

Biotech Software Stack Guide: Infrastructure & Data

2/10/2026 • 30 min read

biotech software stack

r&d digitization

series c funding

cloud infrastructure

data management

lims

eln

bioinformatics

regulatory compliance

life sciences it



Executive Summary

The biotechnology industry is undergoing rapid digital transformation, driving startups to adopt sophisticated software and data infrastructure well before reaching late-stage funding. Biotech ventures preparing for Series C–level investment must build a **comprehensive tech stack** that spans cloud computing, data management, laboratory and clinical systems, and advanced analytics. This emerging biotech software stack integrates four core layers – infrastructure, data, applications, and analytics – to streamline research workflows, ensure regulatory compliance, and enable AI-driven insights ⁽¹⁾ www.mckinsey.com ⁽²⁾ www.mckinsey.com. Key components include cloud/HPC resources for scalable compute; robust data storage and integration (often adhering to FAIR principles); electronic lab notebooks (ELNs) and [laboratory information management systems \(LIMS\)](#) for R&D process management; clinical data and [quality systems](#) (e.g. EDC, QMS) for trials and compliance; and analytics/AI tools for bioinformatics, modeling, and visualization. Each element of this stack must be chosen and implemented during early stages so that by Series C the company can efficiently scale its research and commercialization efforts.

The importance of this software foundation is underscored by trends across the industry. Biopharma leaders report that digital and analytics solutions (including AI) are crucial to boosting R&D productivity and reducing the time and cost of drug discovery ⁽³⁾ www.mckinsey.com ⁽²⁾ www.mckinsey.com. However, surveys find that many organizations still rely on siloed legacy systems: over half of scientists use five or more separate software applications daily, and 84% rely on custom-built code ⁽⁴⁾ www.cloudcomputing-news.net. This fragmentation hinders collaboration and data reuse. For example, a 2023 Benchling industry report found only 28–30% attain organizational-level data interoperability ⁽⁵⁾ www.cloudcomputing-news.net. Companies face barriers such as a shortage of IT talent and lack of built-for-science platforms ⁽⁶⁾ www.cloudcomputing-news.net. As a result, biotechs are increasingly investing in unified data platforms that consolidate workflows, standardize data, and create “AI-ready” infrastructure ⁽⁷⁾ www.benchling.com ⁽⁸⁾ www.benchling.com. For instance, Moderna’s expansion of [Benchling](#) across its R&D organization aims to *consolidate systems, automate workflows, and standardize data* into consistent formats for AI ⁽⁷⁾ www.benchling.com. Likewise, Recursion Pharmaceuticals built an extraordinary AI infrastructure (the “BioHive” supercomputer with >500 NVIDIA GPUs) enabling it to run ~2 million experiments per week and amass a 50+ petabyte biological dataset for ML-driven discovery ⁽⁹⁾ blogs.nvidia.com ⁽¹⁰⁾ blogs.nvidia.com.

Altogether, the pre-Series C biotech must integrate modern software solutions to support complex R&D and regulatory workflows. This entails adopting cloud-first computing, modern data architectures, specialized lab and clinical software, and advanced analytics – all while ensuring compliance and [data security](#). By building these capabilities early, a startup positions itself for successful scale-up and demonstrates to investors that it has the foundation for efficient growth, regulatory readiness, and AI-enhanced innovation.

Introduction and Background

Biotechnology Innovation in the Digital Era

Biotechnology companies today operate at the intersection of cutting-edge life science and advanced information technology. Modern biotechs face unprecedented **complexity** in R&D: discovery often involves high-throughput genomics, multi-omics, automated assays, and large-scale clinical trials, generating massive volumes of diverse data. Consequently, digital infrastructure has become as critical as lab benches. Industry analyses emphasize that **digital and analytics (DnA) solutions** are now essential to boost R&D efficiency and productivity. For example, McKinsey reports that leading biopharma firms leverage AI-powered modeling, digital twins, and automated analytics to accelerate molecule discovery and trial planning ⁽³⁾ www.mckinsey.com. These tools can, in principle, shrink timelines by aiding site

selection, patient recruitment, and generating higher-quality [regulatory documents](#) ⁽³⁾ [www.mckinsey.com](#)). However, overall R&D productivity remains flat; drug development still takes over a decade and costs over \$2 billion per successful asset ⁽²⁾ [www.mckinsey.com](#)).

To overcome these challenges, companies must **rewire their tech stacks**. Traditional industry analyses note that biopharma R&D often operates with “decentralized tech stacks” of legacy systems that hinder scale ⁽¹⁾ [www.mckinsey.com](#)). In contrast, a modern stack integrates cloud infrastructure, unified data platforms, and workflow-specific applications, unlocking seamless data flow and reuse ⁽¹⁾ [www.mckinsey.com](#)). This architecture enables agile incorporation of new methods (e.g. AI), reduces dependency on outdated systems, and boosts interoperability across functions ⁽¹²⁾ [www.mckinsey.com](#) ⁽¹⁾ [www.mckinsey.com](#)). In short, digital transformation is now imperative: a 2025 McKinsey study stresses that a well-designed next-gen tech stack “can unlock the potential of AI, automation, and data” in biotech R&D ⁽³⁾ [www.mckinsey.com](#)).

Funding Stages in Biotech and the Role of Technology

Biotech startups typically progress through multiple venture rounds (Series A, B, C, etc.) as they move from concept toward commercial products. **Series C funding** is a late-stage round—often the point just before or during early commercialization. At this stage, a biotech has generally demonstrated preclinical proof-of-concept and potentially early clinical results, and is now focused on [scaling manufacturing](#), navigating regulatory approvals, and preparing for market. Analysts describe Series C as the pivot from “demonstrating the science” to “demonstrating a scalable business model” ⁽¹³⁾ [bioxlist.com](#)). The capital raised is used to build commercial infrastructure, expand operations, and solidify market position ⁽¹⁴⁾ [bioxlist.com](#)).

This shift has direct implications for technology needs. Early-stage biotechs (pre-seed through Series B) often concentrate on basic research and may use minimal IT infrastructure. But by Series C they must transition to **enterprise-grade systems** that support larger R&D pipelines, clinical development, and business processes. As one analysis observes, Series C is “about showing that the company can create a successful business” ⁽¹³⁾ [bioxlist.com](#)), which includes having the organizational tools (software) to manage complexity. The “valley of death” between research and market requires robust data management, regulatory documentation, and quality systems to scale safely ⁽¹⁴⁾ [bioxlist.com](#)). Regulatory agencies and investors alike expect startups to have formal quality programs and traceable processes even at this stage ⁽¹⁵⁾ [intuitionlabs.ai](#)).

In the digital age, investors also pay attention to a biotech’s IT readiness. A startup gearing up for Series C should present not only scientific data, but evidence that it can **manage data securely and efficiently**, comply with regulations (GxP, 21 CFR Part 11, HIPAA/GDPR), and leverage computational tools to enhance discovery. As one industry report notes, regulators “expect structured quality programs early” in biotech development, and due diligence by investors increasingly demands proof of quality oversight ⁽¹⁵⁾ [intuitionlabs.ai](#)). Similarly, platform providers are building specialized solutions (e.g. cloud-based QMS) targeted at small biotechs, reflecting the market need for out-of-the-box compliance software ⁽¹⁵⁾ [intuitionlabs.ai](#) ⁽¹⁶⁾ [intuitionlabs.ai](#)).

Thus, a biotech approaching Series C must have constructed a **software stack that bridges laboratory science, clinical development, and business operations**. This report surveys the emerging components of that stack, drawing on industry data and expert analysis. We address each layer—computing infrastructure, data management, laboratory and clinical applications, and analytics—highlighting current trends, example platforms, and the implications for growing biotech firms.

Core Components of the Biotech Software Stack

A modern biotech software stack can be conceptualized as four integrated layers: **(1) Infrastructure (Compute & Cloud), (2) Data Management, (3) Applications (Lab/Clinical/Business systems), and (4) Analytics/AI** ^[1] (www.mckinsey.com). Each layer contains essential tools and platforms that enable the smooth flow of biological and clinical information through research workflows. While McKinsey's analysis focused on large pharma R&D ^[1] (www.mckinsey.com), the same layered framework applies to startups, albeit often in more agile, "cloud-first" forms. Below we examine each layer in detail, with emphasis on what startups need before repeatedly raising late-stage capital.

1. Infrastructure Layer: Cloud, Compute, and Security

At the foundation is the **infrastructure layer**, providing on-demand computing power, storage, and network resources. For modern biotechs, this typically means embracing cloud computing (AWS, Azure, Google Cloud) rather than buying extensive on-premises hardware. Cloud platforms offer **elastic scalability** for compute-intensive tasks (e.g. AI model training or large-scale sequence analysis) and big data storage (genomes, imaging, electronic records) without upfront capital expense. Today, virtually all emerging biotechs adopt some public cloud services: life sciences cloud market analysts project the segment to grow from ~\$25 billion in 2024 to over \$100 billion by 2034 ^[17] (www.towardshealthcare.com), driven by digital R&D needs.

Using cloud also simplifies collaboration and security. Top pharmaceutical firms (and partnering biotechs) now use cloud as a "preferred cloud provider" for R&D and core IT ^[18] (press.aboutamazon.com). For example, Gilead Sciences cites Amazon Web Services as its selected cloud to unify ML and analytics across the company, from biomarker discovery to recruitment algorithms ^[18] (press.aboutamazon.com). Gilead's CIO notes that AWS provides the "agility and security needed to deliver new medicines at speed", enabling real-time data analysis (genomics, imaging, etc.) that informs pipeline decisions ^[19] (press.aboutamazon.com). Similarly, big-data biotech Tempus and genomic platforms (e.g. DNAnexus, Illumina BaseSpace) offer genome-scale compute on cloud, reflecting broad industry adoption.

Given stringent data privacy and regulatory requirements in biotech/healthcare, infrastructure choices also emphasize **compliance**. Cloud providers now offer HIPAA- and GDPR-compliant services, audit logging, and virtual private clouds. Biotech startups must ensure that patient data (e.g. sequencing, EMRs) and proprietary research data are encrypted and access-controlled. This often involves identity management (IAM), encryption keys management, and possibly on-premises or hybrid clouds for maximum control. Some companies even use purpose-built solutions (e.g. fully on-prem AI appliances) to satisfy ultra-sensitive use cases.

From an operational standpoint, infrastructure also includes enabling technologies like containerization (Docker, Kubernetes), DevOps tools, and automation/orchestration frameworks. Many startups leverage container clusters (e.g. Amazon EKS, Google Kubernetes) to run bioinformatics pipelines reproducibly. Infrastructure as Code (Terraform, etc.) and CI/CD pipelines are increasingly used to manage updates. While details can vary, most biotech Series C candidates plan their compute infrastructure early: either cloud-native or hybrid HPC architectures that can handle peak loads.

Key takeaway: By Series C, a biotech should have a cloud-first, scalable compute/storage foundation. This layer underpins everything above it. Industry data highlight the trend: North America leads, with companies pushing cloud strategies and even migrating ERP systems (e.g. Gilead running SAP S/4HANA on AWS ^[19] (press.aboutamazon.com)). In practical terms, founders must decide on cloud providers and security architecture, ensuring they meet both scientific and regulatory needs. The cost efficiency and flexibility of the cloud are especially beneficial for startups that face variable workloads.

2. Data Layer: Integration, FAIRness, and Management

Built on the infrastructure is the **data layer**, responsible for aggregating, integrating, curating, and serving data across the organization. In biotech, data can be highly heterogeneous: structured lab measurements, "omics" sequence data, imaging, electronic notes, clinical trial records, vendor reports, and more. As McKinsey notes, this layer "manages the

integration, curation, and accessibility of clinical and operational data, adhering to FAIR principles (findable, accessible, interoperable, reusable)” (^[1] www.mckinsey.com). Implementing FAIR data practices is critical: uniform metadata, ontologies, and APIs allow data generated in one experiment to be reused in others and fed into analytics.

Data integration tools have therefore become vital. Biotechs often deploy data warehouses or lakes (using solutions like Snowflake, Databricks Delta Lake, or AWS S3/Glue) where raw and processed data from various sources are stored. Integration middleware (ETL/ELT) and workflow managers (Apache Airflow, Kafka) automate data pipelines from instruments or labs into these repositories. Platforms like Benchling serve as “scientific data hubs” that ingest experimental results, sample histories, and assay outputs into a unified schema. Data cataloging tools (e.g. Amundsen, Collibra) may also be used for governance, letting scientists discover datasets and understand lineage.

Industry surveys highlight the gap and opportunity in this layer. A large Benchling report found only ~28–30% of companies have achieved data interoperability or reusability organization-wide (^[5] www.cloudcomputing-news.net). Half of IT teams support dozens of distinct lab apps (^[4] www.cloudcomputing-news.net), meaning data often remains siloed. Consequently, many startups view data engineering as an ongoing challenge. For example, Moderna’s digital transformation focused on “*structured, comprehensive*” data capture – standardizing naming schemes and schemas so experiment data can automatically feed downstream pipelines (^[8] www.benchling.com) (^[20] www.benchling.com). Without such curation, building AI models or performing cross-study analysis is infeasible.

Another element of the data layer is specialized databases. Depending on the modality, a startup may use genomic databases (NoSQL for sequence reads), cheminformatics libraries, or image storage. Some biotech firms use domain-specific platforms: e.g. the Broad Institute’s Terra provides a cloud resource and workspace for genomics data and workflows. Others develop internal data lakes enriched with biological context (e.g. Recursion’s 50+ PB database of imagery and multi-modal data built specifically for AI training (^[10] blogs.nvidia.com)).

Security and compliance must also extend here. Sensitive data requires controlled access and audit trails (21 CFR Part 11 compliance). Electronic records and signatures used in R&D need cryptographic safeguards. Many startups integrate compliance tools at the data layer: for instance, ensuring every entry has a recorded provenance (who, when, changes) that meets FDA requirements. Commercial platforms often include these features out-of-box; otherwise, custom logging and versioning systems are built.

Key takeaway: The data layer is about making data usable and trustworthy. Before Series C, biotechs should invest in a robust data architecture: centralized stores, unified metadata, and integration pipelines. This is generally an evolving process, but by Series C investors will expect clear evidence of data governance (e.g. demonstration of FAIR principles and quality data sets). Companies like Moderna achieve this by migrating R&D data into a single platform (^[8] www.benchling.com), while others (e.g. Recursion) crowd-source massive standardized datasets (^[10] blogs.nvidia.com). The long-term effort pays off by enabling higher-level analytics and easier scaling of research.

3. Application Layer: Lab, Clinical, and Business Systems

Sitting on the data are **application-layer systems** that directly support laboratory operations, research workflows, clinical trials, and business processes. These applications consume and contribute data to the underlying layers. Key categories include:

- **Electronic Laboratory Notebooks (ELNs)** and collaborative research platforms. ELNs (e.g. Benchling, LabGuru, SciNote) let scientists document protocols, experimental results, and observations digitally (^[21] www.universitylabpartners.org). Unlike paper notebooks, ELNs provide search, templates, and the ability to share data easily. Studies note that whereas ELNs capture individual experimental details, LIMS (below) handle broader workflows (^[21] www.universitylabpartners.org). Startups often begin with low-cost or open-source ELNs to improve record-keeping; by Series C a more scalable, enterprise-grade ELN is expected to track all lab operations and link to inventory.

- **Laboratory Information Management Systems (LIMS).** LIMS (e.g. LabWare, Autoscribe, CloudLIMS) are software suites for managing samples, reagents, workflows, and instrumentation in the lab. They track where each sample is, link results to experiments, manage standard operating procedures, and often include inventory management. LIMS enable lab scalability by enforcing standardized processes and ensuring traceability (critical for regulatory compliance). For example, a sample's chain of custody (from receipt to analysis) is tracked in the LIMS. As University Lab Partners notes, LIMS focus on "top-down standardized workflows and simplifying the tracking of items," complementing ELNs (^[21] www.universitylabpartners.org). By Series C, companies engaged in large-scale R&D typically have implemented a LIMS to replace spreadsheets and ad hoc tracking. Adoption of cloud-based LIMS (SaaS) is growing, lowering the entry cost for startups.
- **Laboratory Automation & Instrument Integration.** An increasing number of startups use lab automation (robots, liquid handlers, screening platforms) to accelerate experiments. Control software (often custom or vendor-specific) schedules robots and reads instruments. More advanced setups integrate this automation into the data system: for instance, the Arctoris Ulysses platform couples robotic assays with a digital pipeline for data capture. Benchling's R&D Cloud also offers APIs to orchestrate lab workflows. Early-stage companies may only have a few benchtop instruments, but by Series C, automated systems for screening or manufacturing may be integrated into the software stack. Instrument data (raw files, QC metrics) are typically fed directly into the LIMS or databases.
- **Clinical Trial and Regulatory Systems.** If a biotech has any clinical programs by Series C, it needs software for clinical operations. This may include an **Electronic Data Capture (EDC)** system (e.g. Medidata Rave, Veeva Clinical) to collect clinical trial data from sites, and a Clinical Trial Management System (CTMS) to track timelines and budgets. Data from clinical systems often flows into regulatory submissions or analytics. Quality Management Systems (QMS) fall here as well – software for document control, CAPA (corrective actions), training records, etc. Companies like MasterControl or Veeva Vault QMS provide cloud-compliant modules. For biotech startups, a QMS may start as manual spreadsheets, but regulators increasingly expect validated QMS software or equivalent processes by late-stage development (^[15] intuitionlabs.ai). Indeed, one analysis stresses that "even tiny biotechs" need robust QMS aligned with FDA/ISO standards (^[15] intuitionlabs.ai) (^[16] intuitionlabs.ai). Failing to have adequate QMS and e-archive systems before pivotal trials can jeopardize approvals.
- **Manufacturing and Supply Chain Systems** (as applicable). If the startup has at-scale production (or is preparing for it), basic manufacturing software may be introduced. This could be an ERP/MRP system (SAP, Oracle) for supply chain and inventory, especially if the product is a biologic needing GMP manufacturing. Many companies defer full ERP until after Series C, but having a plan for it is important. For example, Gilead uses SAP S/4HANA on AWS to run supply chain and finance globally (^[19] press.aboutamazon.com). Even if not fully implemented pre-Series C, startups should track resources and adopt good warehousing practices (often via LIMS or inventory modules) ready for scale.
- **Collaboration and Knowledge Management.** Beyond scientific apps, general-purpose tools (Slack/MS Teams, Confluence or SharePoint, Jira/Trello, etc.) are typically introduced early to manage projects, documents, and communication. These are not biotech-specific, but they integrate with the biotech stack by storing SOPs, protocols, and linking meeting notes to project IDs. By Series C, structured collaboration (cloud drive, CRM) is usually in place (even if basic) to coordinate cross-functional teams.

Across these systems, **integration API/connectivity** is crucial. Modern biotech software often exposes RESTful APIs, enabling data to flow between LIMS, ELNs, analytics, and other tools. Point-to-point integrations or iPaaS (integration Platform as a Service) solutions may be used. The aim is to avoid data silos: for example, sample IDs generated in the LIMS should automatically tag results in the ELN and update inventory.

Key takeaway: By Series C, a biotech should have implemented a suite of applications tailored to its domain: ELN/LIMS for lab R&D, EDC/CTMS for trials, QMS for compliance, and enterprise tools for administration. These improve efficiency and ensure audit readiness. Many startups find that off-the-shelf SaaS offerings (benchmarked for biotech) are faster to deploy than building custom systems (^[22] intuitionlabs.ai). Stack examples abound: companies like Benchling, Labguru, or QBench offer integrated ELN/LIMS; larger groups may use Veeva for both clinical and quality. The Lab Stack directory (an industry listing) documents hundreds of such platforms across discovery, translational, and clinical stages (^[23] thelabstack.com). Choosing and configuring the right mix of these applications is a defining task before growth.

4. Analytics and AI Layer

At the top of the stack is the **analytics and intelligence layer**, where data is transformed into insights. This includes basic statistical tools, visualization platforms, and advanced AI/ML frameworks. In practice, biotech teams use languages

and environments like Python (with libraries such as TensorFlow/PyTorch, scikit-learn, Pandas) and R (Bioconductor, Shiny). They run Jupyter notebooks, RStudio, or specialized visualization tools (Spotfire, Tableau) for data exploration.

For genomics and bioinformatics, common platforms include Terra/FireCloud, DNAnexus, or cloud-based Nextflow execution. These allow scalable pipeline runs: a startup can run many genome analysis workflows in parallel on cloud clusters. There are also dedicated AI tools: for instance, Illumina's BaseSpace Sequence Hub can run AI models for variant calling, and platforms like GNS Healthcare offer causal AI for clinical outcomes. Machine learning is used for target identification, compound screening, image analysis (as in Recursion), and predictive modeling of ADMET properties. Generative AI is emerging for molecule design or language-based strategies; some biotechs experiment with large models on scientific text or patent data.

This layer relies on the lower layers for data and compute. A critical requirement is well-structured input: AI models demand clean, labelled data. Thus, companies invest heavily at the data and application layers to "train the engine" of analytics. For example, Recursion uses its 50+ petabytes of standardized biological image data (generated by automated assays) as a training set for deep learning (^[10] blogs.nvidia.com) (^[24] blogs.nvidia.com). Moderna's initiative to enforce consistent schema and traceability in its R&D data was explicitly aimed at making the analytics layer feasible (^[8] www.benchling.com) (^[25] www.benchling.com).

In addition to internal tools, many biotechs leverage third-party analytics platforms. These include generic cloud platforms (AWS SageMaker, Google AI Platform) as well as life science-specific solutions (e.g. Databricks for genomics, Dotmatics for cheminformatics). Containerization plays a role here as well: trained models and pipelines are often packaged in Docker images or on platforms like AWS Batch/Kubernetes for reproducibility.

Performance and cost considerations: AI workloads can be expensive. Leading-edge biotechs address this with specialized hardware and partnerships. Recursion's example is notable: their 504-GPU supercluster (BioHive-2) achieves ~2 exaFLOPS, **five times** faster than their previous system (^[9] blogs.nvidia.com). This is an outlier scale, but it underscores a trend: as AI grows, expect more biotechs to invest in GPU clusters or cloud AI instances. By Series C, a startup should at least have access to GPU compute (commonly via cloud) and have run proof-of-concept models.

Key takeaway: The analytics layer turns data into scientific and business knowledge. Startups should plan for data science workflows from early on: hire the right talent (a common industry challenge (^[6] www.cloudcomputing-news.net)) and choose flexible tools. Experimentation with AI (protein folding, phenotypic screens, etc.) often demonstrates possible ROI to investors. But it rests on a solid foundation: without standardized, high-quality data, AI yields little. Therefore, by Series C, the analytics toolbox should be in place, even if not at full scale – typically meaning data science environments are configured, pilot models running, and insight pipelines established (e.g. routine analysis scripts or dashboards for key R&D metrics).

Data and Evidence: Industry Trends and Statistics

Adoption of Enabling Technologies. Recent surveys quantify how biopharma firms are embracing (or lagging in) various tech components. A 2023 Benchling study of 300 R&D/IT professionals found that **70%** of organizations had adopted cloud-based R&D data platforms, **63%** had robotics/automation, and **59%** were using AI/ML in R&D (^[26] www.cloudcomputing-news.net). However, only **18%** reported using SaaS software for most of their R&D/IT work (^[26] www.cloudcomputing-news.net), highlighting that many are still on-prem or legacy solutions. Notably, **84%** of scientists indicated use of some custom-built software (^[4] www.cloudcomputing-news.net), meaning that true out-of-the-box platforms have not fully displaced homegrown tools. The complexity of toolsets is also high: 53% of scientists regularly use five or more distinct scientific apps each day (^[4] www.cloudcomputing-news.net), and 40% of IT groups support more than 20 research apps. These figures underscore the fragmented state of many biotech tech environments.

Market Growth and Spending. Multiple analyses project rapid growth in biotech and life science software markets, driven by AI and digital R&D. For instance, the global **Computational Biology** software market was estimated at \$6.34 billion in 2024 and is forecast to jump to \$26.54 billion by 2035 (CAGR ~14%) (^[27] www.pharmiweb.com). This segment covers tools for simulating biology, analyzing genomics, and designing drugs, reflecting surging demand for bioinformatics and modeling platforms. Similarly, the broader **life science cloud computing market** is expected to expand at ~15% annual rates through 2034 (^[17] www.towardshealthcare.com) (^[28] www.biospace.com), as practitioners in pharma, biotech, and CROs move workflows to the cloud. Importantly, North America — representing nearly half of this market in 2024 (^[29] www.towardshealthcare.com) — dominates due to its mature digital ecosystem. SaaS (cloud software) leads the models in life sciences (^[30] www.biospace.com), aligning with the move towards subscription-based biotech platforms.

Quality and Compliance Expectations. Industry consensus emphasizes that even small biotechs must build quality and compliance into their processes early. Analysts note that **regulators and investors alike demand robust quality management** from the outset (^[15] intuitionlabs.ai). One comprehensive review argues that structured QMS software (often cloud-based, life-sciences–tailored platforms) is now “*rapidly maturing and widely adopted*” in small companies (^[22] intuitionlabs.ai). By contrast, ad hoc or custom-built QMS solutions often fail to save time or costs (^[31] intuitionlabs.ai). These recommendations suggest that startups should either implement a commercial QMS system (aligned with FDA/ISO) or prepare very well-documented manual procedures before late-stage funding.

Case Studies and Real-World Examples

To illustrate how these trends play out, consider a few representative biotech companies at different scales:

Moderna (Notable Biotech with Integrated R&D Platform). Moderna, the mRNA vaccines pioneer, publicly emphasizes technology as central to its mission (^[32] www.benchling.com). To accelerate R&D, Moderna has rolled out Benchling’s integrated R&D cloud across hundreds of scientists globally (^[33] www.benchling.com). The aims are to **consolidate previously siloed systems, automate experimental workflows, and enforce consistent, AI-ready data standards** (^[7] www.benchling.com). In practice, Moderna standardized naming schemes and schema across labs, so that experiment records now feed directly into computational pipelines (^[8] www.benchling.com) (^[20] www.benchling.com). This has yielded benefits such as faster data sharing between teams and built-in audit trails (^[34] www.benchling.com) (^[35] www.benchling.com). Moderna’s case shows that even a large biotech must invest early in a unified lab informatics system to realize AI/ML gains. Benchling’s press release highlights Moderna’s priorities: a single system where “scientists can design experiments, track samples, and analyze results” without manual file juggling (^[7] www.benchling.com) (^[36] www.benchling.com).

Recursion Pharmaceuticals (AI-Driven Discovery). Recursion is a small-molecule biotech that heavily leverages automation and AI. Its **BioHive-2 supercomputer** (a collaboration with NVIDIA) is the world’s fastest in the pharma sector (^[9] blogs.nvidia.com), with 504 NVIDIA H100 GPUs delivering ~2 exaFLOPS. This HPC capability multiplies Recursion’s throughput: its researchers process over 2 million high-content imaging experiments each week (^[24] blogs.nvidia.com). Every experiment (images of cells under various conditions) is fed into an AI pipeline. Recursion has intentionally built one of the largest curated biological datasets (~50 PB) to train its deep learning models (^[10] blogs.nvidia.com). The result is a proprietary AI platform (the “Phenom” foundation models) that can predict promising drug candidates and optimize biological assays. This example underscores the peak of current biotech stack: advanced robotics feeding a massive data pipeline into cloud-scale AI. While typical startups cannot match Recursion’s scale, their model shows the end-state goal: rapid, data-driven hypotheses.

General Biotech Startup (Hypothetical Example). Consider a Series-C-bound biotech focused on a novel biologic. By this stage, it might have achieved successful Phase I/II trials. Its software stack could look like the following:

- **Cloud Infrastructure:** The company runs on AWS, with a VPC that isolates R&D and clinical data. Storage buckets contain raw sequencing data, while an AWS Batch cluster handles bioinformatics pipelines. Critical clinical data is stored in a HIPAA-compliant workflow (e.g. encrypted RDS database).
- **Data Integration:** Data from lab instruments (mass specs, sequencers, plate readers) flows through automated ETL into a Snowflake data warehouse. Benchling is used as the central repository for all experimental notes and sample metadata, ensuring FAIRness. Data governance tools manage access and tracking.
- **Laboratory Applications:** The research team uses Benchling ELN to record experiments. A cloud-hosted LIMS (e.g. CloudLIMS) tracks sample inventory and test results. Laboratory automation robots (e.g. liquid handlers from Opentrons or Hamilton) are scheduled via Benchling's integration, with output files automatically uploaded.
- **Clinical / Quality:** For a small first trial, the company uses a SaaS EDC system (e.g. Medrio) to collect patient data. It also employs a SaaS QMS like Veeva or Qualio for document control (SOPs, batch records). All digital signatures and audit trails meet 21 CFR Part 11 standards.
- **Analytics:** The data science team uses Databricks notebooks (Python/R) on the same cloud, reading from the Snowflake tables. They have developed custom ML models (e.g. predicting patient response) and routinely visualize results in Spotfire. Early-stage AI models operate on on-premises GPUs, with plan to move to cloud GPUs for scaling.

This startup's architecture shows how the theoretical layers become working tools. By Series C, its investors should see that the company's R&D pipeline is supported by a cohesive digital platform, with no walled-off data silos and with compliance built in.

Discussion: Implications and Future Directions

Benefits and Challenges

Implementing a modern biotech software stack yields many benefits. Founders and investors report **accelerated discovery and higher throughput**: automation and AI can replace repetitive lab tasks, allowing scientists to “do more with less” (^[37] [blogs.nvidia.com](#)). Data standardization and integration reduce human error and ensure reproducibility, which both speeds regulatory review and instills confidence in results (^[35] [www.benchling.com](#)) (^[7] [www.benchling.com](#)). Roughly speaking, having well-structured data and cutting-edge tools can dramatically shorten the hypothesis-test-analyze cycles. In Recursion's case, integrating AI allowed them to capture “80% of the value with 40% of the wet lab work” (^[37] [blogs.nvidia.com](#)). In aggregate, the digital stack can lower operating costs per discovery and increase the chance of pipeline success.

On the business side, such a stack signals maturity to investors. Venture capitalists expect due diligence evidence of data integrity and scalability. The absence of robust systems could raise red flags about future delays or technical debt. Conversely, demonstrating a coherent tech stack can be a competitive advantage, showing readiness for rapid growth. Some investors even see healthcare informatics as a differentiator: startups that are “data-driven” often achieve higher valuations.

However, challenges remain. Smaller biotechs often lack the budgets to custom-build ideal systems. There is a trade-off between cutting-edge capability and lean operations. Over-engineering the stack too early can waste resources; under-investing can cause chaos later. For example, stitching together myriad point solutions without a unifying platform can lead to inefficiencies (the “spaghetti” dilemma). Companies also struggle with change management: scientists and clinicians have established workflows, and moving them to new software requires training and buy-in. Vendor lock-in and interoperability are additional concerns. The aforementioned Benchling report cites organizational culture and tech priorities misalignment as hurdles (^[38] [www.cloudcomputing-news.net](#)). Talent is a bottleneck too: finding staff who understand both biology and data systems is hard (^[6] [www.cloudcomputing-news.net](#)).

Critics caution that tech alone cannot substitute strategy. A biotech still needs a valid scientific hypothesis and strong leadership. While Gen AI tools (like protein folding predictors) are powerful, their outputs must be carefully validated experimentally. The promise of AI has led to hype; emerging best practice is to treat it as an assistive layer atop solid fundamentals, not as magic. Therefore, a balanced approach is recommended: pilot new technologies (e.g., test a machine learning model on one project) while retaining manual checks.

Future Trends

Looking ahead, the biotech software stack will continue evolving along several fronts:

- **Advanced AI and Automation.** Generative AI (large language models and beyond) is poised to become integral in research. Tools that can propose experiments, scan literature automatically, or even control lab robots via voice commands are on the horizon. In silico screening using AI (as done in materials science) may mature, speeding early discovery phases. Lab automation will incorporate more IoT and real-time feedback, with software systems dynamically adjusting experiments based on interim results.
- **Standardization and Interoperability.** Industry groups are likely to push common data standards (ontologies, data models) so that platforms can interoperate seamlessly. The FAIR initiative may gain traction, perhaps aided by regulatory recommendations. Startups should watch for emerging standards (e.g. HL7 FHIR for clinical data) and design their systems to be flexible.
- **Cybersecurity and Resilience.** As ransomware and data breaches become more prevalent, biotech must fortify defenses. Future stacks may incorporate zero-trust architectures and encrypted computation (e.g. secure multi-party crypto for collaborative research). Disaster recovery (e.g. automated cloud backups) will be standard practice. Biotechs will need dedicated IT security protocols akin to those in enterprise software.
- **Regulatory Tech (RegTech).** New regulatory tools will emerge. For example, AI-driven QC may automatically flag data anomalies. Submission platforms may move onto blockchain or other tamper-evident records. Some startups are beginning to use platforms that directly assemble eCTD (electronic Common Technical Document) dossiers from linked data.
- **Democratization of Tools.** Low-code/no-code platforms are making inroads, so smaller teams can configure databases or workflows without deep programming. This could allow lean biotech teams to build custom dashboards and reports much more rapidly, provided they maintain governance.
- **Integration of Consumer/Patient Data.** As precision medicine grows, biotech might integrate real-world data (wearables, mobile apps) into their pipelines. Software stacks will extend beyond the lab and clinic to patient-focused platforms, requiring interoperability with healthcare APIs.

In summary, the future biotech software stack will be ever more data-driven, automated, and intelligent. Biotech companies will increasingly resemble software-centric enterprises, competing on data as much as science. Startups positioning themselves for Series C must therefore keep an eye on these emerging capabilities, building extensible architectures today for the tools of tomorrow.

Conclusion

Biotechnology startups on the cusp of Series C funding must have laid a robust software foundation. This **biotech software stack** – spanning cloud infrastructure, data systems, lab and clinical applications, and advanced analytics – is no longer optional but essential. It enables the rapid, cost-effective R&D required to bring novel therapies to market. As documented by industry research and cloud platform case studies (^[3] www.mckinsey.com) (^[19] press.aboutamazon.com), companies that invest in digital tools and AI enjoy faster discovery cycles, better collaboration, and stronger data integrity. Conversely, those that neglect their IT backbone risk inefficiencies and investor skepticism.

Key elements of the stack include: scalable computing (often cloud/HPC), FAIR-compliant data management, ELNs/LIMS for experiment control, clinical data and QMS for trials and compliance, and analytics pipelines for AI/ML. Real-world examples – from Moderna's enterprise R&D cloud to Recursion's GPU-enabled AI platform – show how these elements

coalesce in practice. Credibly, by Series C investors expect to see evidence that such systems are in place. Regulators too expect organized quality processes (^[15] intuitionlabs.ai) by late-stage preclinical work.

Based on current trends, we anticipate that biotech firms will continue to deepen their use of technology. Future Series C candidates may feature integrated digital twins of bio-processes, routinely use predictive algorithms, and collaborate through interoperable data networks. For now, the recommendation is clear: build the digital foundations early. As one Benchling executive summarized, companies that “swiftly adopt new...scientific software built for this new era of biology” – and align R&D and IT – will drive innovation across breakthroughs (^[39] www.cloudcomputing-news.net). A well-chosen, flexible biotech software stack will be a major asset on the path from laboratory benchtop to marketplace success.

Tables

Life Sciences Technology Adoption (2023 Survey)	Adoption Rate (% of companies)	Source
Cloud-based R&D data platforms	70 %	(^[26] www.cloudcomputing-news.net)
Robotics and lab automation	63 %	(^[26] www.cloudcomputing-news.net)
AI/ML used in R&D	59 %	(^[26] www.cloudcomputing-news.net)
Majority of R&D work done on SaaS platforms	18 %	(^[26] www.cloudcomputing-news.net)
Scientists using ≥5 distinct lab software tools daily	53 %	(^[4] www.cloudcomputing-news.net)
IT groups supporting >20 research applications	40 %	(^[4] www.cloudcomputing-news.net)
Organizations using custom-developed software	84 %	(^[4] www.cloudcomputing-news.net)

Software/System Category	Purpose/Function	Example Platforms/Providers
Compute & Cloud Infrastructure	Scalable computing, storage, networking for data and workloads	AWS, Azure, Google Cloud; HPC clusters; Oracle Cloud HPC
Data Management & Integration	Data warehousing, curation, FAIR compliance, ELT workflows	Snowflake, Databricks, Apache Airflow, Benchling R&D Cloud
Electronic Lab Notebook (ELN)	Digital experiment record-keeping, protocol templates	Benchling, Labguru, PerkinElmer Signals Notebook
Laboratory Information Mgmt (LIMS)	Sample/inventory tracking, SOP management, audit trails	LabWare, Autoscribe, CloudLIMS, LabKey
Lab Automation Control	Schedule/control of robotic instruments and assays	Arctoris Ulysses, Opentrons (OT-2), Benchling Automation API
Quality Management (QMS)	Document control, CAPA, training, regulatory compliance	MasterControl, Veeva Vault QMS, Qualio, Greenlight Guru
Clinical Data Capture (EDC/CTMS)	Clinical trial data collection, study management	Medidata Rave, Oracle Clinical, Veeva Clinical Suite
Analytics & AI Platforms	Data analysis, ML workflows, scientific insights	Jupyter/Python/RStudio; NVIDIA GPUs; DNAnexus; Terra (Broad); SAS, Spotfire
Collaboration/Project Mgmt	Team communication, documentation, project tracking	Slack, Microsoft Teams, Confluence, Jira, Notion
Security & Compliance Tools	Access control, encryption, audit logging	AWS/GCP IAM & KMS; Fortanix, CyberArk, Vault; GxP compliance modules
ERP/Supply Chain (long-term)	Manufacturing planning, inventory, financials	SAP S/4HANA, Oracle ERP, Salesforce (for Sales/CRM)

Sources: Compiled from industry reports and vendor documentation. McKinsey’s tech stack model is adapted with illustrative examples (^[1] www.mckinsey.com). Benchling survey data informs adoption rates (^[26] www.cloudcomputing-news.net) (^[4] www.cloudcomputing-news.net). The specific platforms listed are representative of those commonly used in biotech R&D and are mentioned in sector analyses and case studies (^[7] www.benchling.com) (^[9] blogs.nvidia.com). (No single source covers all categories; the table is a synthesis of cited insights.)

External Sources

IntuitionLabs - Industry Leadership & Services

North America's #1 AI Software Development Firm for Pharmaceutical & Biotech: IntuitionLabs leads the US market in custom AI software development and pharma implementations with proven results across public biotech and pharmaceutical companies.

Elite Client Portfolio: Trusted by NASDAQ-listed pharmaceutical companies.

Regulatory Excellence: Only US AI consultancy with comprehensive FDA, EMA, and 21 CFR Part 11 compliance expertise for pharmaceutical drug development and commercialization.

Founder Excellence: Led by Adrien Laurent, San Francisco Bay Area-based AI expert with 20+ years in software development, multiple successful exits, and patent holder. Recognized as one of the top AI experts in the USA.

Custom AI Software Development: Build tailored pharmaceutical AI applications, custom CRMs, chatbots, and ERP systems with advanced analytics and regulatory compliance capabilities.

Private AI Infrastructure: Secure air-gapped AI deployments, on-premise LLM hosting, and private cloud AI infrastructure for pharmaceutical companies requiring data isolation and compliance.

Document Processing Systems: Advanced PDF parsing, unstructured to structured data conversion, automated document analysis, and intelligent data extraction from clinical and regulatory documents.

Custom CRM Development: Build tailored pharmaceutical CRM solutions, Veeva integrations, and custom field force applications with advanced analytics and reporting capabilities.

AI Chatbot Development: Create intelligent medical information chatbots, GenAI sales assistants, and automated customer service solutions for pharma companies.

Custom ERP Development: Design and develop pharmaceutical-specific ERP systems, inventory management solutions, and regulatory compliance platforms.

Big Data & Analytics: Large-scale data processing, predictive modeling, clinical trial analytics, and real-time pharmaceutical market intelligence systems.

Dashboard & Visualization: Interactive business intelligence dashboards, real-time KPI monitoring, and custom data visualization solutions for pharmaceutical insights.

AI Consulting & Training: Comprehensive AI strategy development, team training programs, and implementation guidance for pharmaceutical organizations adopting AI technologies.

Contact founder Adrien Laurent and team at <https://intuitionlabs.ai/contact> for a consultation.

DISCLAIMER

The information contained in this document is provided for educational and informational purposes only. We make no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability of the information contained herein.

Any reliance you place on such information is strictly at your own risk. In no event will IntuitionLabs.ai or its representatives be liable for any loss or damage including without limitation, indirect or consequential loss or damage, or any loss or damage whatsoever arising from the use of information presented in this document.

This document may contain content generated with the assistance of artificial intelligence technologies. AI-generated content may contain errors, omissions, or inaccuracies. Readers are advised to independently verify any critical information before acting upon it.

All product names, logos, brands, trademarks, and registered trademarks mentioned in this document are the property of their respective owners. All company, product, and service names used in this document are for identification purposes only. Use of these names, logos, trademarks, and brands does not imply endorsement by the respective trademark holders.

IntuitionLabs.ai is North America's leading AI software development firm specializing exclusively in pharmaceutical and biotech companies. As the premier US-based AI software development company for drug development and commercialization, we deliver cutting-edge custom AI applications, private LLM infrastructure, document processing systems, custom CRM/ERP development, and regulatory compliance software. Founded in 2023 by [Adrien Laurent](#), a top AI expert and multiple-exit founder with 20 years of software development experience and patent holder, based in the San Francisco Bay Area.

This document does not constitute professional or legal advice. For specific guidance related to your business needs, please consult with appropriate qualified professionals.

© 2025 IntuitionLabs.ai. All rights reserved.