Machine learning and AI are big, scary topics.**

"How do we get our people to agree that our organization needs to pivot into doing this work?" = Fear!

Myth #2: **Data science and analytics are only for data scientists.**

"What if our experts leave?" = Fragility!

Myth #3: **Data-first is the right strategic posture for success.**

"How do we get value from data with so many diverse tools, techniques, technologies, talents, and technical debt?" = Friction!

The result of our natural reactions to these challenges and myths is "3 F's" on our analytics scorecard: Fear, Fragility, and Friction. That spells "Failure" unless we adopt a different mindset.

Adopting the "Analytics by Design" Mindset

Fortunately, there are three better responses to the three challenges and three myths than the three Fs, respectively:

1. Embed the data insights discovery technologies (analytics, machine learning, AI, and automation) within existing (already adopted) enterprise tools and systems.
2. Adopt a culture of experimentation and a "data for all" (data literacy) policy across the organization.
3. Analytics-first is a better strategic posture for success than "data-first". The "Analytics by Design" strategy (mentioned earlier) directs the corporate focus onto the analytics: i.e., the products and outputs (not on the data, the input). The analytics focus explicitly induces the organization's communications, culture, and corporate activities to align with what matters the most: mission objectives, business outcomes, value creation, and competitive advantage. Every organization now has tons of data. But your analytics outcomes and products are your organization's value proposition – its unique, invaluable output.

Empowering all digital workers in your organization to innovate, deliver ROI, and create value from data assets, even on small projects, will inspire greater digital transformation, data analytics adoption, and cultural change. Committing to a clear, analytics-first strategy prepares your entire organization for the larger enterprise implementations (machine learning, AI, and automation) that will come.

Data analytics "busy work" (motivated by hype and FOMO) without a mission outcomes focus can produce another "F" – "Fantom" (phantom) analytics – corresponding to lots of activity around data, with little value or results to show for it.

A STELLAR Framework for Enterprise Analytics

STELLAR analytics can boost analytics performance from early-stage "sandbox" experiments to late-stage enterprise projects. The key is to get moving, keep moving, and accelerate forward progress. If the inhibitors (fear, fragility, and friction) stand in the way of analytics progress, then the project will likely receive a failing grade of "F". A better grade is not just a grade-inflated "A", but a stellar "A", which is enabled by STELLAR analytics. I describe here my definition of STELLAR analytics.

The seven dimensions of analytics associated within the STELLAR framework include: Streaming, Team, Edge, Location, Learning Business System, Agile, and Related-Entity Analytics. Here is a short description of each:

- Streaming Data Analytics: real-time access to, interaction with, and discovery from data,such as detecting POI (persons, patterns, products, processes, or points of interest) and BOI (behaviors of interest for any “dynamic actor”).
- Team Analytics: a culture of experimentation that celebrates and validates the power in diversification, collaboration, data-sharing, data reuse, and data democratization.
- Edge Analytics: locality in time, at the moment of data collection (enabled by the Internet of Things [IoT]) – “What else is happening now?”
- Location Analytics: locality in geospace, within a given spatial context (also enabled by IoT) – “What else is happening at that place?”
- Learning Business System: A learning business system embodies data-driven knowledgegeneration business practices, with performance measurement, continuous feedback, learning, and improvement, which are embedded in daily business practice (example: Learning Health Systems).
- Agile Analytics: outcomes-driven, iterative, builds proofs-of-value, fails fast to learn fast, thinks big, but starts small (with the Minimum Viable Product or Minimum Lovable Product) with continuous integration and delivery through DataOps (DevOps for data analytics).
- Related-Entity Analytics: locality in data feature space – “What else is like this entity/ event in my data collection?”

The STELLAR analytics objectives invoke data science techniques to detect existing, emerging, and actionable patterns in data. Such patterns include: (a) segments (classes), (b) trends (correlations), (c) surprises (anomalies, outliers), and (d) linked entities and events (co-occurring associations).

Moving Business at the Speed of Data

The mission of data-rich organizations is this: to produce successful business outcomes and value from data through analytics. Since the rate at which data flows through organizations is lightning fast, users of analytics applications need strategies, tools, and techniques that quickly leverage those data to extract insights, make data-driven decisions, and take the next best actions – to help a business move at the speed of data!

To achieve those goals, developer teams must be able to provide agile user-accessible analytics, data exploration dashboards for “any user”, and capabilities for user-generated actionable insights – all while avoiding the three F’s (fear, fragility, and friction) in the end user’s experience of their analytics applications and the fourth F (“Fantom” analytics) in enterprise projects.

STELLAR analytics boosts the ability of organizations to deliver big value from big data and to achieve straight A’s on their business outcomes scorecard -- indicators of stellar performance and clear forward progress in enterprise analytics projects. Adopting STELLAR analytics is a good data practice and adopting an analytics-first strategy is a good business practice. With that, you can give your team an A+ for analytics mastery – that will deliver Analytics for the Win through the Data Lens!

Author Biography

Dr. Kirk Borne is the founder and owner of Data Leadership Group LLC. He is also the Chief Science Officer at AI startup DataPrime. Before that, he was the Principal Data Scientist, Executive Advisor, and Data Science Fellow at consulting firm Booz Allen Hamilton. Before those roles, Kirk was Professor of Astrophysics and Computational Science in the graduate and undergraduate Data Science degree programs at George Mason University for 12 years. He taught courses, advised students, and researched data-intensive domains. Before that, he worked for 18 years on various NASA contracts -- as a research scientist, as a manager on a large science data system contract, and as the Hubble Telescope Data Archive Project Scientist. He has applied his expertise in science and large data systems as a consultant and advisor to numerous agencies and firms. He is also a blogger (rocketdatascience.org) and actively promotes data and analytics literacy for everyone on social media, where he has been named consistently since 2013 among the top worldwide influencers in big data, data science, AI, machine learning, IoT, and digital transformation. His PhD is in Astronomy from Caltech.

Chapter 2

The 10 Crucial Elements of a Winning Data & Analytics Strategy

Ganes Kesari

Ganes Kesari, Co-founder & Chief Decision Scientist, Gramener

Imagine you're the captain of a boat moving upstream. Your crew of five struggles to row against the current. You then add a new member to the team. This person jumps into the boat, picks up an oar, and enthusiastically rows —but in the opposite direction. Their efforts aren't just in vain; they're slowing the entire boat. By rowing backward, this new arrival cancels part of the effort spent by the rest of the crew. This happens when your data and analytics (D&A) efforts don't align with your organization's business goals. This typically happens when the data strategy is weak.

Many organizations venture into D&A with no data strategy. The results are equally disastrous. I've often observed this unfold while working with chief data and analytics officers (CDAOs) in my data advisory engagements across industries. To pinpoint why this happens, we'll look at the ten critical elements of a winning data & analytics strategy using examples from industry experts. Why do organizations drive blind on the data highway? McKinsey found that just 30% of organizations align their data strategy with their organizational strategy. By implication, 70% of leaders are burning money in the name of data.

Why do organizations make this fundamental mistake of not having a string data and analytics strategy? “Truly impactful data strategies are only just coming into their own due to the advancement of technology and maturing of things like enterprise artificial intelligence (AI),” says David Benigson, CEO of Signal AI. Signal AI is an AI-powered business intelligence and media monitoring company that aggregates, analyses, and provides business leaders with insights into digital, print and broadcast media, news, and regulatory data.

Despite the buzz around data analytics and AI, leaders often view these elements tactically. They turn to analytics projects to solve localized pain points across the organization. A tactical approach with data delivers marginal results for a business. Benigson has seen such misalignment at legacy organizations that operate with an old-world mindset. They fall short on innovation, fail to implement iterative sprints, and struggle to pivot into an open culture with a data-first mindset. Orchestrating an organization-wide data strategy takes time and effort. “However, failing to plan is planning to fail,” quips Maya Zlatanova, CEO of FindMeCure.

What is data strategy?

Data strategy can mean a variety of things, ranging from the most strategic to the operational. Strategy is defined as a plan to achieve one or more long-term goals under conditions of uncertainty. Similarly, a data strategy is a plan with a set of choices that help achieve long-term business goals. Benigson adds that “a data strategy should enable better decision-making using data. It should define how best to capitalize on, manage, analyze, and act upon data to realize organizational goals.” Against this backdrop, what are the ingredients of a winning data and analytics strategy? Let’s look at the ten critical elements of an effective data strategy framework.

1. Key Business Goals

Since a data strategy is crafted in service of the business goals, there must be absolute clarity on the organizational vision and key business priorities. Ask the leadership team, “what are the longterm business goals?” Review and internalize the organizational strategy. Consider the business goals of Janssen Pharmaceutical Companies owned by Johnson & Johnson: “We aspire to transform lives by bringing lifesaving and life-changing solutions to people who need them. We’re committed to providing safe and effective medicines as well as the services and support that contribute to healthy outcomes.” We’ll next see how data can enable this goal.

2. Data & Analytics Vision:

Once you understand the business goals, find how data and analytics can help you achieve them. This vision for data can guide your choice of stakeholders, selection of initiatives, and validate whether they deliver the intended outcomes. “At Janssen, we’re applying data science end-toend across our portfolio by focusing on finding solutions to big questions that can advance our impact on patients,” says Najat Khan, Chief Data Science Officer & Global Head of Strategy & Operations for Research & Development at Janssen. “It’s helping us better understand diseases and the patients impacted by them. It [helps] us to select the most promising compounds to advance into clinical development; design more efficient, diverse, and targeted clinical trials; diagnose rare and difficult-to-detect diseases earlier; and connect patients with treatment sooner,” she adds.

3. Target Stakeholders:

While picking business objectives, ask, “who do you want to enable through your data initiatives?” While it’s tempting to serve everyone, this isn’t realistic—particularly at the start of your data journey. Pick a set of target departments and roles. For example, you could start with Research & Development and Commercial teams in a pharma firm. The organizational data plan could help accelerate drug development and achieve targeted business growth. Then, you plan to expand coverage to the entire organization.

4. Strategic Initiatives:

Once you define your destination, find which big initiatives will get you closer to your goals. These strategic programs will help your chosen stakeholders achieve their business objectives. (Check out my earlier article on picking initiatives and building a data strategy roadmap.) “Start with the question you are trying to answer and work back since there is no point having data for data’s sake,” advises Benigson. “For example, in the communications function, we now have quantifiable reputation data on the association between brands and topics of interest in the global media.”

5. Measures of Success:

For each strategic initiative, ask, “what will success in these initiatives look like?” People often pick initiatives based on urgency rather than business impact. Documenting the desired outcomes also helps validate whether the chosen initiatives are the most important. “Use key performance indicators (KPIs) and objectives & key results (OKRs) to track progress with your business goals,” recommends Zlatanova. “This could help uncover blind spots along the journey.”

6. Sources of Funding:

Clarity about the initiatives and their outcomes will inform the next question, “who will foot the bills for these programs?” Some organizations make big strategic plans but soon realize they cannot secure the funds midway. Identify your likely sources of funding. For example, some firms assign a fixed percentage of the departmental budget toward organization-wide D&A initiatives. Others allocate a portion of their centralized technology spend for such programs.

7. Top Enablers:

Strategic plans may be easy to make but are tough to execute. Leaders who manage this ask, “what are the strategic ways to support the efforts?” They spot tailwinds within the organization and capitalize on them. For example, a firm with an innovation-friendly culture might see lower resistance to user adoption. Zlatanova shares that “a good strategy should have all the answers for the why, what, and how questions.”

8. Top Challenges:

While identifying the tailwinds, savvy leaders must watch out for the likely headwinds or challenges. Find the biggest roadblocks to anticipate and plan ways to mitigate them. Articulating the support needed with the organizational data strategy is an excellent way to secure the resources. For example, firms often vary in the sophistication and maturity of technology implementation across departments. This could hamper rollouts and must be factored in while planning initiatives.

9. Governance Plan:

To ensure a data & analytics strategy is executed throughout the year, ask, “what mechanisms will track and review outcomes from D&A?” Plan to review the progress not just of business projects but also of technology initiatives such as platform upgrades or upskilling programs. For example, organizations may set a steering committee that convenes quarterly to review the progress of programs, validate outcomes, and greenlight new initiatives—plan for such interventions.

10. Capabilities to Build:

While detailing the data and analytics strategy, you must invest in capabilities across people, process, and technology perspectives. This is crucial for building the strategy execution muscle. Ask how to onboard and empower users, how to rewire your business processes to integrate your D&A initiatives, and how technology strategy can enable D&A efforts. Organizations are often reactive and don't plan for these capabilities. Benigson adds that "companies must make the critical investment of building competencies, adopting new tools, and helping teams keep an open mind to experiment and adopt."

DATA & ANALYTICS (D&A) STRATEGY TEMPLATE

1 **Key Business Goals** What are your long-term business goals	**2** **D&A Vision** How can D&A help you achieve your business goals?	**3** **Key Business Goals** What do you want to enable throught your data initiatives?
4 **Strategic initiatives** What are the big initiatives that will get you closer to your goals?	**5** **Measures of Success** What will success in these initiatives look like?	**6** **Sources of Funding** Who woll foot the bills for these initiatives?
7 **Top Enablers** What are the strategic ways to support the efforts?	**8** **Top Challengers** What are the biggest roadblocks you foresee?	**9** **Governance Plan** What are mechanisms to track and review data & analytics efforts?
10 **Capabilities to Build** People How can you enlist and empower your teams in this effort?	Process What organizational structures and processes will help your initiatives succeed?	Technology How can your technology strategy align with and enable your data & analytics efforts?

Figure 1: Key Elements of Data and Analytics Strategy

(Source: Gramener Data Advisory Framework)

What does it take to execute an organizational data strategy roadmap? Executives play a crucial role in the crafting and realizing an organization's data strategy. Benigson shares that "leaders must help define the most important business problems to solve." They should help set measurable goals for the D&A initiatives.

To nurture an environment conducive to decision-making with data, "[Leaders] must lead by example and show that they use data for their own decisions," advises Zlatanova. Bringing about a cultural shift with data is easier said than done. "But, we've found that early successes generate momentum," reveals Khan. "When you show people what data science can do for patients, it helps them see the possibilities and gets them excited about the potential for impact." She shares an example from the development of their COVID-19 vaccine. "We leveraged data science to identify where the 'hot spots' would be when we were ready to launch our Phase 3 clinical trials. This had a 90% accuracy, down to the county or province level, four months in advance. It helped accelerate our development timeline by six to eight weeks. Examples like these generate significant momentum." To quote the famed military strategist Sun Tzu, "Strategy without tactics is the slowest route to victory. Tactics without strategy is the noise before defeat."

Author Biography

Ganes Kesari is an entrepreneur, AI thought leader, author, and TEDx speaker. He co-founded Gramener, where he heads Data Science Advisory and Innovation. He advises executives of large organizations on data-driven decision-making. His expertise lies in applying data science to solve business challenges and building teams to promote a culture of data. Ganes writes regularly in leading magazines such as Forbes, Entrepreneur, TechCrunch, and The Enterprisers Project. He also runs corporate workshops and teaches data science in leading universities.

Chapter 3

Managing the ML Project Lifecycle

Vidhi Chugh

Data Transformist and AI Strategist

Today many organizations are steering towards building AI (Artificial Intelligence) powered solutions to gain a sustainable competitive edge and stay ahead of the curve. While a lot of it genuinely appears to be the best bet for the organizations sitting on the humongous amounts of data, unfortunately, only data cannot get them to the NorthStar. But often, the organizations rush through AI solutions simply due to the fear of missing out on the AI, including the ML (Machine Learning) wave timely. But lack of proper planning, strategy, and execution roadmap leads to early fallouts and unsuccessful projects. So, if you are also planning to build the gen-next ML-driven solution, here is a list of key elements in your ML project lifecycle.

Data strategy

Organizations often have data but miss out on the strategy and data governance policies. Data Governance Framework requires collaboration from multiple stakeholders to discuss the objective of data, what data is available vs what is needed, where the data would flow from, what transformations it will go through the data pipelines and how it will create actionable insights. If I must sum it in one word, data governance is “all things data”.

A lot of organizations are facing the trouble of how to manage the data and ensure good quality data. But they do not need to learn it from scratch as there exist a lot of guidelines, best practices, and tools that others in the industry are following already. Well, that’s what the frameworks are here for. If you are wondering whether these frameworks can meet the varying data needs of every organization, you guessed it right. There is no one size fits all solution. But an organization can adopt these frameworks and adapt according to their data needs - the keywords here are “adopt and adapt”.

Defining roles and responsibilities play a fundamental role in successful data governance. The data governance council prepares the data policies and strategies that need executive sponsorship and buy-in. These data policies are then baked into database systems to aptly meet the varying data needs of the organization. While Data owners, as the name suggests, own the data assets, the data stewards are tasked with maintaining the data quality and integrity.

Impact and Value Analysis:

Data culture lies on the shoulders of people, processes, and technology. The organization needs to have the people with the right skill set who can not only endorse the processes but also embed them in technology. The skilled team of data-centric developers is responsible for clearly describing the potential business value from the right use of data. While the leaders know data can be monetized in different ways, the data team's responsibility is to show the impact and expected value from the proposed solution that would get the executive sponsorship.

An example could be to demonstrate how a Machine learning model trained on good quality data will generate inferences with high confidence to understand better the customer preferences leading to customer retention and loyalty, which is directly related to the business revenue. Aptly designed data strategies also lead to operational efficiencies bringing in cost savings and improved processes.

Data science Teams:

When an organization is not mature enough to understand the data and its objectives, they rush into hiring a lot of data scientists expecting them to solve all their data-centric problems. This is a big red sign of not being clear with the expectations and offloading the responsibilities with no vision and definition of ownership. The first step is to understand data and infrastructure. Is building an ML solution the only resort, or can heuristics serve the business problem with a more glass-box view and control of the system? What are its limitations that require you to scale up the solution using a probabilistic ML model?

Are there a lot of variables affecting the process so much so the human experts cannot create sufficiently complex rules to capture the phenomenon and learn the data regularities? Know the answers to these questions before adopting an ML model route that can graduate to a black-box algorithm abstracting away the model internals and leaving little to no control to debug the model.

Besides being ML-aware, organizations also need to know what data would be needed, where that data will come from, and what part of it is logged? How would the incoming data be stored, what all customer touchpoints need to be captured, what would be the scale of data, and how would the ETL pipeline of the data look like? Who would be the consumers of the data and how would this data be analyzed? What checks should be in place to ensure that the data is reliable, accurate, and complete enough for the teams to ingest and act upon? How confident would the data consumers be to take the business-critical decisions on this data?

If I ask you the question - do you trust your data? What would your response be? Are you convinced with the data insights and model outcome to call the shots? The underlying theme is “Trustworthy AI Solutions can only be built on Trustworthy data”.

Data Quality:

By now, we have the data. But does that cut the mark? Is possessing data enough of a differentiator between competing organizations? Per research by Experian plc in 2017, most companies lose 15-20% of their revenue due to bad data. This points us to the importance of maintaining good quality data, data governance measures, and the significance of adopting a data-centric culture.

But first, let's understand what is culture? Culture is something that you adopt without being enforced and is an integral part of how your organization operates. If you are entrusted with making a business decision with no explicit instruction, would you trust your intuition coming from years of experience or knock on the door of data to get back with data-driven facts and insights? Your answer to this question surfaces the deep-rooted organizational culture primarily driven by leadership.

What could be potential data issues you need to be watchful of? The data comes from multiple streams and needs to be integrated to build a unified view to be fed as the training data into a machine learning algorithm. It involves a lot of complex business logic coming from the topic experts that need to be well-understood and implemented by the developers. A slight gap in understanding such logic to stitch the data together can lead to erroneous data preparation, which will not yield output any better than the garbage input supplied to the algorithm.

It is well a case of garbage in, garbage everywhere, i.e. the nonsense or flawed input will only generate nonsense output. Considering the already iterative nature of the machine learning algorithms, think of the disappointment that the data science teams go through when they need to iterate multiple experiments and ideas due to the bad quality of the input data. The key characteristics of data quality are:

- **Completeness:** It implies documenting the clear definition of what attributes are critical to declare the data as complete and ready to be used by the downstream consumers. Let's take a reallife example where you want to courier an item to location X, but have not listed the zip code of X. Would the courier agency be able to correctly deliver the courier in the absence of the zip code? If not, then zip code is an important attribute that marks the completeness of the data.
- **Accuracy:** Is the data in the correct shape and format? A large part of data accuracy is governed and driven by business owners. The mandate of how the expected data should flow directly comes from what business objective it should serve.
- **Timeliness:** In the digital world, especially the eCommerce data distribution changes dynamically. It is imperative to keep refreshing the data and make sure that the business actions are not taken on stale data that could include changing customer preferences and behavior.
- **Consistency:** Organizations need to maintain a single source of truth, i.e. there should be a data dictionary that defines the meaning and use of each attribute. If an ‘item’ attribute is present in table D1, and another ‘item’ attribute is present in table D2, then maybe these items are not similar. In the supply chain, an item could be a raw material in one location vs a processed item or a finished product in other locations.

Once a data quality report is published under the data quality characteristics, data consumers need to share whether the data should be continuously served to them with a warning and disclaimer regarding the state of the data, i.e., soft constraints. Or the data serving should be stopped till the new incoming data is vetted to be of good quality (hard constraints).

While evaluating such a decision, care needs to be taken for the rule of 10 aka 1:10:100. The rule, originally developed by George Labovitz and Yu Sang Chang in 1992, is related to the cost of quality, i.e. the cost of prevention is cheaper than the cost of correction which is cheaper than the cost of failure. The sooner you fix the data and put checks at the origin, the lesser is the cost you must pay for the bad quality data propagating through your systems. Once a bad quality has made its way to the database, it takes ten times the effort to clean and correct the data as it would have taken when the data originally would have arrived in a perfect state.

Getting ready for analytics

So, after being data-literate and having good quality data, you are now ready to clean and transform the data to weave the stories that the data has to share. The more time you spend in this data exploration phase, the closer and more connected you would be with the data.

Dissecting the data into different slices, checking correlations among various variables, identifying the missing values - whether it's a systemic miss or randomly missing, defining and detecting outliers in the data, and understanding the data distribution - are various ways to understand the data patterns. If you are working towards the ultimate end goal of building an ML model, then understanding the business objective and defining target variables is the starting point. Deciding on the evaluation metric to judge the goodness of your model comes next.

Is there already a process currently serving the business task, or perhaps a human expert? Understand what is the current baseline that the model should first try to achieve (and beat maybe). If your solution is one of its types with no prior work to gauge the baseline, then deriving the mean or average of the target value from empirical data is generally adhered to as the best bet.

By now, you have all the ingredients to build an ML model ready to get into production. If reaching this point appears to be a rigorous task, then you need to also devise a strategy of how to test the developed model. Would you go all out, perform an AB test or try with a smaller blast radius? What is the business impact if the model doesn’t perform as expected? How long should the burn-in phase be, i.e., the time model first gets launched and learns from the real data distribution?

You also need to consider the benefits of building the first baseline model faster and taking it to production to test the end-to-end pipeline rather than waiting for the best model. The error analysis from the baseline model gives a glimpse of how to improve the current model and passes a feedback loop to the model by:

- **Improving the class distribution**
- **Introducing a new feature**
- **Creating engineered features**
- **Improving the model architecture**
- **Augmenting more data for the misclassified or errored cases**

This leads to the birth of the new model, and this process continues to iterate till the model achieves a reasonable level of performance or meets the business KPI.

Looking after your deployed model:

While planning a blueprint of your first ML model, you need to remember that as soon as your model is in production, the goal post shifts to post-production care of the ML model. There are myriad ways the model will not behave as intended, including but not limited to model drift, algorithm drift and data drift. While model monitoring and maintenance is literature on its own, it is a key element in assessing your readiness to be an AI-centered company.

I hope the pointers shared above helped you fool-proof your strategy and prepare you in advance to ride onto the wave of building AI-pioneered solutions. More groundwork at the beginning of the project makes you more ML-aware and helps you devise the right solution to combat the multiple issues that can show up during different stages of an ML project lifecycle.

Author Biography

Vidhi Chugh is Data Transformist and AI Strategist. She is an award-winning AI/ML innovation leader with a vision to build trustworthy AI solutions. She carries over a decade of experience enabling data-driven solutions spanned across a wide spectrum of domains, including digital advertising, risk analytics, e-commerce, and supply chain. Besides, she is an AI Ethicist and has conducted several workshops demonstrating how to integrate ethical principles in an AI/ML project lifecycle. She is an international speaker and is on a mission to democratize machine learning and break the jargon for everyone to be a part of this transformation.

Chapter 4

Textual analytics and Textual disam-biguation

W H Inmon

Founder of Forest Rim Technology

Initially were simple systems that collected data, wrote data to files and created reports. Mostly, these systems operated on transaction-based data – bank deposits, sales, and telephone calls. An entire infrastructure grew up in support of these essential business systems. But in building these structured systems, there was little or no support for text. All data in these early systems was highly and tightly structured.

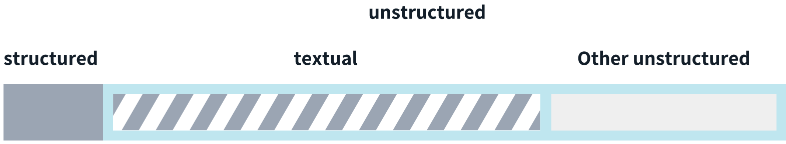
Text was ignored. Sometimes, text was collected. But little else was done with the text other than collecting the text.

However, consider ALL the data in the corporation. All the data in the corporation is represented by the bar -

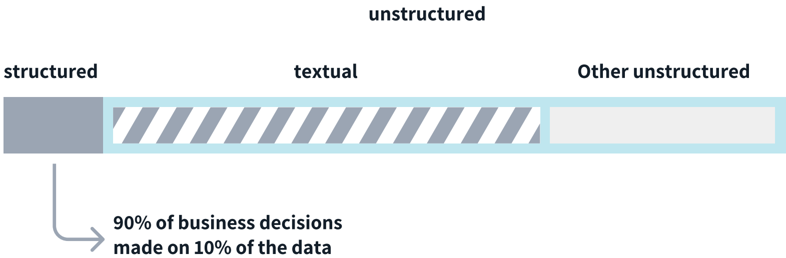
All the data in the corporation

Now of all of that data in the corporation, how much of that data is structured data? The answer is – not much. Depending on the corporation, from 5% to 15% is structured data. The actual percentages vary from one business to the next. The rough proportions of structured data to unstructured data is (symbolically) in the figure -

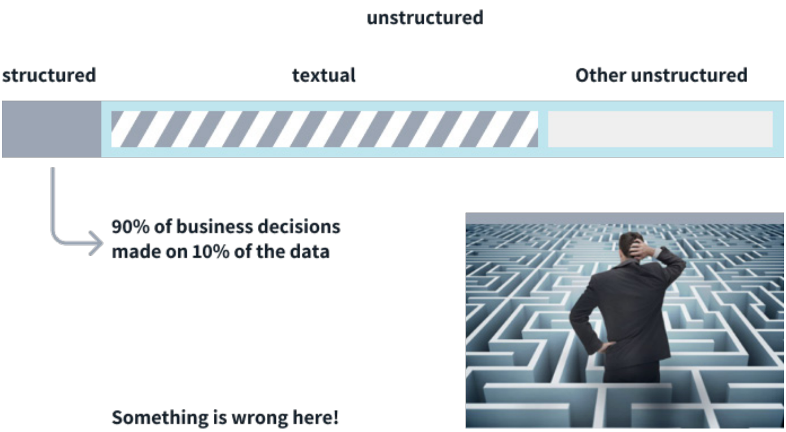
A closer examination of the unstructured data in the corporation shows that some of the unstructured data is textually based and some of the unstructured data is not textually based. The non-textual unstructured data is typically analog and/or IoT data. The following figure shows this delineation.



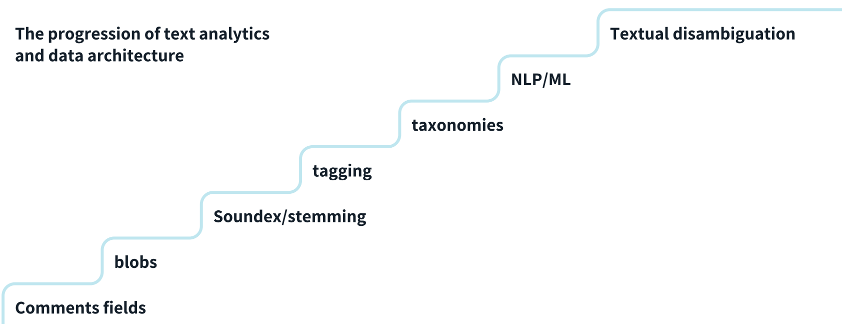
Most of the structured data is transaction-based data. Such business activities as bank deposits, payments made, sales made, telephone calls and so forth are captured in the structured component of corporate data. There is typically great business value in structured data. In most corporations, the vast majority of business decisions are made based on structured data. The preponderance of business decisions being made based on structured data is shown in the figure.



Most corporations exist happily and blissfully in the state in the figure. But forward-looking businesses see there is something askew with this figure. The thing that is wrong is that the corporation is making the majority of business decisions based on a minority of the data. This is like a golfer playing with only a driver and a putter. Or a racing skier using only the left foot. Or a race car driver using only one hand. It simply does not make business sense to NOT be using the majority of the data that exists in the corporation for making business decisions.



For many reasons, text has presented a challenge to the world of technology. The simplest (but hardly the only) reason text presents a challenge to technology is that text just doesn't fit well with a standard configuration of a database. Most dbms are optimized for handling structured data. There has been a progression in technology that has attempted to solve the challenges presented by text. That progression looks like -



Initially, programmers defined a field known as the comments field. And programmers filled the comments field with text. Some people put in long diatribes. Other people put in short sentences and a few words. Some people put nothing in the comments field. In early versions of databases, space had to be allocated for the largest comment anyone could make. And most comments were not as large as the largest comment. So, in the earliest renditions of text management, there was a lot of wasted space

Soon the database vendors created a “blob”. Blobs took care of the problem of variable amounts of space being required and/or wasted. But there was little else that could be done with blobs. Once text was entered into a blob, it was essentially useless data. There was little you could do with text placed in a blob.

Next, analysts tried something called Soundex. Soundex attempted to standardize text based on its sound. Then came stemming. With stemming, words were reduced to either their Latin or Greek word stems. The problem with Soundex and stemming is these solutions addressed only a small part of the issues of text analytics. There wasn’t much business value in stemming or Soundex.

Next came tagging. In tagging, words were selected, and their position in the document was noted. Tagging was the first real attempt to solve the basic problems of managing text inside a computer. But tagging had its own set of problems. The main problem with tagging is that the analyst had to know what words needed to be tagged before the tagging was done. Unfortunately, prescience and crystal balls were not a feature that many analysts had access to.

Then came taxonomies. Taxonomies were classifications of text. Taxonomies were a major step forward. With taxonomies, text could be externally categorized independently of the text. This was a major advance in the move to address the issues of text analytics. After taxonomies appeared, the next step was the movement to NLP – natural language processing and machine learning. NLP and ML took textual analytics to a new level.

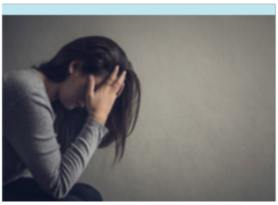
And then there came textual disambiguation. Textual disambiguation built on the work that preceded it. There were many differences between textual disambiguation and NLP and ML. But the primary difference between textual disambiguation and NLP was that NLP and ML focused on text, and textual disambiguation focused on text and context equally.

One way to describe the difference between textual disambiguation and NLP and ML is to understand the difference between answering the question “what” and answering the question “why”. Textual disambiguation answers the question of “why” whereas NLP answers the question of “what”.

Below is a really simply way to distinguish between "what" and "why".

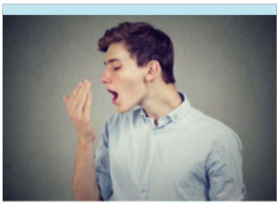
The difference between "what" and "why"

What



Your girlfriend is mad with you

Why



You have bad breath

NLP tells you that your girlfriend is upset with you. Textual disambiguation tells you that your girlfriend is upset with you AND she is upset because you have bad breath. There could be lots of reasons she might be upset with you. You forgot her birthday. You smiled at a pretty waitress. You left a small tip. You had too much to drink.

Now, knowing she is mad is valuable information. But knowing why she is mad is even more valuable information.

Text then presents challenges not found when handling classical structured text. The main challenge of text is that it is not enough to just handle the text. To handle text, you must handle BOTH text and context.

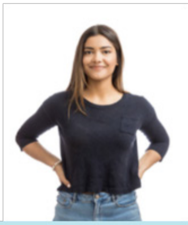
The challenge of text

Text – interesting

Text + context – valuable

To illustrate the value of context, suppose we drop in on a short conversation between two young men. One man says to the other as a young girl passes by, "She's hot."

Now, what is meant by "she's hot"?
What os being said here?



She's hot!

One interpretation is that the young lady is attractive, and the young man wants to have a date with her.

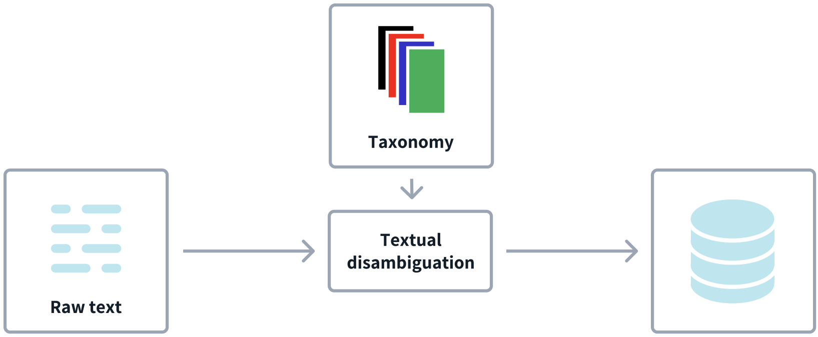
Or maybe the two young men are in Houston, Texas, on a July day, and it is 100 degrees Fahrenheit and 99% humidity. The young lady is pouring sweat. She is physically hot.

Or maybe the two young men are doctors and think the young lady has coronavirus and a temperature of 104 degrees. She is sick, and the doctor has taken her temperature. Her body is hot with fever.

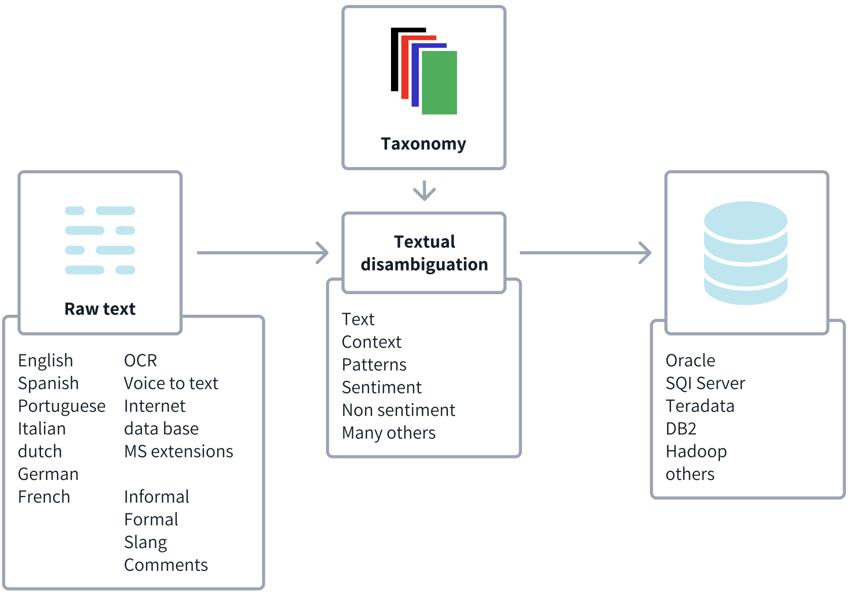
And there are probably lots of other interpretations of "She's hot".

So, the meaning conveyed by "she's hot" depends on the context in which the words are said. You cannot understand the meaning of "She's hot" without understanding the context. Trying to interpret words without interpreting context is a waste of time.

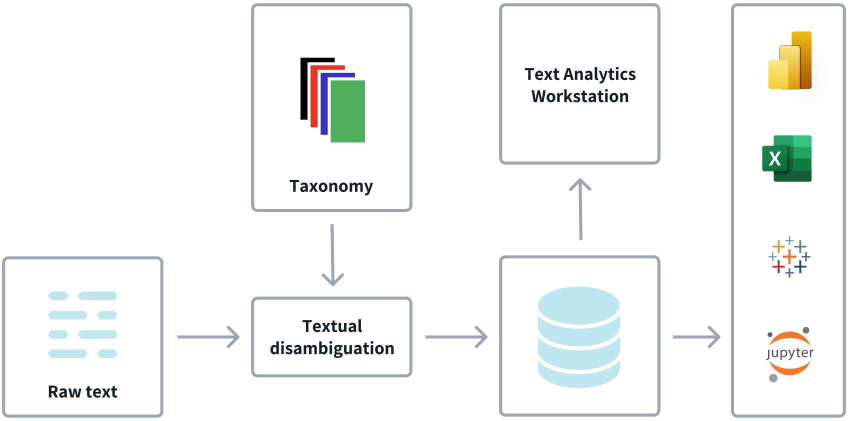
The overall architecture of textual disambiguation looks like -



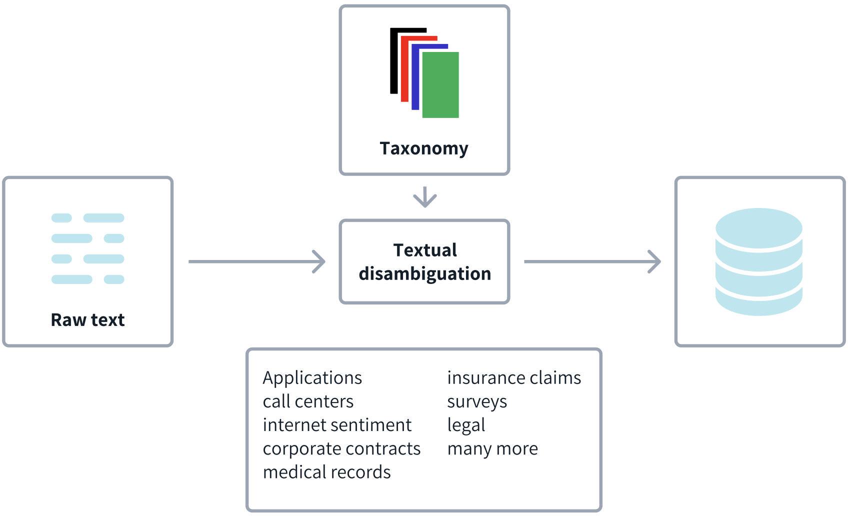
Raw text is ingested from a variety of sources. External taxonomies can be introduced. The raw text is read and analyzed, and a database is created. Once the standard database is created, it can then be analyzed by standard analytical tools. Some specifics of textual disambiguation are shown by -



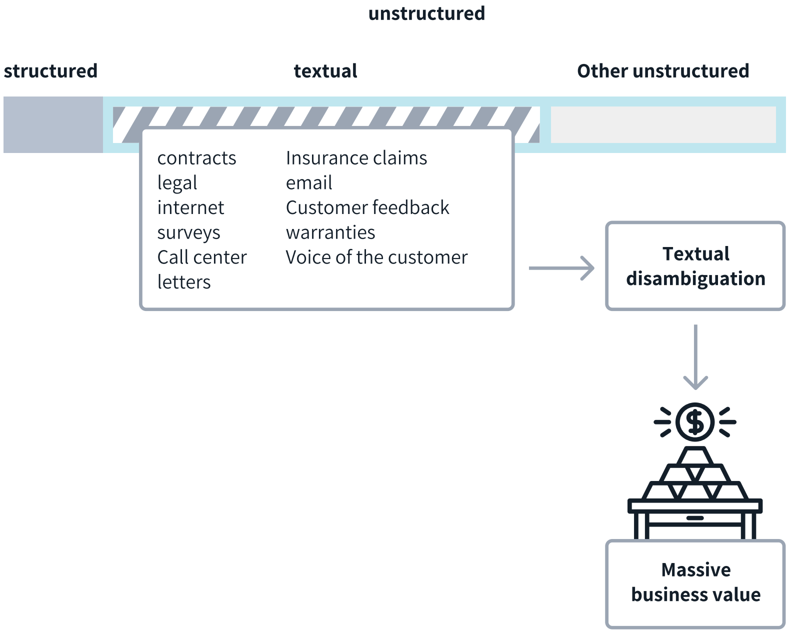
Once the standard database is created, the results can be analyzed. One way to analyze the results is in terms of standard analytical visualization software such as Tableau, Qlik, Excel, PowerBI, and other tools. Another way the output text can be analyzed is by Forest Rim's Text Analytics Workbench.



Now – for the first time – the massive amount of textual information in the corporation can be read and analyzed. The result is the unleashing of massive business value into the corporation.



The ability to not be constrained by volumes of data cannot be overstated. As long as you try to read and process text manually, you will always be limited as to how much analysis you can do. But once you enlist the aid of the computer and the powers of automation, you can handle huge amounts of text. Handling text in an automated manner brings down the cost of processing, enhances the speed of processing, and enhances the accuracy of processing. Those reasons greatly favor the automated approach to handling text.



Author Biography

Bill Inmon is recognized as the "father of data warehousing." He is co-creator of the Corporate Information Factory and, more recently, creator of the Government Information Factory. He has over 35 years of experience in database technology management and data warehouse design. Bill's mission is to educate professionals and decision-makers about data warehousing and the Corporate and Government Information Factories. Because of this mission, Bill has authored close to 700 technical papers about building, using and maintaining enterprise-level data warehouses and operational data stores and how they integrate into a total Corporate or Government Information Factory. Besides the technical papers, Bill has authored at least 46 books translated into nine languages. Bill is one of the most sought-after speakers for every major computing association, industry conference, seminar, and tradeshow around the world and doing private speaking and seminar engagements with businesses throughout all industries.

ATSCALE

Chapter 5

Hallmarks of a True Data-driven Organization Culture

Ram Kumar

Chief Data and Analytics Officer,
CIGNA International Markets

The term “Data-driven Organization” is overused in industry nowadays, and it means different things to different people, organizations and industry. What does it mean to be truly a datadriven organization? Does it have anything to do with the infrastructure, tools and technologies? Or does it mean that data has to play an important role in the functioning and decision-making within the enterprise? What about managing data effectively and efficiently to enable value creation, and is it also a critical component of the data culture? Does democratizing data help in data-driven value creation? Questions like these and (obviously) many more are always a part of boardroom discussions — be it in small start-ups or buttoned-up Multi-National Companies.

Most companies in this age are striving to become data driven. But as we saw from the myriad of questions above, it is impossible to have one definition. There is a general misconception in industry that if business value is created from data using data analytics/data science or intelligent data-driven solutions such as AI and automation, then an organization is data-driven or has a data-driven culture. Data-driven culture is much more than that.

Recent surveys (e.g., Harvard Business Review; New Vantage Partners) have shown that achieving data-driven leadership to drive data culture remains an aspiration for most organizations. The surveys show that organizations continue to struggle to become data-driven, with only 26.5% reporting having achieved this goal and only 19% reporting having established a data culture. Cultural impediments remain the greatest barrier to organizations becoming data-driven, for the 4th consecutive year, 92% of executives point to culture as the greatest impediment to achieving this. The percentage of firms identifying themselves as data-driven is declining for the past 4 years, and the principal challenge of becoming data-driven regarding culture is increasing for the past 4 years. While 80% of the CEOs claim to have operationalized data as a strategic asset, only 10% say that their company actually treats it that way

What is True Data-driven Organization Culture?

Performing analytics/advanced analytics and building data-driven intelligent solutions (e.g., AI solutions) to drive business outcomes using data does not necessarily mean an organization is “data-driven” or has a “data-driven culture”. An organization has a true data-driven culture if it can understand, manage and govern the “lifecycle” of its data assets effectively and efficiently that would enable the organization to organize and democratize its data assets for consumption and drive data-related activities in an acceptable manner to make informed decisions, create value, resolve conflicts and manage risks.

Figure 1 below summarizes the “Lifecycle of Data” supported by Data Governance and Management functions. “Use” component of the data lifecycle is where data science and AI focus is. Close to 80% of the time is spent by data analytics professionals preparing the data for value creation. The reason is because of poor data management practices designed and implemented and followed by people, processes and technology in organizations. Data lifecycle governance and management are, therefore, foundational and fundamental to an organization and should be in the organization’s data strategy. To be data-driven, it is more than the “Use” component of the lifecycle, as shown in the figure. The components of data management, e.g. data quality, metadata, data retention and destruction, acceptable use of data – privacy, ethical and legal, data security, data architecture, data standards, etc., remain the same despite the size of data that an organization collects, manages and uses it.

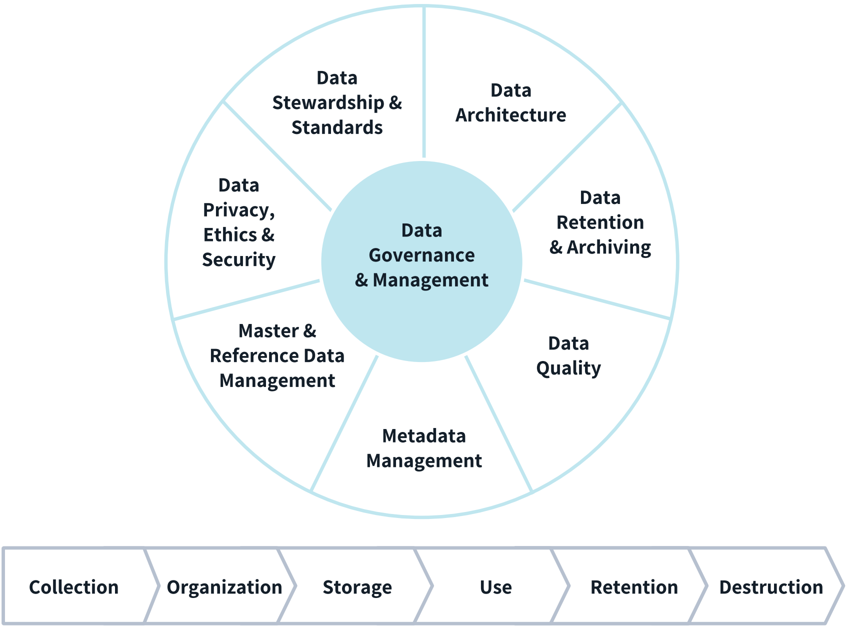


Figure 1. Data Lifecycle Management & Governance

Whether an organization's goal is to achieve digital transformation, "compete on analytics," become "AI-first," or become "Digital first", embracing and successfully managing the lifecycle of data in all its forms is an essential prerequisite to being data driven. Critical obstacles regarding managing data still must be overcome before organizations see meaningful benefits from their big data, analytics, digital and AI investments.

Hallmarks of Organizations that successfully created a true data-driven culture

A true data-driven organization (TDDO) does most or all of these strategies to drive the right culture by being proactive, and that is forward-looking by sincerely and religiously treating data as a "firstclass citizen" or "crown jewel" of their organization.

Accountability from the Top: "Leadership needs to think of ways to reward those who shared data, incentivizing individuals and departments that develop and nurture open, accurate and sharable data and analytics. —Jennifer Cobb 2013”

TDDOs rewards its employees, partners and customers for helping drive a strong data culture that creates a “WIN-WIN” position for all. Therefore, to treat data as the lifeblood of the organization and as a strategic and competitive asset, TDDOs enterprise data strategy to drive data-driven culture and value creation push accountability from the top, namely, the Board, CEO and his/her leadership team by leading from the front and as an example and this is a critical differentiator. The bottomup approach works to an extent and will be short-lived. An organization that thrives on data has its C-suite championing its usage and reliance in the organization, not sponsoring it. In one TDDO, it went one step ahead where the CEO was the “Data Champion” and x% of CEO and his/her leadership team’s annual bonus was tied to data culture-related Key Performance Indicator (KPI) as part of their balance scorecard. Also, all employees who use data had the responsibility to manage the data and had data culture KPIs in their balance scorecard. Figure 2 highlights the set-up. This is one of industry’s best-case studies in driving a true data culture.

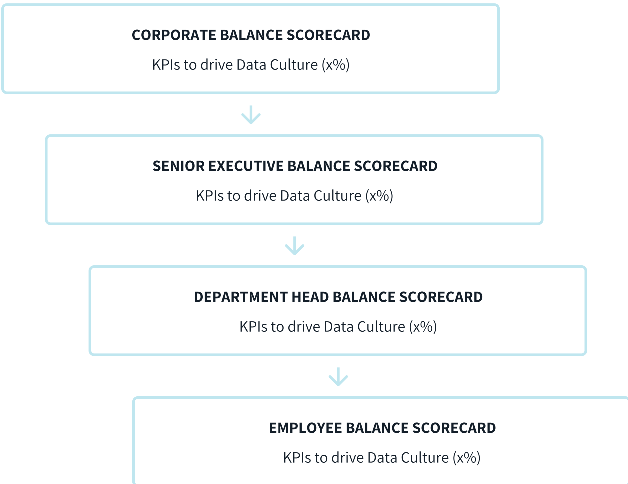


Figure 2. Data-related KPIs in Balance Scorecard of Employees

“Data First” Mindset and Culture:More than installing the right tools and applications, and having the right people for an organization to be data-driven, data should be at the “centre of everything” the organization does; namely, it needs to be integrated into its strategies, innovation, operating model, systems, processes and culture. It’s about creating a mindset in which data-driven value creation supports all fact-based business decisions and is embraced by all levels of the organization.

TDDOs operate with the above "Data-Centric" mindset by ensuring all technology, business processes or products related initiatives are built/transformed by placing data at the "centre". This means the initiatives will have a clear data strategy supported by a solid data architecture. Having a "Data First" or "Data by Design" mindset and culture is the hardest as traditional organizations, unlike TDDOs, focus on people, process and technology by taking data for granted and not realizing that data brings people, process and technology together, and therefore should be given top attention. Today's technology is tomorrow's legacy; people come and go, process changes are continuous, but data keeps growing and is not thrown out. Figure 3 summarizes that data is the "lifeblood" of an organization by fueling people, process and technology

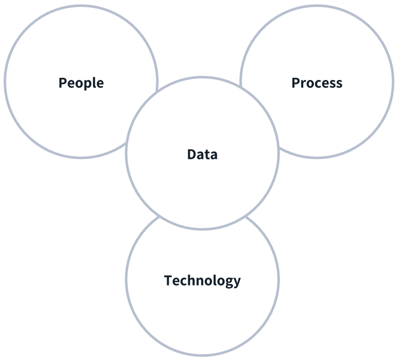


Figure 3. "Data First" Culture

Data and Analytics is a Business Function:Data and analytics function in TDDOs sit under the business as data is a strategic asset of the business. The function reports to a Chief Business Executive. This enables data strategy to be implemented holistically and integrated well as part of business strategy to enable or create value for the business through execution of the business strategy and drive data culture effectively and efficiently. Data management and governance is more than IT operations and business is accountable/owner of data, and IT is the custodian of data and TDDOs understand it well.

Data Strategy that is Executable: TDDOs understand that creating business value out of data requires people, process, technology, data and the right culture must work together in an integrated manner in figure 4 below. They also design and develop their data strategy based on the data capability maturity of their organization and develop a roadmap to build data-driven capabilities by addressing gaps. TDDOs design data strategy by bringing data teams, business teams, privacy legal, risk and compliance teams and technology teams together. The data strategy is holistic for their organization but pragmatic and implementable rather than having a lofty vision. The data strategy links data foundations and data-driven value creation supporting business priorities with clearly defined roadmaps and milestones executed in an incremental manner rather than as a big bang. The roadmaps and milestones are reviewed frequently to ensure alignment with value creation and business priorities.

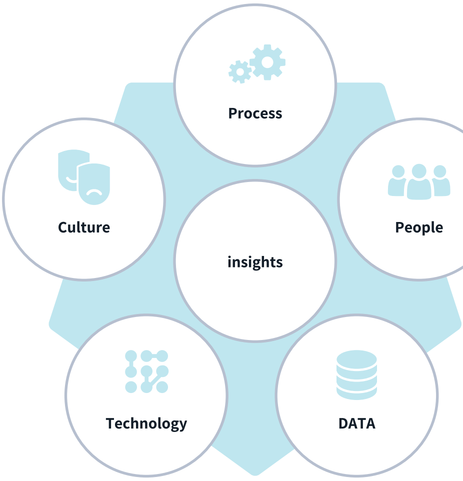


Figure 4. Integrated approach to driving data strategy design and execution

Prioritization of Value Creating D&A Use Cases: TDDOs top priority is to execute D&A projects that create value by operationalizing the outcomes of the projects so the value can be measured. This requires discipline and collaboration between D&A teams, business teams and stakeholders, and the technology team in picking the right use cases to execute. TDDOs have a solid business data and analytics use case prioritization framework to prioritize the use cases for execution based on several key metrics such as value creation in terms of quick wins, size of the prize, alignment to business priorities, difficulty in execution, commitment to the operationalization of use case outcomes, data availability and quality, etc. Business should avoid the temptation of falling into “bright & shiny” or “hype” when implementing D&A solutions, which might provide new tools and new ideas, and there is nothing wrong in seeking them. It is, however, a problem if the solution alienates the user – the business teams – which ultimately affects the adoption of such solutions. Gartner (2021) provides a simple framework for determining use cases based on value creation and implementation complexity as shown in figure 5.

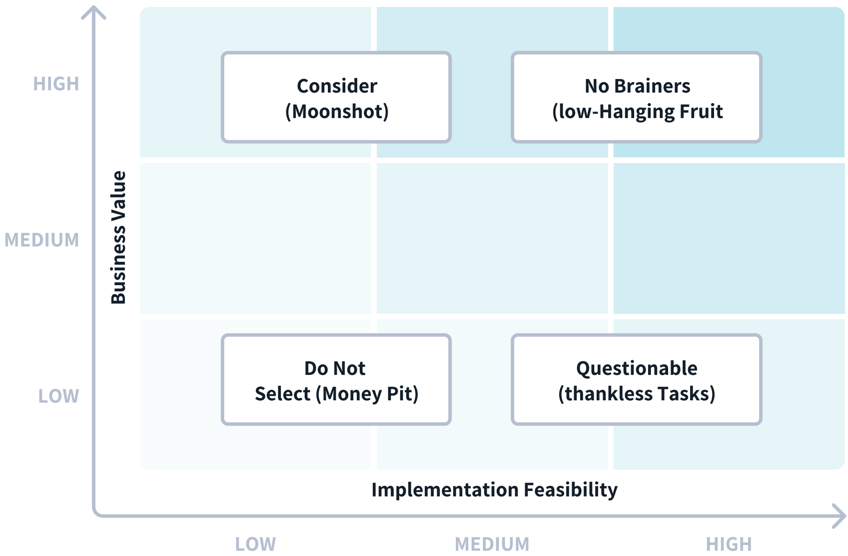


Figure 5. Prioritization of Use Cases

Measuring Data Monetization:"Someday, on the corporate financial balance sheet, there will be an entry that reads, "Information"; for in most cases, the information is more valuable than the software and hardware that processes it." - Admiral Grace Murray Hopper (Inventor of COBOL), 1965. TDDOs focus on implementing KPI to measure data monetization (internal and external data-driven value creation) and work to have the value tracked as part of corporate financial balance sheets and report it. Douglas Laney's "Infonomics" book (Reference: Gartner, 2018) is a great read about this subject.

Data Champions to Drive Data-driven Culture: TDDOs identify and appoint executives to be accountable for data and drive data cultural transformation, evangelize the use of data and drive data-driven initiatives across the organization to support business objectives and priorities. These champions identify people in their departments passionate about data and make them Data Stewards and Data Champions to embed data practices on a day today basis by working closely with various teams across various departments/functions. Data champions and stewards work closely with Data Custodians, who are the IT system owners or the gate keepers of data and manage the data held by the IT systems on behalf of the business.

Democratize Data for Easy Access, Sharing and Use: "Get the right data to the right place at the right time and with the right acceptable use" – Ram Kumar, 1996. Right data includes quality and context

TDDOs are much more open and transparent and continuously invest in the data foundational and management activities as a strategic program that would enable the democratization of data that matters only and not all for easy access and use by users for value creation. This is done by breaking the data silo mindset and increasing data collaboration and interoperability. If more employees have access to the data that they need, the skills to analyze and interpret it, and there is sufficient trust, then more decision-making can be democratized. This involves a huge element of trust. The organization must trust that the data will not be abused, leaked to competitors, or used to fuel political battles but instead will be used in an appropriate and acceptable manner to further the business.

Data Risk Management: Managing business risks associated with data in an organization is critical in managing an organization's reputation. TDDOs are serious about it and embed data-related risk as a key KPI in the Organizations' Risk Profile that gets discussed at the Board. Data-related risks that get measured include acceptable use of data, namely, ethics and privacy, legal and regulatory compliance, data quality, data security, data access, and data sharing. This framework is supported by a strong data ethics and data use framework tied to the organization's core values. All datarelated initiatives get tested against the framework by applying social, privacy and legal lens before commencing an initiative.

Smart Data Governance to Enable Data for Value Creation: TDDOs see data governance as an enabler of data assets for value creation by managing data-related business risks effectively and efficiently and not just from legal and regulatory compliance perspective, which organizations that are not data-driven do. Traditional approaches to data governance, generally huge initiatives and are a big bang, do not work nowadays as organizations do not have the patience, time or budget for it as they are keen on quick ROI of data-related investments. TDDOs, therefore, come up with smarter ways of doing data governance by focusing on what matters, and that would enable value creation for the business – an incremental but focused approach.

Minimum Data Standards Framework: TDDOs have clearly defined minimum sets of data principles, policies, standards, processes and procedures supporting practical data governance frameworksthat are implementable, can be developed and executed as part of data strategy and ensure that they are implemented across the organization in an incremental manner from a value creation perspective. Any initiatives, including technology and business processes must comply with these “minimum standards”. The implementations are regularly audited for compliance.

Data Quality by Design Culture: If an organization has timely, relevant, and trustworthy data, decision-makers have no alternative other than to decide by gut. Data quality is fundamental. No matter how smart and efficient the organization’s business processes are, how advanced, savvy and solid the IT systems that support the processes are, how capable and skillful the employees who use the processes and technology are, and how sophisticated the data analytical models are to produce data-driven insights, and how intelligent AI solutions are, if the underlying data that these processes, technology, analytical models, people and AI use is not good enough in terms of its quality, the expected business outcomes driving benefits such as data-driven insights, digital experience, intelligence and effective and efficient business decision-making will be poor. No silver bullet solves data quality problems. It requires a collective effort of all – people at all levels, processes, and technology.

TDDOs are good in driving data quality by design culture, which means data quality is conceptualized at the planning stage of a project/initiative/program, e.g., business process, technology, business products, and not as an afterthought. Data Quality is embedded as a KPI in Organization’s risk profile and is measured and monitored across the organization as Business As Usual (BAU). Employees have KPIs defined in their balance scorecard and employees and functions/departments are rewarded for capturing and managing quality data, and third parties and partners are rewarded for providing quality data.

Data Literacy Embedded as part of Learning and Development: TDDOs focus on educating/training all employees at all levels of the organization about data, its lifecycle and supporting data-related components e.g., common definitions, data quality, data lineage, data privacy and security, and data-driven business value creation as part of its learning and development program similar to IT security training, code of conduct training etc. They do not focus on data literacy or storytelling only around analytics/AI only, but on the whole lifecycle of data, its management, use and value creation and its governance. This helps organizations respect the strategic asset and drive the right data culture.

Data Analytics Playground to Explore Data: Data science and AI can have a transformative impact on businesses. To achieve that impact, TDDOs harness the power of new (and old) technologies to provide data scientists with an ever-evolving, fit-for-purpose data science and AI workbench. Or, generally called in the industry a “data analytics playground.” This playground is essentially a workbench of the right tools and systems for data preparation and exploration that allows data scientists to explore data to identify the “art of possibilities” with data that would create value for the business and to concentrate their time and effort on the math behind the business issue and drive tangible value. Cloud, data mesh, data fabric, etc., are technology components to support the playground

Pilots/Experiments with Data: TDDOs embrace data in a full-fledged way and start by launching small-scale pilot projects. Inculcating data in a big way needs time, but to understand how it can be done needs smaller projects and experimentation. Test, fail and learn fast — that's the mantra.

Manage HiPPOs: Successful implementation of data analytics in an organization also means driving the culture of decision-making supported by data-driven decisions. “HiPPOs” or “Highly-Paid Person/s' Opinion” can mess up that balance if data-driven decision-making is not supported by them and they just use intuition based on their experience for decision-making. TDDOs ensure that HiPPOs are given the opportunity to transform themselves and are educated to use analytical insights as a supporting tool to inform and influence the influencers and in decision-making. If the HiPPOs do not adopt this approach, their presence will be detrimental to the organization's ambition to create a data-driven culture and TDDOs do not hesitate to replace them.

Evidence Based Culture: TDDOs trust their data and use data-driven insights as critical evidence to help inform and influence its business strategy.

Acceptable Use of Data: TDDOs are mindful of the acceptable use of data when focusing on monetizing data. They are mindful of their social, legal and ethical obligations when using customer data. They treat having access to customer data as a privilege given to them by their customers by placing trust in them that they will not misuse it for their financial benefits. TDDOs find the right balance between data monetization and acceptable use of data which includes ethics, and apply a well-defined and implementable data ethics framework to determine whether to proceed with analytics/AI use case or not. The data ethics framework is linked to the organization's core values. Figure 5 (Reference: Tom Wilson, 2011) is a great reference for the acceptable use of data.

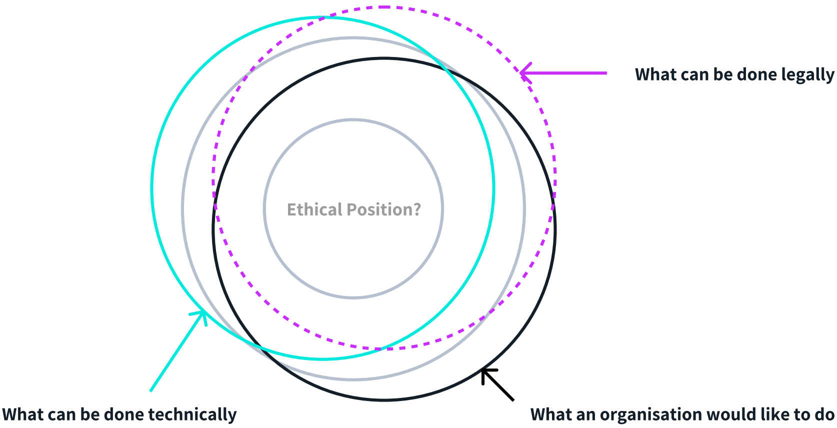


Figure 5. Balancing Data Monetization & Data Ethics

Continuously Testing and Improvement: TDDOs do continuous testing to improve its data-driven solution offerings using a strong feedback loop process in place, both internally and with external customers and partners. Tests may also include user testing—working directly with actual customers or users to obtain direct feedback on possible new features or products. The organizations also have a continuous improvement mindset. It may be involved in repeated optimization of core processes, such as shaving minutes off manufacturing times or decreasing the cost per acquisition. This comes about through careful analysis, crafting mathematical or statistical models, and simulation.

Continuous Model Management and Improvement and Automation: : Industrializing analytical model management and continuous improvement of it is key to getting new intelligence into the business stakeholders' hands, and TDDOs are good at it. Where repeatable processes are deployed, it is critical to automate them and data science can contribute to it too. Keep analytical models fresh and develop design-led applications to make them as relevant and accessible to the broader business as possible. Automated continuous improvement will free up data scientists' time to focus on key business issues.

Author Biography

Ram Kumar is the Chief Data and Analytics Officer of Cigna's International Markets. He drives data and analytics strategy and its execution for 30+ countries covering the Americas, EMEA, and the Asia Pacific. He has held many global executive roles in his 33+ year's career, including as CEO, Group CTO, CIO, and CDAO. Ram has served as a member of the Data Research Advisory Board of MIT Sloan School, published over 150 articles, written chapters for books and is a regularly invited keynote speaker in conferences globally and has spoken extensively. Ram is CDO Ambassador for Singapore, a joint global initiative of MIT CDOIQ program, Institute of CDOs and CDO Magazine. He is a global editorial board member of CDO Magazine. Ram holds a master's degree in Computer Science and Engineering and a bachelor's degree in Electronics and Communications Engineering with AI as a major in both.

Chapter 6

Data Culture and Data Ethics in Business

Meltem Ballan, PhD

Principal and Founding Partner,
Concrete Engine

Data privacy and ethics have been one of the growing topics for organizations to constantly change and improve their data programs to protect their customer's data. The first step to consider for companies is to collect customer consent – The customer must consent to receive the data sharing services. There are increasing standards around the world that govern the data – and organizations must comply with those, as well. Organizations should build a culture and communicate with their teams that everything starts with consent and they must earn and maintain the trust of consumers by protecting their data. Therefore, companies need to formalize data governance and culture programs to hold themselves accountable and responsible.

Importance of the Data Governance for Businesses

To formalize data governance and culture programs stem from the ever-changing forces in businesses and industries. Technology and data is the engine to make or break a business. Having a solid data policy and implementing this policy within the organization presents a competitive advantage for businesses. Day over day, more consumers demand data privacy, and they will seek providers that offer full transparency about their data collection and processing. Data governance and policy should be considered at a CEO level as it is directly affected by overall industry and organizational culture, and it might cause severe damage to the organization's reputation and revenue channels. To accommodate and manage the forces and keep the power engine intact, organizations need a formal program to ensure standards. These standards should be up to date and evaluated regularly.

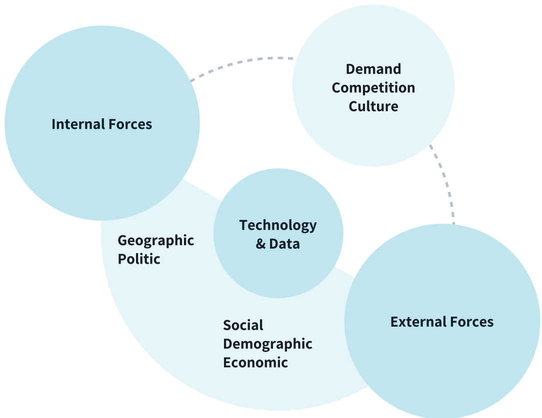


Figure 1. Data Governance Framework

Building Blocks of Corporate Data Governance

To build a case for a holistic data program, the organizations need to make a fine distinction between governance and ethics. Data governance is managing the availability, usability, integrity, consistency, acceptability and security of the data in enterprise systems based on internal and external standards, policies and regulations. It is usually enforced by government, federal and state levels to the senior leadership down to the related groups within an organization.

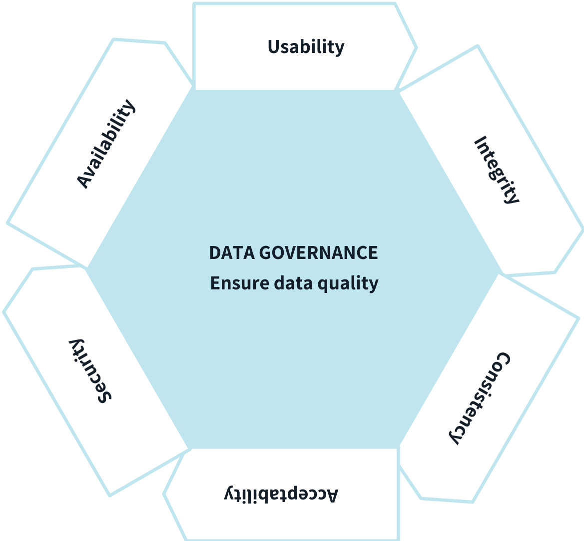


Figure 2: Data Governance for Data Quality

The six pillars of data governance policy are defined to ensure that all business functions have access to the data assets and data is protected:

1. **Availability**: covers the policies to ensure that data is available and easy to use by business stakeholders. Most common metrics to measure availability are frequency, time sensitivity and backup capacity.
2. **Usability**: covers the policies and protocols around data structure, documentation and labels to allow different business functions to use the data and monetize the data both internally and externally.
3. **Integrity**: assures that the data is maintained consistently over its entire lifecycle. These policies also cover the retention of the same standards of data within and across different platforms.
4. **Consistency**: ensures that data is correct, consistent and uniform across business functions. The protocol covers the data quality in format transformations, duplicated data, and missing information. It is usually measured by a unit comparison of two datasets.
5. **Acceptability**:ensures that data is consistent and all the attributes are intact. Metrics are usually created by comparing historical data and compatible industry data.

6. **Security**: ensures that data is safeguarded over the life cycle of the data. The security policies include hardware standards, encryption and transfer protocols between databases and user permissions.

Overall, data governance policies are designed to protect the business data and the privacy of the consumer's data. Data privacy covers compliance with regulations. These regulations can be defined at local, federal and international level. The primary focus is to protect sensitive customer data ensuring the data collection and sharing across platforms (internal and external) don't risk customers' privacy. There are several data protection laws introduced at local, federal and international levels and the new ones have been considered.

TheEuropeanUnion'sGeneralDataProtectionRegulation(GDPR): the GDPR is to protect individuals and their personal data by ensuring that businesses collect and safeguard their customers' data in a responsible manner. GDPR's core concepts are to ensure that (1) Minimum personal data is collected and stored; (2) data is protected against unauthorized and accidental use and loss; (3) Organizations are accountable for usage and storage of the data and informing their customers. (https://gdpr-info.eu/)

California Consumer Privacy Act (CCPA): covers the similar rules; but also gives consumers the right to know how their data is used and delete their data if they desire so. (https://oag.ca.gov/privacy/ccpa)

Most data and ethics programs start with measuring and ensuring the data governance and regulations. Companies use metrics to measure the maturity and requirements to meet the minimum criteria for each step of the curve: (1) In its earlier days, the governance team worked across teams to identify technology gaps, regulations and rules in place within business units and cross-functional teams and training needs. (2) After collecting the current stage governance team acts to implement the regulations and ensure that policies are in place. This team is also responsible to communicate with senior leaders to identify budget and timelines for tooling and cross-functional training. (3) Once the team complies and implements the standard policies, puts in place training and certification and modernizes the technology, the maturity curve and data program are compiled. However, the data programs should also cover the ethics. Ethics is a big part of Governance and it comes into play when technology professionals hit the ground. The red arrow shows where most companies consider data ethics to extend governance and program rules within the company. The main difference between data governance and ethics is that data governance is imposed by regulators and ignited by senior leadership (SLT) whereas ethics are implemented and imposed by practitioners. The SLT seeds the rules and practitioners implement, develop technical road maps, execute and ensure the success of the entire data and culture program.

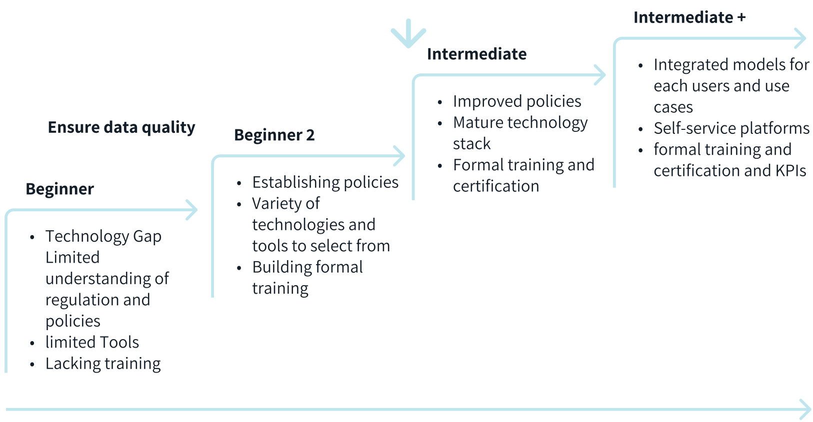


Figure 3: Data Governance RoadMap

Completing Corporate Data Governance with Ethics

The data governance (Level 1) starts with regulators and STL; however, to build robust and continuous data ethics programs the role of mid-level leadership (level 2) becomes critical. Level 2 plays a mediator and translator role for the success of ethical data programs as most ground rules and implementations come from Level 3.

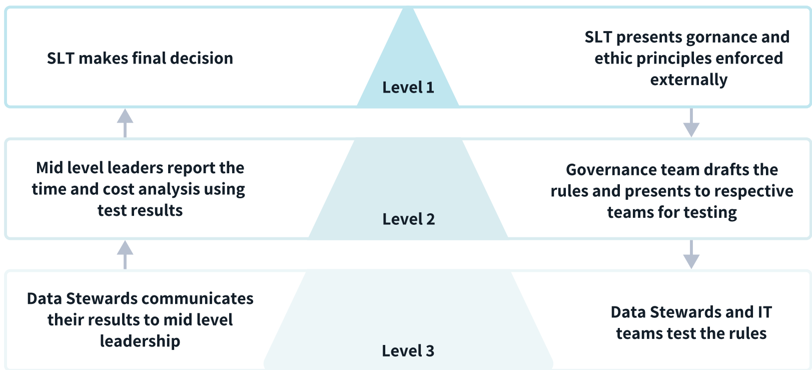


Figure 4: Organization Design for Data Governance

Core Team, Vision Alignment and Value of Data

The core team is the teams who deal with day-to-day data operations- consisting of not only the technical units; but also the business and customer-facing units. The core team should understand that they are accountable and responsible for the successful implementation.

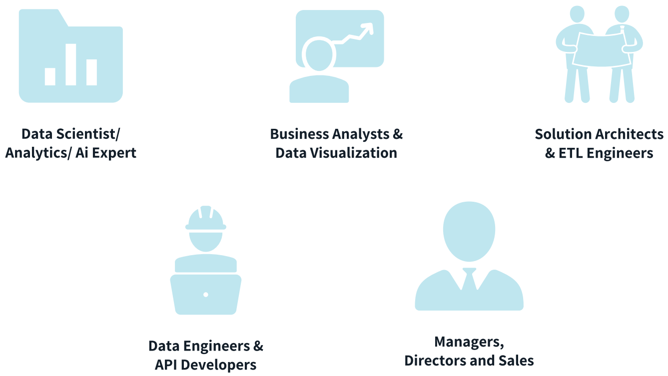


Figure 5: Data Governance Core Team

For core teams to work integrated, the organization should have a clear and concise vision and value statement that everybody shares uniformly. The vision should be set such that at a business level, the value creation and value capturing with data should be clear to everybody. The value should be created at three levels to ensure the scalability, scope and network effect of the data and its value.

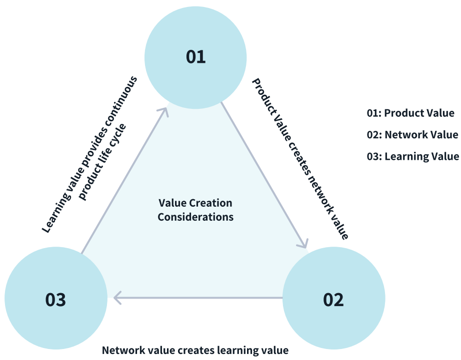


Figure 6: Data Governance Value Creation

The organizations should ask the following questions while aligning the vision and beliefs across units:

- **Product value:** What is the inherent value in the data? Is the value a good one? What kind of innovation underlies the product/service concept?
- **Network value:** How would the data value change as more employees and customers join the platform? Can the data services be offered across business functions?
- **Learning value:** How would the business value change as more data is collected/connected to it? What role does data play in creating value for the product? How might algorithms help improve the product/service being offered?

Data Ownership and Risk Mitigation

Once the value and core teams are informed and vision aligned, it becomes critical to have data owners mitigate the risk. The teams should identify internal and external data responsibilities and ensure that the organization minimizes the data and implements rules and protocols to add/remove customer data. At this stage, the business should honor peer reviews on analysis and data continuity within each team and across teams. It is on each employee to train/mentor each member of the team and comply with the ethical rules. A bonus benefit of peer reviews is that team members can learn from each other. Peer reviews are the best way of training data professionals.

Building Human-centric Data Ethics

The combination in the industry is shifting towards data and technologies used. Corporations should include data policies and data practices as part of their competitive advantage and stare internally and externally with their customers.

Building comradery and training on companies' overall vision and data are critical for companies to compete in the data and analytics era. The field is rapidly growing. Even though most professionals come with some transferable skills in data and presentation, job-related skills are essential to complete the product life cycle. The training framework should include data literacy, data queries and locations, business intelligence, data science, risk mitigation and internal and external communication. So the company can have a uniform identity and transparency about their business and data culture.

Data Culture and Ethics Committee

Data culture and ethics community can be defined as the sounding board of the company's data and training. This community should identify the education materials, and communicate the requirements and budget with the teams and STL. This committee should also input and create data policies and governance rules independent of business functions and products, ensuring those align with the company's values and vision.

Continuous Success of Data Culture and Ethics

One of the important elements of continuous success is continuous growth without changing the core value of the company and its data policies.

There are five levels of continuous success for data practices. The foundation starts with healthy data and technology stacks that can support federated and transparent data and technology sharing. These foundation layers should be compatible with scaling additional, improved data governance and ethics rules. This then allows new products and revenue channels and equal opportunity for teams to innovate and grow.

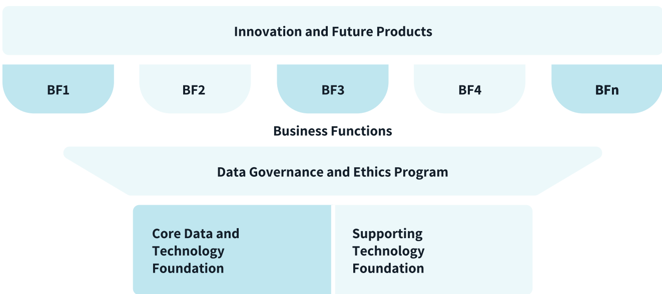


Figure 7: Data Culture and Ethics in Data Governance

References and Future Readings:

1. Brian McCarthy, Chris McShea, and Marcus Roth, Rebooting analytics leadership: Time to move beyond the math,McKinsey Analytics:
2. https://gdpr-info.eu/
3. https://oag.ca.gov/privacy/ccpa
4. Ashutosh Gupta, 7 Key Foundations for Modern Data and Analytics Governance, Gartner IT insights Articles
5. Janiszewska-Kiewra, Jannik Podlesny, and Henning Soller, Ethical data usage in an era of digital technology and regulation,McKinsey Analytics
6. Elizabeth Pike, Defending Data: Toward Ethical Protections and Comprehensive Data Governance, Emory L.J.

Author Biography

Dr. Meltem Ballan uniquely combines analytical, and business expertise developed over 20 years, both in industry and academia. She is a recognized leader and an advisory board member for high technology startups and M&A holdings. She helps technology startups to transform their POCs into scalable and commercialized MVPs. During her career, she has designed complex machine learning models and implemented AI projects, including natural language processing (NLP), linear and logistic regression, supervised and unsupervised learning, neural network, deep learning algorithms, and hybrid approaches to computer vision.

Chapter 7

Business Analytics: A Blueprint for Unlocking Digital Value

Victor Ojeleye

Business Group FP&A Planning & Reporting Manager, Cargill Protein North America

The future of business will demand companies invest in digital technologies that drive value for their customers as a stepping stone to transformative growth. A key area for this investment will be business analytics. In a world where uncertainty prevails in almost every global market, the need for data and insights to help businesses win today and in the future is as high as ever. According to a recent Gartner survey, “eighty-two percent of chief financial officers (CFOs) report that their investments in digital are accelerating, exceeding investments in other areas such as talent, supply chain, business services or fixed assets. But the challenge continues to be how to turn investments for both the enterprise and finance function into digital wins for the organization.”

For many companies, digital investments have been underway for years, and success depends highly on factors such as readiness of the organization, capabilities to deploy, available capital for incremental scope, and willingness to adopt new technologies and ways of working.

This chapter will explore how companies can unlock digital value by designing intentional, insightsfirst analytics and establishing data platforms that serve as strongholds to enable highperforming analytics. The blueprint must be tailored for the user, and it must create value through insights that yield a competitive edge for decision-making in the current and future business environment.

To build a good blueprint, the analytics and data problems must be defined. The challenge in unlocking powerful analytics is that data must be transformed into powerful combinations of information that can be interpreted as trends or patterns businesses want to repeat, take advantage of, or course correct toward desired performance outcomes. Business leaders may also have different perspectives or aspirations for analytics capabilities and how they serve the business. If some define analytics as predictive analytics or artificial intelligence (AI), then the systems and capabilities to build and utilize AI differ greatly from those required for general business or statistical analysis.

Similarly, the data problem companies face is multi-faceted. Data often lives in multiple systems in different formats and may not be easy to consolidate. Data can also be a challenge due to a high volume of records to manage on a database, inaccuracy, or complex warehousing structures. Companies also often consolidate all their data into an ERP system, and then they build webbased or tabular reports informational but fall short of unlocking insights that drive real value. Begin with the end in mind – insights first. Each of these data management hurdles, if not addressed proactively, can lead to poor data product performance and ultimately limit any return on digital investments (RODI). So, what solves unlocking business value from digital investments with analytics?

Understand the users – End users want the analysis to be right, so they need not debate the number. They need it to be on time or sooner and ready at the frequency by which they make decisions. For example, shift operations, weekly sales, monthly inventory, or annual earnings. They also want it to tell them the simple things and the things they did not know. Steve Jobs said, “Some people say, "Give the customers what they want." But that's not my approach. Our job is to figure out what they will want before they do. Henry Ford once said, "If I'd asked customers what they wanted, they would have told me, 'A faster horse!'" People don't know what they want until you show it to them. That's why I never rely on market research.” Our task is to read things not yet on the page.

Building user-based analytics requires an understanding of business and user goals and the questions they are trying to answer regularly. A commercial manager may want to understand what their inventory positions will be during peak demand so they can plan for tough allocation decisions and conversations with their most treasured customers. An operations lead may want to know which production facility is most profitable for producing a specific product and needs overhead and costing allocation insights to make a decision that will guide supply planners. Similarly, a financial analyst may want to gain insight into customer buying patterns or simply want an interactive and drillable report to review financial results and key performance metrics to benchmark against top competitors. Every user has a need. When designing analytics, start with the need and ask many questions to understand what users are driving toward. Then work with them to develop solutions that range from minimum viable products to comprehensive analytics reports.

Build the analytics to be user-friendly – Richard Thaler at the University of Chicago Booth School of Business teaches a course called Responsible Leadership Through Choice Architecture. The objective of the course is to empower students to explore solutions that help companies to use choice architecture to improve their business results by changing how they operate, manage their workforce, or serve their customers. One team was assigned to help a Chicago-based company with a two-fold strategy to improve the productivity of its customer service department while nudging customers toward its self-service mobile app to improve call resolution and save on costs. The course borrowed from Thaler's book, "Nudge: Improving Decisions About Health, Wealth, and Happiness", where he explains the concept of choice architecture by identifying decision weaknesses people have and proposing solutions to counter them. Thaler's research on the power of installing defaults to guide human decision-making relates to business analytics. We see defaults in everyday life. Imagine you walk up to a glass door with a silver handle and letters above it that reads "push". This is a signal that helps most people interact with the door in the desired way. Similarly, if left to choose, many would not opt into tax withholding or retirement benefits programs. Analytics can apply the default concept. How did the strategy turn out? The company improved first-time resolution metrics for all customer service representatives and guided thousands of customers to use the self-service application, which saved costs in 10x multiples.

We must think about analytics considering the default parameters we want to architect for the users and the insights we want to nudge them to uncover. As you build, you get to design everything from role level access, level of detail provided, forms of interaction with the analytics, core insights, and even aesthetics to guide them in a general direction. This must be thoughtful, or you could create choice overload or limit user knowledge of functionality, so there must be work done to ensure that users understand the capabilities of the tools and believe in its ability to help them succeed.

Bring an inclusive mindset to the solution work. Ask questions about whether one part of the solution will work for all users. For example, there are often design principles established by data and analytics teams for analytics products developed for businesses. Ask questions about colorblind users or differentiated business areas that use unique metrics. Analytics engineers must work to consider all the likely possibilities for how users will interact with the analytics. Embrace a design mindset. It's like benefits plans or 401k programs that call for employees to make elections, but there is a likelihood they miss the deadline, so the default selects a set of reasonable opt-in options for them. Similarly, when analytics are published or models are refreshed, analytic product owners often notify users, and then business analysts across various functions will leverage the tools and synthesize core insights so users can dive in. Build for the users and keep an open mind for continuous improvement opportunities.

Build analytics that tell stories - - Analytics should be built to answer questions with multi-level insights. Often, we need facts to answer simple questions, but in financial planning and analysis (FP&A) we are leveraging demand models driven by statistical models and financial statements that reflect historical and current performance. This information is combined with market trends to predict our future business performance against our ambitions and targets. The analytics must be able to provide simple facts while also providing different lenses through which end users can view the patterns that will provide guidance toward the actions the business needs to take. Telling a story will do more to mobilize the actions needed to win.

Help the end-users draw conclusions - - Is there a second or third-order insight in the data not unlocked? Provide data sets or flexibility for users to create their own combinations of insights. Some end users are technically savvy and will download the data behind the analytics (if access is granted) to build their own analytics. This should be encouraged, but companies should be cautious about data distribution. Incorrect conclusions can be reached when data is used in unregulated formats. For example, finance or operations data has specific calculations and definitions governed by financial regulators or corporate accounting and finance teams. If analysis is not performed in line with standard definitions or requirements for operational, commercial, or financial metrics, you will create multiple truths. You never want users to come to incorrect conclusions, so this is often a great opportunity for data governance and upstream data product management to ensure that anytime a metric is utilized, it is utilized consistently. As companies engineer more powerful analytics, they must also build stronger data foundations by removing complexity, installing checkpoints to monitor data products, and maximizing performance.

Remove complexity and build sustainably - Complexity naturally exists across businesses. It is our job to simplify the complex as we build business analytics for insights. But where does complexity manifest itself? It can be found in multi-step manufacturing processes. For example, converting raw materials to finished goods is labor intensive, and shipments contain multiple batches with different inventory values. Complexity can also arise in the data infrastructure. If there are legacy systems mixed with modern systems, tremendous effort is required to bring those data to a common platform before considering usage for analytics. Sometimes legacy systems are left alone following acquisitions and companies pay the price the longer they wait to integrate. Complexity can also be organizational, where the business units and the business activities cross multiple geographies, markets, customers, systems, and products. You can imagine the many considerations that must be accounted for in a global business from exchange rates to regulations to systems. Complexity presents businesses and their data, analytics, and finance functions with difficult choices that often cannot be made in a vacuum. Data sets cannot be changed without downstream impacts. Systems cannot be upgraded without considering ongoing business processes or user impacts. Remove complexity to enable maximum analytics performance.

Maximize performance– Consider Formula 1 drivers. While they may not manage everything on their cars, they control the cars and receive information on vital signals from the machine that, if not looked after, can be catastrophic to the race, especially if they do not act in time. Business analytics are the same. Business or market activity is measured at the source and is then transformed into usable data or semantic data sets. From there, it is routed to various forms of tabular reports or self-service analytics that deliver insights that ultimately inform decisions. When these links are being formed or data is being cut over to new platforms, analytics and data engineers are presented with multiple options for how to ingest data and structure and maintain analytic architecture. The choices they make are directly correlated with the performance level of the analytics tools deployed. Without high-performing analytics, you cannot provide the key metrics that signal the health of the business or positions of markets and competition. And make sure the “car” is quick!

Finally, **data must be well-governed**. There must be an established ownership structure to ensure that data sets are built according to data engineering standards that are most sustainable yet flexible enough to allow for continuous improvement. The better the structure of the data, the easier it is to run robust calculations and troubleshoot when developers need to unwind code or drill into underlying data structures. Data is the foundation to enable powerful analytics!

Analytics must be designed for the decisions businesses need to make. Every day we all make choices, and to analyze every decision-maker or analyst requires a meaningful preset selection of options through which they can put on a different lens to find patterns and signals. If companies can apply some of these principles to their recent and ongoing digital investments, they will give themselves a better opportunity to understand their internal and external business drivers and make better decisions in uncertain markets. They will also unlock value in some of these ways:

- By developing analytics that are role-based, enabling users to interact with the tools, query for specific views and deep dive into details as needed
- By delivering business insights that address the specific business and market questions and decisions that users need to make to deliver the bottom line
- By unlocking workforce capacity to spend more time in analysis, decision-making and execution
- By establishing new certified data sets built to standards that allow for consolidation to a single global data platform that is an asset
- By improving visibility into business processes and activities by levering analytics to measure endto-end process health through leading and lagging indicators

The remaining results will come from the engagement of users to adopt the tools. Like defaults, analytics are opt-in tools that can be leveraged with a starting point to guide the user. If users engage, then there is a starting point from which they can explore various insights to reach desired conclusions. As they increase familiarity with the tools, they may desire changes or enhancements to the tools to improve functionality or to extract more value from analyzing the underlying data differently.

Designing business analytics and supporting data products is an iterative process. Begin with the users and the problems they want to solve or those that the business needs to solve. Then, gather the requirements to meet the needs and experiment with solutions, gathering feedback along the way. Building a house is a good analogy. Understand the clients' vision and goals and then design a blueprint that meets the expectations. There will be constraints and standards for the build that exist, but through intentional discovery and disciplined engineering companies can achieve greater returns from their digital investments.

Author Biography

Victor Ojeleye is FP&A Planning & Reporting Manager at Cargill Protein North America. Victor drives business growth by creating integrated financial planning and analysis through strategic planning, budgeting, and forecasting while driving modernization, simplification, and standardization of management reporting. His deep passion for community advancement, education, and mentoring are evidenced by his service in the Wichita community universities and local K-12 schools. He serves on boards, sponsors YoungLife, and coaches youth basketball. Victor holds an MBA from Chicago Booth School of Business with an emphasis on analytical finance,strategy, and people. His undergraduate degree is Accounting and Finance from Kansas State University.

Chapter 8

Creating Business Value with Data, Analytics & Insights

Michael Taylor

AI Chief Data Scientist: Rail Analytics Center, Siemens

As the ancient Chinese proverb says, the hardest part of the thousand miles journey is the first step. There is no more daunting journey than a mountainous shifting landscape that grows bigger by the hour. Yet, that is an accurate summary of the task involved in managing data and getting the fullest, most up-to-date and informed intelligence possible. It is a truth universally acknowledged that the modern business is awash with data. This data could be the most asset a company has- but only if it is used right. Businesses today are constantly generating enormous amounts of data, but that does not always translate to actionable information.

Why? Overall, as analytics comes of age, there are growing pains. While investments in analytics are booming, many companies do not see the ROI they expected. They struggle to move from employing analytics in a few successful use cases to scaling it across the enterprise, embedding it in organisational culture and everyday decision-making. How do we empower people with analytics? Empowering people with analytics is where the real value creation occurs. And simply having the best data or writing the most cutting-edge algorithm will not make it happen. It requires a wholesale organisational transformation, complete with robust change management and analytics/AI education programs.

Analytics create value when big data and advanced algorithms are applied to business problems to yield a solution measurably better than before. By identifying, sizing, prioritising, and phasing all applicable use cases, businesses can create an analytics strategy that generates value. Capturing value from data depends on the integrity of the entire insights value chain, and the chain is only as good as its weakest component. Organisations looking to succeed in data insight must ensure excellence in all components and steps of the insights value chain.

The organisations leading the way in data analytics are demonstrating an enormous capability to capture, store, process, scale and provide data to their organisations. But is this enough? The challenge is to develop this data into insights that can increase business value – moving from data to insight to action. In the rest of this chapter, we will share thoughts on how to create business value with data, analytics, and insights.

Analytics value from insight comes not from the activity but from the execution

Data and analytics capabilities have made a leap forward in recent years. The volume of data has grown exponentially, more sophisticated algorithms have been developed, and computational power and storage have steadily improved. The convergence of these trends is fuelling rapid technological advances and business disruptions.

Most organizations have a relatively immature understanding of what “data analytics” is, let alone how it creates value. Possibly the best-guarded secret in analytics, according to Evan Stubbs, “is that, in practice, its success comes down not only to organizational culture but also to the ability of managers to successfully sell the value of analytics”.

As researchers such as Thomas Davenport and Jeanne Harris have rightly pointed out, overall success can often be linked to a variety of factors, including organizational structure, management commitment and successful strategic planning. However, it is often “where the rubber hits the road” that the greatest impact can occur.

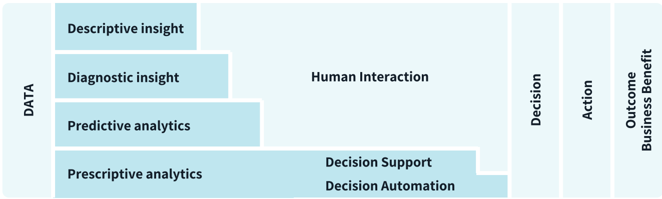


Figure 1: Data Analytics Lifecycle

We can see from the image above that analytics provides the capability to make a better and informed decision. It is acting on these decisions that lead to business outcomes and benefits. Data is only useful if it is actionable. And to make data actionable, it needs to be supplemented with context and creativity. Insights drive processes and actions to impact results. Analytics operationalization requires a robust operating model, an approach that enables driving value to justify major investments, all coordinated through a practical execution plan.

Analytics is a multi-disciplinary activity: the value from insight comes not from the activity but from the execution. Often, this crosses a variety of departments within an organization – few analytics groups have responsibility for both the insight creation and the execution of that insight. Because of this, selling the value of analytics is not just a goal for managers; it is a necessary criterion for success. Insights are great, but without action, they are just facts and figures. To be truly data-driven, you act based on the data and what it tells you. For all the efforts we take to record data fairly and accurately, it means little without applied analysis, revealing new insights, and acting on those insights.

Enabling Analytics through Operationalization: “data-to-insight” and “insight-to- action”

Embedding insights to drive business decisions and actions is one of the key incentives to realizing the business value of analytics. Performing discovery, analysis, and reporting on a snapshot of data in a one-off, stand-alone fashion is valuable. However, we can amplify this value enormously by using the insights gained to inform decision-making. Doing so involves providing feeds of live operational data in a constantly flowing loop. To operationalise analytics, we need to close this data-to-insight-to-action loop, which requires a deep understanding of the applications and integration infrastructure environment.

Despite the significant hype around big data and analytics, success in harnessing them for material business impact is often elusive. This is because a critical aspect is often neglected;

“data-to-insight” and “insight-to-action” are business processes, and to generate material business impact, they require scale and appropriate design, related change management, incentives, and people accountability. Analytics provide insights that enable organisations to act successfully in a reliable manner. This is called “insight-to-action.”

The analytics challenge is generating insight (the data-to-insight process) and embedding (insight-to-action) that insight so it can be used at scale. Successful organisations recognise that analytic models are essential corporate assets that produce and deliver answers to production systems for improved customer relationships, improved operations, increased revenues, and reduced risks. So they seek to create the best models possible. However, few manage all the complexities of the complete analytical model life cycle. It is such a multifaceted task

When you operationalize something, you make it part of a business process. Because analytically driven decisions are better decisions, incorporating analytics insights into your decision-making processes enables you to make the best choices every time. This requires operationalizing analytics at scale. Operationalizing analytics is important because it makes analytics more actionable and thus drives more value. For example, a data scientist might build a predictive model for churn. The model is then embedded in a system, and the model scores customers as they call in. Based on this score, information flows to a call centre agent as part of a business process—say, to up- or cross-sell or take other measures to retain the customer. The agent does not need to know how the model works but can use the output for business advantage. Operationalizing analytics also makes it more consumable.

Being able to collect the right data is one thing but turning it into actionable insight requires a different skill- and mindset. Creating actionable insights is a process like any other, and it is perhaps the most important factor for a business's overall success. The value of analytics lies in its ability to deliver better outcomes. And for selling the value of business analytics, do the following to succeed:

1. Define the value: Analytics creates value. However, the only way to communicate that value is to quantify it. Defining the value need not be a complex process.
2. Communicate the value: The best analytics champions are more than just domain experts; they are also evangelists. A key skill to have been your ability to understand your audience and tailor your message to suit. Making sure your message is relevant is a critical first step.
3. Deliver the value: For any analytics project, the pinnacle of success comes from not only delivering real economic value but also creating and leveraging sustainable competitive advantage. Key to this is understanding how tactical activity maps into strategic outcomes.
4. Measure the value: Measures are only useful if they can be captured. That may seem

obvious, but it remains surprising the number of times people establish outcomes that, despite being intellectually attractive, fail to have the underlying systems for their capture.

Dealing with data deluge requires being smarter. It requires developing the ability to selectively process information based on value, not sequence. It requires, more than anything else, the realization that brute force and manual effort are, eventually, an impossible solution. It requires the effective application of analytics.

Armed with the ability to quantify, communicate, deliver, and measure the value you create, you understand that statistical expertise alone is not enough. Instead, become a change agent, transforming the organization around you. What is needed to make the integrated vision a reality and capture the highest return on analytic investments? A large part of the answer lies in building processes and enablers that operationalize analytics effectively.

So, operationalizing analytics is about building a model focused on building an ecosystem of people, process and technology that enables sustainable value creation. It requires both a repeatable, industrial-scale process for developing the dozens or hundreds of predictive analytic models needed and a reliable architecture for deploying predictive analytic models into production systems. It also requires a more systematic focus on using analytics in operational systems. As organizations expand their use of predictive analytics, it becomes increasingly clear that the long-term value of predictive analytics comes in improving the quality and effectiveness of operational decisions.

Leverage data assets to create a sustainable competitive advantage

Another aspect of creating value with data, analytics and insight is “data monetization”. Data monetization refers to using data to obtain quantifiable economic benefit. The highestperforming and fastest-growing companies have adopted data monetization and made it an important part of their strategy. Let me ask you some questions:

1. **Are you leveraging your data assets to create a sustainable competitive advantage?**
2. **“Are you using the data to differentiate your products and services, which would lead to brand loyalty?”**

Data, as we know, is the new electricity. Armed with it, companies are disrupting established industries, and traditional businesses are transforming the way they operate. Not all organisations, though, are equally adept at translating data into dollars, but their ability to do so is affecting their ability to compete.

Most organizations realize they have a wealth of data -- but not all of them are able to realize its potential value because technological and cultural challenges often stand in the way. Even though more lines of business are better at leveraging their data for their own purposes than they once were, the value of the data from a company perspective may not yet be fully realised. Data quality issues are common. In addition, compliance, privacy, and security issues may limit how the data can be used.

Bruce Daley, an analyst at market intelligence firm Tractica, said: Where knowledge is power, data is wealth. It is not intrinsic in the data, it is what you do with it". The companies that are most progressive in thinking about data differently are the companies changing the economy, like Google and Uber. Most businesses lag way behind in terms of the idea that data could be their primary reason for being.

The data revolution is here, and it creates an investment priority for companies to stay competitive and drive new opportunities. One of the brightest areas is data monetisation, which describes how to create economic benefits, either additional revenue streams or savings, utilizing insights provided by data resources. A strategic approach to analytics will provide the key to unlocking the value in your data – enabling you to gain unprecedented insights, bridge the gap between analysis and business outcomes, and combine the power of big data with analytics to drive business success.

This asset (data) is generated through the deployment of technology; however, monetization of this critical asset is most definitely a business challenge. Technology will help collect, store, and deliver the data; advanced analytics will help explore the data and discover strategic and transactional value. But it is a sound strategy, followed by an effective business model, which helps the enterprise focus on creating value from data and analytics, effectively monetize the asset and create a competitive advantage for the enterprise.

Data is everywhere, and it is changing all interactions companies have with their customers. With data undisputedly fuelling a competitive advantage, leading organizations realize they must leverage data to transform their businesses. Today companies that use data to deliver exceptional customer experiences, build great products and drive operational efficiency are the winners in this new digital economy. Have a structured methodology to harness the power of data. While the deployment methodology is usually agile, having a structured, repeatable program enables you to transform quickly into data-centric organizations. This ensures you stay competitive in a world that increasingly finds itself under a data deluge.

Business intelligence (BI) solutions, tools, and dashboards that deal with packaging and visualizing historical performance data to support decisions are hugely valuable to the day-today management of operations. An effective data monetization strategy is not about tools, underlying technologies, or the tactical management of the enterprise. As an analogy, if we view BI tools and databases as the pipelines and the data as the electricity that flows through it; although an effective pipeline is important for delivering electricity, the electricity is monetized (sold) and not the pipeline network.

Key Takeaway

To create business value from data, analytics, and insights, here are steps you can take:

1. Start a rigorous process with your executive team to decide where the most promising sources of value exist. To start, identify which functions or parts of the value chain have the most potential. When insight is closely tied to your key business goals and strategic initiatives, it is more likely to drive action. If you do not know how to react to a particular metric when it significantly increases or decreases, you might look at an unnecessary vanity metric. Insights based on key performance indicators (KPIs) and other key metrics inherently engender a sense of urgency that other data will not. It is easier to interpret and convert strategically aligned insights into tactical responses because they often relate directly to the levers in your business you control, influence or are focused on.
2. Then come up with possible use cases and how new data and techniques could be applied to them. Using outside benchmarks can be useful to get a sense of how valuable a use case might be.
3. Make sure the insight is Relevant. To be relevant, an insight needs to be delivered to the right person at the right time in the right setting. If insights are not routed to the right decision-makers, they will not receive the attention they deserve. If insights are not timely, they might be too stale for stakeholders to act on. If insights are trapped in an analytics tool that managers never access or deliver to devices they use infrequently, the insights may never reach the intended audience.
4. Make sure the insights are specific and complete. The more specific and complete the insight is, the more likely it can be acted on. Sometimes insights based on KPIs, and other high-level metrics can highlight interesting anomalies but lack sufficient detail to drive immediate action. If an insight does not adequately help to explain why something occurred, it is not yet actionable. Deeper probing may be required before it is ready for primetime.
5. Finally, if people do not clearly understand an insight, why it is important and how it can help them—the insight will be overlooked and forgotten. Communicating insights effectively is important to their adoption and fruition. The right data visualizations and messaging can help explain insights so they are more easily understood and correctly

interpreted. However, poor communication can cause the signal to be lost in the noise. A clearly communicated insight creates a strong signal that is hard to miss or ignore, and it prepares a pathway for action to occur.

To extract value from analytics, you need to focus on improving strategic, technological, and organisational aspects on how you treat data and analytics. Richard Bach said, “Any powerful idea is absolutely fascinating and absolutely useless until we choose to use it.” While the increased actionability of an insight does not guarantee its adoption or application, it should motivate more individuals within your company to think more deeply about the data and encourage them to act on a more consistent basis. Make sure your hard-earned insights are actionable so they are primed to drive value for your organisation.

References

1. https://tdwi.org/articles/2017/04/07/5-steps-to-monetize-your-data.aspx
2. https://sloanreview.mit.edu/article/how-to-monetize-your-data/
3. https://home.kpmg.com/content/dam/kpmg/pdf/2015/10/framing-a-winning-data.pdf
4. http://analytics-magazine.org/the-value-of-business-analytics/
5. http://www.decisionmanagementsolutions.com/wp-content/uploads/2014/11/Operationalizing-Analytics.pdf
6. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Analytics%20comes%20of%20age/Analytics-comes-of-age.ashx
7. https://www.sciencedirect.com/science/article/pii/S0268401217300816
8. https://www.forbes.com/sites/brentdykes/2016/04/26/actionable-insights-the-missing-link-between-data-and-business-value/#573a2ef51e57

Author Biography

Michael Taylor is the AI Chief data scientist for Siemens Mobility Rail Analytics Center in Singapore with a mission to democratize Data Science, reliable AI, and deeply committed to designing innovative solutions to solve customers’ business problems. Michael is a Fellow at the Royal statistical society, Member of the Operations Research Society, and a Guest Lecturer in Big data for Business analytics at Bocconi University in Milan since 2019.

Chapter 9

Defining the Modern Data Landscape with the Semantic Layer

Brian Prascak

Customer Success Director, AtScale

This chapter introduces the modern data landscape, including providing perspective on the data and analytics industry - where we've been, where we are now, and where we are heading, linking the catalysts for improvement to the capabilities being sought, the vendors providing them and the investments being made. The emphasis will be to define the modern data landscape, frame its purpose and direction, with a focus on the fundamental need for actionable, impactful insights and analytics that are delivered with speed, scale and cost-effectiveness. The rise of the semantic layer will be featured, including new research that affirms the value of using a semantic layer to deliver increased speed, scale and success for AI and BI.

Defining the Modern Data Landscape - Catalysts

It's hard to believe that just 15 years ago, big data and cloud technology emerged with the introduction of Hadoop and cloud vendor offerings. Over the past 5 years, most enterprises have moved to the cloud, motivated by the dual need for digital transformation of their business, coupled with embracing cloud-based data platforms and tools to realize the benefits of advanced insights and analytics. Most companies have now implemented migration to the cloud, with most, if not all, of their data in a cloud-based data lake. Also, given the ever-increasing number of success stories across a wide variety of industries, companies have bought in to using data and analytics to significantly improve business performance, with many implementing an initial set of use cases and planning for continued expansion. The resulting investment in big data technology reveals the scope of this transformation: according to research firm International Data Corporation (IDC), worldwide spending on big data and business analytics (BDA) solutions in 2021 was forecast to reach $215.7 billion, an increase of 10.1% over 2020 - with IDC forecasting that BDA spending will gain strength over the next five years as the global economy recovers from the COVID-19 pandemic. The compound annual growth rate (CAGR) for global BDA spending over the 2021-2025 forecast period will be 12.8%, much larger than most every category of IT spending. Per IDC, total BDA spending is expected to be split evenly between services and software solutions.

The Need for Speed, Scale and Cost Effectiveness

Let's start with aligning the key drivers of value for implementing data and analytics capabilities and selecting vendors. The market has moved from an emphasis on basic infrastructure and tools, primarily the bookends of the modern data stack, representing cloud data lake providers such as AWS, Google and IBM, to basic data wrangling and BI tool providers like Alteryx, PowerBI, Excel, Tableau and MicroStrategy and open-source resources such as Python and R to additional emphasis on capabilities and vendors. The focus now is to improve performance, productivity and impact by establishing core processes, resources, strategy, planning and governance and integration focusing on delivering improved speed, scale and cost effectiveness:

Actionable Insights -Fundamentally, data is a means to an end - and that end is an insight that leads to a better answer. And that answer must be actionable to deliver impact. Data is nothing if it is not actionable and impactful, so companies are seeking to make their data more actionable, and the way to do that is by addressing the four (4) A's of Actionable Insights - Availability, Accuracy, Actionability and Automation.

AVAILABLE	ACCURATE	ACTIONABLE	AUTOMATED
Data are easy to locate and access	Data are accurate and complete	Data address key questions / needs	Data processes are automated
Accelerated – Provide Semantic Layer, Self Service, Governance, COE's			

Scale - Now that most companies have all of their data in one place, they are seeking to scale the number of data sources, users and use cases they can support. Scale is the most critical challenge that companies face today, and the reward should be worth the effort - as a recent McKinsey study Tipping the Scales in AI indicates, companies that scale insights and analytics achieve, on average over 8% points higher EBIT (3.4x improvement) than companies who have not achieved scale. **Achieving scale with Data and Analytics has four core Elements**

DATA SOURCE	USES	USERS	CONSISTENCY
Expand number of integrated data sources for analysis	Expand number of use cases that can be addressed: AI and BI	Expand number of insights creators and consumers, including self-service users	Implement semantic layer,data catalogs and feature stores for reusability
Data Governance - Data Access, Pipelines, Master Data, Data Products, Metrics/Features			

Speed - Speed and Scale are really two sides of the same success coin: scale means nothing without speed, but speed means nothing without scale. The most important characteristic of actionable insights is that they are relevant - and relevant means timely. All too often, companies take way too long to make desired data sources available and even longer to make that data actionable for AI and BI. Why? The process for turning data into actionable insights - what we call the “Last Mile” - can take as many as seven steps (accessing, profiling, preparing, integrating, analyzing, synthesizing, presenting) - often these steps are done manually with multiple resources requiring multiple handoffs and reviews with frequent refinement loops - most companies say it takes an average of 4 months or more to launch a new data source. Recent research reveals that using a semantic layer can reduce this time by 1/4th to just 4 weeks or less. Achieving speed-to-Insights with Data and Analytics has the following four core Elements.

RAPID ACCESS	RAPID PREPARATION	RAPID REFINEMENT	PUBLISHING/SHARING
Enable rapid, governed access to analysis-ready data	Provide self-service data preparation and modeling tools	Use semantic layer and feature stores for consistency and reuse	Use Semantic Layer to automate data product publishing
Capabilities - Data Strategy, Capability Roadmap, Tools, Skills, Literacy, Delivery			

Before we dig into the modern data landscape, let's also review the major areas being transformed to deliver actionable insights for AI and BI with speed, scale and cost-effectiveness. These address the people and process aspects. We see these four capabilities being implemented by companies successfully implementing data and analytics at scale.

DATA LITERACY	DATA AS PRODUCT	DATA DEMOCRATIZATION	DECISION INTELLIGENCE
Understanding how to improve using data and analytics	Managing data as a product across the enterprise	Decentralizing insights and analytics with central support	Understanding how decisions can be improved and scaled
Technology - Data Lakes, Virtualization, Semantic Layer, Catalogs, Feature/Metric Store			

Defining the Modern Data Landscape

Now that we have defined what we want to achieve and what we want to transform, let's define the modern data landscape, including key vendors. We will also consider the areas rapidly emerging, including those most suited to support achieving increased speed, scale and cost savings for AI and BI. The modern data landscape consists of seven (7) major capability areas, representing fifteen (15) individual capability components. Let's briefly review each of the capability areas:

1. Raw Data - Raw data represent sources and storage of raw data sources. There are two major categories of raw data:

    a. Data Lakes managed by cloud providers

    b. SaaS applications where data is managed by the vendor for clients, who can access via web protocols, including APIs.

2. Data Preparation, Integration, Workflow (DPIW) - The DPIW capability area enables data to be extracted and prepared: cleaned/transformed and provided as a ready-to-use set of data. There are two major categories of DPIW:

    a. Data Transformation and Preparation Tools - These are tools to profile the data, assess it, cleanse it and transform to it to make it ready for analysis, including as a single data source or integrated with other data prepared data sources. Often, these tools create data pipelines and automate the process of data preparation.

    b. Customer Data Platforms/Event Tracking - With the increased maturity and confluence of ecommerce and digital marketing, companies now must use a plethora of data sources and vendors to manage customer data across a myriad of channels. Customer Data Platforms, which offer purpose-built capabilities to manage customer identification, hygiene as well as rapid access and integration of data between multiple marketing data vendors and channels have increased exponentially in popularity. According to IDC, the worldwide customer data platform software market will grow at 19.5% CAGR from $1.3 billion in 2020 to $3.2 billion in 2025.

3. Data API - The Data API Layer is rapidly emerging as another accelerator for companies to more rapidly access data from source systems, including data warehouses in the cloud and process it at the source (rather than move it). There are three (3) major categories of Data API vendors:

    a. Cloud Data Warehouse - A cloud data warehouse is a database stored as a managed service in a public cloud and optimized for scalable BI and analytics. Cloud data warehouses typically offer three major services: secure access, compute or query processing and storage.

    b. Data Lake Engines - A data lake engine is an open-source software solution or cloud service that provides critical capabilities for a wide range of data sources for analytical workloads through a unified set of APIs and data model. Data lake engines address key needs in terms of simplifying access, accelerating analytical processing, securing and masking data, curating datasets, and providing a unified catalog of data across all sources. Data lake engines simplify these challenges by allowing companies to leave data where it is already managed, and to provide fast access for data consumers, regardless of the tool they use.

    c. SaaS APIs - These providers offer rapid, software-as-a-service (SaaS) data integration service for companies to extract, load and transform (ELT) data from different sources into data warehouses. Often these providers create a standardized data model and framework to move data from standardized sources, including other SaaS-based data providers/sources.

4. Logical Data Models - This capability is critical to ensuring that data is consistently available to the consumption layer for AI and BI applications. With the number of applications consuming data, it is critical to ensure that the data is consistently defined, modeled, aggregated and optimized for presentation and rapid query response. There are three (3) major categories of Logical Data Model providers:

    a. Semantic Layer - The Semantic Layer improves the time to insights for AI and BI by simplifying, automating, standardizing, and optimizing how data products are created, consumed, and queried for AI and BI. Semantic Layer leaders like AtScale offer a comprehensive set of capability components, including Consumption Integration, Semantic Modeling, Data Preparation Virtualization, MultiDimensional Calculation Engine, Performance Optimization, Analytics Governance and Data Integration.

    b. Metric/Feature Stores - Another fast-growing area within the data landscape is the use of metric and feature stores. Metric stores are typically used to support business intelligence whereas feature stores support data science uses. Both metric stores and feature stores address common needs and provide common benefits - namely to support the consistent definition of metrics and features, and provide a single, centralized source for consistent reuse across the enterprise.

    c. Data Virtualization - Data virtualization provides a logical data layer that presents and enables integration of data that may be siloed across the disparate systems, manages the unified data for centralized security and governance, and delivers it to business users without having to physically move the data. Data Virtualization is often used with a semantic layer, where the data virtualization speeds access to the data whereas the semantic layer speeds the ability to access the data consistently (and refine it) through multiple AI and BI consumption tools without creating multiple versions for each tool.

    d. Data Governance - Data governance (DG) is managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage. As the number of data sources, users, uses and consumption tools increase for both the data, but also the data products (refined data sets created by the semantic layer models and metric/feature stores), data governance becomes increasingly important. Companies like AtScale provider governance capabilities built into the semantic layer to govern data as a product.

5. Data Consumption - This capability is critical to ensuring that data is structured and presented effectively for business intelligence and analytics. There are two major categories of data consumption vendors:

    a. BI Tools - Business intelligence (BI) tools are application software that collect and process large amounts of data from internal and external systems, including books, journals, documents, health records, images, files, email, video, and other business sources. BI tools provide a way of amassing data to find information primarily through queries. These tools also help prepare data for analysis so you can create reports, dashboards, and data visualizations. The results give both employees and managers the power to accelerate and improve decision-making, increase operational efficiency, pinpoint new revenue potentials, identify market trends, report genuine KPIs, and identify new business opportunities.

    b. AI/ML Tools - These are tools designed to speed up the process of creating AI/ML models. Often, they offer workflow automation, data preparation, access to models/algorithms and support training and operationalization.

6. Data Catalogs - A data catalog is an organized inventory of data assets for access within the enterprise. Data Catalog uses metadata to help organizations manage access to their data, including collecting, organizing, accessing, and enriching metadata to support data discovery and governance.

7. Data Observability - Emerging as a newer area within the modern data landscape, Data Observability refers to an organization's ability to understand the health and reliability of the data in their system. Traditionally, data teams have relied on data testing alone to ensure that pipelines are resilient. However, as companies ingest ever-increasing volumes of data and the data pipelines become more complex, testing during deployment is no longer sufficient. Continuous monitoring of data to determine if changes are taking place is critical to ensuring tracking of data quality, lineage, consistency, usage, governance, and refinements across the entire ecosystem - all part of what is now being called "data operations" - ensuring that all data sourced, created, transformed, synthesized, summarized and consumed used to support multiple applications are consistently defined and delivered as needed.

Figure 1: Modern Data Landscape - Capability Areas and Vendors

Modern Data Landscape - Fast Growing Areas

As the modern data landscape continues to evolve, focusing on delivering actionable insights and analytics via improved speed, scale and cost savings, we see these areas accelerating growth in investment, customers and market coverage:

- Semantic Layer - Although AtScale was the first to introduce a Semantic Layer over 10 years ago and has over 50 clients, client interest and investment has accelerated in the past 3 years as companies have migrated to the modern data platforms, and have realized the importance of achieving speed, scale and cost savings at the actionable insights level. As more companies move more of their data to the modern data platforms, the importance of the semantic layer becomes even more important - recent research by Ventana Research reveals organizations that have successfully implemented a semantic model are more than twice as likely to report satisfaction with analytics (77%) compared with a 33% overall satisfaction rate.
- Metric/Feature Stores - Supporting the accelerating interest using a Semantic Layer for AI and BI, enterprises are also embracing the complimentary capability of using centralized metric and feature stores to ensure consistent definition of metrics and features, and provide a single, centralized source for consistent reuse across the enterprise. Companies embracing the use of the Semantic Layer typically also embrace the use of metric and

feature stores to ensure that both existing and new datasets/data products are consistently defined and productively shared.

- Data Governance, Data Catalogs and Data Observability - As companies embrace the use of cloud-based data platforms, and as data sources and applications that consume data expand, companies are embracing the use of the Semantic Layer and Metric/Feature Stores supported by the increased emphasis on Data Governance to manage data privacy, access and usage, Data Catalogs to support data discovery and data observability to monitor data moving through the entire data ecosystem.
- AI/ML Tools - As companies increase their embrace of AI/ML, they are also looking to improve the productivity of their data science teams, including analysts using AI/ML automation tools. As more companies do more with AI/ML, interest in tools to increase productivity through workflow and automation, including self-service increases, as does improvements in their capabilities to support self-service for both data scientists and analysts.
- Customer Data Platforms (CDP) - Investment in CDP's is growing exponentially due to the combination of cookies going away (companies having to manage customer data more directly with explicit permissions), digital transformation and use of analytics to improve customer engagement and deliver more personalized recommendations.
- Data API Layer - All of the capabilities within the Data API layer are growing rapidly as companies seek faster ways to access, integrate and compute data from multiple sources within the cloud, including many new data sources from existing sources (not analyzed before) and new sources (new vendors).

Semantic Layer Rising

Over the past year, there has been a tremendous resurgence in the Semantic Layer among large enterprises. This traces to their recent experience migrating to modern data platforms and now experiencing the need to improve speed, scale and cost savings for AI and BI - generating actionable insight from newly available data sources for many new users and use cases. Fortunately, recent research affirms the value of using a semantic layer. The research points to companies realizing the promise of successful, impactful data and analytics programs using a semantic layer - and in stark contrast to those that don't use a semantic layer.

According to recent research from Ventana Research, based on 300 respondents, organizations that have successfully implemented a semantic model/layer:

- Are significantly more satisfied with analytics (77% compared with 33% overall)

- Have more of the workforce engaged in analytics (43% compared with 23% have more than one-half the workforce using analytics)
- Find analytics capabilities adequate (62% vs. 33% overall)
- Are more comfortable with self-service: (54% very comfortable vs. 14% overall)

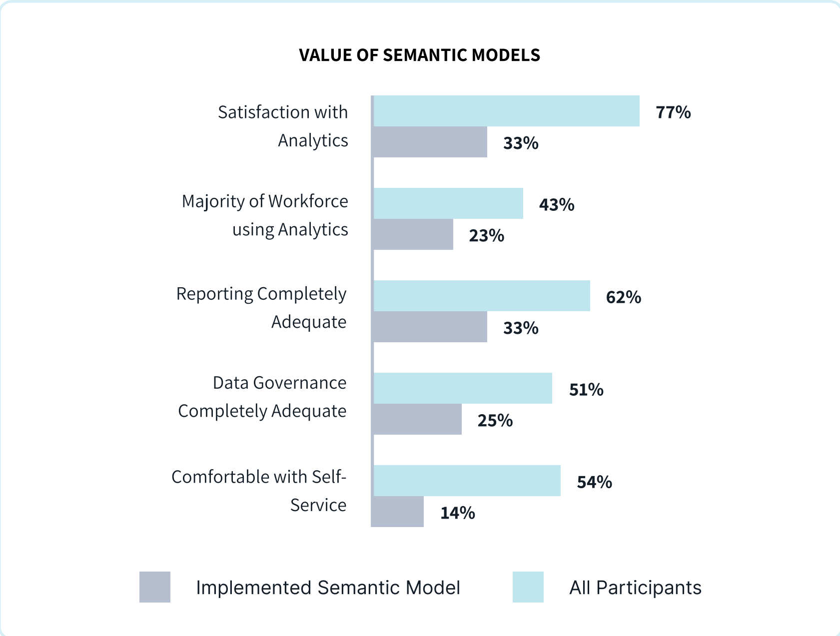


Figure 2: Value of Semantic Models

Further recent research from DBP Institute, over 100 respondents cited that companies using a Semantic Layer cite a 4.2x improvement (i.e., a magnitude of 4.2 times improvement over the base level of performance from not using a semantic layer) in performance with less than half the effort required (e.g. savings in both numbers of resources, hours, project time/duration, and overall cost). This is a significant order-of-magnitude improvement in performance as well as a reduction in effort and cost. It means that a typical project taking 4 months to complete could be done in just 4 weeks using a Semantic Layer. The performance improvement was significant and consistent across every measure.

- 4.4x improvement in Time-to-Insights (e.g., insights and analytics development)

- 4.4x improvement in a number of self-service users, data sources, metrics consistency
- 4.2x improvement in Cloud Analytics performance
- 3.7x improvement in cost savings

AtScale Semantic Layer: Enabling Actionable Insights for Everyone

AtScale provides a Semantic Layer, which sits between the Data Source Layer and the Insights Consumption Layer (e.g., AI, BI and Applications). The Semantic Layer converts data into actionable insights via Automation (self-service data access, preparation, modeling, and publishing), Alignment (centralized data product management and governance with a single, consistent metric store) and Acceleration (cloud analytics optimization - BI query speed optimization, multidimensional OLAP in the cloud, AI-based data connectors, and automated PDM tuning). This supports insights and analytics creators, enablers and consumers without requiring data movement, coding, or waiting.

AUTOMATION	ALIGNMENT	ADVANCEMENT
Self-service data access, preparation, modeling, publishing for AI & BI	Centralized Data Product Management with Single Enterprise Metric Store	10x Increase in Query Performance, Automated Tuning, Cloud OLAP

Author Biography

Brian Prascak is an expert in data, insights and analytics, with extensive experience helping companies realize the benefits of using AI/ML and BI across a wide variety of industries and functions in the US and globally. Brian is the Customer Success Director with AtScale. Before supporting AtScale, Brian was Director, Advanced Analytics, Platforms and Data Services at Wawa. Brian has worked with many global organizations, including previously at IBM, Diageo, Mastercard, JPMorgan and ACNielsen

Chapter 10

Demystifying BI with AI

Tobias Zwingmann

Author and Founder of RAPYD.AI

The biggest misconception about the relationship between business intelligence and artificial intelligence is that you first must master the one before you can tackle the other. Let's make it clear: AI is not the next stage of BI. AI should rather be recognized as an enabler for modern BI systems right from the start. Let's discover how to make this work.

What is BI? BI is difficult to define because interpreting this term depends on whom you ask, when and in what context. For the context of this article, we use business intelligence as a technical term: it describes a system or software that allows business users and analysts to look at data from multiple sources within an organization to make better-informed decisions. So, for example, a BI system is software such as Power BI, Tableau, Looker, etc., and the underlying infrastructure that feeds these systems.

So, what is AI? There are two broad concepts: One concept is general artificial intelligence or strong AI. It describes a system able to act intelligently on a human level by solving a new, previously unknown problem on its own using logic and reasoning. Strong AI is still a research case and we do not know when it will actually ever exist. The other concept is narrow AI. Narrow AI describes an apparently intelligent system capable of solving a specific problem for which it has been trained.

All business use cases involving "AI" today use narrow AI. The main driver behind these narrow AI use cases is a technology called machine learning (ML). Machine learning is essentially a programming paradigm that allows you to automatically extract patterns from data at scale. The most popular modern AI use cases are actually powered by ML are Search engines, churn prediction, demand forecasting, document processing, and predictive maintenance. Thus, I will use the terms AI and ML interchangeably in this chapter.

Why should we care for AI in BI?

The total data continues to grow at a rapid pace with no end in sight, as recent studies show. This data growth is putting the traditional systems of BI to the test: data growth is fueled not only by well-structured tables, but above all, by unstructured data in text, images, videos, documents, audio, or log files that come from an ever-increasing range of different sources. To make matters worse, most companies face a severe shortage of data-savvy employees skilled in handling highdimensional data sets and the corresponding tools (in this case, not Excel). While these circumstances are already challenging the organization, there are other trends at work:

- **The need to get quick answers from data**: To stay competitive, companies need datadriven insights to help them grow. Data analysts are inundated with requests to examine this or that metric or review this or that data set. Business users need to get answers from data quickly and easily. If they can ask Google or Alexa for the current stock price of a particular company, why can they not ask their professional system BI for yesterday's sales figures?
- **Democratization of insights**: Business users have become accustomed to using selfservice BI solutions to get insight from their data. However, today's data is often too large and too complex to be left to the business for pure self-service. To continue the democratization of insights across the enterprise, BI systems are needed that are easy to use and automatically deliver insights to end users.
- **Accessibility of ML services**: Low-code or no-code platforms make it easier to bring machine learning technologies to non-data scientists and pressure the BI team to incorporate predictive insights into their reports.

When these factors interact, the question is not whether you will improve your BI with AI but when. Let us look at a framework for applying AI to BI.

The AI For BI Framework

The AI For BI framework extends the analytical insights model published by Gartner1

. At its core, it describes four types of analytics. Descriptive and diagnostic analytics of historical data are core functions of any BI system to provide hindsight and insight. These two methods are critical because all other analysis processes build on them. Reliable and structured reports remain the most important backbone of enterprise data analysis, as they support further concepts and actually trigger questions that lead to further analysis. Predictive and prescriptive analytics seek to fulfill the desire to not only understand the past but also predict the future. If we knew what would happen tomorrow, we could make better decisions today. These tasks are much more complex because they do not deal with factual data from the past but with uncertainties, assumptions, and probabilities about future events. The AI For BI framework extends these types with three AI impact areas in the following figure:

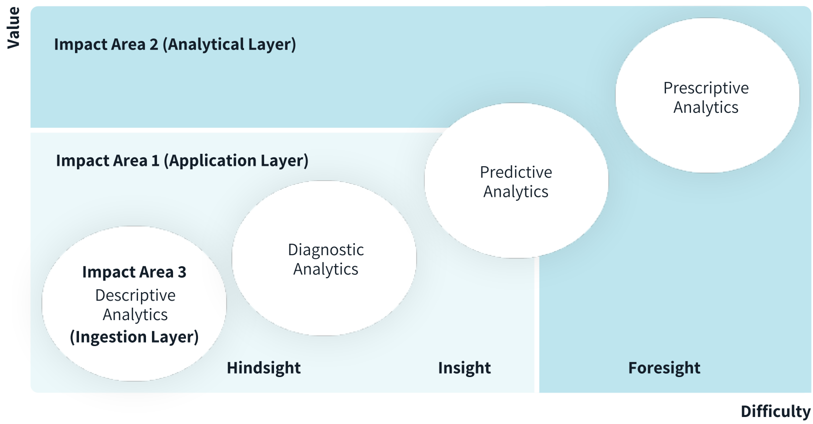


Figure 1: AI For BI framework

AI can typically empower BI systems across all stages at three layers:

- **The application layer:** AI can help to make the BI system easier to use and reduce the overall time to insights (TTI).
- **The analytical layer:** AI can help to make better predictions and forecasts at scale.
- **The ingestion layer:** AI services can unlock new data sources

Let's explore these three layers in more detail.

1 https://blogs.gartner.com/jason-mcnellis/2019/11/05/youre-likely-investing-lot-marketing-analytics-getting- right-insights/

Impact Area 1: The Application Layer - Improve the BI User Experience with AI

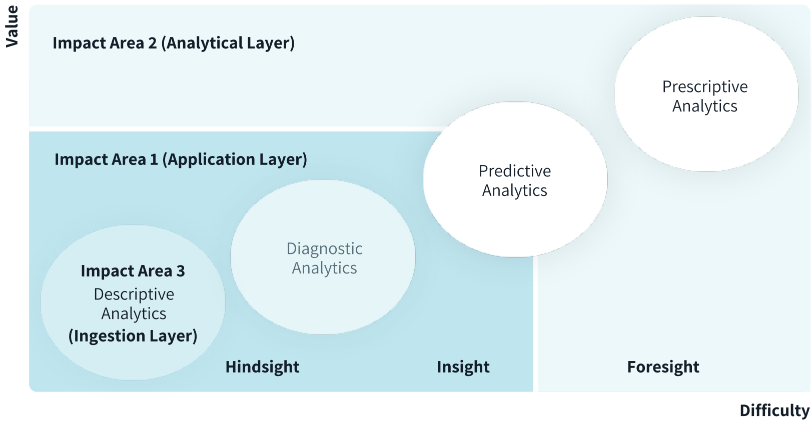


Figure 2: Impact Area 1 in AI For BI framework

Traditional self-service systems BI, which allow users to drill down into complex data on their own, have a lot of technical friction. To allow as many drilldowns as possible and anticipate almost every possible question, BI engineers have created complex data models with a wealth of different attributes and dimensions. This complexity makes it difficult, especially for nontechnical users, to gain the desired insights without the help of an experienced data analyst. AI can help reduce friction and make the user interface more intuitive for professionals. Here are three examples:

Use Natural Language to get insights

If there is one digital skill that people have learned in the last decade, it is probably how to use a search engine. Google has become ubiquitous, and for many users, it is easy to get the information they need with a simple search. Whether you want to learn more about tomorrow's weather, Apple's stock price, or when the French Revolution took place, you can get the answers with a single search. Why should the same not be possible for BI? Thanks to major advances in natural language processing (NLP) and the introduction of large language models like GPT3, technology is better than ever at interpreting human text and optimizing it for specific tasks like searching for questions and answers.

While many of these technologies are still in the experimental stage, some BI systems have put these features into production. In Power BI, for example, you can use the Q&A tool to ask queries about your dataset, such as:

"Show me the sales in the U.S. last year."

The software will understand these queries and respond with an appropriate diagram displaying the information. For most users, it is much easier to work with a simple text entry area and perhaps a handful of suggested questions than it is to put together a report from a complex pivot table. Using NLP techniques can greatly increase the adoption of BI and relieve analysts of mundane tasks.

Detect patterns in data automatically

Many diagnostic questions in BI systems revolve around finding out why something happened or comparing different data distributions (e.g., Which product line contributed to the sales increase in the last quarter?).

Intelligent algorithms make it possible to sift through mountains of data in seconds and provide business users or analysts with interesting patterns or insights, reducing time to insights. In addition, AI can help discover interesting correlations or unusual observations between many variables that humans might otherwise miss. Often, AI can better look at multiple metrics in combination, whereas humans typically focus on a single metric at a time. These data segments can then be displayed to a report user for further exploration or validation.

For example, in Power BI, report designers can use the Key Influencer and Decomposition Tree visualizations to allow users to quickly explore their data using ML-based techniques.

Summarize analytical results

The third example where AI can help make the BI system more user-friendly is not about discovering insights but about communicating them. Even if a diagram seems self-explanatory, it is advisable to summarize the key findings in one or two lines of natural language to reduce the risk of misinterpretation. But who likes to write self-evident descriptions under diagrams in reports or presentations? Most people do not, and that's where AI can help.

AI-powered natural language processing can create summaries based on data. These automatically generated paragraphs of text include descriptive features about the data and notable changes or streaks. Here is an example of an automatically generated chart label from Microsoft Power BI:

Sales $ for Texas increased for the last 5 years on record and it experienced the longest period of growth in Sales between 2010 and 2014.

These little snippets of text generated by AI can make the lives of business users and analysts easier and save a lot of time. They also help make reports more accessible to screen readers.

As you can see, AI-powered techniques can make BI easier to use overall and, therefore, more user-friendly. Making BI systems more accessible to non-technical users further democratize data and takes work away from analysts. These capabilities are embedded in modern BI platforms and are sometimes even so hidden that you do not even realize AI is involved. Be aware of these capabilities and use them effectively across your organization.

Impact Area 2: The Analytical Layer – Improve Forecasts and Predictions with AI

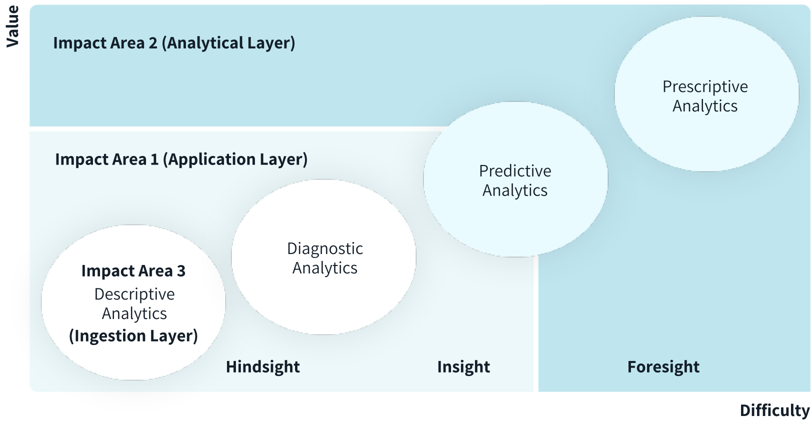


Figure 3: Impact Area 2 in AI For BI framework

Forecasting has always been a challenge in traditional BI systems. Forecasts have typically been based on aggregated data. When business users wanted to forecast sales for the next month, they likely looked at previous months' sales data and drew a trend line based on that. These forecasts were too often rough heuristics and did not allow for any actionable insights because they could not be broken down further.

At its core, AI uses old familiar techniques like linear regression. But AI can apply these techniques to very large and complex data sets and use stochastic approaches to find an optimal solution given the data in a short time, with no extensive human supervision. Here are two examples of how this plays out in practice:

Predict data on a micro level

Using AI, you can calculate a churn probability for each customer in your database based on actual past data. Not only can you use this information to determine which customers are likely to churn in the next month, but you can also optimize your decision-making, such as: Which of the customers who are likely to churn next month should be targeted with a marketing campaign according to their monetary value? The combination of machine learning and BI offers potentially great value to a business.

Capture more complex patterns

Specialized time series prediction algorithms, for example, detect patterns in larger sets of time series data. This can lead to better or more accurate predictions within a short time horizon or to an attempt to make more accurate predictions over a longer period. More complex, non-linear models can lead to more detailed and ultimately better prediction results.

AI capabilities for improved forecasting or better predictions can be deployed both as an integral part of existing BI software (application layer, online prediction), or independently, directly at the database level (analytical layer, batch prediction). So they are always available, regardless of which BI tool you use.

With the advance of new techniques like AutoML and AI-as-a-Service, companies can overcome the bottlenecks of not having enough Data Scientists or ML practitioners to take advantage of these AI potentials – as long as the users solidly understand the underlying data.

Impact Area 3: Ingestion Layer - Unlock new data sources with AI

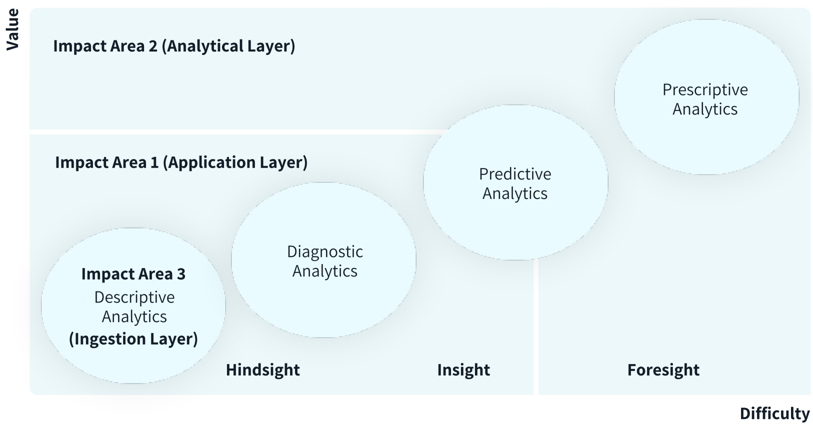


Figure 3: Impact Area 3 in AI For BI framework

BI systems typically work with tabular data from relational databases such as enterprise data warehouses. But with increasing digitization across all channels, we're seeing a dramatic increase in unstructured data in text, image or audio files. It has been difficult for BI to analyze this data at scale. AI brings opportunities to change that.

Extract structured data from unstructured data

AI can increase the volume and depth of available and machine-readable data by using technologies such as computer vision or natural language processing to access new, previously untapped data sources. Unstructured data such as raw text files, PDF documents, images, audio files, etc. can be converted into structured formats that conform to a specific schema, such as a table or CSV file, and can then be consumed and analyzed through a BI system. Since this happens at the level of data input, this process ultimately affects all levels of the BI platform. By including these files in your analysis, you can obtain even more information that can lead to better predictions or a better understanding of your data.

Summary

AI is changing the landscape of BI. The main reasons are the need for business users to get faster answers from data, the growing demand for democratized insights, and the general availability of machine learning tools. We've seen how AI can improve BI systems by helping people make better and faster decisions through automation and better usability, improved forecasting, and access to new data sources.

The most effective AI-powered BI application exists when we combine automated and human decision-making. By now, you should solidly understand the strategic impact areas AI can have on BI to structure your thought process and kick off the next steps for developing your own AIpowered BI use cases.

Author Biography

Tobias Zwingmann is the Co-Founder of the German AI startup RAPYD.AI and is on a mission to help companies implement machine learning and AI faster while delivering meaningful business value. He is the author of the book AI-Powered Business Intelligence and has more than 15 years of professional experience working in a corporate setting, where his responsibilities included building data science use cases and developing an enterprise-wide data strategy.

Chapter 11

Translating analytics products into business value

Megan C. Brown

Ph.D., Director, Global Center of Excellence for Advanced Analytics and Data Science, Starbucks.

As someone who spent a great deal of time studying cognitive science (for someone in industry), the phrase “It’s just semantics!” is confusing. Semantics is the heart of it all – it’s the meaning we create for ourselves from our experiences, training, and exploration. Communicating meaning is even more challenging – your ability to communicate meaning from your experience can be unsuccessful for many reasons. Perhaps the other person doesn’t have the same experience as you, describes the same event differently, or doesn’t understand all the phrases you’ve used. Perhaps you haven’t communicated your thoughts clearly. Semantics and your ability to communicate meaning is the heart of it all, isn’t it?

For example, let’s say this morning I’m a decision scientist and I’ve glanced at a personal dashboard I made to monitor regional weekly sales and transactions KPIs. I noticed a strange dip in sales for this week, considering the week hasn’t closed yet. It seems like something the business should know. I have the dashboard (that I made quickly to sate my own curiosity) and my interpretation. How do I communicate this with my stakeholders in my business? What would happen if I sent my interpretation and a link to the dense, scientist-oriented dashboard to my friends in the business? Would they take the time to read it, open the link, and confirm my interpretation?

Next, let's say I clean up the dashboard – I make it more visual and streamline the number of metrics on the page. I set a meeting with one of my main stakeholders and start the conversation with the dashboard. Three days later we meet, and I talk through the dashboard point-by-point until they have the same interpretation I do. We likely all agree that insight democratization (a.k.a. self-service) is the ideal state. It's inefficient to have a person interpreting and then advocating for their interpretation when our stakeholders could check the dashboard regularly and derive the insight themselves. To live in a future where self-service is an effective way of sharing information, we must move from hard-to-interpret, dense analytics products (built by quants for quants) to products built for the humans in our business. What can we do as analysts, data scientists, and consultants to make it easier for our business partners to see the same patterns we see in the data?

As I wrote in our previous edition of this book, business folks need to grow their data literacy skills and data folks need to grow their storytelling skills. But what about our products themselves – can we get our analytics products closer to what our business friends need for easy interpretation? The Analytics Value Proposition is shown below:

Figure 1: Analytics Value Proposition

We all want to use analytics to drive business value. However, it is easy to mistakenly believe data in a data lake is the source of data-informed decisions and business value. It is not. That skips many steps that rely on the skill of the folks in your business. Instead, a simplified analytics value chain is often more like this:

- An analyst or data scientist gathers data, calculates metrics, and creates analytics
- That same person builds an automated dashboard for their stakeholders
- In a self-service environment, a businessperson looks at the dashboard and creates an

interpretation.

- They use that interpretation to inform or advocate for a decision or solution to a problem.
- If they advocate effectively, their unit of the business acts.
- Ideal state that action drives incremental business value (or reduces costs).

Many business skills can either support or break this chain, including data acumen, analytics skill, visual design skill, data/analytics literacy, communication, persuasion, and skill of implementation. Where is your Analytics Value Chain likely the weakest? Many democratization and self-service launches fall apart right around interpretation. Self-service means you are leaving interpretation (and everything downstream) to your stakeholders. Are you confident in your stakeholders' collective data literacy and analytics interpretation skill? How hard do they have to work to move between the dashboard interface and interpretation right now? The more effort it requires, the less likely they are to spend that effort. Let's close that gap!

At a minimum, we would start with iteration and user research. When your teams build analytics and data science products, do they take the humans who will use it into account? The first thing we'd learn from user research is that different leaders and business units have varying amounts of data literacy and analytics interpretation skill. You'd also learn that people want dashboards that are visually friendly and easy to read. They want as few steps as possible between the dashboard and the decision. In addition, they'll have practical needs and opinions on how the visuals should look and act.

We also try to get the language of the dashboard as close to business language as we can. While some of our main KPIs have a single name (and sometimes multiple calculations), other business metrics may have the same calculation and different names depending on the leader or business unit. [The many-to-many problem is the worst-case scenario: many calculations with many similar-sounding names. Ideally, you'll resolve this as you build out and solidify your metadata.] Metadata means data about data. It includes information about the data in your data lake, data quality, metric names and calculations, and dashboards. Established business definitions for KPIs and metrics are metadata.

The process of creating your organization's metadata is also the creation of your data semantic layer. You are tracking and organizing the meaning of the knowledge your business uses. Knowledge management is one way to pull "tribal" information out of individual employees, organize it, and use it to make navigating your organization's information easier.

Perhaps, as a data scientist, I call one of our metrics "accumulated daily store net sales," in operations, they call it "daily sales," and in finance, they call it "net sales per day." If I show them a dashboard with "accumulated daily store net sales," will they take the time to translate that to net sales? Once they do that, will they have the patience to then look at the metric itself? It's likely they'd rather send me an email asking for daily sales rather than wading through the layers of meaning to interpret the data and worry about whether they did it properly. The closer you can get analytics products to the language of your business, the more likely you are to inform decisions and actions that drive business value. Adding a semantic layer (built out of your welldocumented metadata) translates metrics names for your users and reduces stakeholder effort.

Our goal is to make our dashboard user interfaces as easy to use and as immediately useful as possible. To achieve this goal, we should use lightweight, iterative processes to build our analytics products. These processes need a focus on the folks who will use this dashboard to support datadriven decisions. Frameworks like Agile, CI/CD, human-centered design, and user experience research are helpful here. Well-maintained metadata can become a translator between technical and business terms for your KPIs, metrics, and analytics. That becomes a semantic layer with many benefits to your business, including making it easier for non-technical folks to understand and interpret analytics.

Ultimately, our business stakeholders need to easily interpret our analytics products, connect their interpretation to the next decision they must make, and motivate others to take action. They rightfully demand relevant, intuitive, and easy-to-use interfaces that match their data literacy skills. When we succeed, we are rewarded with great business returns, nimble decisionmaking, high levels of product adoption, advocacy for our products, and greater analytics investment.

Author Biography

Megan Brown is the Director of Starbucks' International Analytics and Data Science team. Their team is responsible for advanced analytics delivery and data science strategy for the Starbucks brand worldwide (outside of North America). Megan considers data literacy and storytelling to be necessary foundations for data-informed decisions and strategy. In addition, Megan was a 2021 Aspen Institute First Mover Fellow focused on creating career pathways to help Starbucks partners grow from stores to analytics roles. Before Starbucks, they were an elementary educator, a research psychologist, and data scientist. They enjoy diving into challenging problems from both tech and business origins. Megan has two kids (ages 2 and 4) and spends their free time writing, dancing, and reading apocalyptic Sci-Fi.

ATSCALE

Chapter 12

A Product Management Approach to Data Monetization

Douglas B. Laney

Data & Analytics Fellow, West Monroe

Central to treating data as an asset, data monetization should align with familiar research and development (R&D) and product management/marketing approaches. Not to oversimplify the many challenges and activities involved in monetizing data, certain basic concepts will reap significant rewards if executed well.

Evolve from Data Project Management to Data Product Management

Although you may already have a data leader, such as a chief data officer (CDO), or an analytics leader, the first step toward data monetization is to designate a team tasked with identifying and pursuing opportunities for and generating demonstrable economic benefits from data assets. They may report to a data and analytics executive, into the enterprise architecture group, a chief digital officer, or perhaps even a business unit head.

Creating a distinct, dedicated data product management role is vital, especially when business and data leaders agree on pursuing direct data monetization by generating revenue or other financial benefits from licensing or exchanging their data. Typically, companies already have a defined approach for managing and marketing products. Analogously, if you are considering licensing data in any form, you need someone whose job is to define and develop the market for the data asset and to productize it.

Finding qualified talent for this role can be difficult. Traditional product managers may have an advantage over other candidates, even without significant knowledge of data and analytics. But why not consider hiring individuals with experience at a data broker such as Experian, Equifax, D&B, IRI, LexisNexis, Nielsen or J.D. Power?

Ideally, the data product manager reports to the CDO (itself, an emerging role for data-savvy organizations) or into a new data product line of business head. This chain of command, askew to the IT organization, underscores that data is a business asset, not an IT asset. A data product manager provides a counterweight to data scientists, who can get seduced and obsessed by intriguing problems that may be tangential to the business objectives.

Speaking of CDOs, Gartner's most recent Chief Data Officer Survey finds that a CDO's success is 3.5 times more likely when they've met data monetization objectives versus only 1.7 times more likely when they've demonstrated return on investment (ROI) from data & analytics investments, and 2.3 times more likely when they've successfully reduced time to market. The reason to hire a dedicated data product manager.

Borrow from the Traditional Product Management Playbook

The data product manager can and should borrow liberally from existing product management disciplines:

- Conceiving and planning new ways to monetize data,
- Identifying or developing information markets among partners and others, and
- Coordinating with IT, marketing, finance, legal, and other product management groups to execute information productization objectives.

Pythian CEO Paul Vallée, the former CEO and current board member of Canada-based Pythian, said company executives spoke about their experience in taking more of a product management approach. They determined a committee approach wasn't getting things done and that the company required a single owner to drive the process: "We needed somebody who understood exactly how the business works. We needed someone who had been with the business a long time and had been involved in establishing our practices. That was what we needed to do in order to break through that inertia and to get rid of the committee for day-to-day decisions. Although a group of stakeholders should always be consulted throughout the project, at the end of the day, one person needs to be a leader."

Similarly, Samir Desai, Chief Digital and Technology Officer at Abercrombie & Fitch, said the key is getting the right individual into the role: "Not everybody is cut out to be an innovator. I think you need to choose someone who understands the business and the technology and who has the right kind or personality fit to play that role."

You May Already Be a Data Product Manager

Many data and analytics professionals believe they've been doing data product management for years without being officially anointed. "The title may or may not matter, depending upon the organization," offered Steve Prokopiou, Data Product & Proposition Lead at First Central. "It's about engaging with the business and delivering what they're looking for by acting as a translator, asking sensible, structured questions about data usage and benefit. And perhaps adopting the language of product management in doing so." Prokopiou also suggests that having the formal moniker might give one a mandate to get involved earlier during requirements specification, rather than waiting for incomplete or difficult to translate requirements to land on their desk.

"A data product manager does need to be entrepreneurial but doesn't necessarily have to have a product management background," says Lillian Pierson, who calls herself a data product manager within her own firm, Data Mania, a creator of educational content. She believes that treating almost everything you produce as an actual product compels you to take a more disciplined approach. Pierson advises that a data product manager should have a multidisciplinary skillset, including:

- An understanding of analytics or data science and data strategy
- Knowledge of how systems and processes operate
- Able to anticipate what technologies work well together
- Knowing how to design features and functions
- Experience with performing market or stakeholder research
- And a penchant for people.

Refine the Data Product Vision by Working Backward

Legendary golfer Greg Norman says he plays each hole backward in his mind. “As I step onto the tee, my mind goes to the green. Before I decide which club to hit or how to play my tee shot, I want to know the exact position of the flag - once I know that, I play the hole backward in my mind.” Similarly, as a data product manager, it helps to start with a vision of what you want to produce. This is the approach companies like Amazon take.

Ian McAllister, the former director of Amazon Day, says that working backward begins by “[trying] to work backward from the customer, rather than starting with an idea for a product and trying to bolt customers onto it.” For each new initiative, a product manager writes an internal press release announcing a finished product. “Internal press releases center around the customer problem, how current solutions (internal or external) fail, and how the new product will blow away existing solutions,” commented McAllister. “If the benefits listed don’t sound very interesting or exciting to customers, then perhaps they’re not, and shouldn’t be built.” And if not, then the product manager should continue revising the press release until they’ve come up with something better.

It may seem to be a lot of work for an idea that may never see the light of day. But as McAllister explains, “Iterating on a press release is a lot less expensive than iterating on the product itself... and quicker!”

A Meta Analysis of Organizations Driving Value from Data

Certain patterns emerge across the spectrum of 100s of real-world stories I have compiled of organizations squeezing value from data assets. These meta tendencies in-and-of themselves are instructive, e.g.:

- An organization need not be big to leverage big data, and an organization doesn’t have to have big data to do big things with it.
- Many organizations make use of data beyond their own four walls, including syndicated data, open data, social media, web content, or data from partners.
- Many organizations find valuable uses for data they’ve collected, used for a single purpose, and forgotten about or archived, i.e. their “dark data.”
- Many stories involve the use and mining of unstructured content, not just structured customer or transaction data.

- Most use cases cite multiple benefits, sometimes even multiple measured benefits.
- High-value implementations do not necessarily require enterprise-wide data warehouses or data lakes but are targeted and vocational, focusing on a single problem or opportunity with a limited set of data.
- Most use cases focus on improving revenue or margin, but the cleverest ones and those with the most ancillary benefits focus on other drivers like customer experience or agility.
- Almost none of the examples involve hindsight-oriented dashboards or scorecards. Real value comes from diagnostic, predictive and prescriptive solutions—or new business models altogether.
- Most implementations involve integration, not just data integration, but integrating the analytic output or data streams directly into business processes or operational systems. Directing analytics at eyeballs is so last century and generates little ROIA (return on information assets).
- Most stories are driven by business leaders, not IT leaders.
- Productizing and licensing data, creating digital solutions, and using data as a form of collateral are emerging forms of data monetization, fast introducing new value streams for businesses.

Also, I have noticed through my research and consulting those certain data are more monetizable than other data. When working with clients on monetization initiatives, I typically start with identifying data with as many of these eight monetizable characteristics as possible:

- Data that originates with your organization or for which you have some claim of provenance and control
- Data proprietary in nature because others do not have it or data like it
- Data of a general or accepted context, rather than data that may not be meaningful outside your organization or industry
- Data that is secure. If it is easily accessed or hacked can result in a black market for it which may cannibalize your own efforts to monetize it
- Data not restricted in the way you can collect or deploy it (e.g., Privacy regulations may be an inhibitor, but not preventative if you're clever enough.)
- Data sufficiently accurate and precise for the intended purpose(s)
- Data that is complete because it represents a significant slice of the known universe of such activities or entities and for which any record is complete (e.g., without missing fields)

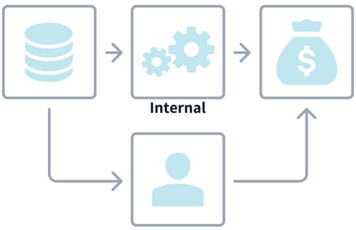
- ▲ Data that is and will be available for the foreseeable future or for which alternatives are available inside your organization

Within these data monetizable sources identified, then you can dig deeper to identify what Rockwell Automation's enterprise data leader, Karan Dhawal, calls "monetizable data elements" (MDEs).

Data Monetization Patterns

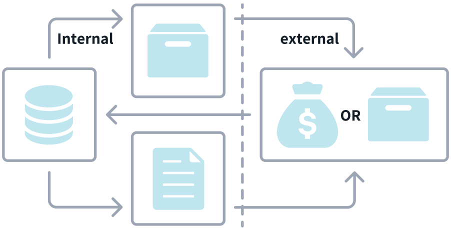
As I discussed in Infonomics, monetizing data can take a variety of forms. Several patterns have emerged when assessing the hundreds of stories I have collected. Data monetization isn't about selling data. More broadly and concisely, I define data monetization as generating new, measurable value streams from data assets. This definition opens a whole world of possibilities, including using data that's not necessarily your own as discussed above and generating a wide variety of economic or societal benefits—albeit ones that can be quantified.

Data monetization falls into two broad categories: Indirect Data Monetization and Direct Data Monetization. Indirect data monetization involves solutions focused more so on internal business processes that generate measurable returns, while direct data monetization involves externalizing data in return for some commercial consideration. See Figure 2.



Indirect Data Monetization

- Improving process performance or effectiveness
- Reducing risk / improving compliance
- Developing new products or markets
- Building and solidifying partner relationships
- Assetizing data via special corporate structure
- Publishing branded indices to promote data products/services



Direct Data Monetization

- Bartering/trading with data for non-cash commercial considerations
- Enhancing products or services with data
- Licensing raw data through brokers or data markets
- Selling insights, analyses and reports
- "Inverted" datal monetization via referral arrangements
- Collateralizing data to securitize loans

Figure 1 - Data Monetization Methods

Productizing and Packaging Data

But why get fixated on just one data monetization method? Data is what economists would call a non-rivalrous, non-depleting, progenitive resource—one that can be used multiple ways simultaneously, doesn't get "used-up" when consumed, and that typically generates more data when it's used. No, data is not "the new oil." Oil has none of these incredible economic characteristics. And winning businesses are the ones taking full advantage of data's unique qualities. In fact, according to infonomics research I performed, businesses demonstrating datasavvy behavior have a 200% greater market-to-book value than the market average, and for data product companies (i.e., those that primarily sell data or data derivatives), this ratio is 300% higher. There's no doubt that investors positively love data-driven companies.

Because of these unique aspects of data, it can be packaged in numerous ways. (See Figure 3.) These range from publishing raw data, to rolling it up into analyses, to creating custom and integration solutions. As with other raw materials, however, the more you process them, the more expensive and exclusive they become. Consider wheat. Wheat itself is quite the commodity—inexpensive but with massive global demand. Then it can be processed into different kinds of flour with specific but still a wide range of uses, then into breads which are more expensive and with a lesser range of uses, and finally integrated with other resources into sandwiches or cakes with more premium price tags and smaller, more localized markets.

Similarly, each subsequent step of processing your data assets can result in them delivering more specific and greater value, yet to a smaller market (user base).

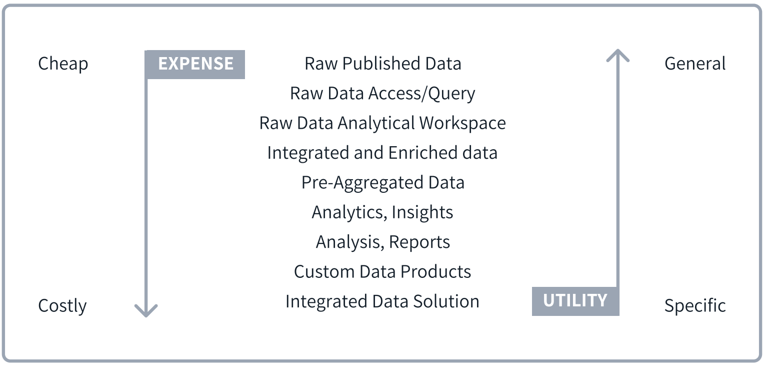


Figure 2 - Data Packaging Options

Your Data Monetization Journey

Although data may have incredible characteristics that other resources and assets like physical materials, financial assets, human capital, and even other intangible asset do not, we need not consider it anomalous with data monetization initiatives. Instead, I recommend borrowing from well-honed product management approaches to lay out your data monetization function. (See figure.

Central to treating data as an asset, data monetization should align with familiar research and development (R&D) and product management/marketing approaches. Not to oversimplify the many challenges and activities involved in monetizing data, certain basic concepts will reap significant rewards if executed well.

Phase I

1. Generate and refine ideas for innovating with data assets
2. Assess and prioritize data monetization ideas based on a spectrum of feasibility characteristics

Phase II

3. Develop and test markets, identify potential buyers/users of data products, and assess data packaging and licensing options
4. Identify, curate and prepare internal and external data assets
5. Specify technical, data, analytics, governance and organizational requirements
6. Define support, maintenance, monitoring and reporting requirements

Phase III

7. Architect and engineer the infrastructure, data products
8. Implement and test new data products and required APIs, search and other components
9. Introduce and support the new data products
10. Provide program-, project- and/or data product management

Figure 3 - Data Monetization Roadmap

The details, specific methods and tools behind each phase and steps are beyond the scope of this book but are available via the courses and workshops I teach and the services my colleagues at West Monroe and I provide

Overcoming Data Monetization Myths and Roadblocks

Along the road to conceiving and creating new data-driven value streams for your organization, you will invariably bump up against many data monetization myths and potential showstoppers. Be prepared for them, or work with someone who can help your organization get better educated and prepared via instruction and workshops, a data literacy program, and/or a change management initiative.

Many myths about data monetization persist today and will continue to do so until business leaders, legal counsels, CFOs and controllers, company executives and boards become better aware and educated on the possibilities and realities of data.

Myth	Reality
Data must be sold to be monetized.	Data can be used indirectly or internally to generate measurable economic benefits.
Data monetization requires an exchange of cash.	Data can be exchanged with others for goods, services or favorable commercial terms.
Data monetization only involves your own data.	Integrating others' data, even freely available data sources, with your own improves its utility and value. You can even collect and productize freely available data itself.
One can only monetize raw data.	Data can be packaged and productized in a wide variety of ways.
Only data that's used has any value.	This is inconsistent with how other assets are valued. Simply having and holding data can have certain economic benefits.
One must be in the data business to monetize information.	At least 30% of non-data-product organizations are monetizing their data today.
Our data is specific to us and no value to others.	Your partners and suppliers, especially, would beg to differ.
It's best for us to share our information with our suppliers and partners.	This becomes more expensive than it's worth. Set the bar at what they get for free versus what you offer for a premium. And consider your extended business ecosystem, e.g., your partner's suppliers, or customers' partners.
Due to privacy regulations, we cannot monetize our customer data.	You can't sell or even share your customer data with others, but you can sell others' offerings to your customers for a referral fee or commission. (Think: Facebook)

Even once you've moved your organization beyond these myths, you'll still find a range or roadblocks along your journey that must be moved or navigated around, including:

- Core business priorities making data monetization a low(er) priority
- Mental blocks due to data not (yet) being a recognized asset because of archaic accounting standards
- Legal, regulatory, or ethical roadblocks (perceived and real)
- Already giving away data for free
- Insufficient foundational capabilities such as data integration, masterdata management, analytics, storage/computing capacity
- Insufficient data quality (accuracy, completeness, timeliness, integrity, etc.)
- No culture of R&D, especially with data
- Lack of organizational experience and skills

Final Thoughts

Economists have long touted the four factors of production as the key components of industry. They are Land, Labor, Capital, and Entrepreneurship. I contend that not only does (should) data qualify as an actual balance sheet asset, but it has ascended to become a fifth factor of production. Data become a production factor but often is a substitute for other production factors. Data and algorithms have replaced labor, machines and shrunk on-hand inventory needs. And in doing so, data has reduced the real estate footprint required to do business, as have digital solutions based on data and analytics.

Yet, even with all this evidence that data meets the criteria of an asset, property, and a production factor, too many businesses and business leaders treat data to create pretty pie charts and dashing dashboards...or worse, as a mere business by-product. And they continue to conflate data and technology. These enterprises are getting lapped by those treating data as an asset and capitalizing on its unique economic qualities. The art of the possible with data has no limit. Or it is limited only by the imagination of your business leaders and the ability of your organization to take a product management approach to become data driven.

Author Biography

Doug Laney is the Data & Analytics Strategy Innovation Fellow at West Monroe, where he consults with business, data, and analytics leaders on developing new value streams from their data assets. He originated the field of infonomics and authored the best-selling book, “Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage,” Doug is a three-time Gartner annual thought leadership award recipient, co-chairs the annual MITCDOIQ Symposium, and is also a visiting professor at the University of Illinois Gies College of Business and Carnegie Mellon University’s Heinz College.

ATSCALE

Chapter 13

Achieving Executive Engagement in AI Programs

marketing sciences

John K. Thompson

Author of Building Analytics Teams, Data for All

Working across an organization is a critical element of the success of an analytics team. Engaging with and gaining the support of executives is a requirement for success. We will delve into the details of capturing the imagination, attention, and support of executives across the organization. We will examine the relevant factors that will guide and shape your communication and engagement with your direct reporting line executives and those executives who own and manage related corporate functions.

Executives and senior managers are important to the ongoing success of any company, and they have a significant role to play in the early and ongoing success of the advanced analytics and artificial intelligence teams and your personal success. Working with executives to support their needs for them to support your needs is a fact of life. Embrace their needs, support and feed what they need, and you and your team will have a much easier time navigating the organization and obtaining the support and funding required to do the work you love.

Executives may not understand what you do or what motivates you, but that doesn't matter. They do not need to understand you, your team and the work you do. You need to illustrate to them in terms they can understand that you and your team are a good investment that will deliver valuable results and impact at a faster and higher rate than other investments. It is that simple. Do this, and your life and work will be much easier. While I do not advocate spending too much of your time selling up, you must engage in selective and strategic selling and sales activities to convince the upper echelons of the firm to secure and maintain funding for your personal efforts and to support the efforts of your advanced analytics and artificial intelligence teams.

We will examine how to develop and deliver a clear and compelling message to the executive team that will engage their imagination and pull them into the analytics journey. We will also discuss how to position the types of work you and your team can and will do and how to handle the discussion about which part of the analytics workload can be outsourced to external parties. We will conclude with information on how the daily operations of your team can be presented to the executive level of the organization. Remember that an organization is a collection of competing strategies, projects and functional areas. You need sponsorship and money, and you need to ensure that the executive team or at least your sponsor on the executive team continues to support and fund the efforts of your team.

You are not the only game in town

Executives have many teams reporting up through the organizational structures, and those teams have numerous initiatives underway, and those initiatives contain a diverse set of projects. Just because you are living and breathing the portfolio of analytics projects doesn't mean that your senior management team and sponsoring executive have more than a cursory understanding of the projects and initiatives you and your team are engaged in.

This organizational separation is part of the reason we have been discussing the need for the analytics team to be connected to a larger overall process of organizational improvement and transformation. Executives can only keep a certain amount of complexity and breadth of activity in mind at any one time. When they see that your efforts, and the efforts of your team, are supporting strategic initiatives, then you are aligned with their thinking, and you will have their support

I took years to get my head wrapped around the fact that not only did they not know what I was doing, but they didn't understand how it was being done and, sometimes, were dubious about the probability of success and the scale of the impact.

Let me help you not make the same error in judgement. It is not their job to understand what you and your team are doing. It is their job to lead the company and make strategic decisions, and to be accountable for those decisions.

It is your job to ensure that they understand how the efforts of you and your team align with where they are attempting to take the company. You must work harder at keeping them informed than you would believe, and to you, it may seem like a waste of time, and you may also think that it is not "work" since it is not building something of value. Well, it is an important, no, it is a critical portion of your primary work duties. You need to support your team in securing the support of the executive team or at least your executive sponsor. Without their support, you will not have a team or the resources to do the work you love.

In this respect, executives are like everyone else; they need to have messages put into a context and language they understand and connect with. You need to put those messages in a language they can hear and comprehend quickly. I know that it is exciting to tell people how hard you worked and how smart you and your team are, but that will not help you in this task. It actually detracts from your effectiveness.

Be assured that the organization and the executives hired you and your team because you are intelligent, can solve complex multifaceted problems, build incredible models and improve the organization in a multitude of ways. You do not need to prove that again and again. Your intelligence and ingenuity are the price of entry, and you have paid that price. The executives and senior managers who have met you and know you respect your abilities and talents. This is without question, and you are viewed as a valued member of the organization.

Your focus and efforts, at least, need to be on refining and honing your ability to connect with senior managers and executives on their level in a voice they can hear. Speak their language, and your success will come much more easily.

Know what to say

Executive communications are easy to understand but can be challenging to navigate. Mostly, executives want simple, fast, easy-to-understand answers that convey that the work you are undertaking has a significant positive impact realized in the corporate or financial reporting period they have on their mind at the time of your discussion. What does that mean?

Your answers should be short and convey the strategic area in which you and your team are working in and that the quantified positive impact will be above the financial or operational threshold they care about in the next reporting period, which is typically a calendar quarter or possibly in time for the next board meeting, or both.

I was in a meeting recently where an executive told a team leader he was not interested in the project if it did not deliver a positive $20 million impact in the next 16 weeks or in the current quarter. Within the next month, the project was sidelined and eventually shut down. This turn of event was devastating for the team leader and as we could tell from the facial expressions, unexpected. This person had joined the firm within the previous year with a publicly stated purpose and mandate to undertake the initiative to establish this new corporate function. Not only was this a seemingly great opportunity, but it was also a passion project for the team leader. The team leader had moved across the United States and worked diligently on the project.

You can argue for failure on the team leader's part to gather the required data from the executive to understand what the level of significance had to gain executive sponsorship, but perhaps the fault lies primarily with the organization that hired the team leader. The organizational structure had not changed significantly in over 5 years. Hence all the senior managers knew what projects raised to the level of interest for this C-level executive. But no one told the team leader, or the team leader was not astute enough to discern this. This person had taken a new job, moved across the country, threw significant energy into the project, and used much time and energy to develop a plan for this new function, only to have it eliminated in a few minutes. Within four months of this meeting, the team leader resigned and left the company. Truly a waste of time, resources, employee engagement and motivation.

The message was delivered bluntly, typical of some corporate executives, but it was callous. The tone and tenor of the message were unnecessary and hurtful. Corporate executives are not known for being compassionate messengers. I doubt we will see a dramatic change in this communication dynamic in the short term. It will change over time, and it has changed a great deal in the past 3 decades, but be ready for harsh criticism of your efforts and projects.

You need to know the expectations from the executives and senior managers you will engage with. If the expectations are in line with the scenario above, you need to be certain that you, and your team, can deliver at the required scale and speed. If you cannot, for whatever reason, find a different functional area to engage with or project to undertake.

Know what to say it

People ask me, with fair regularity, why analytics teams, analytics projects and all things related to analytics are unique. As we have described along the way in our discussion, analytics initiatives rely on a mind-boggling array of moving and interrelated parts - technologies, corporate strategy, functional tactics, data, external suppliers, executives, senior managers and the analytics team. This constantly moving collection of elements is complex to understand, manage and drive in a coherent direction.

It is even more complex for people with little to no understanding of the components, let alone the whole of the ecosystem. You need to assume that the executives you are engaging with have little to no knowledge of any part of this environment, and they have little interest in the topic. They may have a strong interest in the possible outcome(s) and changes that can be driven through analytics, but it may be dangerous to assume anything more than a passing fancy for the topic.

Complexity combined with little to no knowledge and a little interest mixed with a healthy dose of executive ego makes for a volatile mix. Never underestimate the cycle of - lack of understanding/interest, followed by confusion, quickly leading to embarrassment, concluding with lashing out and condemnation. As we all know and have experienced, it is a near-universal reaction – if I am confused and you are presenting or representing something that I cannot understand, and the choice is between me being unable to understand the topic or you being wrong or incapable, then the second option is an easy choice. Executives are exquisite in making this snap judgement in a matter of seconds. Being embarrassed or confused is kryptonite for the vast majority of executives.

To be clear, you are probably much smarter than the vast majority of executives you will encounter. If you engage them in a battle of wits, you will win, but you will never receive the funding and support you need to succeed. Be clear, concise and on message. If you can say it in 45 seconds, that is great; 30 seconds is even better. You can do it. Think of the famous quote from Blaise Pascal, "If I Had More Time, I Would Have Written a Shorter Letter."1 Being concise can be thought of as a gift that some have and others do not. That is not true. Being concise is the product of focus, effort and refinement of what you want to communicate. Work on refining this skill, it will pay off.

Shape and Direct the Narrative

Early in my career, I attended an annual sales meeting for Metaphor Computer Systems at the Silverado resort in Northern California. There was a senior manager, Jay McGrath, leading the discussion. He was addressing a group of system engineers, and I was one of them. I was approximately two years into my consulting career. I had completed my MBA, and I thought I knew it all, forward and backward. Jay ended a section of the discussion with the statement, "Telling is not selling." I was perplexed by this statement. I assumed that most people we were talking with were smart, informed, reasonable people, and we had only to tell them of the benefits and value of our proposition, and they could make the best decision possible.

All of that is true, but it leaves much unsaid or undone when arriving at the destination you desire. When presenting to executives, you need to show them what is good and explain why it is good. I was recently in a meeting reviewing a presentation illustrating the initial results of a predictive application. The first pass of the model produced an error that was less than

4% when compared to the actual operational results; we were ecstatic. We knew that the model would improve and produce even lower error rates, but when I spoke with one of the operational executives in an impromptu hallway conversation, he looked at me with curiosity. I asked what he thought. His response, “Is less than 4% good?” I smiled and said, “Yes, very good, and it will get better.” He smiled in response and said, “That sounds great”. Do not assume that executives and senior managers understand what you are talking about. If it is good news, tell them, and show them why, and frame the story in the way you want to define it.

Know before you go

Cultivating, assessing and obtaining executive support starts even before you join the organization. When interviewing for a new role and when new executives join the organization with a direct or indirect impact on your team size, funding and plans, you need to evaluate the executives and senior managers to understand their history, perspective and experience with analytics teams.

I have heard executives express these views:

- We need to see significant changes resulting from the analytics teams in the primary operating functions in a matter of weeks
- The analytics team should be staffed and running efficiently by the next quarter
- Our company has never been good at innovation, the analytics team will change that
- We will have the advanced analytics and artificial intelligence team build dashboards and reports in their “spare time”
- Data and analytics are not core, we can outsource all analytics work
- This is so simple; we can have two interns and a dog do it
- Listening to the functional teams is a waste of time, the analytics teams need to tell them what to do. The executive that said this, thinks he is Steve Jobs…you, sir, are no Steve Jobs…
- Why should analytical models be updated? If they are right when implemented, why aren’t they right all the time in the future
- We have numerous smart people in the organization, including scientists, doctors, and PhDs: have them do the data science work
- The advanced analytics team looks like a technology team, have them report into the technology function

- Budget? The advanced analytics team needs a budget, why?

These viewpoints indicate that your role, and that of your team, is not well understood, and it is more than likely, that neither you nor the team will be valued and will not be funded.

If you are considering joining an organization where the executives say the things I have listed above, you are probably talking with them because someone at the executive level is advocating for advanced analytics, and that is good, but if other executives hold the views expressed above, it is only a matter of time before the sponsor and the other executives will be infected with and succumb to the same views.

It can go the other way too, but I am a firm believer in planning for the downside. The downside is easy to describe. The executive sponsor advocating for the funding and hiring of an analytics team will fight for the funding and hiring of a leader and a team. The executive sponsor wins the ability to fund a small team. The analytics team leader and the analytics team work diligently to engage and illustrate success. The skeptical and unconvinced executives work consistently to undermine the efforts of the analytics team. The executive sponsor losses the will to continue to fight with the other executives. The funding is reduced for the advanced analytics and artificial intelligence team, the best team members leave, the leader leaves, and the entire effort becomes part of the company history, a footnote of a failed effort. This failure could have been easily avoided if the analytics leader had convinced and conveyed the project scope and efforts to the executive team.

What are you hoping to accomplish?

I have spoken with executives who believe that all data and analytics projects and efforts can be outsourced. That is possible, I suppose. I have never seen this effort deliver strategic impact and value, but that does not mean it cannot happen. One problem of outsourcing anything, not only analytics, but any project, is that you need to know what you want to have in the end before you start. How can you expect an external vendor to deliver what you want when you want it at the expected cost when you don't know what you want or how to achieve it? Sounds straightforward, right? You would be surprised at how many executives, firms and teams have little to no idea of what they want to do.

This is an acute problem in analytics. With the land grab mentality that some firms have, it is challenging to define what they want. Remember the statement, "We need some machine learning." That is an actual quote from an operating executive in a Consumer-Packaged Goods company; shockingly ill-informed.

One challenge in defining what you want is the multidimensional nature of the problems we seek to solve or challenges we seek to overcome with analytics. A related challenge is the

Gordian knot of whole problems or related factors that are bound. Often the challenges that the analytics team attempts to undertake were considered impossible to solve. Decomposing these interrelated elements into their source elements or first principles enables the analytics teams to formulate solutions that are not obvious at first glance through a creative process.

In one project, we started with the stated goal of having a better market-based and competitive forecast. That was too broad of a project mandate. We re-scoped to address the demand side of the market. We then found there was no reliable, easily accessible core data on the activities of many of the smaller competitors. We then decided that we would focus on a demand forecast for the firm and the top 3 competitors. We built a solid and accurate forecasting model for the described entities and over time, built on that platform to move toward the original objective.

The definition of objectives and goals has always been challenging when collaborating with subject matter experts from the functional areas of any business. The process described above works well, but it can be fraught in the effort to bring together the analytics, functional and technology teams to arrive at a consensus regarding the problem to be solved. And once you have reached a consensus on the objectives, then the discussion turns to the ability to actually formulate a workable and practical solution based on the technology, data and analytical skills available to the organization. Just because you can describe the problem doesn't mean you can solve it.

This raises an issue you need to be aware of and can address in a concise and cogent manner. Executives and senior managers that embody the characteristics, mental constructs, personal idiosyncrasies and more also have an issue in that since they do not understand the data, analytical approaches and/or technologies, they do not understand what is possible. For some executives and senior managers, analytics is akin to alchemy or magic.

Many executives have no basis for discerning the possible from the impossible. You need to be aware of and know the practical and possible and be able to discern between the possible and the impossible. You need to possess the wherewithal and composure to explain what can and cannot be done. You want to respond in a tactful and respect manner to what you know to be an impossible request, and you need to avoid embarrassing the source of the request or questions, but you must also not set yourself and your team up for failure by agreeing to work on an intractable or impossible problem or challenge.

Outsourcing

We are moving away from the "not invented here" model where organizations believe that they need to build everything themselves, but there are still pockets of this view. It is the opposite view of the situation described above, where some executives think that everything can be outsourced. As with most things in life, the ends of the continuum are the extreme cases and are typically found in a smaller number of organizations.

Usually, you can outsource portions of the analytical process. Let's break it down for clarity:

- **Easy to outsource**: You can outsource these activities with the lowest risk. Data acquisition, data integration, data profiling, data loading, and data visualization steps are the easiest to outsource and can lower the overall cost of operations.
- **More difficult to outsource and success is difficult to achieve**: Feature engineering can be executed by an outside vendor if the company you are outsourcing to has not only technical skills but also skills in the business or functional domain related to the analytical models to be produced, but this combination of skills is scarce in a service provider.
- **Hard to outsource and hard to achieve repeatable success**: The modelling phase. I have had little success and have seen even less success when other companies attempt to outsource the model-building function. Numerous firms present and represent that they can build models, and I am sure that they can build models, but the question is, are they effective models, are the models easy to update and do they contain a solid foundation that can be extended to include a broader operational domain as you learn more and more about the data, operations and strategic direction planned by the executive team. This can work if the domain you are focused on is one that is generic, well known and widely applicable. The need to build an analytical or predictive system is probably driven by the need that all other relevant competitors have this system, and the company needs to implement the system to simply remain competitive. In those cases, outsourcing may work.
- **Nearly impossible to outsource**: The integration of predictive models into production processes is unique and individualistic and outsourcing nearly always fails or is more expensive and time-consuming than staffing and undertaking the effort internally. suppose that if a vendor is running the production process in a cloud environment outsourced to the external vendor and the vendor is an expert in managing and extending the process, this is possible, but that is a specialized case. I have not seen success in outsourcing the modification of proprietary processes managing a factory or a supply chain or an operating room. Typically, the modification and changing of these processes are unique and require the expertise of multiple groups within an organization. Also falling into this category is model management. It takes significant skill and expertise to know when to change models, and it takes even more skill and knowledge to know which model is the best to move into production next. Outsourcing these tasks are done at your peril.

These are the steps to consider outsourcing – data acquisition, data loading, data profiling, data integration, and data visualization. The steps that carry more risk if outsourced are - model selection, modelling, model tuning, hypothesis definition/testing, pilot/prototype/

production, and model management. Another way to make this point is to think of it from the core vs. context framework. Areas that deliver competitive differentiation for a specific company would not work well for an outsourcer as these are core business factors. Core business processes are typically unique and cannot be generalized across many organizations. Rather, areas aligned with a common commodity function or context are opportunities for working with an outsourced partner. Corporate functions (i.e., HR, Legal, etc.) and analytics related to those functions are a good place to look for this partnership. Unless you're doing something special with those functions, pushing it out to a group that can do it cheaper would free up resources for focusing on core functionality. To sum this up concisely, when asked to outsource a portion of the analytics process, you need to know what you can easily outsource without compromising quality.

Elephants and squirrels

Executives are like other people in many respects; in other respects, not so much, but remember that executives are people too.

One characteristic of people that inhabit the executive ranks I find interesting is that they are typically split into two groups on the ability to retain and remember facts. Perhaps it is my selective memory at play here, but it seems from my decades of experience that executives either are like elephants because they remember everything. They remember where the conversation was had, what was said, who was there, the weather outside, and the stains on the carpet, all in great detail. Or, they are like squirrels in that they can't remember much. They say the same things repeatedly and repetitively. They ask people to do the same or similar projects within days or weeks of each request. They seem surprised when most people in the room understand immediately or seem to have the same idea. Well, it's easy for people to come to an agreement when you heard the same thing a week ago. It is intriguing.

I am like an elephant. I remember conversations for years and can play them back almost verbatim to what was originally said. Funnily enough, I have trouble remembering names. It is an odd memory I have, but it works well in recounting what was said and what was agreed to.

Elephants are easy to work with. You come to an agreement, and you work toward that agreement. Squirrels are harder. You never know what they remembered or how they remembered it or how much they will change the memory. Be alert, be humble and ask lots of questions of the squirrels. Typically, they remember with enough accuracy when they have been prompted and probed enough. Ask lots of questions, and they will remember bits and pieces of the previous conversations the more you talk and move them in the right direction. The squirrels do not like to be reminded that they cannot remember past conversations. Ask the executives questions and bring them along slowly. Do not say, "We agreed to x, y and z, last week, don't you remember?" Not a winning formula with the squirrels.

Programs or Projects

Usually, organizations and in the minds of most executives, you will find an interesting and subtle delineation between short-term expediency and long-term value. Let me explain. In my initial meeting with executives, I let them talk extensively and outline their vision and hope of what data and analytics can do for them and their organization and operation. Once they are finished with their description, and after I recount to them what they said to ensure that I have it right, I describe the difference between a project and a program.

This is what I say...

A project is time bounded, typically relatively short in duration (maybe a month or two) and produces a defined result. A result might be an indicator or an observation. Usually embodied in a single metric in time or over time. The result is a one-time measurement. They may want to know who the most profitable group of customers are, or the region of the world that is the most attractive to launch a product, or the optimum location to open a new factory or all the individuals in a population that may have a condition. These projects can be done reasonably quickly and produce a one-time result.

A program is an ongoing effort and focuses on building an environment that includes automated data feeds, data integration, data quality, analytical application(s), analytical models and is integrated into the production systems and processes to produce continuous analytical outputs used to deliver continuous improvement. This may take six months, or it could take multiple years, depending on the scope and scale of the objectives of the company and the ambitions of the executive.

More often than not, the executives choose the project approach. That is fine. The analytics team collaborates with the staff in the functional business areas, working with the topic experts to produce the result desired from the project.

If the project succeeds, and it should be, the analytics team planned it for success, within 4 to 6 months, it is common for that same executive to set up a meeting to discuss the program approach. In that discussion, the program and objectives are explored. We discuss the time to results, the effort, and the ballpark estimates of the magnitude of improvement that can be delivered. Discuss the magnitude of impact in operational, efficiency, and monetary terms. Money is a good way to focus the dialog. I have used projects to gain entry to organizations and executives who are not well versed in data and analytics. The project/program approach may take longer, but it produces greater success and stronger relationships.

Celebrating Learning (some call it failure)

You have undoubtedly heard, and have been part of, discussions about failing fast or the corollary expression, succeeding quickly. You and your team will undertake projects that fail. This is a fact and one you need to understand and manage. You need to manage the process of failure for your team and the reputation of the group, and your personal reputation. How failure is handled, tolerated or celebrated depends on the organizational culture, and culture varies widely.

In one of my previous roles, failure was expected and embraced. There was a widely known process for failing. Failing in a project, failing to meet the sales numbers, failure to complete any process or project within the expected timeframe, or not achieving the stated objectives or reaching none of the relevant project or process goals was expected.

The steps in the process were and are:

- Acknowledgment of the failure
- Description of the failure
- Description of the chance of recovery or remediation
- Discussion of the best path forward
- Decision on whether to regroup and work toward obtaining the original objective or to move onto the next challenge
- Back to action

This process was well known. People coming in from the outside of the organization either could understand and execute the process, or they could not. In many organizations, failure is followed by condemnation or rejection.

At first, when I encountered the structured and overt process for managing failure, it was uncomfortable and unfamiliar. Typically, in organizations, failure was understood but rarely acknowledged and, even rare still, discussed openly and constructively. Once I understood the process and how the organization managed it, I enjoyed having a failure mode or process. The process was overt and known. Everyone fails; if they are trying to push the envelope, they encounter failure more than others. The real question is how you deal with failure, how you recover and how the organization around you and your team support you in that process.

If your organization does not have a process for failure, I suggest you set one up in your team and use it with your management. It is refreshing. No one who sees and experiences the process is disappointed to see you and your team acting in the manner and working the process. Try it; you will like the results, and so will your team and your management.

Summary

This chapter has been about engaging with the broader organization in a constructive and productive manner with a focus on gaining the attention, understanding and support of the executive team. The executive team is critical to the success of you and your team. The executive level in every organization is accountable and responsible for the overall strategic direction, operational health, and long-term viability of the organization. Working with the executive level team is unique and requires that you adopt and exhibit a certain – attitude, style, stance, tone and tenor in your actions and in your written, verbal communications and presentations.

In outlining how you and your team can and should engage with the executive team, this chapter has provided a guide to productive engagement with executives as a group and as individuals. As with any skill, you will improve with time and practice. You will be more comfortable in this mode of operating the more you engage with the executive level.

One last piece of advice in entering into the executive suite, go in with confidence and go big. Executives, mostly, like big ideas, and they want to hear about ideas that will change the course of the company. Go big, plan for, and present, initiatives, and programs that will require risk, investment, and time. By going big and presenting a vision, you will have their attention, and, in many cases, they will fund and support your plans.

References

1. If I Had More Time, I Would Have Written a Shorter Letter, Blaise Pascal, "Lettres Provinciales", 1657, https://quoteinvestigator.com/2012/04/28/shorter-letter
2. Gordian Knot, https://en.wikipedia.org/wiki/Gordian_Knot

Author Biography

John is an international technology executive with over 35 years of experience in the fields of data, advanced analytics and artificial intelligence (AI). John has been responsible for the advanced global analytics and AI function at a leading biopharmaceutical company, where he led a team that developed and deployed over 25 analytical applications in 3 years. John was an Executive Partner at Gartner, where he was a management consultant to market-leading companies in the areas of digital transformation, data monetization and advanced analytics. Before Gartner, John was responsible for the advanced analytics business unit of the Dell Software Group. John is the author of the best-selling book – Analytics Teams: Leveraging analytics and artificial intelligence for business improvement. The book was published in June

2020 and outlined how to hire and manage high-performance advanced analytics teams. The book outlines how to engage with executives and senior managers. How to select and undertake analytics projects that change and improve how a business operates. John is co-author of the bestselling book – Analytics: How to win with Intelligence, which debuted on Amazon as the #1 new book in Analytics in 2017. Analytics is a book that guides non-technical executives through the journey of creating an analytics function, funding initiatives, and driving change in business operations through data and applied analytical applications.

Chapter 14

Modern Data Management and Engineering

Ujjwal Goel

Director, Data Architecture and Data Engineering, Loblaw

Data management is ingesting, storing, organizing, and maintaining the data created and collected by the organization. Data management was/is largely driven by IT professionals to solve inaccurate or inadequate data issues.

In the 1960s, mainframe-based databases became available, followed by relational databases in 1970s, Data Warehouse was conceived in the late 1980s with early adaptors in the 1990s. These data warehouses worked well and customers are still using them to date. Hadoop systems, along with Spark, hive became available in mid-2000s, and a range of NoSQL databases also started during the same time frame, which promised to solve big data issues, and this gave birth to Data Lake in the late 2000s. Now there were a lot of challenges with big data lakes due to the nature of no relational databases, which eventually started the adaption of cloud data platforms in the mid-2010s, which is a mix of relational and non-relational databases.

Now, with the advent of technology like Cloud, data lakes, Lakehouse and Data management paradigms like Data Mesh, Data Fabric, Hub & Spoke model. The data space is buzzing like never, and although all the technology and paradigms are there, still organizations are confused than ever to decide which is the best data management strategy to go with. Basically, the best data management is the one that best suits your company culture and data literacy maturity level. With this article, I am trying to break these technologies and paradigms and try to tailor-make the best of what is available as per any organization data maturity level (average).

Let's start with Data warehousing. Data warehousing is continuously evolving, as organizations are collecting and analyzing large amounts of disparate and diverse data, they are continuously looking to modernize their data warehousing environments. This often includes converging the data warehouse and data lake, centralizing data in cloud environments or unifying datasets. However, succeeding with modern data management is not just about the technology, it included new organization models, new technology paradigms and new ways of governance. When you talk about data warehousing, there are two approaches for getting value out of data: top-Down (Data Warehouse) & Bottoms-up (Data Lake)

In a high-level simplistic form:

- Use Data Warehouse – if you know your data and you know your questions
- Use Data Lake – if you do not know your questions and you do not know what all data is required for it

Both data warehouse and data lake have their use cases, and the best approach is to use them together.

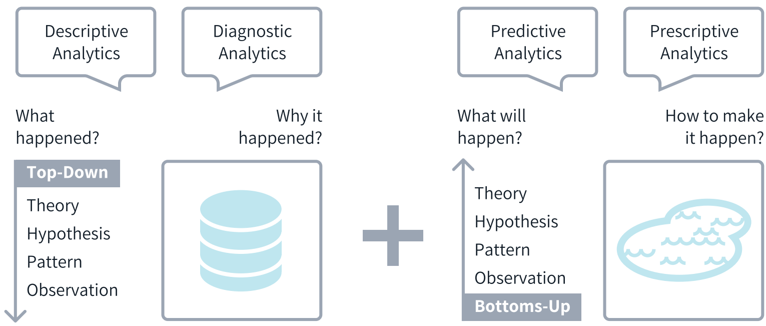


Figure 1: Data Warehouse + Data Lake works better together

Now, let's dig deeper into the mix of Data Warehouse & Data Lake data management architecture. Technology evolution over the last 10 years is dramatic, we understand if not presently, then the future will force decoupling, and data architecture and data engineering should be handled holistically and not piecemeal. Data management should enable us to build repeatable data solutions, that's where **modular, multi-layer** data architecture plays a big role. Data in its journey always must be stored as per data governance rules, e.g., the classification of data might change while you are joining the data from one stream with another, that's where **modular architecture** helps, it allows us to switch domains and sub-domains as the data travels through pipelines. With ever-evolving data definitions throughout the pipeline, if you don't have a modular architecture, then data governance becomes an issue, and now as you are tied to one zone, so access becomes an issue, and data taxonomy becomes an issue. Different zones/layers to perform different functions.

- Raw zone/layer – is the data lake; at a high level, this is the place for all structure, and unstructured data comes in, this data resides in both buckets in their original file format like json, csv, others, and this can also be democratized it by saving it in relational DB like Google BQ.
- Consume zone/layer is the data warehouse, where in consumption, we only model what we need; it has curated generic datasets which is curated as per data standards and policies, this data is pristine, Taxonomy is harmonized

The beauty of the modular architecture is data as a service model and data as a delivery model,

- Data as a service Model – this enables business & data science teams to come to either lake or warehouse and pull the data themselves
- Data as a delivery model – this enables data to be moved to any other data repository which needs the data either for reporting or analytical purposes.

Semantic layer - This plays a crucial role, as data is distributed in multiple layers & systems, and more analytical products will be created in different domains, and they all want to be shared or exposed to the whole organization, for e.g. – an ML feature store and model prediction need to be tied to your universal semantic layer to expose & share this data across organizations.

Now, let's move on and talk about some new frameworks & paradigms, and how they can help organizations in their data journey.

Data Mesh: Data mesh is a shift of paradigm from a centralized architecture to Decentralized architecture. As per Zhamak Dehghani - "The data mesh platform is an intentionally designed distributed data architecture, under centralized governance and standardization for interoperability, enabled by shared and harmonized self-serve data infrastructure." Where

Domains are a first-class concern, creating self-serve data infrastructure and treating data as a product. As per Zhamak Dehghani Data Mesh has primarily four principles.

#1 Domain-Driven Data Ownership Architecture	#2 Data as a product	#3 Self-Serve Data infrastructure as a platform	#4 Federated Computational Governance
Decentralize and distributed responsibility to business domains, to support scalability and change	Analytical data is considered as a data product and their users are cusctomers	Abstraction of infrastructure that removes complexity of provisioning and managing life cycle of data	Global governance which is interoperable while maintaining autonomy of local domains

Figure 2: Data Mesh

One advantage of data mesh is to democratize data and skills of data science in business and overcome traditional deficiencies in data teams where they don't have data skills. In this regard, the key challenges Data mesh is trying to solve are:

- Ownership – it's the domains and not the infrastructure & data teams who own the data
- Quality – this is again Domain responsibility, as the data is closer to the domains
- Scaling – Centralized team & tools are a bottleneck

Data mesh is Decentralized & Distributed innovation where Data mesh is not only about the technology but how you organize your teams, it's an Organization Centric solution where data is a product. The most important part of data got the context – The Domain, it's a Domain-based decentralization, & Domain oriented – Decentralized Data Ownership, Data mesh as it sits in the domain as it gets closer to the business.

Concerns with Data mesh

- One of the big ones is the absence of Infrastructure as a platform technology
- Scaling issue - although it is trying to solve the scaling issue with centralized team/ infrastructure, how to scale is also a problem with data mesh as more technical experts are required in multiple verticals or domains, as domains get quality engineering teams for each of their domains.
- In the classic mesh model, as domains create their own pipelines, so Data duplication &

technical debt is a big problem, resulting in inconsistent technical implementations

- Pitfall of silos – due to domains not being able to join data mesh
- Data Mesh is for the companies which have the highest data literacy across the organization, and data culture is at its core.

To overcome the existing challenges the industry is facing, my recommendation is a mix of Hub-&-spoke & Data mesh with baked-in Data Observability framework is the preferred approach rather than decentralized, I call it as Centralization to Enable Decentralization. This consist of a central core team, distributed Business domains teams and a centralized governance, where all work as partners towards a common goal.

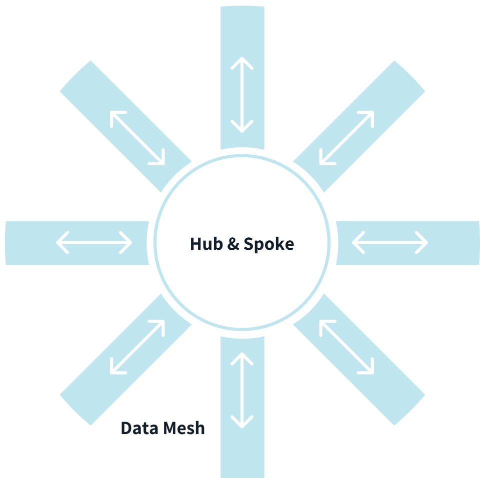


Figure 3: Hub & Spoke + Data Mesh

- Central Core team (Hub) –
  - Lays out the foundation, defining new standards & patterns, through which robust, repeatable, and scalable solutions are created, and by following an orchestrated approach consistency of capability and practice will be encouraged.
  - Manages data observability framework, which relies heavily on metadata to drive recommendations, which continuously identifies, connects, cleanses, and enriches real-time or batch data & discover relationship, and needs continuous harmonization and de-duplication as it integrates with Different Data systems, with different formats

- Distributed Domains (Spoke) (Mesh) –
    - Execution & delivery is owned by domains
    - Business Domain teams own the execution & delivery responsibility where the data ownership, speed, and scale are decided by the domains
    - Domains can use the standards & patterns defined by the core team and concentrate on reusing the repeatable and scalable solutions for the delivery rather than spending time in defining new architecture patterns and how to make them scalable.
- Centralized governance distributed as per domains.

Also, the data team being closer to data technology can continuously evolve and evolve the frameworks to make them faster and scalable, which domain teams or verticals should not concentrate on. They should concentrate on developing products using already defined standardized frameworks.

Data Platforms should act as a **Contextual Knowledge Center**, and the above approach should support what I call 5's theory – Speed, Scale, Secure, Simple & Sound

- Speed – Ability to delivery data at the speed required
    - This can be easily achievable as the business domains in data-mesh delivery patterns are re-using the repeatable and scalable solutions created by the core team.
- Scale – Ability to scale both infrastructure and delivery
    - selecting Infrastructure, creating solutions & continuously evolving them, so it can scale is core team responsibility (as in Hub & spoke model).
- Security – – Managed securely, governance is not a one-off activity, it must be embedded in all aspects of data
    - Centralized governance with domain distribution, helps in streamlining governance at organization level, with scale and agility as per business domains.
- Sound – Reliability (trust) of the data is important.
    - This is shared by both the core and domain teams, where the quality of the data is the domain responsibility as they understand data the best, but the setting up the observability framework, baking DQ checks in the pipeline itself, and monitoring and alerting the whole system is core team's responsibility.

- Simplicity – Complexity to simplicity is smart, superior intelligence
    - The data strategy and approach should be tailored to fit organization culture and workforce.

Now, as per Gartner, the rate of shift of data and analytics work is accelerating in federated line of business and is shrinking in IT, and the rate of investment in line of direct business work is accelerating. A shift in CDO's spending can be seen towards a more decentralized approach, and cost & optimization is big on CDO's list. So, we must be more efficient, and in today's cloud world, rigor and discipline are important, else it will become too chaotic with costs skyrocketing.

No single technology or paradigm can solve all the problems, so the best approach probably is an orchestrated approach where it is not a single platform like either a data lake or a data warehouse, nor is it a point-to-point operations environment. It is a coordinated approach implemented through metadata and governance. It's not about building centralized stuff, the platform is not necessarily individual monolithic, it's a more distributed approach where coordination, orchestration & interoperability is the key and not control. And governance priorities & mandate should be to create an enabling entity rather than controlling one. Data management is not about silver bullets, it's a journey of evolution, and organizations will keep on adapting to new technologies and paradigms to stay competitive.

Author Biography

Ujjwal Goel is the Director of Data Architecture & Data Engineering at Loblaw. He leads a talented team in the development of a petabyte-scale modern data stack in the cloud. This platform powers Loblaw's' Analytics and ML capabilities, accelerates value creation and helps Canadians Live Life Well. Before Loblaw, he has served in multiple Leadership and Architect roles across the globe. Ujjwal has over 19 years of experience in the data and analytical space, holds a Computer Engineering degree, and frequently shares his knowledge as a thought leader and a data expert in the data and analytics space.

Chapter 15

Gung-Ho Analytics

Mark Stern

VP of Business Intelligence & Analytics, BetMGM

Nowadays, everyone in the organization attempts to do data analytics in their role. Many prescribe data and use it to make decisions. Often that data has been shoveled to them by an analyst, or data scientist, without the non-analysts ever asking themselves, am I doing this right? or explaining to the analysts what they want.

I run a webinar, Analytics for Non-Analysts, and I ask the following simple question:

“You meet a couple who have two children, the one standing with them is a boy, what are the chances that the couple have two boys.”

Many people get the answer wrong. The answer is 33%. Even when you give them the answer, they remain puzzled rather than say ahh, yes! It’s easy to conclude the wrong answer, even when you have the data. What you think is the obvious way to interpret the data is often incorrect.

We are awash with data, and in many digital organizations, it's the only way you can see and tell the story about the company. It's the only way you can answer critical questions. But we must select the most important questions to answer. Not every decision must be analysed. Good Analysts are in short supply and in high demand, use them wisely.

Organisations that learn how to structure and Govern analytics as a high-profile organisational unit, rather than small individual and unconnected units pushed down into depths of different parts of the company, will drive the most value from data and the ones that can call themselves self-data driven.

I recall years ago when I had to justify the salary of some analyst roles. I found the HR processes were based on the tradition brick-and-mortar business whereby salary was linked to how many people you had underneath you in the OD and the size of the budget you were custodian of.

Your value was determined by how big your organisation hierarchy was.

My team was small but smart. We looked at the business in a way no one else in the organisation could look at it. We optimised decisions that affected budgets of tens and hundreds of millions of dollars. In addition, simultaneously, we are required to have the skills to influence and convince non-analysts (at every level of the organisation, CEO downward) who were custodians of those budgets that there is a better way to make your decisions.

The data and the analytics techniques we use give us the confidence to recommend and prescribe what to do as opposed to getting the confidence from the size of our department budget or the number of people we have reporting to us.

Carlson Raiders

The rallying cry of the Carlson Raiders was Gung-Ho, derived from a Chinese phrase meaning "work together". Lieutenant Colonel Evens set up and led a new, first-of-its-kind US Raider Battalion, spurred by President Roosevelt.

Carlson was perhaps ahead of his time. He implemented an egalitarian approach, rather than a strict military discipline and hierarchy approach. Amongst other things, he got his Marines to understand and feel what they were fighting for, he installed humility, and perhaps most importantly, he gave them the confidence to speak up, challenge others and make their own decisions. This last point enabled them to adapt rapidly when the field operation was not going according to plan.

Carlson Evans' approach is one I take influence from when building and running my analytics teams, and the reason I have named this chapter Gung-ho Analytics. A highly skilled team whose objective is to get its organisation to excel needs a distinct leadership approach to deliver

The Analyst

For organizations to maximize the benefits of its analysts and data it must do two main things.

- Endower them with the confidence to speak up and challenge people who are more senior than them.
- Provide them with big, interesting problems or opportunities to solve, which they feel have a good chance to be adopted and make a difference to their organization.

Without those two things, an organization will struggle to retain good analytic talent and get true value from data.

Most analysts I know and have managed will work late into the night through curiosity and enjoyment of working on meaningful problems. Thinking and then exploring their trail of thought. Wading through the data trying to find the answer is an enjoyable experience for most analysts, and one in which they can lose track of time.

Trying to then explain and influence the non-analyst regarding the insight and what they should do differently as a result is a different skill set and one many analysts will find difficult and exhausting by comparison. Especially if the Snr. stakeholders are not open to change or are not data or statistics savvy.

The worst thing you can ask an analyst is: can you create a report for me? Analysts need to be better at asking Why? It's important for the analyst to understand the motivation of the person asking such a poor question. Reports or dashboards are technical solutions for giving people a narrow and small view of the data logged in the database. The analyst needs is an articulated business problem or opportunity they can help solve.

Another one I have heard is, "I need more insight?" This is vague and shows little thought by the non-analysts. Many non-analysts find it difficult to structure and articulate their question but saying I want some insight is lazy and unhelpful.

The question should never be "can I have a report or a dashboard". These questions are analogous to a soccer team spending many dollars on a team of players but playing with an approach whereby the goalkeeper kicks the ball long each time directly to the loan striker, missing out on all the talented skills in the field of play. Learn to use the analytics talent by thinking about the business exam questions rather than punting the question “can you give me this report” right up the field.

Data and the Analyst

Most, if not all, reports give the same answer to the same question each day, and most questions they answer are out of date before the report has been built. If you audit all the reports in your organization, you will likely find that many are not looked at.

Companies that focus on churning out reports are report rich and insight poor.

To perform great analytics, we need not ask for reports. We must ask for the data that represents the real-world events pertaining to the process we are interested in, those events need to be logged and recorded in an analytics database. As analysts, we need that level of flexibility if we are to move at speed and be allowed to think for ourselves.

We don’t know what questions we will be asked tomorrow, but we know there will be new questions we have never been asked or asked ourselves before. Without having the breadth and depth of data available and cleaned, it will take too long (or impossible) to answer those questions

The Analytics leader

I have a saying:

“I should be the only Analyst in the organization that should report into a non-Analyst.”

I don’t always get my way, but I believe in the saying.

Data and knowing how to use and interpret it is so important in today’s organizations that if you devalue analytics leadership by having small pockets of individual analysts reporting into nonanalytics leaders, you have little chance of becoming a data-driven company. The analytics Leader must have a good grasp of the business and understand how the data needs to be interpreted to describe and change the organization.

There are so many specialties of Analytics, so see your analysts as an ensemble operating like an orchestra. Like the best soccer teams or the Broadway Theatre Swings, I now look for Analysts who can perform and step into play multiple roles.

At BetMGM we provide on-demand technical training with Datacamp across a range of skill sets; every analyst gets access to training. We invest in platforms like Data Robot, kyvos, Power BI, Trino, Excel …. That makes it easier to move team members across roles. Just recently, one analyst built her first Machine learning model in Data Robot, and it's now been monetized and adopted to improve decision-making in customer conversion. She's now back performing a more commercially focused analyst role generating insights for the Marketeers.

What differentiates one analyst from another is less about their job title, whether we label them Data Scientist or Analyst, but more about their character, the way they think and their business domain expertise.

Analysts also enjoy being part of a bigger group of analysts. It allows them to have career progression. They then benefit from specific technical and non-technical training, learn from their contemporaries or senior analysts and gather their knowledge and skill much faster than they would as an isolated analyst. A Non-analyst often thinks differently to analysts and have different needs and expectations.

Probably most importantly, analysts need to feel they are independent. Many business leaders are invested in the questions they want the answer to and are sometimes too eager to want to know their thing succeeds. The value of Analytics is to get the truth, not the answer you want or the answer going to make your current idea look successful.

Motivated Reasoning can be a real problem, and without analyst independence, you are failing to mitigate the risk. Independence and objectivity are a strong argument for having analysts reporting up into a single Snr. Analytics leader.

I found it hard to stand back and take myself away from the data so I could direct and shape rather than do. It took me time, but when I did, I found I spotted new opportunities and problems I would not have seen when I was working so closely with the data and analysis each day. It's enabled me to see things from a different perspective.

I still enjoy getting my hands dirty in the data, but now I run bigger analytics teams, it important to step away. The things I uncover now would not be impossible for a non-analytics business leader to see or find. For someone to take on an Analytics leadership role, they should grow up as an Analyst, not as IT Developers, Marketeer, accountant, product owners or other nonanalytics specialists.

For most Non-Analytics leaders with analysts in their direct OD, it's usually a small part of their headcount in a much bigger team, and most of these non-analytics leaders don't have a background in directly working with statistics. So, check in to see who your real analytic leader (in your company) is and see how far up the hierarchy they are.

Epilogue

Gung Ho analytics is about getting the leadership right for your analytics team, positioning the Analysts with enough authority and confidence to speak up and challenge based on what they uncover in the data. It about enabling thought and adapting to the situation rather than following a receipt book to develop a recommender ML algorithm. It's thinking of Analytics as a Leadership function rather than a support function that the whole organization aligns around and works with as a partner.

Author Biography

Mark Stern is a business leader in the chosen field of Analytics and Business Intelligence. HE has demonstrated the ability to build high-performing Analytics and BI teams that can influence and help change the direction and decision-making of the business. He has over 20 years of background in Analytics, Business Intelligence, Machine Learning & AI, IT, CRM, project portfolio management, project management, post-merger integration, Big Data. Proven ability to formulate Analytics, Business intelligence and IT Strategy; set up and run a program office and CRM capability/team.

ATSCALE

Chapter 16

Do you really want to be DAD?

Gokula Mishra

Vice President, Data Science, Direct Supply

Everyone wants to be DAD (Data Analytics Driven) these days, but do you really know WHY? Before you can answer that you need to know WHAT becoming a DAD means. You need to ask the question: what I am trying to accomplish being DAD. Build a visual picture of what it will look like. You also must have clarity at the most fundamental level WHY are you doing what you are doing.

Next step - while you know where you are today and what and why of being DAD with the end picture in mind, you must think through how are you going to get there? During the journey course correction is ok but what is not ok is not having your hypothesis / assumptions well understood and documented. Do not try to over engineer and try to be perfect before getting started, many make this mistake. Have the clarity of WHAT and WHY, draft view of how to get there and GET STARTED. Assume that you will learn along the way and readjust the course. I have seen many teams trying to get everything perfect before starting and sometimes wait long enough to restart the effort again because of changes in underlying market conditions. enough to restart the effort again because of changes in underlying market conditions.

For a consumer driven company WHAT might be delivering orders on time, for some it might be delivering on time as well as at the lowest cost to the customer, for some it might mean also anticipating the customer's needs. All these are driven by data and analytics. WHY might mean in this case delivering the utmost customer experience to reduce customer churn.

Our ability to produce data far surpasses our ability to use data. Not only our internal systems produce a large amount of structured and unstructured data, our customers, partners, and market players generate a lot more data that is outside our control. Not only do we not use the data we internally produce to ask questions, also we do the same with external data. Albert Einstein said: ""The World, as we have created, is a process of our thinking. It cannot be changed without changing our thinking." In order for us to leverage all our data we need to change our thinking - a DAD thinking which starts with asking different types of questions of our internal and external data.

A big percentage of our data is never looked at - we call it Dark Data, and many companies have found tremendous value in this dark data when they shine light on it. For example, once I consulted with a manufacturing company whose manufacturing plants used to generate a ton of data about the machines that was stored, never looked at and at some point, simply deleted from the systems. By analyzing that data we were able to predict machine failures and were able to minimize unplanned downtime of the plant.

We generally ask most of the time, 'What happened?' type of questions from our data and create most reports and analytics, some of us also ask 'Why did it happen?' type of questions and try and understand causes of past results. We need to be asking questions about 'What will happen? - and 'What should I do?' - give me options for decision making. That will allow us to get even more value from being DAD. And with advanced AI/ML we can also ask what we not know - discovery of hidden patterns and insights in the data, as well as we can run experiments to test many hypotheses we may have.

Data analytics is key to improving quality and speed of decision making, and becoming DAD requires five key focus areas:

1. **Sponsorship**:

We need both Top-Down and Bottom-up sponsorship. We tend to spend lot more time in getting the Top-Down sponsorship, which definitely is critical, but many tend to ignore the Bottom-up sponsorship, which is key to the change management success, key to the cultural change that will be needed. Top-Down sponsorship will get us financial support as well as strategic alignment with key business objectives, it also gets of the visibility and political capital needed to succeed. Bottom-up will get us the support we need during the execution from the tactical and operational layers of the organization. So please do not underestimate its importance.

2. **Maturity and Culture:**

Maturity in data analytics mindset requires investment in people, focus on the right decisioning process and being data literate, Investment in people is primarily driven by the strategy of creating the skills and knowledge of right tools needed to become DAD. Also, proper attention must be paid to understand how decisions are made in the company and what needs to be done to have the decisioning process be DAD. We also need to improve the ability to read, write and communicate data in context, which according to Gartner is improving Data Literacy. This includes an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, the application, and the resulting value.

3. **Data Analytics Strategy & Management:**

When generally we talk about Data Analytics Strategy, we spend a lot of time talking about the Data strategy and how to treat Data as a corporate asset, which is the right thing to do. But I would highlight another key component to its success – having a dedicated Data Analytics Leader who will create as well as manage the Data Analytics Strategy and management.

4. **Data Quality & Governance**

Data Quality is essential to driving the value from all data assets via analytics, AI & Machine Learning. And Data Governance is key to keeping the data quality at the highest level, Data Governance also includes aspects around data security, privacy as well as life-cycle management. Here also I would emphasize keeping scope simple, get started, learn, and iterate. I am a big believer of what I call the Circle of Value. Focus on the innermost circle – the few numbers of data attributes that impact 80% of key decision making (this should not be a large number of attributes), and make sure you get that right from quality/governance perspective, and then move on to the next circle of value.

5. **Technology Investment:**

There are three key components to technology investments. The first obvious one is being ROI driven. Typically, we look at the ROI in five potential areas: (only one dominant at a time) Revenue, Growth (new Revenue), Cost/Efficiency/Productivity, Risk, and Brand (may not apply to every company). The value we should capture preferably should be quantitative but also, we should capture the qualitative benefits as well. Examples for quantitative are X millions of dollars in savings, 10% improvement in operational efficiency. Examples of qualitative ones are improving customer satisfaction, Brand perception, employee engagement.

The other two aspects are more related to architectural aspects of the technology investment. The investment should go towards flexible and integrated architectural components especially in data analytics. Also, the architecture should be loosely coupled. These attributes of architecture are very important to get most of the technology investments over a longer time period. Designing a semantic layer is a great example towards this concept.

Realizing full potential business value of Data Analytics requires support and collaboration from many different groups:

- First is the maturity level of the LoB in fully embracing usage of Data Analytics in their business, a data analytics culture helps in true adoption of Data Analytics
- Change management is critical to make sure the impact of data driven culture on the business is understood and the changes to business roles and processes are planned from the beginning and implemented.
- Support from sponsors is essential to get the resources needed for data analytics projects driven by a solid business case and ROI.
- SME from the LoB need to work with Analysts and data scientists to make sure that the Analytics & AI Solution is designed and built to solve the right problem. Also, SMEs will help in testing and rollout of the solutions.
- Engineering is critical to provide the right data engineering support as well as in helping implement the Analytics & AI solution in the production environment and monitor it over time
- Analysts and Data Scientists are involved throughout the process and in building/testing/tracking/enhancing the Data driven Analytics & AI solution.

When data driven decision making culture starts taking roots in different parts of your organization, you will see signs for designing for data showing up and then usage of that data back into the design. A great example of this is when you are designing mobile app, if you design the app for collecting a lot of data around how the user uses the App, then that data that was designed to be collected can now be used to improve the design of the app.

Another sign of becoming DAD is around people starting to do experimentation to have a data driven answers where we did not have an answer (rather than go by just gut feel answers). Experimentations can be used not only to test hypotheses about customer behaviors but also these experiments can generate data that are helpful in further analytics.

When you are DAD, you also tend to try what-if scenario simulations as well as discovery into unknown (what do i not know that data can tell me), which is the ultimate level, this emerging area is manifesting itself into the world of Digital Twin that brings together the digital maturity with DAD maturity.

THAT'S HOW YOU KNOW YOU HAVE IGNITED THE DAD FIRE.

Author Biography

Gokula Mishra is a Data Analytics thought leader and is VP, Data Science at Direct Supply, delivering AI/ML products to create business value in the senior living industry. He was the Global Head of data & analytics, and supply chain strategy and implementation team for McDonalds Corporation in 100+ countries. He is a co-author of Oracle Big Data Handbook published by McGraw-Hill, and he speaks frequently at conferences on how to use and leverage enterprise data and open data to generate business value through analytics. He was a keynote speaker at MIT CDO & IQ conference Aug 2019. He received his bachelor's engineering degree from BITS, Pilani, India and a Graduate degree from Northwestern University, USA.

ATSCALE

Chapter 17

The Need for Semantic Layer in Financial Services

Ramdas Narayanan

VP, Product Management, Bank of America

In today's business world, data is coming in at a rapid pace, in different formats. There is a lot of value to be derived from the data by the business to execute and stay ahead in a competitive landscape. Business and Technology work in their own silos. This causes lot of data duplication, lots of siloed applications, this also leads to businesses not having a holistic understanding of the data. In addition, there are also very demanding SLAs on when and how the data needs to be available. How do we approach these issues and come up with a framework to solve this problem? While there are many ways to improve the speed, scale, and impact of data and analytics projects, one solution used effectively by enterprises is a Semantic Layer. Specifically, the Semantic Layer is used when,

- There is data available, how best to bring all the players together (Business, Tech, Stakeholders, Vendors, customers). Very critical to have a complete end-to-end vision to enhance the customer experience.

- There is a need for Data to be presented to be understood by the Business. We need to have a Data Strategy part of the Overall Business Strategy. Ultimately, we need to solve Business Problems/Help Businesses make decisions.
- There is a lot of data/information aggregated in a pre-determined manner in BI Tools/ Value Added Process, which may not be understood by the Business/Stakeholders. These aggregations are built based on the assumption that the business users need. In reality, some of those metrics might not be needed.
- Provide a layer/Space for the business users to work with the data and bring business closer to the data that is available. A semantic layer will need to account for the data landscape that includes people, process and technology
- One challenge today that prevents getting value from data is the lack of access. The data is in control with a few people. Allow end users to work with data in their tool of use and use self-service capabilities.
- In larger organizations, there are lots of data warehouses, data marts that have trusted data. There are lot of insights and value to be derived from this. A semantic layer would be beneficial to derive value and actionable insights.
- There is text, unstructured data, documents, and images being captured by various data systems. The true value of data in certain businesses could be achieved by merging the unstructured and structured data in the Semantic layer.
- Use Technology today (such as Atscale/Starbust/Zetaris/Donado) to build a semantic layer that will map complex data into familiar business terms and provide a consistent provisioning layer.
- In regulated industries, there are compliance requirements, how certain metrics are calculated and represented. A Semantic layer can bring in the much-needed consistency and credibility for the metrics and reduce/eliminate errors/fines.

So, what are the options to build the Semantic Layer? Organizations have different options to build a Semantic Layer, and have a good mix of the long-term vs Short Term objectives when using an option. It is critical to keep the choices in sync with the Business and Data Strategy. It is imperative to bring all the key stake holder, Business and Tech together to devise the plan for the Semantic layer. People and Business Culture are key underpinnings to making a strategy successful. The key factors part of devising a Semantic Layer implementation are:

1. **Business Strategy/Outcomes**
2. **Data Strategy**
3. **People Involvement/Process**

4. **Data Fluency/Literacy**
5. **Data Supply Chain**
6. **Modern Data Catalog**
7. **Technology Stack**
8. **Funding/Resources/Talent Availability**

Orchestration of the key factors listed above is critical; having a well-established data supply chain with trusted sources for different domains is important.

- Use a Semantic Layer Tool, and utilize already available tech to build the abstraction layer. The question is doing this layer scale, can it keep up with the increasing volume of data.
- Build Value Added Process that do the calculations and aggregations. Store these values in tables/DW. These metrics could be provisioned to Consumers via Web Services. In this approach, would there be too much time spent on maintaining the code base, constant changes to keep up with the change in Business Rules
- Use Reporting/Visualization tools to house the aggregations/calculations. Would the users end up having metric definitions within multiple BI tools? How does the business user have a broad view of all the metrics?

This brings us to the need for Modern Semantic Layer tools to handle the calculations/abstractions. This tool could pull data from multiple data sources/execute queries at the different data sources. Once that is completed, the modern semantic layer would be a pass-through to provision data to multiple BI tools. The Modern Semantic layer tools provide the ability to be part of a Data Mesh Architecture. Semantic layer helps with concepts such as Data as a product. This is a key aspect of Data Mesh. One of the Key aspects that need to be factored is the availability of the Modern Data Catalog as one is constructing the Semantic layer. In this approach, one could achieve BI at scale, where the layer continues to evolve as data/Data Sources keep getting added.

With the advent of the Data Mesh framework, which is a decentralized set of different data products with the ability to orchestrate them via centralized provisioning layer, the semantic layer has become important to organizations. The Semantic layer should be able to provide both exploratory and ready-to-use data. Exploratory Data includes curated data, AI Models, APIs, Data Models. Ready-to-use Data includes Insights, benchmarks, reports and dashboards

When implementing a semantic layer, it is critical to provide the context of data to users, along with valid definitions and classification. A modern data catalog is a critical component of semantic layer implementation and execution. Users today must navigate different tools to get data context and definitions. There is a critical need to provide contextual data information. This is a topic. https://atlan.com/platform/data-catalog/. This is by the vendor Atlan in the modern data catalog domain. One consideration when implementing a modern semantic layer is performance and the ability to provide data to consumers within the agreed-upon SLAs. It is important that the Semantic layer does not turn into another data warehouse/data lake. This would cause another set of costs to maintain the layer, and the ability to scale AI/BI could become more complicated.

Other features would need to be part of the semantic layer:

1. Ability to Search the Semantic Layer
2. Work with Recommendation Engines
3. Conform to Data Governance Standards
4. Provide Data for AI/ML Models/experiments

Reference for the above features: https://enterprise-knowledge.com/what-is-a-semantic-architecture-and-how-do-i-build-one/ by Lulit Tesfaye.

When we look at the technology landscape, there are a lot of points where data collection is possible. With IOT, Edge Computing, Computer Vision, Images, Smart devices all growing and providing data capture abilities, synchronize all this data and make it valuable for the Business. Some of these would be more critical for some businesses than others. This brings up an important set of questions. What are the Challenges/Opportunities for Semantic Layer Tools? Below are some of the key challenges.

1. Semantic Layer tools need to be adept at consuming unstructured data. There is a lot of unstructured data being collected today in documents, images.
2. How will semantic layer tools handle IOT/Edge Computing data.
3. Semantic Layers need to be plugged in Customer-facing Applications – The idea here is to use these applications and apply AI/NLP techniques in understanding usage patterns. These patterns could enhance the search features within the semantic layer
4. Tools that provide Semantic Layer need to have constant monitoring of data usage to develop optimized querying techniques.
5. How will the Semantic layer handle data drifts, in some businesses,the data changes more rapidly than others. What would be the techniques put in place to handle such changes.

Now let us look a little deeper into Semantic Layer and AI/ML, regarding AI/ML, there is a need to feature selection/feature stores. Typically, in AI/ML Scenarios to quote **Nick Handel (CoFounder**

at Transform) one likes to get as close to the raw data as possible. But BI/Analytical type of tools/applications take the raw data and present it in a curated fashion for the business to process. One of the key aspects of building innovative products is experimentation This will be an important area to address, would the semantic layer provide such capabilities or would organizations need to have separate sandboxes to download raw data and do experimentation. Reference: https://humansofdata.atlan.com/2022/05/metrics-layer-drew-banin-nick-handel/ (Atlan Podcast/Webinar)

There are other important aspects as part of the data supply chain critical to having a proper/ useful semantic layer. These include:

1. Data Observability/Data Quality – In Semantic Layer, we are attempting to standardize ad provide consistent metrics. To provide this, data quality is key. Can this be addressed at source? What type of data observability framework does the organization have. How well are data pipelines functioning, in case of failures, how does one recover?
2. Tagging Capabilities – This would greatly facilitate in doing metric searches for authorized users.
3. Role-based Access Controls – This is to handle the management of confidential/sensitive data.
4. Automate Lineage – As part of the modern data catalog, can the creation of lineage be automated, and manual effort be reduced.
5. Vendor Dependency – Is one going to have too many vendors in the mix for the semantic layer. This will be an important aspect to focus on; with so many tools being developed and launched, how would the organization navigate through this ecosystem.

In Summary, the Semantic layer/Metrics layer will be a key component of an organization's Business/Data Strategy. Adoption of a semantic layer would also help in driving the concept of Data Mesh. The Success of a Semantic Layer relies on the following key Pillars

6. Definition/Semantics - How do I define this metric
7. Performance – Response time when data is needed
8. Governance – Business rules and do leaders agree on the definition.
9. Query Capability – How do I get the data, what type of queries/tools are supported to get the data.
10. Ease of Access – Easy to get access to the data
11. API Integration – How does the layer work with API's. This integration is critical, especially for applications wanting to access the semantic layer.

When a Semantic layer is implemented and executed that can address all the capabilities above, then it would enhance the productivity of the users/organization and build a sense of trust in using the data. It can also bring the business and tech teams to collaborate better to derive valuable insights from the data.

Author Biography

Ramdas Narayanan is working as a Vice President - Product Manager at Bank of America with a focus on Data Analytics, DataOps, Oracle Exadata Platform (12c), Data Lake Architecture, Data Warehousing/Data Marts, Composite (Data Virtualization), Microsoft SQL Server BI Stack and .NET Framework. He has managed and implemented projects relating to Data Modeling, Data Integration and Data Provisioning projects for Mortgage Servicing as part of the Home Loans Business at Bank of America. Experienced in handling Development teams both onshore and offshore.

ATSCALE

Chapter 18

Understanding and Mitigating AI Risk

BGV

Anik Bose

General Partner at BGV (Benhamou Global Ventures)

Along with the massive value creation potential of trillions of dollars that AI will unlock, according to McKinsey, these explosive technologies also stir fears of robots displacing humans, massive concentrations of assets and power, and exacerbating inequalities in society. A thought leadership vacuum has emerged around the responsible development and deployment of AI technologies. Given the fledgling regulations, a clear lack of awareness around AI risks, and Big Tech's disproportionate influence on AI supply and demand, many AI practitioners are left struggling to identify clear standards and best practices as they develop and deploy AI in their businesses and introduce these systems into the world.

In this backdrop, how can an enterprise assess AI risk? Demand for AI governance tools has surged as global enterprise businesses anticipate GDPR-caliber shockwaves to sweep the tech industry. The European Union has put forth legislation to codify the development and use of artificial intelligence, and noncompliant companies could face fines up to 6% of their global turnover or 30 million Euros. Yet, McKinsey research suggests that companies lack the capacity to address the full range of AI risks they face, and many businesses are unclear on the extent of their risk exposure or the harm their AI could cause society and individuals, such as unintended discrimination, privacy intrusion and social exclusion.

As AI adoption accelerates, the technology poses a broad swath of ungoverned risks. Consider:

1. Starbucks' "just in time" automated scheduling algorithms, used by retail outlets around the world, left workers struggling to manage their calendars and other obligations, such as school, family commitments or other jobs.
2. Zillow's gamble on an automated and frictionless home-flipping business line lasted only 8 months and led to a $304 million inventory write down and the loss of 2,000 jobs. The lesson highlights the difficulties in using algorithmic real estate pricing to predict expensive, real-world decisions in a dynamic and fast-changing market.
3. YouTube discovered troves of videos with profanity and violent themes in its YouTube Kids service, the video-sharing sites' kid-friendly platform. A video called "cocaine pancakes" reached nearly one million views without being flagged as inappropriate content.

These cases are the tip of the iceberg with scandals associated with enterprise adoption of AI systems. Because AI is self-scaling and self-learning with insufficient governance to control for ethical and societal risks, these eventualities, if unaddressed, can grow exponentially and lead to unintended ethical and legal breaches, biases, and other pitfalls on a much larger scale than we've seen. To translate these challenges into risk categories, we delineate three concrete examples:

1. Privacy Intrusion – using mobility data offers new opportunities for product innovation, while leaving the company at risk of violating customer privacy. Legal breaches, damaged reputation and large investments were all at risk.
2. Discrimination – without an ethical filter on AI solutions, an agency risked discriminating against certain groups due to data biases. Transparency and organizational alignment with values stood at risk.
3. Social Exclusion – as a recruitment agency, AI drastically increases the client's capacity to profile, recommend and match job seekers with jobs at a faster rate. Yet this practice requires effective guardrails to ensure explainability and avoid social exclusion.

There are many more risk categories a company may encounter in deploying AI, but the challenge lies in operationalizing the risks so it can be assessed, externalized and dealt with appropriately. The European Union's AI Act (AIA) follows this risk-based approach to AI systems. High-risk AI systems face more restrictive regulations and transparency requirements than lower-risk ones and are likely to have global implications. With the AIA, policymakers in Brussels seek to present the first comprehensive framework to regulate artificial intelligence. Many advocates openly admit that legislators are attempting to mimic the impact of GDPR and set a worldwide standard in the race to regulate AI.

Against the rising tide of regulation, Swedish startup Anch.ai has released a horizontally integrated ethical AI governance platform for businesses to accelerate responsible AI adoption across their organization. The B2B SaaS startup emerged from the AI Sustainability Center, a Swedish think tank focused on the ethical dimensions of AI deployments. The assessment methodology stems from multidisciplinary research to detect and remedy ethical risks in AI solutions. The diagnostic tool seeks to uncover an organization's exposure to ethical and legal breaches while equipping enterprises with the ability to screen and assess risks, obtain mitigation recommendations and report on their progress. Further, customers can receive independent validation of their ethical AI practices to strengthen trust among their clients and stakeholders, comply with regulations and ensure conformity with corporate values. As AI becomes more and more pervasive in the enterprise, tools like the Ethical AI Health Check will mushroom in popularity and help steer enterprises to both leverage the disruptive powers of the AI while doing so in legally compliant and human-centric ways that society can trust.

While most entrepreneurs understand the risks associated with algorithmic biases, privacy breaches and model drift, they lack rules of the road and struggle to identify best practices or ideal resource allocations to properly address these risks in a responsible way while building their businesses. Against this backdrop, investors are seeking to integrate AI risk into their investment theses, their due diligence processes and their board rooms.

To aid in this process, EAIGG, alongside Ethical Intelligence and BGV, developed an Ethics Maturity Continuum to help quickly assess a company's level of ethics maturity and identify areas for improvement. It prioritizes agility and action, enabling users to build concrete strategies for sustainable AI systems and track development overtime. Most importantly, it empowers startups to embed ethics from the beginning, resulting in stronger products, happier customers and more favorable exits. The Continuum is designed for startups and investors to assess the level of maturity a company has achieved in operationalizing AI Ethics using five key criteria.

1. Accountability – When someone is accountable, it means they are answerable for the results of an action after it has been performed. AI accountability means that a company deploying AI systems has designated roles that are both answerable for the impact of the AI systems and responsible for AI governance within company processes.

2. Intentional Design – Successful AI design focuses on creating products that serve

humancentric needs, either on the individual or societal level. Intentional design goes a step further by ensuring significant thought and consideration has gone into understanding potential intended and unintended consequences of designing an AI product to serve such needs.

3. Fairness – Unwanted bias occurs when system-based decisions are made using individual traits that should not otherwise correlate to the outcome (i.e. gender used as a deciding factor for job applicants). Fairness seeks to minimize instances of this unwanted bias and instead promote inclusive representation in AI development.

4. Social Impact – AI has the potential to impact not only vast numbers of individuals but also shape the societies in which we function. It is, therefore, essential to consider the short- and long-term effects the introduction of an AI product will have, giving particular attention to the well-being of end-users.

5. Trust & Transparency – Data is information on individuals and collective behavior, which means users must be able to clearly understand how their data is being handled and protected. Besides transparent communication and robust security, the user must feel that their information remains as private as they so want, the combination of which results in strong user trust.

As the policy is created, refined, and defined, consulting firms may decline in popularity, and MLOps/GRC platforms may rise due to the ability to programmatically enforce compliance given concrete metrics. The modern "tech stack" will likely evolve into an "ethics stack" in which the flow of data is carefully monitored, documented, and analyzed through products provided by companies in the categories. For example, a company might employ Gretel for data privacy and synthetic data, Akira for MLOps management and continuous bias detection, then Saidot for model versioning and governance.

The demand for ethical AI will perpetually increase because performing AI services correctly is domain-specific, context-specific, and very unstable (i.e. always needs to be monitored, checked for quality, etc.). There's no "one-and-done" approach to AI. The "boom" for ethical AI is estimated to be somewhere from the mid-to-late-2020s and will follow a curve similar to "ethics predecessors" like cybersecurity in the late-2000s and privacy in the late-2010s. There will be a time in which policy, real case studies of AI gone wrong, and new discovery of biased AI (and a genuine desire to fix it) nudge companies in the right direction (through fear or will), leading to large demand and more inclusions in EAIDB. This is a nascent but rapidly growing ecosystem and is expected to increase in momentum. Meanwhile, we hope this research has shed some light on what is undoubtedly a fascinating and critical area of startup innovation.

Realizing the full promise of AI will require market education, smart regulation, innovative solutions from startups, savvy investors, and best practice role modeling by Big Tech companies. Smart global regulations will need to incorporate incentives and fines and should vary by use case: more stringent for regulated industries/high-risk use cases and more product labeling for unregulated/low-risk use cases. This will require collaboration between the public and private sectors. Also recognize that bias is both a human and an AI problem. While large data set-driven algorithms can introduce consistency and reduce bias, humans in the loop may be required for life-critical applications like those in healthcare or emergency response. Automated AI audits (AI on AI) are on the horizon and could offer a creative way to build trust in AI models before explainability can be addressed.

Ultimately, the responsibility for ethical AI development and deployment cannot be left to Big Tech influencers or regulators alone. However, grassroots communities and entrepreneurial innovation will be needed to present a bulwark against the polarized standoff between the unconstrained self-interest of Big Tech on the one hand, and the regulatory overreach of activist policymakers, sensationalist journalists, and hacktivists, on the other.

Author Biography

Anik Bose has 18 years of active venture capital and corporate development experience, including 7 years as SVP, Corporate Development at 3Com, and 11 years as General Partner at BGV. As a General Partner at BGV, Anik is active in every portfolio investment, including being a board member of Webscale Networks, Constella Intelligence, Drishti Labs, Spyderbat and Blue Cedar. Anik focuses on investments in the areas of AI and Cybersecurity. He is also responsible for leading BGV's Customer Advisory Board and for implementing ESG within the firm. Anik is the founder of EAIGG, a diverse community of AI practitioners focused on democratizing the growth of Ethical AI governance through best practice innovations around AI governance, Data privacy and AI climate change tech. Anik Bose, enjoys activities such as hiking, traveling, and studying ancient wisdom teachings. Anik holds an MBA from Boston College and a BA in Economics from the University of Delhi, India. He and his wife live in California.

Chapter 19

Actionable Insights for Improved Business Results

Prashanth Southekal

Managing Principal DBP Institute
and Professor at IE Business School

The data economy is increasingly embraced worldwide in every industry. According to the World Economic Forum (WEF), by 2025, it is estimated that 463 exabytes of data will be created each day worldwide – that is the equivalent of 200 million DVDs of data created per day [WEF, 2019]. Faced with overwhelming amounts of data, organizations across the world are looking at ways to derive insights from data analytics for improved business results.

But why do data insights matter for business? Insights help companies gain better visibility to make faster decisions for improved business results. Research by McKinsey Consulting found that companies that are insight-driven report above-market growth and EBITDA (earnings before interest, taxes, depreciation, and amortization) increase between 15 and 25 percent [Mckinsey1, 2022].

What is insight? According to the Cambridge dictionary, insight is the ability to have a clear, deep, and sometimes sudden understanding of a complicated problem or situation [Cambridge, 2022]. In terms of data analytics, insight is the unknown element (such as relationship, patterns, categorization, inferences, prediction, averages, correlations, variations, quartiles, outliers, confidence levels and intervals, and more) that will influence the decision if it becomes known.

Technically, insights can be classified as hindsight (understanding a past situation), near-sight (interpreting the current situation) and fore-sight (predicting a future situation) [Southekal, 2020]. However, from the data and analytics value realization perspective, there are two main insights.

1. **Performance Insights**. . Performance insights provide visibility on the measurement entity. Examples are the top five products by sales quantity, the top three customers by MRR (monthly recurring revenue), and so on. Effective data storytelling techniques hold the key to managing the two options or choices that performance insights offer: (a) to know or (b) to act.
2. **Actionable Insights**. . Actionable insights based on performance insights provide the visibility that can be actioned. Actionable insights involve three elements (a) decision (b) commitment to consume resources like time, money, labor and so on associated with the decision (c) and the business impact and consequences of the decision. Examples of actionable insights are product A is the best option to invest in, customer X can be given a credit, and so on.

Figure 1: Types of Business Insights

So, what can enterprises do to get value from insights? In data and analytics, the "last mile analytics" is considered the missing piece between the data analytics output and actual business results. According to McKinsey Consulting, 90% of organizations significantly outperforming peers are devoting more than half of their analytics budgets to bridging the last mile of analytics [Mckinsey2, 2018].

Business enterprises must focus on actionable insights – or specifically focusing on converting performance insights into actionable Insights. Below are the three key steps companies can take to implement actionable insights.

1. **Derive the performance insights based on business objectives, questions, KPIs (Key Performance Indicators), and data**. These insights can be derived using a combination of descriptive, predictive and prescriptive analytics techniques or models which provide visibility into the past, current, and future states. Given that data is a critical component in this step, the data related to the question and KPI should be sourced from the right data source, typically the transactional SoR (System of Record). The semantic layer can be leveraged for standardized data definitions and rapid data access to derive faster insights.

To ensure that these performance insights are reliable, it's important to factor in different stakeholder perspectives, time frames and locations and avoid framing bias. Framing bias refers to the way the question is framed and can be addressed by reframing the problem in at least three ways. Basically, having the right data means mapping the data source subjects and understanding the data attributes (i.e., features) to align data source content to the question/answer being sought. The relationship between business objectives, questions, KPIs (Key Performance Indicators), semantic data, and models to derive insights is as shown.

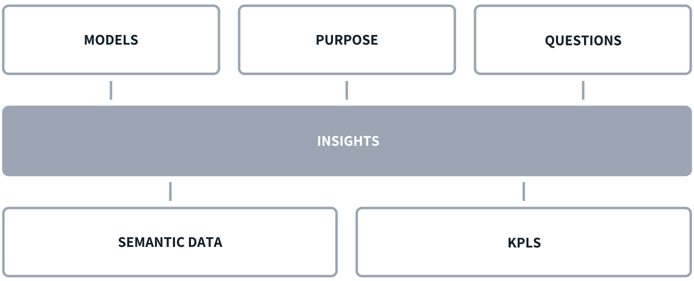


Figure 2: Insight Value Chain

2. **Once the performance insights are derived, we formulate the decision problem**. A typical decision problem has four key elements: objectives, alternatives, outcomes, and payoff.

    a. The **objectives**, which are based on performance insights, are the things the business plans to achieve from the decision.

b. The **alternatives** are potential actions or strategies considered based on different performance criteria such as profit margin, cost, time, quality, service, and more. It is recommended to keep the number of alternatives to three, and the Pugh Matrix or Decision Matrix can narrow down the available alternatives.

c. The **outcomes**, which are usually probabilistic, are the resulting situations that arise by pursuing the selected alternatives.

d. The payoffs or **benefits** are the values placed on the outcomes associated with each alternative. The payoff values combine tangible and intangible benefits. These four elements of the decision problem (objectives, alternatives, outcomes, and payoffs) should help the business select the best or optimal alternative to implement the decision using the appropriate decision science techniques.

These four elements of the decision problem (objectives, alternatives, outcomes, and payoffs) should help the business select the best or optimal alternative to implement the decision using the appropriate decision science techniques.

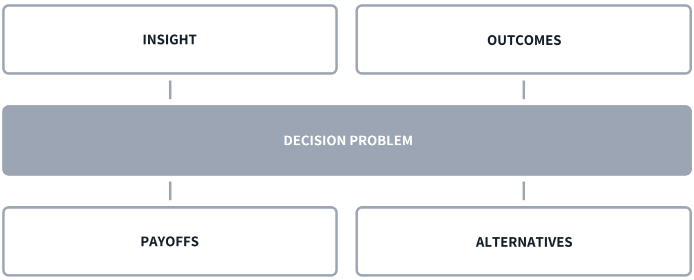


Figure 3: Elements of a Decision Problem

3. **Once the decision to implement the alternatives is made, the next step is to identify the resources needed to execute the decision**. The resources could be time, skills, budget, equipment, and even data. The key step in executing the decision is to manage change with the right ownership or accountability. Successful change initiatives are often associated with strong accountability or ownership. This means having an accountable leader close to the objective and KPIs being tracked for business performance. For example, if the KPI is on “Days Payable Outstanding (DPO)” to improve the cash conversion

cycle (CCC), it is advisable to have the Account Payable (AP) Manager track and improve the DPO KPI. Basically, identifying the right and competent leader will help in mobilizing the support, including the desired resources and culture, to convert the insights and decisions into action.

The effectiveness of a decision can be validated with feedback mechanisms, given that drifts in models and data are bound to happen as the business evolves and adapts to change. The best way to manage model and data drift is by continuously measuring the performance of data and models using the right KPIs [Southekal, 2021] and feedback mechanisms. The insight value chain or the process flow and its components discussed above are shown in the figure below.

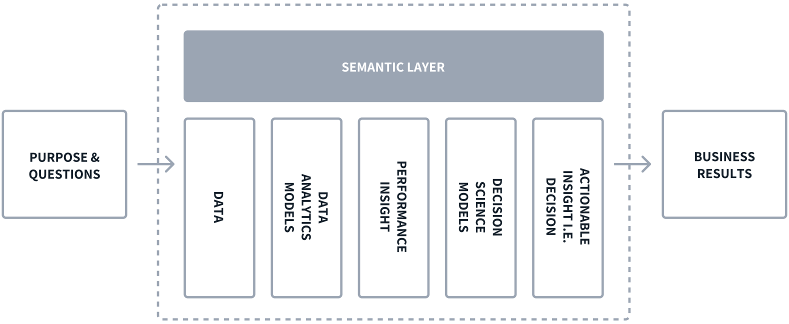


Figure 4: Insight Process Flow

Today, actionable insights are at the heart of every business decision to help companies increase revenue, reduce expenses, and manage risk. The purpose of insights is not just to know; it is also to know and act make informed decisions and improve business performance. The three steps described above can help organizations formulate the right business and data strategy for turning data into performance insights and then into actionable insights for improved business results and performance.

References

- Cambridge, https://dictionary.cambridge.org/dictionary/english/insight, 2022
- Mckinsey1, “Insights to impact: Creating and sustaining data-driven commercial
- growth”, https://www.mckinsey.com/business-functions/marketing-and-sales/our-

insights/insights-to-impact-creating-and-sustaining-data-driven-commercial-growth, January, 2022

- Mckinsey2, "Breaking away: The secrets to scaling analytics", https://www.mckinsey.com/business-functions/mckinsey-analytics/our- insights/breaking-away-the-secrets-to-scaling-analytics, May 2018
- Southekal, Prashanth, "Analytics Best Practices", Technics Publications, April 2020
- Southekal, Prashanth, "Model Drift in Data Analytics: What Is It? So What? Now What?", Forbes, Sept 2021
- Southekal, Prashanth, "Demystifying the Semantic Layer for Smarter, Faster AI and BI", Demystifying the Semantic Layer – White Paper AtScale, Mar 2022
- WEF, "How much data is generated each day? ", https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day- cf4bddf29f/, Apr 2019

Author Biography

Dr. Prashanth Southekal is a Consultant, Author, and Professor. He has consulted for over 80 organizations, including P&G, GE, Shell, Apple, FedEx, and SAP. Dr. Southekal is the author of two books — "Data for Business Performance" and "Analytics Best Practices" — and writes regularly on data, analytics, and machine learning in Forbes and CFO.University. His second book, ANALYTICS BEST PRACTICES was ranked #1 analytics book of all time in May 2022 by BookAuthority. He serves on the Editorial Board of MIT CDOIQ Symposium, Advisory board member at BGV (Benhamou Global Ventures) a Silicon Valley-based (Menlo Park) venture capital firm, and a Data and Analytics Advisor at Evalueserve (CH), Grihasoft (IN), uArrow (SG), Astral Insights (US), and Miles Education (IN). Besides his consulting and advisory pursuits, he has trained over 3,000 professionals worldwide in Data and Analytics. Dr. Southekal is also an Adjunct Professor of Data and Analytics at IE Business School (Madrid, Spain), and CDO Magazine included him in the top 75 global academic data leaders of 2022. He holds a Ph.D. from ESC Lille (FR) and an MBA from Kellogg School of Management (US). He lives in Calgary, Canada with his wife, two children, and a high-energy Goldendoodle dog. Outside work, he loves juggling and cricket.

Chapter 20

The Semantics of the Semantic Layer

David Mariani

Founder and CTO, AtScale

I co-founded AtScale to focus on the challenges of supporting many data analysts working on disparate sets of data managed in a massive lake. We borrowed the term "semantic layer" from the folks at Business Objects, who originally coined it in the 1990s. The term was over 20 years old when we adopted it. So, what is a semantic layer exactly? If you Google the term, the following definition will pop up, which is a darn good definition (Google's highlighted words, not mine):

A semantic layer is **a business representation of corporate data that helps end users access data autonomously using common business terms**. A semantic layer maps complex data into familiar business terms such as product, customer, or revenue to offer a unified, consolidated view of data across the organization.

https://en.wikipedia.org › wiki › Semantic_layer

Semantic layer - Wikipedia

Wikipedia defines a semantic layer as a business representation of data that allows end users to access data autonomously. Everyone can agree that a business-friendly view of data that provides users with self-service access to analytics is desirable — true data democratization. It's easy to see why it is fundamental to scaling data and analytics. The challenge is implementing a semantic layer in a way that just works.

We built the AtScale semantic layer after working on big data from the trenches. We had to deal with the basic challenges of data scalability, query performance, metrics sprawl, complicated data pipelines and shadow business intelligence (BI). While the challenges seemed obvious to us, most of the industry was preoccupied with shifting data gravity to the cloud. With cloud data replatforming in full swing, we are finally seeing attention turning to the last mile of enterprise analytics with the semantic layer topic surging in popularity.

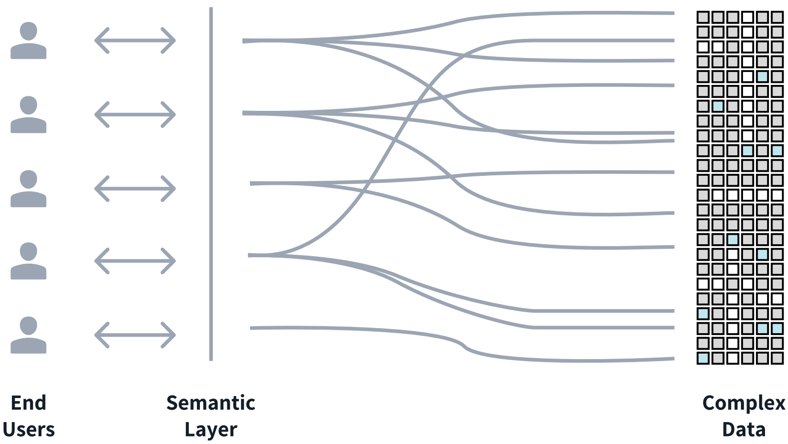


Figure 1: Semantic Layer

The Semantics of a Semantic Layer

Cloud giants like Google and Snowflake, the unicorns like dbt Labs and a host of venture-backed startups are now talking about this critical new layer in the data and analytics stack. Some call it a "metrics layer", or a "metrics hub" or "headless BI", but most call it a "semantic layer". I can't tell you how happy it makes me that the industry is finally recognizing the importance of the semantic layer in a modern, cloud-first analytics stack. I couldn't agree more that a logical, business-friendly view of data is what's needed to make analytics accessible to everyone, not just data engineers and SQL jockeys.

While it might be a matter of semantics, I prefer "semantic layer" over "metrics layer", "metrics hub", or "headless BI." The term "semantic layer" best describes this business-friendly data interface because it covers all types of data and use cases.

For example, the terms "metric store" and "metric layer" ignore the concept of "dimensions" altogether. Look at every BI tool on the market (i.e., Tableau, Looker, Power BI) and they all include measures (or metrics) and dimensions in their interfaces. Metrics measure something, but dimensions (i.e., "product", "time" and "location") categorize data by grouping or aggregating metrics. So, terms using "metric" are confusing and don't map to how these layers will be consumed.

The term "headless BI" is also problematic because it only covers business intelligence use cases. A universal data layer is useful to more than just business analysts and BI. Data scientists need to access a consistent business-friendly interface to data for building and training their models.

More Than Just Metrics

There's a reason independent semantic layers have taken time to come to market — building a semantic layer is hard. Yes, a semantic layer serves as a common metrics store or single source of truth, but there's much more to it than that. For a semantic layer to be viable, it needs to:

1. Support any query tool, interface or protocol with a live connection to data
2. Express the most complex business logic (serve as a digital twin) using a semantic model
3. Deliver queries in under 2 seconds
4. Govern access to data for every query
5. Connect to any backend data store

The 7 Requirements of a Semantic Layer

The following diagram illustrates the core capabilities of a semantic layer:

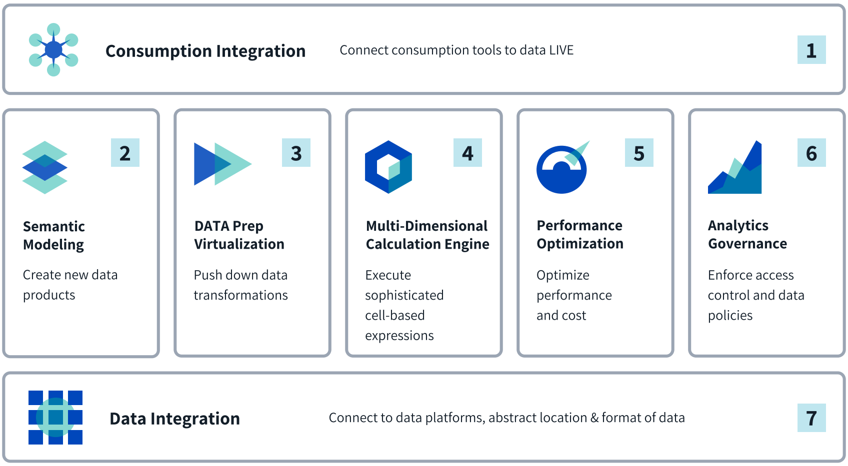


Figure 2: 7 Requirements of a Semantic Layer

1. Consumption Integration

For a semantic layer to be universal, it needs to support "live" query connections for everyone. This means a semantic layer must meet these requirements for connecting users to data:

a. **Multimodal**: It supports a variety of use cases and personas, including the business analyst, data science and application developer.

b. **Open**: It supports a wide range of query tools using their native protocols, including SQL, MDX, DAX, Python, REST, JDBC and ODBC.

c. **Lightweight**: It has a "zero footprint" on end users' machines, so there is no client-side software or plugins necessary to access the semantic layer.

2. Semantic Modeling

The core of the semantic layer is the data model. A business-friendly data layer cannot exist without a map of the logical elements (dimensions, metrics, hierarchies, KPIs) to the physical entities of databases, tables and relationships. To deliver a digital twin of the business, a semantic layer must meet these requirements:

a. **Object-oriented**: It supports reusable models and components to drive a hub and spoke analytics management style.

b. **Multi-source**: It supports the ability to blend data from multiple sources in a model

- Smart push down to data platforms for optimal performance.
- Plug & play pre-built models for known schemas (i.e. SaaS apps, 3rd party datasets)

c. **Programmable**: It supports a CI/CD compatible markup language and shares its metadata via APIs with data catalogs & other tools.

3. Data Prep Virtualization

Data transformations are a necessary requirement for a semantic layer. A semantic layer platform should support virtualized calculations for expressing business logic and be capable of generating multi-pass SQL queries to handle calculations that require different levels of granularity like ratios and weighted averages. The semantic layer engine must deliver these capabilities:

a. **Open**: It supports transformation expressions using the native platform's SQL dialect & MDX expressions.

b. **Multipass**: It supports the ability to perform pre-query & post-query calculations for handling calculations at different levels of granularity.

c. **Virtualized**: It supports inline transformations using direct queries without data movement.

4. Multi-dimensional Calculation Engine

The semantic layer data model must be backed by a scalable, multi-dimensional engine to express a wide range of business concepts in a variety of contexts. The semantic layer engine must deliver these capabilities:

a. **Cell-based**: It supports matrix-style calculations (time intelligence, multi-pass, etc.,) without pre-calculation using a multidimensional expression language like MDX or DAX.

b. **Graph-based**: It supports thousands of dimensions, attributes and metrics using a graphbased query planner.

c. **Dynamic**: It supports "anything by anything" queries with constraints and filters applied at query time.

5. **Performance Optimization**

A semantic layer must accelerate the performance of the underlying data platform to deliver "speed of thought" queries. Without acceleration, a semantic layer will be bypassed using BI tool extracts and imports, defeating the purpose of a semantic layer. A semantic layer must include these capabilities:

a. **Autonomous**: It automatically tunes and improves performance using machine-learning and user query patterns.

b. **In Situ**: It improves performance without moving data outside the native data platform or requiring a separate cluster for managing aggregates.

c. **Adaptive**: It is always learning and improving performance based on system performance and user query behaviors.

6. **Analytics Governance**

There's a range of tools and products that act as a query layer to govern data access. Data governance is a core requirement for a semantic layer and thus must be a core requirement rather than a separate layer. For a semantic layer to satisfy a wide range of data governance, it must deliver these capabilities:

a. **Integrated**: It integrates with corporate directory services (i.e. AD, LDAP, Okta) for user identity management.

b. **Row & Column**: It applies row-level filtering & column-level masking to every query based on user, group and role-based (RBAC) data access rules.

c. **Real-time**: It enforces governance continuously and in real-time at the time of query.

7. **Data Integration**

In modern data and analytics ecosystems, data lives in multiple silos, including on-premises, legacy data warehouses, data lakes, cloud data warehouses and SaaS applications like Salesforce. A semantic layer must be capable of accessing and modeling data across multiple sources with these capabilities:

a. **Data Platform Optimized**: It works with a variety of data silos equally well by supporting native platform dialects and optimizations.

b. **Virtualized & Federated**: It supports the blending of data across multiple data platforms and minimizes data movement with query push-down. Where necessary, a semantic layer may work with data virtualization engines that make it possible to federate large data sets that reside on physically disparate platforms.

c. **Extensible**: It supports a variety of data types, including nested data structures like JSON and supports native data platform extensions.

Not As Easy as It Looks

As you can see, building a semantic layer platform is not simply a matter of defining metrics with a cool new markup language. For a semantic layer to be practical and usable, it needs to:

1. Be capable of expressing your **most complex** business constructs
2. Be able to **perform better** than your underlying data platforms
3. Be able to **connect live** to all your data platforms
4. Be able to **connect to all** your data consumption tools
5. Be able to **govern every query** at the user level
6. Be able to **scale to everyone** in your business

If any of these semantic layer requirements are missing, working groups or individual users will inevitably create a localized data set with a localized semantic model that fits their purpose. And then the semantic layer crumbles. It's binary — it either works 100% or it doesn't work. Therein lies the challenge for anyone building a universal semantic layer from scratch. It's not good enough to deliver an MVP that sort of works and can be enhanced as you go. The MVP is not an MVP — it just needs to work completely on day one, or no one will bother using it.

Author Biography

Dave Mariani is the co-founder and CTO of AtScale. Before AtScale, he was VP of Engineering at Klout and at Yahoo! where he built the world's largest multi-dimensional cube for BI on Hadoop. He is a hands-on technology executive with over 25 years of experience in delivering Big Data, consumer Internet, Internet advertising and hosted services platforms, creating nearly $800M in company exits. David founded AtScale in 2013 to provide access to Big Data to everyone who needs it, regardless of data format, size or the type of tool business users want to use. Enterprises like JPMC, Visa, Bloomberg, United Healthcare, Cigna, Kohls, Home Depot, Wayfair and Toyota all use AtScale to democratize access to data for their employees and partners.

Glossary

- **Aggregation:** Searching, gathering and presenting data.
- **Algorithm:** A mathematical formula or statistical process used to analyze data.
- **API (Application Program Interface):** A set of programming standards and instructions for accessing or building web-based software applications.
- **Artificial Intelligence:** The ability of a machine to apply information gained from experience accurately to new situations so a human would.
- **Big Data:** Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. Big data sets are characterized by 3Vs, i.e. volume, velocity, and variety.
- **Business Intelligence:** The general term used to identify, extract, and analyze multidimensional data.
- **Change Management:** Change management is the discipline that guides how we prepare, equip and support individuals to adopt change to drive organizational success and outcomes.
- **Cloud Computing:** A distributed computing system hosted and running on remote servers and accessible from anywhere on the internet.
- **Correlation.** Correlation is a statistical technique that shows how strongly two variables are related. For example, height and weight are correlated; taller people are heavier than shorter people.
- **Cube.** A data structure in OLAP systems. It is a method of storing data in a multidimensional form, generally for reporting. In OLAP cubes, data (measures) are categorized by dimensions. OLAP cubes are often pre-summarized across dimensions to drastically improve query time over relational databases.
- **Dashboard:** A graphical representation of KPIs and Visuals.
- **Data:** Data is a set of fields with quantitative or qualitative values in a specific format.
- **Data Analytics:** Answering business questions using data. Businesses typically use the three analytics: Descriptive, Predictive and Prescriptive Analytics.
- **Data Architecture:** It is the mechanism in which data is collected and how it is stored, arranged, integrated, and used in data systems and in organizations.
- **Data Governance:** A set of processes or rules that ensure data integrity and that data

management best practices are met.

- **Data Integration:** Combining data from different sources and presenting it in a single view.
- **Data Lake:** A large repository of enterprise-wide data in raw format – structured and unstructured data.
- **Data Mart:** The access layer of a data warehouse used to provide data to users.
- **Data Mining:** Finding meaningful patterns and deriving insights in large sets of data using sophisticated pattern recognition techniques. Data miners use statistics, machine learning algorithms, and artificial intelligence techniques to derive meaningful patterns.
- **Data Product:** A data product is the application of data for improving business performance; it is usually an output of the data science activity.
- **Data Science:** A discipline that incorporates statistics, data visualization, computer programming, data mining, machine learning and database engineering to solve complex problems.
- **Data Warehouse:** A repository for enterprise-wide data but in a structured format after cleaning and integrating with other sources. Data warehouses are typically used for conventional data (but not exclusively).
- **Database:** A digital collection of data and the structure around which the data is organized. The data is typically entered into and accessed via a database management system.
- **Descriptive Analytics:** Condensing big numbers into smaller pieces of information. This is like summarizing the data story. Rather than listing every single number and detail, there is a general thrust and narrative.
- **ETL (Extract, Transform and Load):** Extracting raw data, transforming it by cleaning/enriching the data to fit operational needs and loading it into the appropriate repository for the system's use.
- **Hypothesis.** A hypothesis is an assumption, an idea, or a gut feeling proposed for validation so it's tested to see if it might be true.
- **Insight.** It is the understanding of a specific cause and effect within a specific context. In this book, the terms insight and information are used interchangeably.
- **KPI.** A key performance indicator (KPI) is a measurable value that demonstrates how effectively the entity achieves key objectives or targets.
- **Machine Learning:** A method of designing systems that can learn, adjust and improve based on the data fed to them. Using statistical algorithms fed to these machines, they

learn and continually zero in on “correct” behavior and insights and they keep improving as more data flows through the system.

- **Metadata**. Any data used to describe other data — for example, a data file’s size or date of creation.
- **Multicollinearity**. It is a state of high intercorrelations among the independent variables, indicating duplicate or redundant variables in the analysis. It is, therefore, a type of disturbance in the data, and if present in the dataset, the insights derived may not be reliable.
- **Online analytical processing (OLAP)**. Analyzing multidimensional data using three operations: consolidation (the aggregation of data?), drill-down (the ability for users to see the underlying details), and slice and dice (the ability for users to select subsets and view them from different perspectives). OLAP systems are used in BI reports.
- **Online transactional processing (OLTP)**. Providing users with access to large amounts of transactional data so they can derive meaning from it. OLTP systems are used in transactional reports
- **Predictive Analytics:** Using statistical functions on one or more data sets to predict trends or future events.
- **Prescriptive Analytics:** Prescriptive analytics builds on predictive analytics by including actions and making data-driven decisions by looking at the impacts of various actions.
- **Regression Analysis:** A modeling technique used to define the association between variables. It assumes a one-way causal effect from predictor variables (independent variables) to a response of another variable (dependent variable). Regression can explain the past and predict future events.
- **SQL (Structured Query Language):** A programming language for retrieving data from a relational database.
- **Semantic Layer:** The semantic layer represents data that helps different business endusers discover and access the right data efficiently, effectively, and effortlessly using common business terms.
- **Systems of Insight (SoI)**. It is the system used to perform data analysis from the data combined from the SoR or transactional systems.
- **System of Record (SoR)**. The authoritative data source for a data element. To ensure data integrity in the enterprise, there must be one — and only one — system of record for a data element.
- **Stakeholder:** Individuals and organizations who are actively involved in the initiative,

or whose interests may be positively or negatively affected because of execution or successful completion of the initiative.

- **Structured Data:** Data organized according to a predetermined structure.
- **Unstructured Data:** Data that has no identifiable structure, such as email, social media posts, documents, audio files, images, videos, etc.

Acronyms and Abbreviations

- **3DM** – Data-Driven Decision-making
- **AI** – Artificial Intelligence
- **API** - Application Programming Interface
- **CDO** - Chief Data Officer
- **D&A** - Data and Analytics
- **DLC** – Data Lifecycle
- **EDG** - Enterprise Data Governance
- **IT** - Information Technology
- **KPI** - Key Performance Indicator
- **LoB** - Line of Business
- **ML** - Machine Learning
- **MLR** – Multiple Linear Regression
- **NLP** – Natural Language Processing
- **OLTP** - Online Transaction Processing
- **OLAP** - Online Analytical Processing
- **PII** - Personally Identifiable Information
- **RBAC** – Role-Based Access Control
- **SaaS** - Software-as-a-Service
- **SoI** - System of Insights
- **SoR** - System of Record
- **SSA** – Self Serve Analytics

"The authors in the chapters ahead are colleagues and thought leaders in the field of data with well over 450 years of cumulative experience."

Kamayini Kaul

Global Head Information Insights
and Analytics, CSL Behring

"This book covers foundational principles and examples of the world class capabilities that enable and accelerate the value of data and analytics."

Andy Keller

Vice President of Enterprise Data,
Cardinal Health

Brought to you by

atscale.com

In partnership with

cdo magazine