

22 May 2024 | Analysis

The R&D Technology Stack: How Pharma Can Gain A Competitive Edge With AI, Data And Tech

by Margaret Ayers, Alexander Böcker, Julia Wiesermann, Christoph Meier

The use of Artificial intelligence, data analytics and digital technologies hold the potential to transform pharmaceutical R&D and massively improve productivity. However, to achieve game-changing impact, the right technology infrastructure is required.

Research and development is the engine that powers the pharmaceutical industry and is an area in which companies have long sought competitive advantage. In the last decade, digital, data and AI have started to make inroads into this function and have started to transform how R&D is conducted. This includes new AI-powered workflows in discovery, especially in small-molecule research, as well as increasing use of analytics in clinical trials. Recent advances in Generative AI have only heightened expectations, especially in experiment and trial design and documentation.

To succeed in the new world of digital and AI-powered R&D, many companies realize they need a technology infrastructure that is fit for purpose. This is reflected in substantial technology investments in R&D. From experience, BCG has found that pharmaceutical companies spend up to 40% of their overall IT budget – typically tens to hundreds of millions of dollars – on R&D.

But despite this investment, many companies struggle to achieve value and the expected return-on-investment from their

Key Takeaways

- Industry experts discuss key features of a modern R&D technology stack: reviewing core transactional systems in research and in development, and how these can be enhanced with data platforms for exploratory analysis.

technology investments. Recent research reported by the Financial Review showed that up to 70% of such digital transformations fail to deliver on their objectives.

By making the right technological choices, pharmaceutical companies can massively increase the likelihood of success, and thereby achieve strong positive return on their technology investments in R&D. Industry experts from BCG have collated key insights based on their experience supporting large pharmaceutical companies as well as midsize companies in the R&D space.

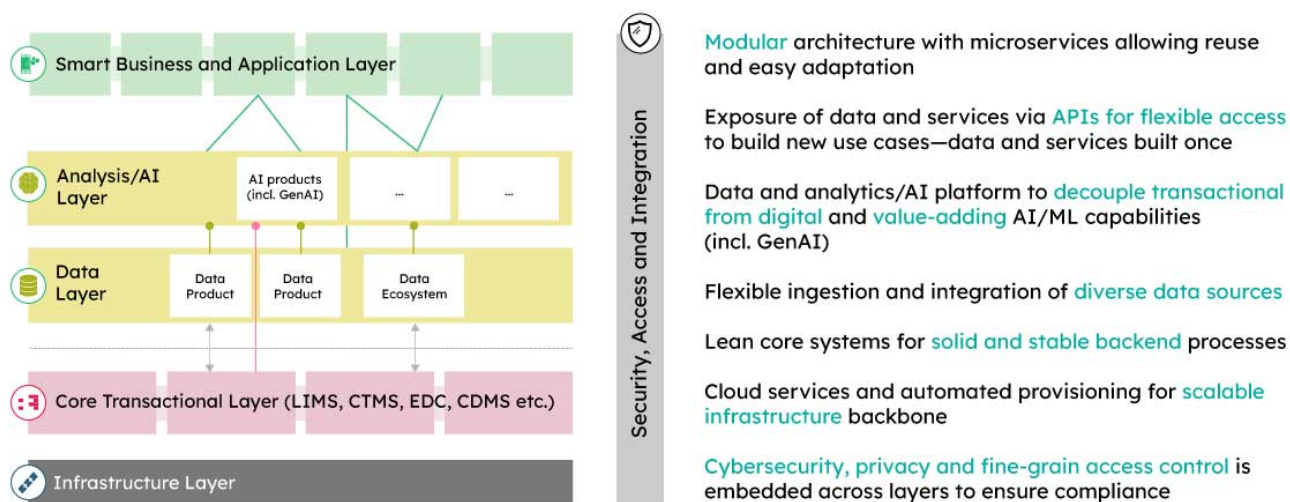
- Exploration of the critical role of data management, ingestion and integration in R&D. Including the applications of AI and Generative AI across R&D, and how the R&D tech stack can become “AI-ready.”
- Showing that by making the right technology investments pharmaceutical companies can achieve strong return on investment.

An Architecture To Amplify Value From R&D Tech

Historically, most R&D technology investments have focused on the transactional systems that support core processes and serve as systems of record, such as the CTMS (clinical trial management system), CDMS (clinical data management system), RIMS (regulatory information management system), ELN (electronic laboratory notebook), and LIMS (laboratory information management system).

While these systems are critical and some of them required by regulation, and while they can sometimes streamline processes, companies cannot usually achieve competitive differentiation through systems alone. Instead, competitive advantage is achieved through better and more holistic use of the data that sits within these systems – whether it be clinical data, operational data or other domains – and combining with external data sources to drive new insights. For example, when studying disease mechanisms, the combination of OMICs and phenotypic data in a knowledge graph can unlock new understanding. Similarly, during clinical development, the combination of randomized control trial data (RCT) and real-world evidence (RWE) in a synthetic control arm can lead to massive acceleration.

Exhibit 1: Illustrative overview of a modern R&D technology stack, consisting of multiple layers

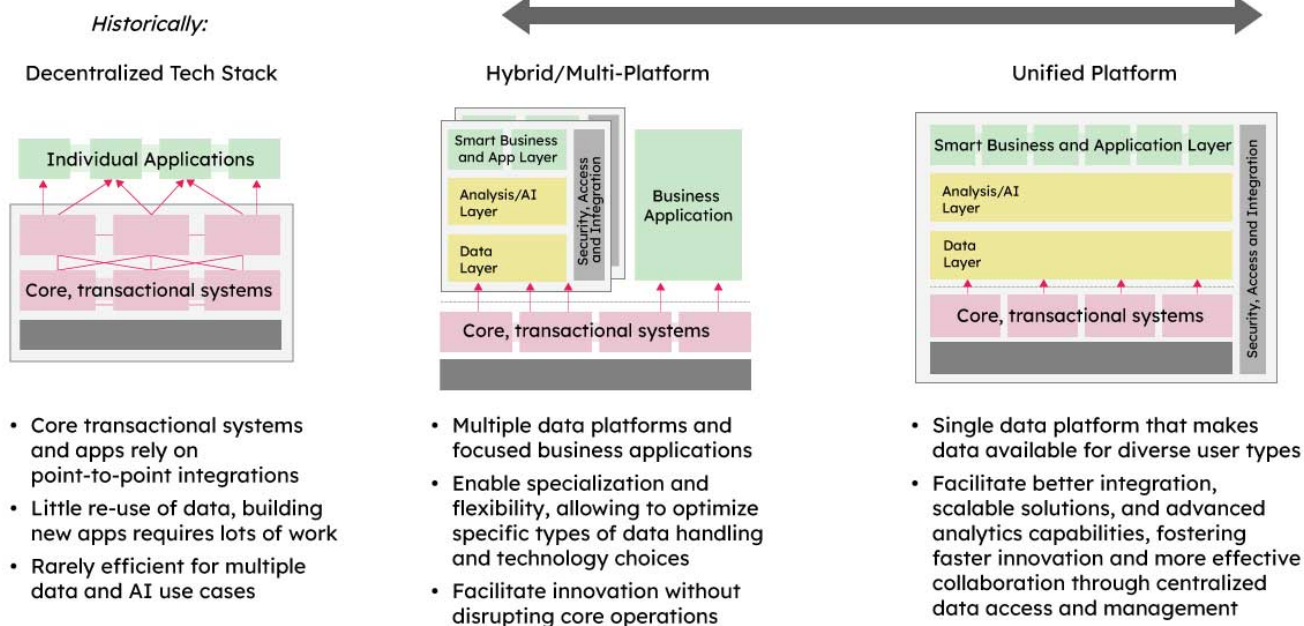


Across pharmaceutical companies there is a spectrum of architecture used. Many organizations have gone for a decentralized technology stack in which core transactional systems and applications operate separately and are connected by point-to-point connections. Whilst this design is relatively simple to implement in the early stages, it is not easy to maintain and scale as the number of point-to-point connections increases exponentially.

More recently, most companies have started building layered technology stacks, as outlined in Exhibit 1. This layered architecture has multiple benefits: it decouples use-case delivery from the time-consuming process of legacy modernization. It also makes data and AI services re-useable which accelerates use-case build. And importantly R&D, this architecture enables both GxP and non-GxP analysis. Lastly, the set-up facilitates interoperability by managing data and services both in and out of transactional systems.

One of the key strategic design questions is whether to aim for a federated or multi-platform approach, e.g. data mesh or data fabric, or for unified platform, such as data lake. These designs often differ between research and development, and they will depend on business requirements (see Exhibit 2). BCG has observed that, particularly in research and discovery, federated multi-platform architectures are increasingly common, often combining multiple cloud vendors.

Exhibit 2: Spectrum of potential technology architectures for R&D



Drug discovery: A Modern Tech Stack Enabling AI-Powered Research

AI and machine learning are transforming drug discovery and research, from a predominantly experimental process towards a hybrid of *in silico* and laboratory work. This promises to improve efficiency, to deliver scientific innovation by opening new chemical and biological spaces, and ultimately to improve the odds of success and bringing new medicines to patients faster.

Consider the standard small molecule drug discovery process. Biological research first identifies a potential protein target and experimentally validates its role in a disease. Thousands of compounds from known chemical libraries are then screened and, where there are hits, scientists embark on multiple cycles of design, synthesis, and testing. AI and data analytics are revolutionizing this process by reducing the number of experimental cycles needed, accelerating the overall process, and enabling simultaneous optimization across more parameters than is possible in an experimental approach.

However, AI and expanded data sets bring their own set of technology challenges. For example, biological and chemical databases come in a broad range of formats with little standardization. Also, historical experimental data often lacks the standardization and contextualization required to be valuable training data for AI.

How can modern technology address these challenges? A future-ready technology stack allows researchers seamless access to data. It is designed to handle large volumes of data across different modalities, ensuring that data is easily accessible while reducing redundancy and

unnecessary data transfers. This improves both efficiency and cost-effectiveness. One key feature of this tech stack is the availability of high-performance computing resources, which can scale according to the program's needs.

At the core of modern research data platforms is the ability to quickly ingest and integrate data, often using AI-powered tools for master data management. These platforms are increasingly employing semantic layers, data dictionaries, common ontologies, and data access management frameworks to facilitate self-service data access for scientists. While many companies build bespoke data and AI platforms for research, there is also a growing number of vendors that extend or integrate with transactional systems in research, such as the lab information management systems (LIMS) or lab equipment software, to enhance automation, interoperability, and data harmonization.

Drug development: A Modern Tech Stack Powering Evidence Generation

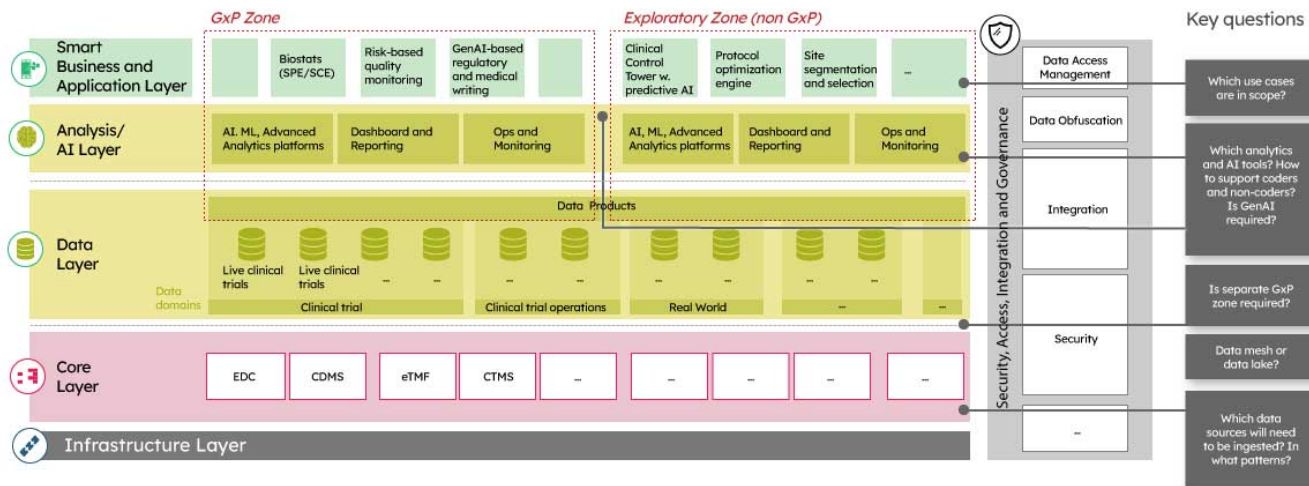
The goal of drug development is the generation of clinical evidence, in an effective and efficient way. This requires extensive collaboration across multi-disciplinary teams. For example, a protocol developed by clinical development team must be executed by clinical operations teams. Patient data gathered in clinical sites must flow to data management to biostatistics and ultimately into regulatory documents, all in a GxP-validated environment.

AI, digital and data analytics is transforming clinical evidence generation end-to-end: from devising evidence generation strategies, to optimizing studies and protocols, to accelerating patient enrollment and optimization of trial in-flight, to using Generative AI to draft regulatory dossiers and other documents. Trials are also increasingly using wearable devices, patient reported outcomes, and imaging to complement centralized electronic data capture (EDC). Clinical data is also increasingly leveraged for secondary, non-GxP purposes, such as disease understanding, translational medicine, and biomarker identification.

What does this mean for technology infrastructure? Modern technology stacks in development must facilitate seamless, near-real-time data flows and support both GxP and non-GxP analysis of clinical data. To accomplish this, the typical architecture include (*see Exhibit 3*): first, a set of transactional systems for core R&D operations; secondly, a set of data, AI, and application layers that provide GxP and non-GxP environments for biostatistics and secondary analysis of clinical data.

Exhibit 3: Technology stack for drug development and clinical trials

Illustrative view with substantial core systems, a platform with distinct GxP and non-GxP zones



While most companies will have the same set of core systems, they can choose whether to opt for a ‘platform’ vendor that offers productized integrations or use best-of-breed vendors and build an integration layer.

For clinical data, a harmonized set of data, AI, and application layers can deliver value for both GxP and exploratory purposes. It should focus on efficient data handling and sharing across many user groups while minimizing the diversity of platforms and technologies. To enable both GxP and non-GxP analyses in an efficient way, companies should consider a unified and integrated infrastructure that allows for seamless data sharing and migration of data pipelines, analyses, and algorithms from exploratory to GxP validated environments. One enabler is fine-grained data access control ensures appropriate shareability by user type or attribute.

Within this model, there are still several strategic choices: Should companies go for a vendor-provided clinical data repository, or build a bespoke platform? To what extent will they need to ingest and process data from sources such as devices? Should they build their own control tower to manage clinical operations, or will off-the-shelf clinical trial management (CTMS) functionalities suffice? Does the AI layer need to provide for generative AI builds? These choices are determined by R&D needs, level of sophistication, and budget.

Maximizing Value: Lessons learned

The cost of the R&D technology stack is often substantial. Transactional system budgets can be tens of millions of dollars per year during the build phase, and the entire stack (including run cost) can exceed \$100m. To ensure that companies are investing wisely, BCG recommends following tried-and-tested rules.

Firstly, pharmaceutical companies should distinguish between three types of investments: “license-to-operate” (LTO), “industry-standard”, and “differentiating”. LTO includes capabilities that are required or considered best practice for GxP (such as eTMF, RIMS etc.); these systems typically provide limited acceleration or cost reductions, so spend should be calibrated.

Industry-standard capabilities provide workflow efficiencies, but only in line with other pharmaceutical or biotechnology companies. Examples include clinical operation dashboards included in many CTMSs, or content management as part of . In our experience, these typically provide broad workflow improvements but overall trial acceleration of no more than 10%.

Differentiating capabilities are those that lead the industry – these often need to be built or customized in-house, leveraging unique combinations of data and existing systems. Examples include protocol complexity optimization, clinical trial simulations, and generative AI-enabled medical writing. In our experience, such systems can accelerate trials by 20%-30% or more by transforming specific critical path activities, depending on how extensively they are utilized.

To achieve good return-on-investment, pharmaceutical companies need to balance across LTO vs. industry-standard vs. differentiating systems. In our experience, companies that ensure investment across all three categories in parallel achieve the highest impact and ROI.

Key Lessons

Lesson One: Be business-led and define overall ‘north star’

Select use cases based on what will deliver impact for R&D, not what vendors offer. Given the highly interconnected nature of R&D, this cannot be done through individual projects. Consider a short but sharp exercise, together with R&D leadership, to define the key business priorities and how digital, data and AI will deliver these over the next two to three years.

Lesson 2: Take a use-case approach to the data platform

Pick a limited number of differentiating use cases (typically no more than 5-7) to guide the initial data platform build. The value delivered from these can often fund the longer digital transformation and create a blueprint for follow-on use cases.

Lesson 3: Be deliberate in platform design and vendor selection

For transactional systems, make a deliberate decision how to design the tech stack and select vendors. Single-vendor approaches are often preferred by IT and procurement departments since they streamline data flows and enable competitive price negotiations. In contrast, R&D teams sometimes prefer best-of-breed approaches. And whilst customization is critical for differentiated capabilities, only a selected number of systems should be customized.

Lesson 4: During the build and implementation phase, move fast and ensure that incentives are aligned

Implementing a new R&D technology stack is a complex undertaking. To move at pace, it is critical that all stakeholders collaborate efficiently. When using external parties, it is critical to incentivize for effective delivery and speed. Fixed-price and defined-scope contracts, with clear acceptance criteria and backloaded payment structures can often be very helpful.

Lesson 5: Build for where R&D is headed, not for where it is today

When investing in the data platform and transactional systems, consider technologies that enable the forward-looking trends in R&D. For instance, the proliferation of data sources and formats beyond the EDC call for a more automated, AI-supported data management process.

Looking Ahead

Digital, data and AI offer exciting opportunities to transform drug discovery and development.

To unlock and take full advantage of these opportunities, pharmaceutical companies will need to evolve their technology infrastructure. Those organizations that get it right will improve R&D efficiency, support decision making, deliver scientific innovation, and ultimately bring new medicines to patients faster.