

WHITE PAPER

# Harnessing AI for Breakthroughs in Pharma and Biotech



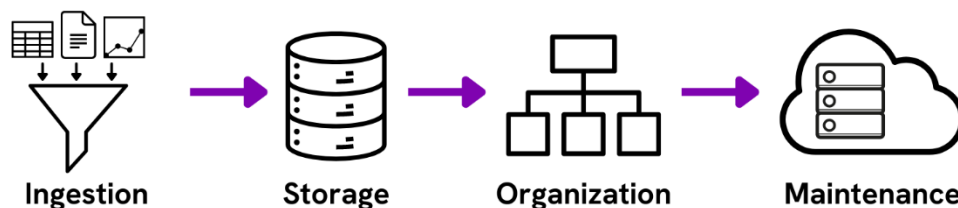
revvity  
signals



## The Crucial Role of Data Management

The latest artificial intelligence (AI) and machine learning (ML) methods are transforming the research and development process within drug discovery, delivering on the promise to accelerate the journey from laboratory insights to lifesaving therapies. But AI techniques aren't enough on their own; AI is only as powerful as the data it consumes, and it works best with data that has the proper quality, detail, and context. While there is abundant scientific data in pharmaceuticals and biotechnology, it is often siloed, poorly modeled, and ultimately inaccessible. A meticulous approach to capturing and curating R&D data is indispensable to unlock the full potential of novel AI methodologies.

This white paper explains how optimizing data management – the ingestion, storage, organization, and maintenance of data – helps leaders in drug discovery leverage AI in their mission to deliver innovative drugs to patients in need.



## AI Fuels Scientific Discovery

While the growing volume, variety, velocity, and complexity of data today is overwhelming for scientists to sift through, AI rapidly finds patterns and discovers hidden insights that humans would otherwise miss. AI's capacity to consume and analyze complex datasets and ML's ability to learn from and predict based on that data can lead to exponential improvements in drug discovery and patient outcomes. In accelerating the search and analysis of data, tools like generative AI and ML algorithms not only give back quality time to scientists, they also fuel what scientists do best: think, experiment, and discover new therapies faster.

Use cases for AI vary widely, from target identification and validation to clinical trial design and patient recruitment to drug formulation and streamlining regulatory filings. It's been applied to analyze complex biological data such as genomic data sequences, clinical trial results, and drug compound properties. It's also used to predict which compounds in dense chemical libraries have the best potential for therapeutic effect or to suggest the conditions or a particular catalyst for a chemical reaction that produces a desired result. ML helps researchers test and improve the design of antibodies or compounds.

---

*AI works best with data that has context*

---



Another benefit of AI is the speed it brings to finding, analyzing, and summarizing large volumes of external literature. It can surface emerging trends, popular research topics, and shifts in scientific paradigms over time. AI plumbs the depths of scientific publications and papers faster than humans possibly can, helping researchers to quickly grasp the main findings and significance of studies much faster – leading researchers to stronger hypotheses and more promising discoveries. Techniques like natural language processing (NLP) extract specific information from scientific texts, such as outcomes, methodologies, and statistical results, while semantic search provides the scientific context that delivers more relevant results based on the meaning of queries compared to general keyword matches.

Collectively, these AI/ML capabilities inform the drug discovery process by identifying potential therapeutic targets, predicting drug efficacy, and optimizing treatment strategies, while predictive models help researchers anticipate adverse effects, streamline experimental designs, and prioritize resources for maximum impact.

In the future, researchers could potentially use mathematical techniques to make discovery even more efficient. They could use generative AI methodology to develop entirely novel drug concepts based upon existing data and pair that with ML/predictive techniques analytics around synthetic pathways. Alternatively, researchers could use data and AI to avoid drug design areas where prior work was unfruitful due to lack of efficacy or unacceptable toxicity, thus focusing on more promising design spaces. Another potential use of data and AI is [drug repurposing](#), where researchers find new applications for previously approved drugs.

The use of generative AI in predicting novel therapies is still an area of active research and development. While there are challenges to overcome, such as ensuring the safety and efficacy of AI-generated therapies and navigating regulatory approvals, the potential benefits of faster, more efficient drug discovery processes are substantial. As AI technology continues to evolve, it will play an increasingly significant role in the future of medical research and healthcare innovation.



### Use Cases for AI in Drug Discovery:

- **Drug Design and Molecule Generation:** Generative AI models can design new molecular structures with desired properties for treating specific diseases. These models learn from existing databases of chemical compounds and can generate novel molecules that are predicted to have the desired efficacy and safety profile.
- **Target Identification:** AI models can analyze biological data to identify new targets for drug therapy. This involves understanding the complex interactions within biological systems and identifying points where therapeutic intervention could be beneficial.
- **Biomarker Discovery:** Generative AI can also aid in discovering new biomarkers for diseases, which can lead to the development of more precise and targeted therapies.
- **Personalized Medicine:** By analyzing patient data, including genetic information, AI models can predict how individual patients might respond to certain treatments, leading to more personalized and effective therapy plans.
- **Simulation and Prediction of Drug Effects:** AI models can simulate the interaction between drugs and biological systems, predicting drugs' efficacy and potential side effects before real-world testing.
- **Enhancing Clinical Trials:** Researchers can use AI to design clinical trials, identify suitable participants, and predict outcomes, thus speeding up the development process and reducing costs.



### AI Improves Use of Scientists' Time

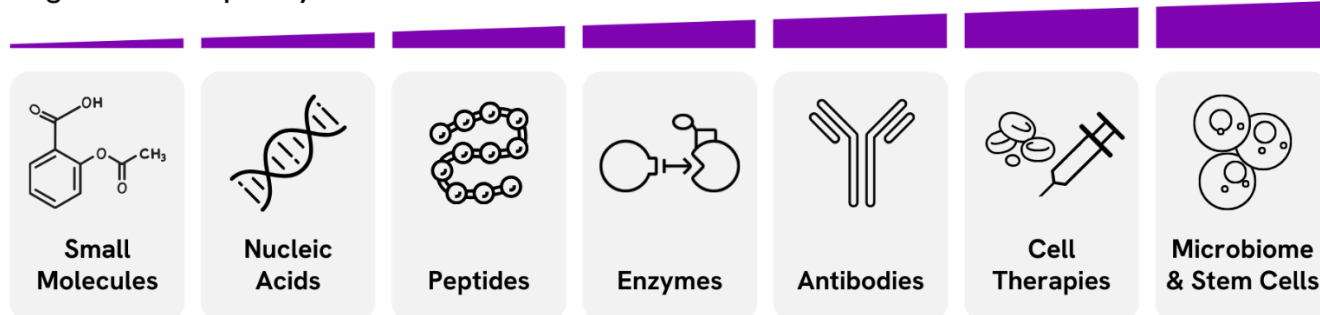
- Just like in other industries, AI can replace repetitive, mundane tasks that get in the way of higher-value work. Researchers can use AI as a helper or copilot:
- Deploy a large language model to summarize a peer-reviewed paper
- Use generative AI to create templates for protocols and more, instead of manual processes
- Use AI to enable more intuitive search by using natural language
- Build dashboards for advanced analytics and assistance during the drug discovery journey



## AI Success Relies on Effective Data Management

All industries are exploring AI, and many are already experiencing the benefits of automating tedious, data-heavy tasks. As a data-intensive enterprise, drug discovery must embrace AI today to efficiently deliver innovative treatments that cure or treat illness and extend lives. Modern drug therapies based on new drug modalities are more complex and face more precise oversight, than those developed by pharmaceutical scientists in previous decades. These newer drug modalities (antibodies, cell and gene therapies, etc.) comprise large, intricate molecular entities, and potentially terabytes and petabytes of data, that humans are challenged to interpret.

Degrees of Complexity →



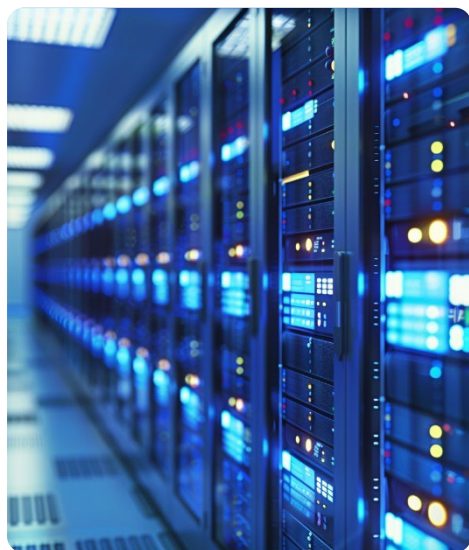
Furthermore, biological assay testing is elaborate. There are numerous types of tests, each producing complex datasets that are inherently difficult and time-consuming to understand. Therefore, there is a higher chance for humans to overlook crucial elements and insights that could lead to a more comprehensive understanding of the mechanism of action with the end goal of novel therapies.

As AI advances, its impact on scientific discovery deepens. To maintain this momentum, organizations must optimize data management. A robust data management system helps ensure that data is not only captured comprehensively and as early as possible, but that it is also curated in a manner that preserves its scientific context and enhances its utility for AI and ML processing today and in the future.

---

*Early capture and curation of data, preserving its scientific context, enhances its utility for AI and ML processing.*

---



## Crucial Aspects of Data Management

Research organizations face a common challenge: They are eager to use the latest AI technologies to discover insights among abundant data, but when they attempt to use AI/ML models, they often run into problems because the datasets aren't well managed. Awash in experimental and results data, they find it difficult to organize, structure, tag, and otherwise prepare the data in the best way to apply AI/ML algorithms. Data is often collected in laboratories without useful context or metadata that describes what the data entails. A file with tabular results, for example, is not useful if it does not provide the context to describe what was measured. In other words, data should have a level of enrichment that helps facilitate drawing meaningful insights without researchers having to struggle to understand what the data means.

## The Importance of Metadata

Data is not enough on its own; it needs to be associated with the appropriate experimental context to describe the data. Metadata, which can be described as data about data, contextualizes each piece of information, detailing its origin, nature, and specifications. This includes information about the methods used to collect the data, the conditions under which experiments were conducted, the parameters measured, and the protocols followed. The FAIR principle applies here: data should be findable, accessible, interoperable, and reusable. This requires sufficient metadata and unique and persistent identifiers; that data is both human and machine-readable; that it share a common structure; and have strong governance for its use.



**F**indable



**A**ccessible

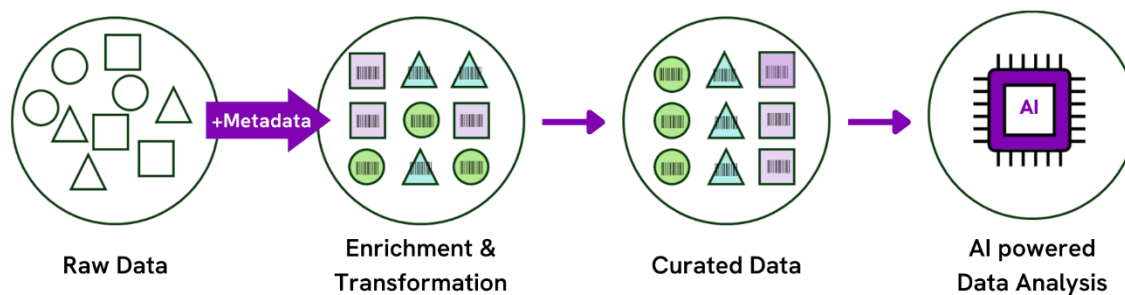


**I**nteroperable



**R**eusable

For AI to be successful, data management practices must enrich and contextualize data with metadata and subsets of data. Tagging and metadata make the data effective for training ML models, which continue to improve – or learn – over time. For example, having more information on candidates' toxicity could better inform ML models and steer researchers away from false routes.



Without this context, the resulting data is susceptible to misinterpretation and can lead to inaccurate conclusions, potentially derailing the research process and compromising the development of new therapeutics.

Ensuring the accuracy and completeness of the metadata is no small challenge. It requires robust validation mechanisms within the data management system. When intuitively integrated into the data collection process, at the time it is top of mind for the researcher, metadata capture is better contextualized (or enriched) and primed for analysis.

Contextualization via metadata ultimately gives users the intended results and benefits of AI. Properly enriched data at scale enables algorithms to perform the analytics and to reveal meaningful patterns.

#### Data without context lacks value:

#### Best practices for data enrichment

- Capture and organize machine-readable metadata that contextualizes scientific results.
- Use tools that prompt scientists to fully and effortlessly document all details that inform captured results.
- Incentivize scientists to capture and attach metadata early for success with scientific AI/ML analytics.
- Enable flexibility with late binding of schema, leaving room for restructuring the data for analysis in new, unimagined ways in the future.



### The Limitations of Rigid Data Models

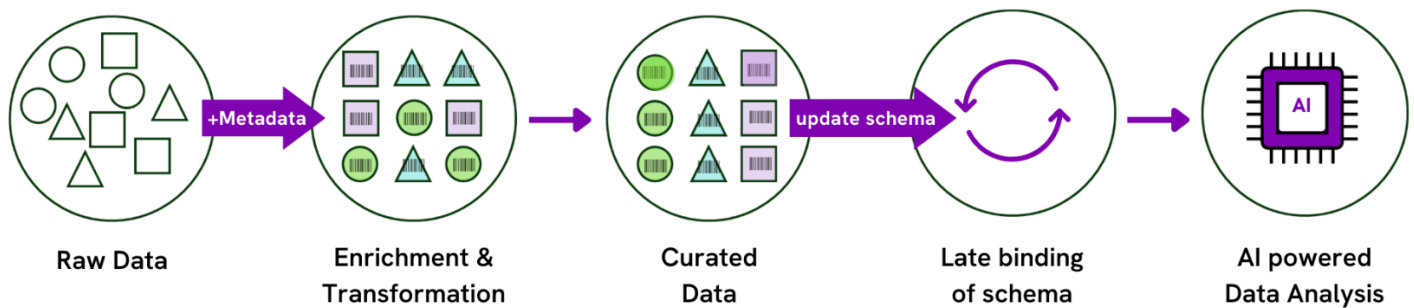
Many scientific organizations use rigid data schemas with predefined structures and strict parameters. This type of data model forces scientists to fit their rich, multifaceted data into limited input boxes, often at the expense of nuance and detail that could be crucial for future discoveries. Scientific discovery is inherently unpredictable, and data captured today may need to answer questions that arise tomorrow.

Data bound too early into a rigid schema may lose its potential to inform future analyses or be re-interrogated under a new scientific lens. Additionally, rigid data models can impede the integration of disparate data types, which is an increasingly critical capability as research becomes more interdisciplinary. Combining biochemical data with clinical observations, patient demographics, and real-world evidence is essential for a holistic understanding of therapeutic outcomes.

In the context of AI, where the richness of the training dataset directly influences the performance of the models, the limitations imposed by rigid data schemas can be particularly crippling. ML models trained on incomplete or inadequately structured datasets may yield suboptimal or biased predictions, undermining the validity and utility of the insights they generate.

### Late Binding of Schema: A Methodological Innovation

A different methodology called “late binding of schema” involves capturing data in a way that leaves room for changing the data’s presentation and restructuring it flexibly when needed. This is a more flexible and future-proof approach that acknowledges the unpredictable nature of scientific discovery and the diverse needs of data analytics, including AI. By allowing late binding of schema — where the data structure can be defined and modified closer to the time of analysis — data can be restructured as new hypotheses arise.



This methodology not only preserves the full informational content of the data – the metadata – but also provides the flexibility needed to adapt to the evolving landscape of scientific research. It facilitates the integration of varied data types and enables data to be repurposed for unforeseen analytical needs. In doing so, this methodology ensures that data remains a living entity within the research ecosystem, capable of continual growth and re-examination in the quest for scientific breakthroughs.

---

*Late binding of schema ensures data remains a living entity, capable of continual re-examination in the quest for scientific breakthroughs.*

---





## Pioneering AI-Ready Data Management with Signals Research Suite

Revvity Signals is committed to developing cloud-native software solutions that are responsive to the current needs of scientists, including and anticipating the future landscapes of discovery shaped by AI.

The company's Signals Research Suite, a complete SaaS solution to increase scientific productivity and collaboration, embodies this understanding by providing an ecosystem that is engineered to optimize data for AI workloads. The suite's prowess in managing and enriching data extends beyond mere storage and retrieval; it is a comprehensive platform that facilitates the capture, processing, modeling, and analysis of scientific data.

Revvity Signals has crafted a system where the intricate details of experiments—often the kind that hold the keys to innovative breakthroughs—are captured with precision and clarity within the scientist's workflow. This attention to detail ensures that every data point can be leveraged to its fullest potential when subjected to AI/ML algorithms.

In the era of large data sets, where the volume of information can be overwhelming, Revvity Signals offers a solution that transforms this deluge into a structured, contextualized, navigable, and analyzable resource. Signals Research Suite is equipped with tools that enable researchers to annotate data with rich metadata, employ flexible data modeling, and perform complex data analytics—all within a single, integrated environment.

Moreover, Revvity Signals understands that the scientific process is a collaborative endeavor. Therefore, Signals Research Suite is designed with collaboration in mind, allowing for seamless, secure data sharing and integration across teams and disciplines. This collaborative feature ensures that data is not siloed but can be enriched and expanded upon by the collective expertise of the scientific community.

The company's forward-thinking approach is reflected in its adoption of the late binding of schema. As a result, it is simple and easy to reorganize data, correct errors in data modeling, or address an unforeseen area of scientific research, e.g., by adjusting the schema, so that users meet their goal of having the right data organized correctly. Flexibility is essential because metadata schema relationships may change or evolve, and having a system that is too rigid will cause data analysis and metadata capture issues, leading to poor usage and data capture while slowing overall R&D progress.

Researchers save time with Signals Research Suite because the software includes the ability to correct mistakes without starting over from the beginning of the process. By allowing the data structure to be defined and refined as close as possible to the time of analysis, Revvity Signals ensures that the data remains agile and adaptable—qualities that are essential for feeding the ever-evolving AI techniques that drive modern scientific discovery.



## How Signals Research Suite Sets the Stage for AI Success

Signals Research Suite addresses the end-to-end R&D workflow needs of research scientists. As a unified SaaS software solution, it enables life science, pharmaceutical, and biotechnology users to capture context effortlessly, early, and often. For instance, Signals Notebook, the electronic laboratory notebook within Signals Research Suite, is configured with customizable and requirable experiment properties to incentivize scientists to capture metadata as early as possible — when it's top of mind — and attach it to the data.

With Signals Research Suite, organizations get robust tools to manage the data schema with which their data is captured and maintained. The schema ensures the data are well structured and annotated for efficient downstream analysis and interrogation — for example, enabling the software to quickly deposit a subset of data with full context into the algorithms used for training ML models. The ability to segment data is essential because training AI with general data that does not include the proper context and annotation dilutes the quality of results. Signals Research Suite enables algorithms to be trained on the organization's specific data so results are precisely focused.

Revvity Signals equips researchers with the right tools at the right part of the workflow. Intuitive and straightforward, Signals Research Suite reduces barriers and makes it easier for scientists to adopt the latest AI/ML technology, leveraging it to gain analytical insights while freeing them to focus on moving their science forward.

*The ability to segment data is essential because training AI with general data dilutes the results.*

### Remaining challenges with AI

While AI is poised to transform the industry by reducing tedious tasks and uncovering patterns that lead to insight, there are some remaining challenges.

- Large language models such as generative AI tools are susceptible to hallucinations—incorrect or nonexistent patterns that create inaccurate or useless outputs.
- Scientists need education regarding AI strategy and how to use the new generation of tools.
- Human effort is still necessary, such as inputting metadata to provide context or opening a new browser to use an AI tool.

Much of the recent focus has been on generative AI (large language models), but the benefit for organizations requires pivoting to applying those technologies to specific datasets instead of using generalized data that will not produce the desired results.



[revvitysignals.com](https://revvitysignals.com)  
940 Winter Street  
Waltham, MA 02451 USA  
P: (800) 762-4000 (+1) 203-925-4602

 Revvity Signals

 RevvitySignalsSoftware

 revvitysignals

 Revvity\_Signals

 RevvitySignals