BARC

BARC Spotlight

A Blueprint for Modern Data Science:

Unifying Open Source and Enterprise Governance

Author: Shawn Rogers

Publication: August 2025

Research provided by

posit®

Table of Contents

Executive Summary

Enterprise data science has evolved from a challenge of talent acquisition to a comprehensive effort to optimize, scale, and govern workflows across the organization. While open-source tools like R and Python remain essential for their flexibility and broad ecosystem support, they are often used in disconnected environments with limited integration into enterprise governance frameworks.

These silos introduce operational risks and inefficiencies, particularly in regulated industries. To overcome these challenges, companies need to adopt centralized, collaborative platforms that embrace open-source tools like R and Python while embedding data access controls, model management, and security oversight. BARC Research findings show that a significant number of organizations still lack formalized governance and data use policies, underscoring the need for smarter, more integrated solutions that better scale enterprise data science.

The State of Enterprise Data Science

Data science has delivered deep insights and value to enterprise companies for decades. Data-intense industries identified the value of data science insights early on. Financial services, pharmaceutical manufacturing, and healthcare differentiated themselves competitively by gaining significant advancements via data science, combined with traditional analytics.

Data scientists were once a difficult skill set to hire, as outlined in the 2012 Harvard Business Review article, *Data Scientist: The Sexiest Job of the 21st Century*[1]. The challenge has since evolved from hiring to optimizing, integrating, collaborating, and scaling data science practices. Historically, data scientists functioned in silos, leveraging powerful open-source solutions, but often operated outside of the enterprise governance and oversight landscape.

Eliminating these siloed work environments and addressing the governance gap are the next maturity stage for data science programs. Foundational to this strategy is a mandate to enable data scientists to continue relying on open-source tools and frameworks. Providing a platform that seamlessly surrounds their work with governed and secure access to data, integrated tools, and collaboration brings the data science team closer to the business and aligns them to governance guidelines.

[1] Thomas H. Davenport and DJ Patil, "Data Scientist: The Sexiest Job of the 21st Century," *Harvard Business Review*, October, 2012, https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

Understanding the Challenges to Optimize Data Science

Disconnected Environments, Unaligned Governance

A significant challenge for successful data science practices is the prevalence of siloed environments and governance disconnects. Open-source tools play a critical role in the daily life of a data science professional, offering flexibility, lower costs, and native support for critical programming languages like R and Python alongside integrated development environments (IDEs) such as RStudio, VS Code, and Jupyter. While these tools are often highly customizable and provide easy access to powerful libraries and methods that streamline development, they are not always well integrated for complete end-to-end workflows, hindering efficiency and delaying projects. Additionally, in these specialized environments, organizations are often forced to build bespoke governance and security frameworks, creating process gaps and increasing risk.

Bringing data science teams together on a unified platform that maintains the flexibility of open-source solutions while wrapping the work in a collaborative, secure and governed environment is critical. This approach reduces many manual processes, promotes reuse, standardization, and integrates with enterprise governance, enabling data science to scale in a valuable way.

Operational Risks and Governance Gaps

Working outside the scope of enterprise governance processes creates significant operational risk, particularly for highly regulated industries. According to BARC research, 46% of respondents have yet to formalize security and compliance standards, and 47% struggle to formalize their data access/use polices[2]. These critical challenges drive a need for smarter, more unified approach to data science.

Integrating data science with centralized governance layers is a seamless way to support creativity without slowing the overall process. Think of your childhood school playground: everyone knew there was a fence in the distance, but it didn't get in the way of having fun and it kept everyone safe. By staying connected to enterprise governance tools, it's easier to control updates, monitor system performance, and configure models. This non-restrictive environment keeps data scientists happy and produces better work, while IT has oversight to ensure the work is secure.

[2] BARC Research, *Preparing and Delivering Data for AI: Adoption Trends, Requirements, and Best Practices* (May 2025).

Artificial Intelligence (AI), Machine Learning (ML) Creates New Challenges

AI innovation is growing faster than any technology trend in the last 30 years. This surge in innovation has instantly applied pressure to integrate data science tools and workflows seamlessly with the data on platforms like Azure ML, Amazon SageMaker, Snowflake, Databricks and others, fueling the demand for connecting open-source data science tools and application development environments in a single, governed environment. Staying siloed impedes the innovation process, slows production, and exposes companies to greater operational risks. These siloes require data scientists to find their own path to integration of AI that often results in undocumented, proprietary processes that lack security, governance and standardization.

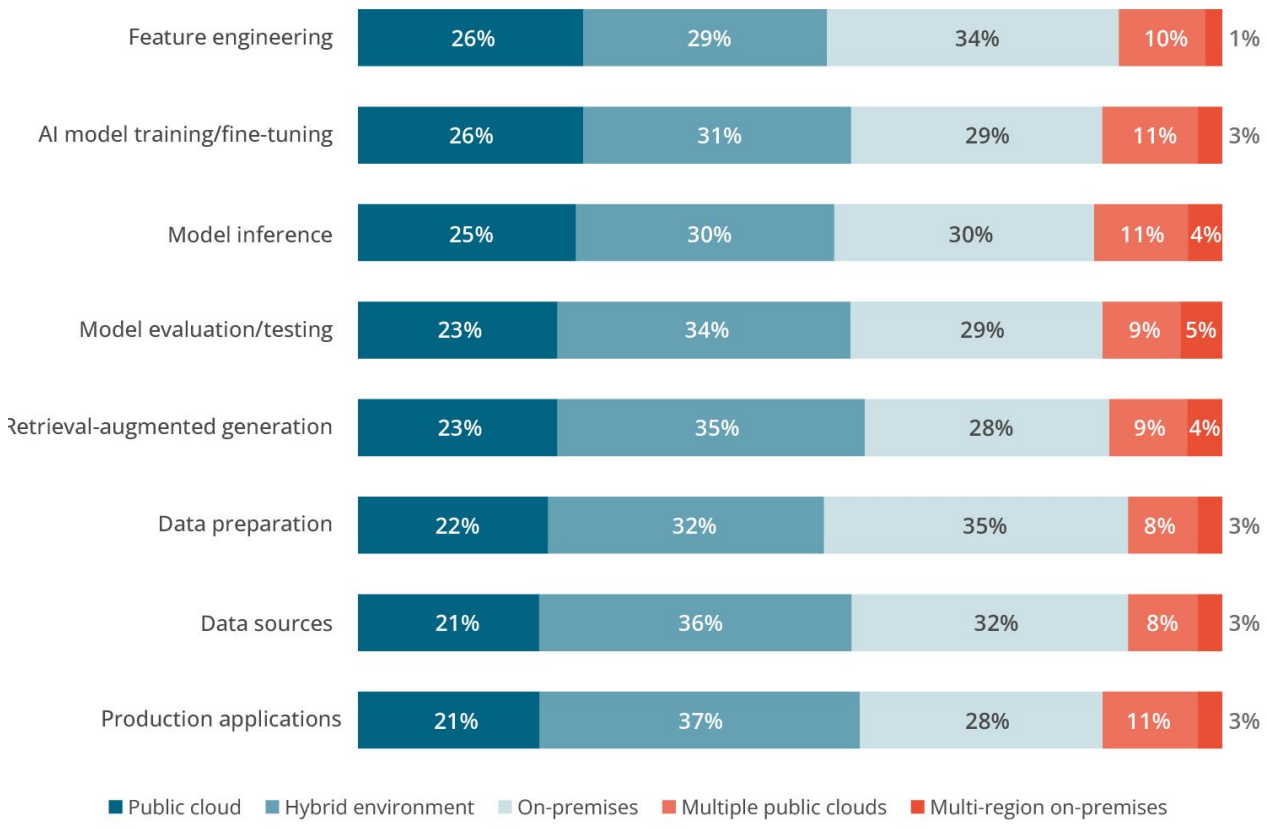
Streamlining Data Driven Application Development

Application development is a core competency for data science organizations. Working independently often highlights the challenges of disconnected environments that data science professionals rely on. Merging development efforts with a wider ecosystem of solutions will drive efficiency and production. Platforms that support and integrate with popular frameworks like Shiny, Dash, and Streamlit will speed efforts and allow production to take place alongside an enterprise governance architecture. These technologies enable rapid application development and engage business stakeholders with powerful and consumable solutions. Equipping data science teams with these tools leverages open-source innovation, delivers trusted applications and faster return on investment (ROI).

Extensible and Connected: Cloud Data and AI

Data and compute ecosystems have grown well beyond the confines of corporate data centers making data access, orchestration, and governance increasingly difficult. These distributed data ecosystems include AWS, Azure, Google Cloud Platform, Databricks, Snowflake, and a long list of others. The diversity of sources and platforms is a catalyst for delivering deeper integration of data science tools and processes with these environments. In recent BARC research, 318 respondents illustrate how widely varied data workloads are on hybrid, on-premises, and cloud platforms. These patterns make it clear that data science environments need to easily connect to and leverage these systems via their open-source tools, while being governed.

Figure 1: Where does your organization host the following components of its AI projects?



Use Case Example

Below is an excellent example of how financial services (FSI) are eliminating silos and iterating quicker, while leveraging open-source technology in governed environments.

Accelerating Transparent Credit Modeling in a Regulated Financial Environment

A global financial services firm specializing in credit ratings and research sought to deliver transparent, repeatable, and explainable models within a regulated environment. Data scientists were tasked with evaluating the probability of default across factors such as company cash flow and balance sheet strength. Their conventional data science approach, which encompassed offline model development, static presentation preparation, and feedback sessions, proved inefficient and hindered decision-making.

The quantitative modeling team adopted a code-first workflow and a suite of integrated data science tools to enhance collaboration, reproducibility, and real-time model delivery. This enabled the team to build and publish models as APIs and interactive applications, thereby reducing turnaround time and augmenting engagement with internal stakeholders. Analysts gained the ability to modify assumptions and promptly observe results, facilitating expedited validation and review processes.

Previously, model development was conducted across individual environments utilizing tools such as RStudio and Jupyter. In the absence of a centralized infrastructure, the team encountered challenges in managing model versions, enforcing governance standards, and consistently sharing results. The environment produced capabilities supporting multiple languages and frameworks, including R, Python, Flask, and Streamlit. It also facilitated onboarding for staff with diverse technical backgrounds.

Interactive applications fostered stakeholder trust by rendering model behavior that was visual and easy to understand. Stakeholders could explore the impact of input modifications on outcomes, thereby enhancing communication and promoting confidence in the results. This transformation not only enhanced the technical workflow. but also underscored the team's value in delivering timely and reliable insights.

The revised workflow enhanced compliance, accelerated the delivery of credit models, and strengthened the role of data science in business decision-making.

Recommendations

- **Embrace Open-Source as a Centralized Platform.** Data science solutions that leverage R and Python are non-negotiable tools in successful practices. Augment them with a centralized platform that enables collaboration, easy access to data, and enterprise security and governance processes.
- **Promote Reuse and Standardization.** Encouraging the reuse and standardization of data science models, methods, and libraries speeds up innovation and reduces project costs. Quicker application development delivers a better ROI and allows for greater scale
- **Encourage Collaboration:** Collaboration brings data science teams out of their silos and closer to business partners and the broader data science community. Sharing applications and insights across the enterprise is the smartest path to success.
- **Embed Governance:** Embedding governance into data science workflows without restricting access to tools data scientists prefer provides the necessary guardrails to meet regulatory and IT requirements without restricting innovation. This approach encourages safe experimentation while ensuring compliance.

Conclusion

The time has come to optimize the data science investment at your organization. Ignoring siloed practices results in lower ROI, higher project costs, and missed competitive opportunities. This strategy is extremely relevant for companies in highly governed sectors that require alignment with regulatory and compliance guidelines, while empowering their data science organization with open-source tools. Executives that invest in this new level of data science maturity will immediately see it as a force multiplier that increases project velocity, enables stronger stakeholder collaboration and alignment with enterprise security and governance standards.

Supporting teams with powerful open-source solutions while enabling collaboration, security and governance is the key to bringing your teams together and maximizing your data science investment.

About Posit Software

Posit, formerly RStudio, is dedicated to advancing open-source software for data science, scientific research, and technical communication.

Trusted by millions of users, including 25% of the Fortune Global 100, Posit empowers organizations to drive innovation and informed decision-making.

We focus on making data science more open, intuitive, accessible, and collaborative, offering tools that enable powerful insights and smarter, data-driven decisions.

Contact info

Global Headquarters
Posit Software
250 Northern Ave.
Boston, MA 02210

Phone: +1 844-448-1212
Posit.co

Posit: Enterprise Data Science: Eliminating the Speed vs Control Trade-off

Every data science leader knows this pain. Your best analysts work in R and Python on laptops while IT demands enterprise governance. The result? Data scientists can't iterate fast enough because they're constantly switching between tools, waiting for data access approvals, and working with IT to recode their models to meet production standards.

The business costs are mounting. Teams can't own the entire development lifecycle, so they waste months in handoffs between exploration, development, and deployment. Top talent walks away when forced to abandon the tools that make them productive or when simple iterations take weeks instead of hours. Critical insights arrive too late because teams can't accelerate prototyping and gather feedback quickly. In an AI-driven world, this development friction becomes a competitive liability.

Posit solves this by bringing enterprise capabilities directly to R and Python workflows. We're the team behind RStudio, Shiny, and tidyverse, with 38% of our engineering team working full-time on over 350 open source projects used by millions daily. Our platform provides a seamless workflow for fast iteration where R and Python developers get a seamless, AI-powered path from prototyping to development through exploration (Workbench and Positron IDE), deployment (Connect), and package management. This creates flexibility in the choice of tools preferred by data scientists and statisticians balanced with governance and security required by IT. You don't have to sacrifice speed and choice of tools while meeting the security needs of the business. No handoffs, no rebuilding, no compromises.

The impact is immediate. Microsoft cut capacity planning cycles from three days to one hour without sacrificing enterprise controls. NASA moved from two-week analysis cycles to real-time workforce scenario modeling. Unity Health Toronto deployed an AI early warning system that reduced unexpected patient deaths by 26%. These wins are repeatable and evidenced by over 1,500 organizations across finance, manufacturing, healthcare, and government who rely on Posit to scale their data science impact.

While other platforms ask data scientists to adapt to enterprise tools, Posit adapts enterprise governance to data science workflows. Our commitment to choice means you can deploy on-premises, in the cloud, or in hybrid models that fit your existing tech stack. Your models, dashboards, and analyses remain yours with no vendor lock-in because everything runs on open source foundations. As a Public Benefit Corporation, we're legally committed to balancing community impact with business success, ensuring the tools your teams rely on remain free and widely accessible while giving you the flexibility to choose what works best for your business.

Bruno Trimouille, CMO Posit