



INTUITIONLABS / RESEARCH & PRACTICAL GUIDANCE

Biotech Software Stack Guide: Infrastructure & Data

Build practical AI for pharma and biotech with IntuitionLabs. We help life-science teams turn complex information and workflows into useful software, governed knowledge systems and AI tools.

Published by IntuitionLabs · 2026-02-10 · biotech software stack · r&d digitization · series c funding · cloud infrastructure · data management · lims · eln · bioinformatics · regulatory compliance · life sciences it

Original article: <https://intuitionlabs.ai/articles/biotech-software-stack-series-c-guide>



In this article

1. Executive Summary
2. Introduction and Background
3. Core Components of the Biotech Software Stack
4. Data and Evidence: Industry Trends and Statistics
5. Case Studies and Real-World Examples
6. Discussion: Implications and Future Directions
7. Conclusion
8. Tables

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 can use a **comprehensive tech stack** spanning cloud computing, data management, laboratory and clinical systems, and analytics as a roadmap, selecting components for their actual workflows ([McKinsey](#)). This emerging biotech software stack integrates four core layers – infrastructure, data, applications, and analytics – to streamline research workflows, support compliance controls, and enable AI-driven insights ([www.mckinsey.com](#)) ([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. Prioritize systems according to assay volume, trial phase, regulated activities, manufacturing plans, and staffing; defer components without a current use case.

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 ([www.mckinsey.com](#)) ([www.mckinsey.com](#)). However, Benchling's 2023 survey of 300 R&D/IT professionals found that 53% of R&D respondents used more than five scientific applications in day-to-day work and 84% of respondents used some custom-built software. These are sample findings, rather than an industry census. Organization-wide interoperability was reported at 28%; reusability was a separate measure at 30% ([original report](#)). This fragmentation can hinder collaboration and data reuse. Companies face barriers such as a shortage of IT talent and lack of built-for-science platforms ([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 ([www.benchling.com](#)) ([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 ([www.benchling.com](#)). Likewise, NVIDIA's May 2024 account describes Recursion Pharmaceuticals' BioHive-2 supercomputer with 504 GPUs, alongside a laboratory operation running more than 2 million experiments per week and a database exceeding 50 petabytes ([blogs.nvidia.com](#)) ([blogs.nvidia.com](#)).

A growing biotech should choose software to support its R&D workflows and applicable regulatory obligations. Depending on those activities, this may involve cloud computing, modern data architectures, specialized lab and clinical software, and analytics, alongside governance 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 (www.mckinsey.com). These tools can, in principle, shrink timelines by aiding site selection, patient recruitment, and generating higher-quality **regulatory documents** (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 (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 (www.mckinsey.com). In contrast, a modern stack integrates cloud infrastructure, unified data platforms, and workflow-specific applications, unlocking seamless data flow and reuse (www.mckinsey.com). This architecture enables agile incorporation of new methods (e.g. AI), reduces dependency on outdated systems, and boosts interoperability across functions (www.mckinsey.com) (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 (www.mckinsey.com).

Funding Stages in Biotech and the Role of Technology

Series C funding is a financing round, not a clinical or commercial milestone. For example, Terremoto’s April 2026 Series C financed Phase 1 clinical development (**company announcement**). Depending on development plans, priorities may include scaling manufacturing, coordinating outsourced production, or preparing trial systems. Software decisions should follow these activities and their data volumes, rather than the funding label.

McKinsey recommends tailoring the stack to organizational needs (**technology framework**). Use assay volume, trial phase, regulated record types, manufacturing plans, and staffing to define requirements. Assess quality procedures and electronic-system controls for their intended use rather than assuming out-of-the-box compliance.

This guide surveys infrastructure, data management, laboratory and clinical applications, and analytics as a possible implementation roadmap. Adopt the components needed for current work and plan integrations before adding complexity.

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** (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 (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, as a roadmap to adapt to the company’s development activities and resources.

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. Life sciences cloud market analysts project the segment to grow from ~\$25 billion in 2024 to over \$100 billion by 2034 (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 (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 (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 (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.

Security and privacy obligations depend on the organization and processing activity. Cloud deployments can support compliance controls, but encryption alone does not establish HIPAA compliance. Where a provider maintains ePHI on behalf of a HIPAA covered entity or business associate, an appropriate business associate agreement and applicable safeguards remain necessary ([HHS cloud guidance](#)). Assess risks, service coverage, access controls, logging, backups, and operating procedures before deployment; review other applicable privacy obligations separately.

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. Plan compute infrastructure around measured peak loads, data sensitivity, cost, and available operational support.

Key takeaway: Choose a scalable compute/storage foundation when workload and security requirements justify it. 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 (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)” (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.

Benchling’s 2023 respondent sample reported organization-wide data interoperability at 28% and reusability at 30%. Among IT respondents at companies with 1,000 or more employees, 40% supported more than 20 scientific applications ([original report](#)). 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 (www.benchling.com) (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 (blogs.nvidia.com)).

Security practices and legal requirements should be assessed separately. Use access controls, provenance, and appropriate logging for sensitive research data. [21 CFR Part 11](#) applies to specified FDA-required or submitted electronic records; sensitivity alone does not establish applicability. Identify the relevant predicate-rule records and intended uses. For applicable open systems, [section 11.30](#) calls for additional measures such as encryption and digital-signature standards as necessary under the circumstances, rather than a universal cryptographic-signature requirement for all R&D records. Assess closed-system controls separately and document validation and operational controls appropriate to intended use and risk ([FDA guidance](#)).

Key takeaway: The data layer is about making data usable and trustworthy. Start with consistent identifiers, metadata, ownership, and recoverable storage. Add centralized stores and integration pipelines when data volume and reuse needs justify them. Companies like Moderna achieve this by migrating R&D data into a single platform (www.benchling.com), while Recursion describes generating proprietary datasets through automated, standardized, repeatable experiments and accessing patient data through partnerships ([Recursion](#)). Its publicly released [RxRx research datasets](#) represent less than 1% of its full proprietary dataset. 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 (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 (www.universitylabpartners.org). Startups often begin with low-cost or open-source ELNs to improve record-keeping; consider a more scalable ELN when user count, record volume, and inventory integration warrant it.
- **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 (www.universitylabpartners.org). Consider a LIMS when sample volumes and traceability needs exceed what the current tracking process can reliably handle. 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, and can integrate automated screening or manufacturing systems as throughput requirements grow. Instrument data (raw files, QC metrics) are typically fed directly into the LIMS or databases.
- **Clinical Trial and Regulatory Systems**. For a clinical program, evaluate an Electronic Data Capture (EDC) system (e.g. Medidata Rave, Veeva Clinical) and, where coordination warrants it, a Clinical Trial Management System (CTMS). Quality workflows need document control, corrective actions, training records, and clear process ownership. Selecting a vendor does not establish regulatory conformity: FDA places responsibility for applicable Part 11 requirements on regulated entities ([FDA guidance](#)). Define intended use, relevant records, validation, access controls, audit trails, retention, and oversight before selecting software.
- **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. Defer full ERP until production planning, inventory complexity, or financial coordination warrants it. For example, Gilead uses SAP S/4HANA on AWS to run supply chain and finance globally (press.aboutamazon.com). Track resources and plan inventory integrations before expanding production.

- **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. Choose collaboration tools and access rules that match the team's coordination needs.

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: Prioritize applications for actual work: ELN/LIMS for laboratory records and sample tracking, EDC/CTMS for trials, quality systems for controlled processes, and administrative tools as staffing grows. Assign an owner to each system and verify its controls and integrations before relying on it. Many startups find that off-the-shelf SaaS offerings (benchmarked for biotech) are faster to deploy than building custom systems (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 (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 (blogs.nvidia.com) (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 (www.benchling.com) (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 chemoinformatics). 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. In NVIDIA's May 2024 account, Recursion's 504-GPU BioHive-2 delivers approximately 2 exaflops of **AI performance** and is nearly five times faster than BioHive-1 (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. Evaluate GPU access and model pilots when a defined scientific task and suitable data justify the cost.

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 (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. Start with reproducible analysis scripts and the metrics the team needs. Expand to model pilots and managed analytics when data quality, expected benefit, and maintenance capacity justify them.

Data and Evidence: Industry Trends and Statistics

Adoption of Enabling Technologies. Benchling's 2023 survey sampled 300 R&D/IT professionals; these respondent findings are not an industry census. In that sample, 70% reported R&D-data-platform adoption (not exclusively cloud platforms), 63% robotics/automation, and 59% AI/ML adoption. The 18% SaaS measure meant that at least three of five scientific applications used by R&D or supported by IT were cloud-based, rather than a share of workload. Some custom-built software was used by 84% of respondents. Among R&D respondents, 53% used more than five scientific applications in day-to-day work; 40% of IT respondents at companies with 1,000 or more employees supported more than 20 scientific applications ([original report](#)).

Market Growth and Spending. Multiple analyses project rapid growth in biotech and life science software markets, driven by AI and digital R&D. For budgeting, prioritize the workflows, data volumes, and organizational needs the stack must support ([McKinsey](#)). Similarly, the broader **life science cloud computing market** is expected to expand at ~15% annual rates through 2034 (www.towardshealthcare.com) (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 (www.towardshealthcare.com) — dominates due to its mature digital ecosystem. SaaS (cloud software) leads the models in life sciences (www.biospace.com), aligning with the move towards subscription-based biotech platforms.

Quality and Compliance Planning. Define quality procedures and system controls according to the company's activities and applicable rules. FDA states that regulated entities are responsible for applicable Part 11 requirements ([FDA guidance](#)). Compare commercial quality software and documented manual procedures against those requirements and the team's capacity; use [build-versus-buy considerations](#) as planning input, rather than a financing-stage compliance deadline.

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 (www.benchling.com). To accelerate R&D, Moderna has rolled out Benchling's integrated R&D cloud across hundreds of scientists globally (www.benchling.com). The aims are to **consolidate previously siloed systems, automate experimental workflows**, and enforce **consistent, AI-ready data standards** (www.benchling.com). In practice, Moderna standardized naming schemes and schema across labs, so that experiment records now feed directly into computational pipelines (www.benchling.com) (www.benchling.com). This has yielded benefits such as faster data sharing between teams and built-in audit trails (www.benchling.com) (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 (www.benchling.com) (www.benchling.com).

Recursion Pharmaceuticals (AI-Driven Discovery). Recursion is a small-molecule biotech that heavily leverages automation and AI. NVIDIA's May 2024 account describes its **BioHive-2 supercomputer** as a major biopharma AI system with 504 NVIDIA H100 GPUs delivering approximately 2 exaflops of AI performance, nearly five times the speed of BioHive-1 ([NVIDIA](#)). That account also reports more than 2 million laboratory experiments per week; laboratory throughput is separate from the supercomputer's AI performance. 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 (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 historical example illustrates a large-scale 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 hypothetical biotech focused on a novel biologic and preparing for a clinical trial. This illustrative stack is a set of options to evaluate against its needs ([McKinsey](#)). 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. For HIPAA-regulated ePHI, assess covered services and execute an appropriate business associate agreement; encryption alone is insufficient, and risk assessment and operational safeguards remain necessary ([HHS](#)).

- **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). For records within Part 11's scope, the regulated entity must verify applicable requirements; vendor selection alone does not establish conformity (FDA).
- **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 a plan to move to cloud GPUs for scaling.

This startup's architecture shows how the theoretical layers become working tools. Use the architecture to plan data flows and compliance controls. Verify integrations, applicable obligations, and the deployed systems before relying on the workflow; responsibility for applicable electronic-record requirements remains with the regulated entity (FDA).

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" (blogs.nvidia.com). Data standardization and integration reduce human error and ensure reproducibility, which both speeds regulatory review and instills confidence in results (www.benchling.com) (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" (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 (www.cloudcomputing-news.net). Talent is a bottleneck too: finding staff who understand both biology and data systems is hard (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. Track emerging capabilities and adopt them when they serve a defined scientific or operational need, with enough flexibility to extend the stack later.

Conclusion

Biotechnology startups should build a software foundation proportionate to their development activities. The **biotech software stack** – spanning cloud infrastructure, data systems, lab and clinical applications, and analytics – offers a roadmap to adapt as needs change ([McKinsey](#)). 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 (www.mckinsey.com) (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.

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 (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.

A Prioritized Implementation Sequence

Treat the following as a planning recommendation, rather than a financing-stage requirement:

1. **Secure and organize the essentials.** The IT/security owner and scientific lead should define access, recoverable storage, shared identifiers, and experiment records. Test recovery and exports before depending on the system.
2. **Add laboratory tracking when volume warrants it.** The lab operations owner should introduce or expand ELN/LIMS workflows when sample reconciliation, handoffs, or assay volume strain the current process. Agree on sample IDs, metadata, and instrument interfaces first.
3. **Prepare systems for regulated activities.** Quality and clinical owners should map applicable record requirements and controls before trial or manufacturing workflows start. Define validation, vendor oversight, retention, and operating procedures before procurement.

4. **Integrate and automate proven bottlenecks.** The data engineering owner should add pipelines or a warehouse when cross-study reuse or manual transfers justify them; use stable identifiers and documented mappings.
5. **Expand analytics after demonstrating value.** The data science owner should pilot models on fit-for-purpose data and compare their benefit and cost with existing analysis. Defer dedicated GPU clusters, full ERP, or extra platforms without a named owner, current use case, and maintenance budget.

Tables

Benchling's 2023 survey sampled 300 R&D/IT professionals. Percentages describe respondents or the specified subgroup, rather than all industry companies. See the [original report \(readable copy\)](#).

Measure (2023 respondent sample)	Reported proportion
R&D data platform adoption (any deployment model)	70%
Robotics and lab automation adoption	63%
AI/ML adoption	59%
At least 3 of 5 scientific applications used (R&D) or supported (IT) are SaaS/cloud-based	18%
R&D respondents using more than 5 scientific applications in day-to-day work	53%
IT respondents at companies with 1,000+ employees supporting more than 20 scientific applications	40%
Respondents using at least some custom-built software	84%
Organization-wide data interoperability	28%
Organization-wide data reusability (separate measure)	30%

Software/System Category	Purpose/Function	Example Platforms/Providers
<i>Compute & Cloud Infrastructure</i>	Scalable computing, storage, networking for data and workloads	AWS, Azure, Google Cloud; HPC clusters; Oracle Cloud HPC
<i>Data Management & Integration</i>	Data warehousing, curation, FAIR data practices, ELT workflows	Snowflake, Databricks, Apache Airflow, Benchling R&D Cloud
<i>Electronic Lab Notebook (ELN)</i>	Digital experiment record-keeping, protocol templates	Benchling, Labguru, Revvity Signals Notebook
<i>Laboratory Information Mgmt (LIMS)</i>	Sample/inventory tracking, SOP management, audit trails	LabWare, Autoscribe, CloudLIMS, LabKey
<i>Lab Automation Control</i>	Schedule/control of robotic instruments and assays	Arctoris Ulysses, Opentrons (OT-2), Benchling Automation API
<i>Quality Management (QMS)</i>	Document control, CAPA, training, compliance support	MasterControl, Veeva Vault QMS, Qualio, Greenlight Guru

Software/System Category	Purpose/Function	Example Platforms/Providers
<i>Clinical Data Capture (EDC/CTMS)</i>	Clinical trial data collection, study management	Medidata Rave, Oracle Clinical, Veeva Clinical Suite
<i>Analytics & AI Platforms</i>	Data analysis, ML workflows, scientific insights	Jupyter/Python/RStudio; NVIDIA GPUs; DNAnexus; Terra (Broad); SAS, Spotfire
<i>Collaboration/Project Mgmt</i>	Team communication, documentation, project tracking	Slack, Microsoft Teams, Confluence, Jira, Notion
<i>Security & Compliance Tools</i>	Access control, encryption, audit logging	AWS/GCP IAM & KMS; Fortanix, CyberArk, Vault; GxP compliance modules
<i>ERP/Supply Chain (long-term)</i>	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 (www.mckinsey.com). Benchling's [original survey report \(readable copy\)](#) informs the respondent measures. The specific platforms listed are representative of those commonly used in biotech R&D and are mentioned in sector analyses and case studies (www.benchling.com) (blogs.nvidia.com). (No single source covers all categories; the table is a synthesis of cited insights.)

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. <https://www.mckinsey.com/industries/life-sciences/our-insights/boosting-biopharma-r-and-d-performance-with-a-next-generation-technology-stack>
2. <https://intuitionlabs.ai/articles/lims-vs-elms-sdms-cds>
3. <https://intuitionlabs.ai/articles/deviations-capability-change-control>
4. <https://www.scribd.com/document/961643401/Benchling-2023-State-of-Tech-in-Biopharma>
5. <https://www.cloudcomputing-news.net/news/2023-state-of-tech-in-biopharma-report-reveals-tech-strategies-in-era-of-data-and-ai/>
6. <https://www.benchling.com/news/benchling-and-moderna-collaborate-ai-driven-research>
7. <https://www.benchling.com/customer-stories/moderna-building-ai-driven-research-and-technical-development-for-tomorrow>
8. <https://intuitionlabs.ai/articles/gxp-elms-comparison-benchling-idbs-labarchives>
9. <https://blogs.nvidia.com/blog/drug-discovery-recursion-supercomputer/>
10. <https://intuitionlabs.ai/articles/iso-27001-life-sciences-compliance-guide>
11. <https://intuitionlabs.ai/articles/ectd-publishing-best-practices>
12. <https://terremotobio.com/press-release/terremoto-biosciences-closes-108-million-series-c-financing-to-advance-selective-akt1-inhibitors-in-oncology-and-hereditary-hemorrhagic-telangiectasia/>
13. <https://www.towardshealthcare.com/insights/life-science-cloud-market-sizing>
14. <https://press.aboutamazon.com/2021/11/gilead-and-aws-collaborate-on-development-and-delivery-of-new-medicines-for-patients>
15. <https://www.hhs.gov/hipaa/for-professionals/special-topics/health-information-technology/cloud-computing/index.html>

16. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11/subpart-A/section-11.1>
17. <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11/subpart-B/section-11.30>
18. <https://www.fda.gov/media/166215/download>
19. <https://www.recursion.com/news/since-its-inception-recursion-has-been-building-the-foundation-for-the-first-virtual-cell>
20. <https://www.rxrx.ai/>
21. <https://www.universitylabpartners.org/blog/laboratory-information-management-systems-lims-for-biotech-startups>
22. <https://intuitionlabs.ai/articles/biotech-qms-build-vs-buy>
23. <https://thelabstack.com/search>
24. <https://www.biospace.com/press-releases/life-science-cloud-computing-market-accelerates-as-ai-and-digital-transformation-redefine-r-d-and-clinical-operations>
25. <https://ir.revity.com/news/investor-news/news-details/2023/Revity-Launches-First-Scientific-SaaS-Platform-to-Accelerate-Drug-and-Materials-Development/default.aspx>

THE PUBLISHER

About IntuitionLabs

IntuitionLabs is an AI consulting, custom software development and data engineering firm serving pharmaceutical, biotechnology, medical-device and other life-science organizations. We work with clinical, regulatory, medical-affairs, commercial, quality and IT teams to connect technology decisions with the work people need to accomplish.

AI consulting and adoption

Our AI enablement services cover readiness assessments, use-case selection, governance and policies, team workshops, adoption measurement and ongoing advisory support. We help organizations structure the information layer behind AI: source material, context, permissions and maintained knowledge that make generated answers useful and reviewable. Private LLM inference and hosted AI options support teams evaluating how to operate AI with appropriate control over their data and infrastructure.

Software, data and life-science workflows

IntuitionLabs develops custom software for pharma and biotech, integrates enterprise systems, and builds data engineering and business intelligence solutions. Areas of focus include AI agents, regulatory research, medical writing, medical affairs, CMC information, competitive intelligence and clinical-document workflows. Our eTMF intelligence work includes cross-system reconciliation and inspection-readiness support.

Enterprise platforms and regulated delivery

We provide Veeva services, application support, managed services, integrations and custom applications, alongside enterprise content work involving platforms such as Egnyte. For regulated workflows, our services include GxP enablement, computer-system validation and software development addressing 21 CFR Part 11 requirements. The applicable controls, validation responsibilities and acceptance criteria are defined for each engagement.

Work with IntuitionLabs

Explore [AI enablement](#), [pharma and biotech software development](#), [data engineering and BI](#), and [Veeva services](#). [Contact IntuitionLabs](#) to discuss your workflow, information sources and implementation needs.

IntuitionLabs publishes educational research to help life-science teams make informed technology decisions. Coverage of a product or organization does not imply a client relationship, endorsement or partnership.

Website: <https://intuitionlabs.ai>

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.