Benchling

What to Look for in an Informatics Platform for Life Sciences R&D

benchling.com

Introduction

In silico molecular biology tools, electronic lab notebooks, and systems for registration, sample management, workflow management, and request management are the foundation of data management for modern life science R&D.

While many R&D organizations have adopted individual systems that address these needs independently of one other, the industry is recognizing the need for a more unified approach. For many IT departments, integrating data systems is a top priority, and more and more of them are looking to informatics platforms as a solution. Informatics platforms offer a wide range of capabilities – which would typically be provided by separate software solutions – in a single interface.

The informatics platform you adopt and develop over time is an important consideration with serious consequences for both R&D and IT. It impacts the research productivity of your scientists, the decisions you make as an R&D leader, and your success in bringing products to market — as well as the ability of IT to strategically support R&D and ensure the longer-term relevance of your infrastructure.

The five critical platform capabilities to look for when evaluating an informatics platform for modern life science R&D are:

Purpose-Built for Modern Life Sciences

Modeling complex entities and processes

Unified Data Infrastructure

Harnessing organization-wide data to accelerate R&D

Codeless Configuration

The benefits of custom, with the ease of off-the-shelf

Open Integration

Not all APIs are created equal

Easy Extensibility

Designed for IT innovation

Conclusion

Transforming Life Sciences R&D

Purpose-Built for Modern Life Sciences

Modeling Complex Entities and Processes

Over the last decade, industries from pharmaceuticals and biotechnology, to energy and food production, have seen a big shift toward biologics and processes based in biology. With this shift have come increased R&D complexity and uncertainty.

By comparison, the processes of traditional chemistry-based R&D are well-characterized. But cell-based processes must be rapidly iterated upon. R&D organizations that rely on them aren't just developing a single drug or product: they're developing a discovery platform.

Any software platform for modern life science R&D needs to accommodate **rapid process changes** and capture, synthesize, and report on **countless types of multi-dimensional data** from across the organization. At the same time, it has to accomplish **seamless, often bidirectional, handoffs** between numerous teams.

As legacy informatics providers expanded into modern life science, many of them simply re-purposed their small molecule software, with disappointing results. Fundamentally, platforms that are effective for modern life sciences R&D must be built from the ground up to address the field's unique, complex needs.

Model the Complexity of Biological Entities

While many life science R&D organizations have some chemistry-based workflows, the majority of their work centers around complex biological entities such as cell lines, plasmids, and proteins. Digitally expressing these entities is a challenge in and of itself, but expressing the relationships between them adds an additional layer of complexity.

Model Any Type of Biological Entity

An informatics platform for modern life science should enable you to model any biological entity – from antibodies, to viruses, to yeast strains – and its unique attributes.

Structure Linkages between Entities

Flexibly structure relationships between any type of entity. For example, link plasmids to viruses, and then to cell lines.

Surface Parent Data on Child Entities

Directly display relevant characteristics of certain entities onto related entities. For example, surface a tag from a plasmid up to a protein, so you can easily see what tag to use for purification.

Rollup All Relevant Assay Data and Results to Lots

Centralize data and results from individual lots up to their parent lots. Create a rich history of experiments across your organization's samples.

Model the Complexity of Life Sciences R&D Processes

The processes of modern life sciences R&D take place across numerous teams, branches, and cycles, all of which need to be modeled using an informatics platform's process management capabilities. Additionally, since life sciences workflows are nonstandard and not always well-defined, R&D needs to quickly iterate on and analyze the results of these processes.

Model Any Type of Workflow

Given the nonstandard nature of life sciences R&D, look for a platform that's flexible enough to adapt to your organization's particular workflows for processes such as assay development and screening.

Built for Rapid Optimization

Swiftly altering workflow structure is crucial to optimizing processes and identifying the most promising candidates.

Combine Process Flexibility with Structured Results Capture

When processes change frequently, it's critical to have a system that provides both flexible workflow design and the structured results capture that allows for easy data analysis.

Power Multi-Team, Iterative Processes

Look for a platform that can assign tasks to teams and individuals, equip them with complete experimental context, and surface their progress as tasks are carried out.

Manage R&D Programs

The platform should give managers a central dashboard on which they can track the progress and performance of all your organization's projects.

Unified Data Infrastructure

Harnessing Organization-Wide Data to Accelerate R&D

The years of organizationally complex research that go into a single candidate scatter results and institutional knowledge across different systems. This makes it difficult for teams to synthesize data from across the organization and use it to accelerate R&D.

Connecting disparate systems, one-by-one, in an attempt to centralize data is extremely expensive and time-consuming. Oftentimes, companies can end up with a rigid system that breaks often and is difficult to use.

The only way to make sense of all this data and harness its full potential is with an out-of-the-box platform that **automatically centralizes and interlinks data** across your organization.

Your platform should consist of unified in silico design tools, an electronic lab notebook, registration, sample management, workflow management, and request management modules that work together to streamline scientists' work. Your platform should also integrate with instruments such as liquid handlers, so that the large quantities of data your instruments produce will be centralized and contextualized. With all of these applications working together on a single platform, you can make comprehensive, fully-informed business decisions.

Automatic Data Linkages Minimize Busywork

- **In-Line Registration**
  Look for a platform that lets you register entities directly from the ELN or molecular biology design applications.

- **Link Results Data with Samples**
  From a particular result, directly access the record of the sample that produced it.

Auto-Linking between Samples and Upstream Entities

Configure relationships that automatically link physical samples with parent entities, such as cell lines and sequences.

Reliable Reproducibility

Completely interlinked data and conditions mean any experiment should be reproduced with minimal effort.

Bidirectional Linkagess

Any link between two files should be automatically bidirectional. For example, linking in a lab notebook entry to a sample will create a link on the sample record to the lab notebook entry.

Collaboration and Cross-Learning for Your Entire Organization

Full History of Any Piece of Data

Your platform should surface the origin of any notebook entry, protocol, process, entity, sample, or sequence, as well as everywhere it has been used or mentioned.

Eliminate Re-Research

With easily accessible experimental records, scientists should be able to quickly identify prior efforts relevant to their current research.

Seamless Team Hand-Offs

Give colleagues links to samples and a complete history of upstream activity.

Submit Requests to Teammates

Structure tasks with desired protocols, inputs, and outputs, and send them to team members.

Institutional Knowledge Base

Deepen retained institutional knowledge over time by gradually building up your organization's records.

Tracking Every R&D Sample, Process, and Result from Start to Finish

Real-Time R&D Progress

For every project across your organization, dashboards should display step-by-step progress and up-to-date timelines.

Measure Output in Real-Time

Configure live dashboards that report process output and efficiency.

- **Lineage Tracking**
  After tracing from physical lots up to parent entities, compare the functional data of lots across different parent entities.

- **IND Filings**
  Complete IND filing in days by exporting the complete experimental history of candidates.

- **Pinpoint Locations of Lots from Parent Entities**
  On any entity, see the locations of every physical lot of it that has been produced.

- **View Progress Across R&D Groups**
  Directors and VPs should be able to get a 360-degree executive view of all processes and results across different groups.

Data-Driven Decisions That Accelerate R&D and Increase Efficiency

- **Compare Results Across Different Conditions**
  Drive process optimization by identifying the conditions that produce desired results.

- **Identify Successful Parent Entities**
  Compare the functional data of different lots to identify which parent entities lead to desired results.

- **Measure Tradeoffs**
  Across process iterations, measure tradeoffs between quality of results, methods used, and resources.

- **Eliminate Bottlenecks**
  Identify the processes that take the most time or resources.

- **Resource Utilization and Planning**
  With centralized experimental records, easily track resource utilization and forecast future needs.

- **Mine Success Patterns**
  Across candidates, determine upstream predictors of effectiveness to streamline the research process.

- **Identify Promising Technology Platforms**
  Single out platforms showing the most promise to advance more effective candidates faster.

Codeless Configuration

The Benefits of Custom, With the Ease of Off-The-Shelf

Traditional R&D informatics platforms have required heavy, ongoing vendor involvement for configuration. While it's important to choose a vendor that can provide guidance on setting up a data model that meets your unique needs, you shouldn't be dependent on a vendor when you need to make adjustments in the future. Oftentimes, even simple changes that should be done with the press of a button can take months and additional fees for a vendor to accomplish.

Your platform should offer a point-and-click interface for configuration changes, rather than require changes to code. The available changes to configuration should range from modeling new entities and designing new workflows, to simpler changes such as adding new fields to entities.

> Look for a platform that can be **configured and re-configured at the discretion of IT** or designated R&D users, without the need for extensive coding.

Custom Permissions

You should be able to structure permissions based on business criteria. For example, define access to entity and data types based on a user's role, team, function, or any other user characteristic. Also configure permissions to limit the data accessible by external collaborators.

Custom Entity Types and Relationships

Look for a platform that can model any biological entity, from patient samples to yeast cells. Require specific entity types to link to other entity types; for example, require that every expression plasmid have a link to the antibody that it expresses. You should be able to add and interlink new entities just as easily.

Custom Fields

Design custom workflows with trackable stages that can be assigned to teams. In a screening workflow, for example, configure stages to capture hits on individual wells. Then, link those hits to the lots of antibody libraries used.

Custom Workflows

Design custom workflows with trackable stages that can be assigned to teams. Designated users should be able to model new workflows with a point-and-click interface. In a screening workflow, for example, configure stages to capture hits on individual wells. Then, link those hits to the lots of antibody libraries used.

Custom Formulas

Across interlinked entities, surface data from one entity onto an entity that it links to. For example, for each lot of expression hosts, show the average yield and purity of the antibody lots it has produced.

Open Integration

The Benefits of Custom, With the Ease of Off-The-Shelf

The standard for integration of R&D systems used to be hard-coded, one-to-one connections between different software products. As the number of these connections grew unmanageable, it became clear that a platform approach could offer distinct benefits. Informatics platforms promise an end to this architectural complexity, since a platform can serve as a centralized hub for many stable, API-driven integrations. But not all platforms' APIs are created equal. If the right endpoints aren't exposed in the right way, then certain integrations simply aren't possible. Other times, a platform's APIs can just be slow.

Completely exposed APIs ensure that when you're integrating your platform with other software, you're only limited by the APIs of the other software. IT can then link every relevant application to the central platform, enabling them to exchange data in real-time. The platform should standardize the outputs of each application, making the data of each application accessible to every other application.

For modern life science R&D, the ideal is a unified data platform with **completely exposed REST APIs,** described with accessible documentation.

Integrate Instruments

Integrate any instruments, such as liquid handlers, smart fridges, and plate readers, to automate data capture. For example, use a matrix scanner to scan a plate and open it in your informatics platform, along with the full lineage of the contents of every well in the plate.

Integrate Software Applications

Beyond instruments and data stores, an informatics platform should be able to integrate with other software when necessary. For example, directly integrate data analysis software such as Spotfire to automatically run and visualize complex queries. As another example, integrate niche point solutions to maintain necessary functions that a platform might not perform out-of-the-box.

Integrate Databases

Alongside its APIs, your informatics platform should have a built-in data warehouse to accomplish bidirectional integrations with your other databases, warehouses, and lakes. For example, link your manufacturing and R&D databases to streamline data transfer according to business and compliance rules.

Easy Extensibility

Designed for IT Innovation

No matter how rich the functionality of your R&D informatics platform for notebooking, registration, sample tracking, workflow management, and other needs, R&D will always need additional functionality that no off-the-shelf software can accommodate. Building custom software for these needs typically involves spinning up a separate hosting environment, dedicating developer resources, and spending extensive time on boilerplate.

By minimizing IT overhead and producing a smooth user experience, this approach opens up wholly new development possibilities.

> Rather than invest in a separate development platform that would only leave scientists with yet another distinct piece of software, look for an R&D informatics platform that **accommodates custom development** within itself.

Custom Calculations

When an organization has its own SOPs for calculating particular properties, they often have to export their samples, run the analysis, and then re-upload the results. Look for a platform that lets you embed code for any proprietary calculation so that you can, for example, calculate the molecular weights of your proteins with whatever standards you desire.

Build Validation Rules

Put complex constraints on the platform's data validation, such as on entity types. For example, require that an antibody entity links to a unique set of light and heavy chain entities. Or, define success thresholds so that only antibody lots above a certain purification value move to the next stage of a workflow.

Build New Applications

On top of the platform, build completely new applications to support your niche workflows. For example, if you have complex SOPs around labeling, you could develop an application that prints barcode labels for a certain set of samples that you know a particular workflow will generate. By developing applications such as these within your platform, you can ensure adherence to SOPs while streamlining scientists' work.

Conclusion

Transforming Life Science R&D

Across different organizations, the needs of modern life science R&D are complex, variable, and fluid. To manage this complexity, you need a life science informatics platform that's built to interlink R&D data, track R&D progress, and drive R&D decisions. In addition, to accommodate the evolving R&D needs of your organization, an informatics platform should offer codeless configuration, open integration, and easy extensibility.

Put another way, R&D organizations need a platform that tames the complexity of modern life science with unified, off-the-shelf functionality. But they also need a platform that's flexible and easy to customize. Platforms that combine the best of both worlds eliminate vendor dependence and minimize IT overhead while aligning with R&D's needs.

By evaluating platforms based on these capabilities, you will transform your R&D organization. Instead of troubleshooting unwieldy software, IT will be able to proactively focus on more strategic projects. And rather than be hindered by hard-to-use software, R&D will work more efficiently to bring drugs to market faster.

Benchling

The Life Sciences R&D Cloud

Ready to learn more? **Request a Demo**

benchling.com