



Organizing Enterprise Data Science

Recipes from the Global 2000 and Beyond



Contents



Participating companies

The contents of this report reflect conversations conducted with data science leaders across numerous industries and geographies. Conversations occurred by phone and in person from 2019-2020. Domino would like to thank the following individuals and companies for their time and input to this paper.



Brian Loyal
Cloud Analytics Lead



Naveen Singla
Vice President, Data Science



Amanda Waldo
Enterprise Data Science Talent & Community Lead, Data Science Center of Excellence



Michael Xiao
Divisional Vice President, Enterprise Data Science



Sebastien Conort
Chief Data Scientist



John K. Thompson
Global Head, Advanced Analytics & AI



Randi Ludwig
Senior Manager, Applied Data Science



Bruce Campbell
Director, Data Science



Vince Stuntebeck
Senior Director, Enterprise Data & Analytics



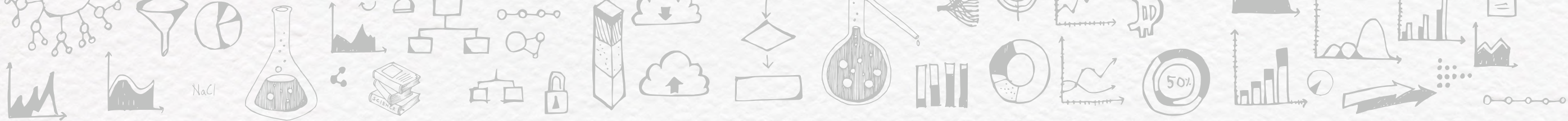
Antoine Ly
Head, Data Science



Matt Cornett
Director, Data Science



IT Technical Director
Fortune 500 Insurer



Executive summary

Everyone is talking about machine learning and AI. Most companies are dabbling in it. But few are deploying it on an enterprise scale. In fact, [PwC](#) found that only 18 percent of companies it surveyed had implemented AI in multiple areas.

For those that do, the rewards can be game-changing. Leading adopters achieve [higher revenue increases, and more significant cost decreases than other companies using AI](#), according to McKinsey & Company.

While the necessity to embed AI into the business is clear, the road to get there isn't. One question many data science leaders wrestle with is how to organize data science teams to achieve the greatest impact.

Do you centralize the function to build out economies of scale?

Should data scientists be federated, working in the trenches with business staff where they can solve real problems faster?

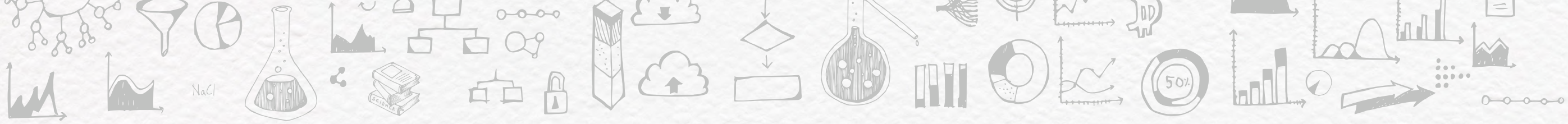
Or is a hybrid model that falls somewhere in between the best path to success?

In our conversations with nearly a dozen industry leaders building model-driven businesses, we found that there's no one-size-fits-all answer. But there are a set of practices that leaders use to successfully:

- **Create a discipline** for managing the people and cultural changes necessary to embrace data science
- **Establish scalable and repeatable processes, including metrics to measure success**, across the end-to-end data science lifecycle
- **Ensure data science teams have the right technology foundation** to foster productivity and collaboration

Ultimately, these leaders demonstrate that regardless of how you organize data science, you need a strategy for scale and a path to get there quickly. Those lacking one will fall behind and likely struggle to deliver business impact.

In this report, we break down best practices across all three areas (discipline, process, technology) that these leaders of high-performing global data science teams shared with us. Whether you're early in your journey or well underway and seeking to strengthen the impact of existing efforts, their insights can help you chart the right course for your organization.



Creating an enterprise data science discipline

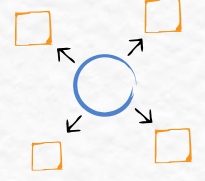
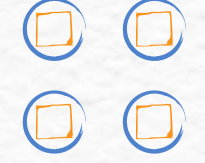
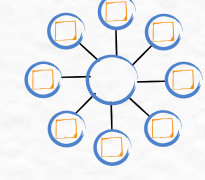
To scale data science, organizations need to establish a discipline for aligning people and effectively managing change that embeds data science within the enterprise’s DNA. But the “right” way to structure teams, and whether or not a formal Center of Excellence (COE) should be created, may be different for every company.

Bayer Crop Science initially used a decentralized organizational structure, and then shifted to a hub-and-spoke model using a COE to provide teams with foundational capabilities and to improve governance. Red Hat has several “centers”—including an Enterprise Data and Analytics team, an AI COE, and a Data Science Council—all of which work in concert to help the business build data science at scale. And Dell Technologies found success in establishing a Data Science Council to scale data science across 11 business units within its Support and Deployment Services Division.

"What we learned was there were a lot of advantages to having a group focused on problems that span the org and best practices."

Brian Loyal, Cloud Analytics Lead, Bayer Crop Science



Data science orgs come in all shapes and sizes — and they all benefit from an enterprise data science strategy			
Organizational Structure	Primary benefits	Main drawbacks	What an enterprise data science strategy can do
<p>Centralized</p> 	Greater consistency and knowledge sharing with single leader overseeing talent, infrastructure and tooling, and best practices	Increased potential for being disconnected from business priorities	Ensures development of programs and processes for regularly connecting with business teams and deepening domain knowledge
<p>Federated</p> 	Strong alignment with business processes and priorities, and deep domain expertise	Increased likelihood for duplication of effort, isolation of practitioners, and silos of expertise	Provides center to drive community building, portfolio management, and best practices for achieving greater economies of scale
<p>Hub-and-Spoke</p> 	Offers best of both centralized and federated approaches	Requires significant coordination and agreement between COE and embedded teams on what each area’s role should be	Unites business and data science leaders to strengthen capabilities and optimize efficiency

Yet, despite the differences, we found these organizations had much in common in how they determined their strategy.

First, even the most distributed of organizations we spoke with viewed consolidation of some data science capabilities and an enterprise data science strategy as critical to becoming model-driven.

Second, while their approaches varied, most leaders did four things consistently when choosing how (and what) to consolidate and assembling their teams: aligning design with corporate culture, identifying a tipping point, convening the right resources, and choosing their home wisely.

1. Aligning organizational design with corporate culture: Corporate culture significantly influenced how leaders approached their data science strategy. For example, global reinsurer SCOR emphasizes knowledge and the ability to anticipate future challenges as a core component of its culture. This led Antoine Ly, the company's Head of Data Science, to build an enterprise COE that could serve as a bridge between local and global knowledge across distributed data scientists in the U.S., Europe, and Asia. "We are always building on top of what has been done before," he said.

In the case of Red Hat, the company's fierce commitment to open source technologies places a high value on community-led efforts. "Trying to formalize something like a COE and defining the responsibilities is a lot of work at Red Hat," explained Brue Campbell, the company's Director of Data Science. "To do this work organically is much easier."

As a result, rather than having a single COE guiding the business, Red Hat has several central teams working in partnership to transform data science into a core capability. When multiple hubs exist, leaders say coordination is critical. At Red Hat, for instance, the CIO (who owns the Enterprise Data & Analytics team) and the CTO (who owns the AI COE) meet regularly to discuss how the teams can complement each other.

2. Identifying a tipping point. Getting support for building a "central team" can sometimes be challenging. As Red Hat's Senior Director of Enterprise Data & Analytics Vince Stuntebeck said, "When you fire one of these things up, often one of the first questions is: 'Are you going to tell me how to do my work?'"

As a result, many leaders we spoke with used a burning issue, inflamed by rapid growth, as the tipping point to start the conversation about consolidating some data science responsibilities into a central hub. By doing so they found they could more easily gain support for strategy changes, demonstrate value quickly, and build trust with the business teams—all crucial to long-term success.

Many leaders we spoke with used a burning issue, inflamed by rapid growth, as the tipping point to start the conversation about consolidating some data science responsibilities into a central hub.





What's in a name?

In some cases, we found companies were hesitant to refer to an enterprise team as a COE. One concern was the negative perceptions that practitioners have when COEs are put in place—such as lack of alignment with business goals or having someone tell them how to do their work—and whether the name might slow acceptance. Others viewed the name itself as unnecessary. As Matt Cornett, Transamerica's Director of Data Science, put it: "The phrasing seems there's something old about it, like you're hanging a banner in a hospital lobby."

Red Hat's "centers" for data science/AI excellence



Enterprise Data & Analytics

Provides centralized tooling, data warehousing capabilities, and services, such as building analytical outputs and establishing datasets for embedded practitioners.



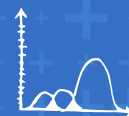
Data Science Council

Participants include data science managers and senior data scientists from the business units along with leaders from the Enterprise Data & Analytics team. Meets regularly to drive DevOps excellence, build academic collaborations, and address other cross-functional needs.



Analytics Community of Practice

Includes hundreds of practitioners who focus on community building and project sharing across divisions.



AI COE

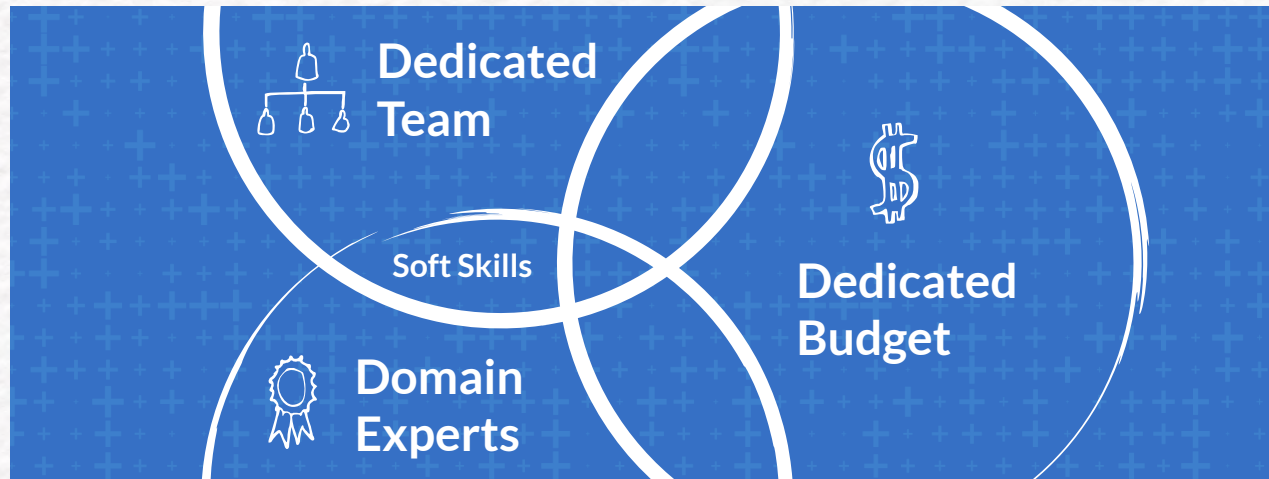
An innovation effort using Red Hat, open source, and certified Independent Software Vendor (ISV) Ecosystem Partner technologies to demonstrate to customers Red Hat's advanced capabilities for analytics and AI.

That tipping point typically fell into one of the following categories:

- **Talent gaps.** BlueCross BlueShield of Illinois, Montana, New Mexico, Oklahoma, and Texas, one of the largest non-profit health insurers in the United States, launched its COE to address the need to hire 35 new data scientists and engineers. The COE created a hiring committee, secured a central recruitment budget, established relationships with key universities, and standardized interview and onboarding processes to attract the best talent. The business units actively participated in the process, helping build trust and drive support as the COE expanded its efforts to address process and technology challenges.
- **Inefficient processes,** such as duplication of effort or a lack of a clear path to production, that were holding the business back from realizing the impacts of data science. For example, at Dell Technologies, Randi Ludwig launched their Data Science Council after discovering commonalities across challenges that multiple business units were trying to solve. At BNP Paribas Cardif, early efforts focused on packaging reusable assets to help local teams more rapidly deliver new models.
- **Growing risk** as distributed teams manage governance and compliance independently. Take Bayer Crop Science, for example. As pockets of data science began sprouting up across the company five years ago, executives wanted to ensure they had the right guard rails in place.
- **New business opportunities.** When CSL Behring hired John K. Thompson to build out an enterprise data science capability, he met with more than 1,000 people (from C-level executives to frontline factory workers) across North America, Western Europe, and Australia to discuss how advanced

analytics and AI could drive new business opportunities. He also launched the company’s first Data Science Summit. “We had people from all over CSL talking about what they were doing, how they were doing it, and what they wanted to do,” he said. “We made it very clear that we were happy to help do projects, as consultants or a peer level organization, but we were also happy to support the people already doing projects.” Ultimately, as expertise and interest grew across their business, the team moved from a fully centralized approach to a hybrid model with a formal COE.

Your strategy should determine...	Key options to consider...
Reporting Structure	<ul style="list-style-type: none"> • C-Suite • IT • Line of Business/Function
Charter	<ul style="list-style-type: none"> • Create an enterprise strategy for scale • Define standards and practices • Develop talent and community • Manage infrastructure, tooling, and vendor partnerships • Engage in model development
Team Composition	<ul style="list-style-type: none"> • Hire externally vs. upskill internally • Domain skills • Soft skills
Scope of Authority	<ul style="list-style-type: none"> • Select an executive sponsor • Determine rules of engagement for: <ul style="list-style-type: none"> • Funding • Tone (when to lead by example versus institute formal guidelines and standards)
Success Metrics	<ul style="list-style-type: none"> • Identify metrics to track <ul style="list-style-type: none"> • Top-line impact • Bottom-line impact • Use case specific metrics • Team performance • Set regular cadence for reviewing and refining target metrics



3. Pulling together the right resources: Staffing and budgeting requirements will ultimately be dictated by many factors, including a company’s overall data science maturity level, the type of “center” being built, and the work the center will do. For example, will it take a hands-on role in model development or only develop partnerships with data science practitioners embedded within business units? However, most leaders emphasized the need for the following:

- **A dedicated team to fulfill the organization's mission.** Many teams had between 20 and 80 staff members. (And in fact, outside of these conversations we’ve identified Domino clients with more than 100 team members in their COEs.) Building out the team can be accomplished by both hiring externally as well as upskilling internally. In hiring externally, CSL Behring’s John Thompson collaborates with leading colleges and universities to identify rising stars to join their COE. All candidates first participate in several internships with the company to confirm there’s a good fit for both the individual and the team. To upskill internally, the global team at SCOR is creating a curriculum that identifies a minimal skill set an employee would need to become a core team member.

- **Domain experts who can effectively interface with business units.** For example, a key job profile within Transamerica's Enterprise Data and Analytics team is an “engagement manager” who builds relationships with business units. Bayer Crop Science's COE has a horizontal team that builds foundational capabilities and a vertical team with experts aligned to key business units, such as Research and Development and Supply Chain. And BCBS’s Michael Xiao looks for practitioners who understand the daily challenges of those working in data science across the business.
- **"Soft skills."** Everyone, leaders say, from the head of the COE on down, needs to have collaboration and communication skills to deliver on their enterprise data science strategy. "It's more about the culture that you build as opposed to the things that you do," explained Transamerica's Matt Cornett. "If you don't have the right people with a collaborative mindset, then you're not going to get them to collaborate after the fact."
- **Dedicated budget to support initiatives.** Unfunded programs can spend a great deal of time seeking funding to support critical initiatives. Many operated with both a dedicated budget along with additional funding from Business Units to support specific projects. For example, at SCOR, the global team has a budget for global tools and infrastructure and funds staff salaries with a mix of global and local funding. "Our core group of resources is managed at the global level, so it brings us a lot of flexibility in terms of headcount," said SCOR's Antoine Ly. At CSL Behring, the COE has dedicated funding but also looks to the Business Units to fund the purchase of data specific to their areas.

"You need to find the right balance for your organization in terms of speed, agility, and flexibility versus consistency, standards, and scalability."

– Vince Stuntebeck, Senior Director,
Enterprise Data and Analytics, Red Hat



4. Choosing a home wisely: Often when starting a central data science team, there's debate about where it should report: the C-suite, IT, or even a specific line of business or function. Companies we spoke with represented a mix:

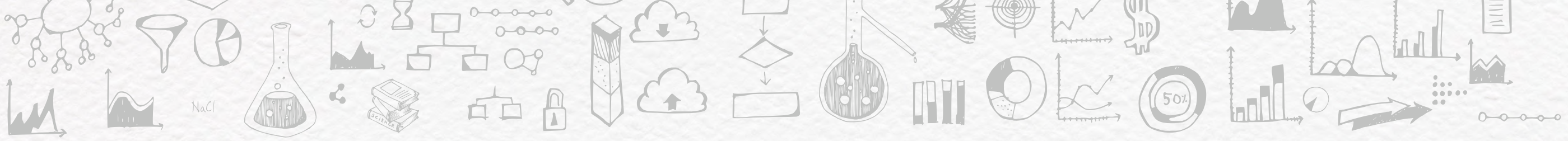
- **The C-suite:** Transamerica's Enterprise Data and Analytics team reports, for instance, to the company's Chief Administrative Officer. A Fortune 500 insurer's COE reports to the company's Chief Financial Officer. BNP Paribas Cardif's COE reports to the Chief Analytics Officer. And CSL Behring's COE reports to the Chief Digital Officer. For these organizations, alignment with the C-suite ensures executive support and independence. As Transamerica's Director of Data Science, Matt Cornett, put it: "You're not politically connected or beholden to any specific functional area."
- **IT:** Several organizations we spoke with, including Red Hat and Bayer Crop Science, placed their data science hub or COE under IT. According to Red Hat's Vince Stuntebeck, having a similar mandate and being near the data team are two key advantages. "IT is already providing technology to the organization and serving a central role which lines up well with our team's mandate," he said.

- **Specific lines of business (LOBs) or functions.** One multinational company we know placed its COE under its Software business. Similarly, Dell Technologies' Randi Ludwig created Data Science Councils for both practitioners and managers within the company's Support and Deployment Division to serve as its "center". For these companies, placing a hub within the LOB or function ensures greater accountability to the business and alignment with business needs.

How do you decide? Most leaders acknowledged there were pros and cons with each approach. For many, the decision comes down to support and sponsorship. "At the end of the day, it doesn't matter where the Center of Excellence is as long as whoever you report to has full buy-in into giving you resources that you need to execute on everything," said Michael Xiao, Divisional Vice President of Enterprise Data Science at BlueCross BlueShield.

The case for reporting directly to the C-suite

CSL Behring's John K. Thompson advises companies that having a central AI team or COE report to the C-suite is critical for enabling breakthrough innovations. "The COE is an environment for creative professionals," he said. "It's not a technology team. We're not building the next ERP or CRM system. We're taking many sources of data, and we're mashing them together, and we're trying to come up with new insights and what the possibilities are. Thinking about data science this way allows data scientists to investigate things that can change the way the business operates."



Establishing scalable and repeatable processes

To transform data science into a core capability, leaders need to ensure it operates like a well-oiled machine. This spans every part of the data science lifecycle—from effectively supporting the people who will build models, to providing a framework for innovation and speed while ensuring a smooth path to production.

While it's a significant undertaking, leaders shared six key practices that had helped them achieve their goals faster.

- 1. Secure executive support.** Executive sponsorship, in particular, can help minimize resistance, open doors, and ensure funding. At Dell Technologies, for instance, executive support has opened doors for the Data Science Council, in one case enabling the team to bring on additional data engineers during a time of constrained budgets, recognizing the critical importance of these skill sets.

Many advised securing executive support (both at the C-level and within each line of business) from the outset. One multinational company we know shared that they hadn't involved key partners across the organization when they created their COE and said this would be its downfall.

- 2. Focus on evolution, not revolution.** A consistent theme we heard from leaders was also to start small, think about value creation, and expand over time. Bayer Crop Science's COE expanded its charter to implement collaboration platforms, codify best practices for creating robust data products, and provide community building activities only after it first addressed governance—a pressing and high profile issue for executives. BlueCross BlueShield's COE widened its aperture over time beyond talent recruitment to address community building, collaboration, and technology and infrastructure, promoting preferred technologies and providing a standard infrastructure that teams could use as a launch pad for development. At Red Hat, leaders regularly examine how their different data science “centers” work together, deliberating on what additional efforts might drive further economies of scale.
- 3. Lead by example, not by edict.** It's not uncommon for a centralized team to take an authoritarian stance, dictating how things should be done. But this is an approach that's ripe for failure. "I've never actually seen that work," said BlueCross BlueShield's Michael Xiao.

Instead, many leaders found that they were most successful when they led by example. BlueCross BlueShield structures delivery teams the way they would like the business units to eventually set them up, with a product manager, data engineering, and data scientists working hand in hand alongside

analysts. At SCOR, rather than dictating model development best practices, COE staff actively engage in data science initiatives in different regions, sharing best practices and knowledge as they work side by side with colleagues in each market. This approach has enabled SCOR's COE to build a strong reputation as a valuable partner, further the adoption of data science within the company, and help teams develop models in a quarter of the time of previous efforts.

4. **Build consensus.** When you do need to set guidelines, leaders advise that it's easier to propagate best practices throughout an organization when everyone feels they've had an opportunity to weigh in. As a result, leaders emphasized the need to communicate regularly and actively gather feedback from the business. For example, at Red Hat, the Enterprise Data & Analytics team regularly engages with the business to hear what challenges they face and processes they would like standardized. SCOR's global team asks business owners to provide feedback on any recommendations the global team is developing. BlueCross BlueShield's COE chose a data science leader from one of the company's business units to lead the hiring process. "Share power and make everyone feel like they're involved and then they're much more willing to work within whatever framework you create," said Michael Xiao at BlueCross BlueShield.
5. **Measure success.** As with any business initiative, the ability to show quantifiable results, especially revenue increases or cost savings, is critical to both gaining and maintaining support. Beyond traditional top- and bottom-line benefits, leaders recommend the following:
 - **Regularly review and refine how you measure success.** For example, Bayer Crop Science's COE conducts an annual "state of the world survey" with anyone doing or enabling data science work in the company to assess

what's working and what's not. According to Amanda Waldo, Enterprise Data Science Talent and Community Lead, the survey has generated surprising feedback that led to changes in how they did things.

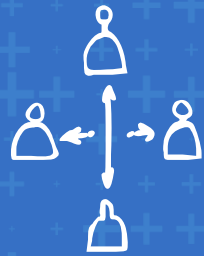
- **Think about the different ways to track value.** At Bayer Crop Science, COE leaders found tracking return on investment for models within its Research and Development division can be difficult as it can take up to seven years before a product makes it to market and delivers a financial return. Instead, they focus on "hit rate" — how many products that enter the pipeline are successful at each step. "We want that number to go up year-over-year, and we've seen a positive trend over the past few years," said Naveen Singla, the company's Vice President of Data Science.

"Make everyone feel like they're involved and then they're much more willing to work within whatever framework you create."

—Michael Xiao, Divisional Vice President,
Enterprise Data Science, BlueCross BlueShield

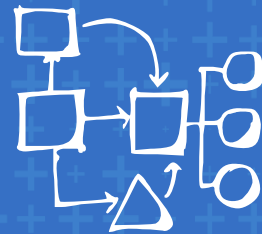


Bayer Crop Science's COE focuses on four areas



Talent and community

Attracting, retaining, developing and rewarding data science talent



Portfolio management

Tracking data science projects enterprise-wide and identifying synergies for better collaboration and resource utilization



Technology

Providing platforms to facilitate development and collaboration at scale, track progress, and streamline processes for moving projects through the pipeline



Best practices

To reduce duplication of effort and eliminate process bottlenecks

- **Assess team performance.** For example, the COE for a Fortune 500 insurer clocks speed to deploy models and the number of "bad models" to understand the impact of its governance and model lifecycle management efforts. Likewise, Red Hat tracks metrics related to adoption of best practices and tools, such as usage of executive dashboards, decreased business spend on local tooling, and reduction in data latency, to measure the success of its technology and infrastructure initiatives.

6. Set parameters for engagements. In many cases, enterprise teams took a hands-on role, dispatching data scientists to business units to help build models. While budget and sponsorship were sometimes factors, most leaders emphasized that they typically decided which projects to take on and how to prioritize work based on a use case's strategic importance, global impact, and business value and feasibility.

At BlueCross BlueShield, Michael Xiao's team is trying to build up its expertise in consumer-focused marketing as the company expands its focus beyond its traditional base of large employers.

Bayer Crop Science's COE supports foundational data science projects in predictive pricing, environmental classification, genomics, and operational efficiency that multiple business units can benefit from. "We work on the projects that enable business partners to meet their revenue and cost reduction goals," said Brian Loyal, Cloud Analytics Lead for Bayer Crop Science.

At SCOR, Antoine Ly's team takes an active role to support use cases outlined in the company's strategic plan, global initiatives affecting all SCOR markets, and client-facing services that can be used by multiple clients. "We have to make sure that the COE can benefit not just one region but other regions at the same time," he said.

What projects don't enterprise teams take on? There were some companies where everything was fair game. But most leaders we spoke with drew the line at use cases with a very narrow scope and usage, like local simple marketing scores. Indeed, even then, leaders made exceptions. Bayer Crop Science's COE, for instance, typically doesn't work on LOB specific projects, but will in cases when the business unit doesn't yet have a data science team in place. At SCOR, even when the global data science team can't provide development support, Antoine Ly makes sure to still share any tools that might facilitate the work and offer guidance on best practice approaches his team has used.

"To share knowledge and practices,
we have to share tools."

—Antoine Ly, Head, Data Science, SCOR



According to Dell Technologies' Randi Ludwig, an enterprise data science platform has enabled data scientists across diverse teams to more easily share best practices and "level up" their performance.

Likewise, at CSL Behring, ensuring reproducibility and reuse is a priority. "We are always designing models and applications with reuse in mind," said CSL's John Thompson. "To enable reuse, you have to have a method for storing and managing application code and models. We are refining and developing the infrastructure to manage these assets now."

Conclusion

How will you scale data science? Designing a best-in-class enterprise data science organization requires a strategy and vision for bringing together the people, processes, and technologies that can maximize productivity and efficiency while enabling innovation and speed.

As these organizations show, data science leaders must be prepared to align their efforts with their company's culture, effectively manage change and unite silos, and build trust and buy-in across the business. It's a tall order, but in our conversations with global leaders we find these practices, along with a commitment to flexibility, can enable organizations to more rapidly and successfully set their strategies for enterprise scale.

Moreover, these leaders' work serves as an important reminder for those just getting started: Time is of the essence. Leaders are already well underway in transforming their organizations to become model-driven. Those that don't move quickly can easily lose their competitive edge.

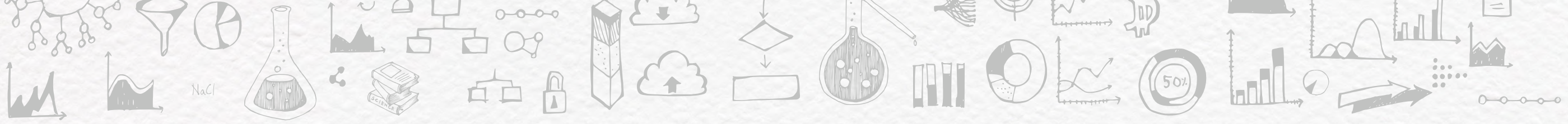
Get started now

1. Use the **enterprise data science strategy checklist** on the next page to help you start thinking about the best approach for your organization.
2. Take the **Data Science Lifecycle Assessment** to receive a detailed appraisal of where your organization currently stands and to identify opportunities to improve data science processes.
3. Read the **Practical Guide to Managing Data Science at Scale** for more insight on managing data science projects and portfolios at high volume and velocity.

About Domino Data Lab

Domino is the enterprise data science management platform trusted by over 20% of the Fortune 100. Our products enable thousands of data scientists to develop better medicines, grow more productive crops, adapt risk models to major economic shifts, build better cars, improve customer support, or simply recommend the best purchase to make at the right time.

Many of the companies profiled in this report are Domino customers.



Enterprise data science strategy checklist

1

Discipline: Choose the right format and focus

- Does your proposed strategy align with your company's culture?
- Is there a pressing issue business leaders need solved?
- What prevailing Business Unit sentiment might impact your strategy?
- Have you identified who can provide the support and resources you need?

2

Process: Implement your strategy effectively

- Have you identified ways to quickly deliver value to the business?
- Have you secured executive support?
- Are you giving your business a chance to weigh in?
- Do you have a plan to measure your team's success?
- Have you set parameters for engagements with Business Units?

3

Platform: Deliver the right foundation

- Are you providing tools and technologies that drive economies of scale across the enterprise?
- Do data scientists have the freedom and flexibility to choose the tools best suited to their needs?
- Can data scientists easily reproduce models, share knowledge and build off each other's work?

