

CAS BIOFINDER®

Does your drug
discovery budget
fund novel ideas or
data crunching?



A division of the
American Chemical Society

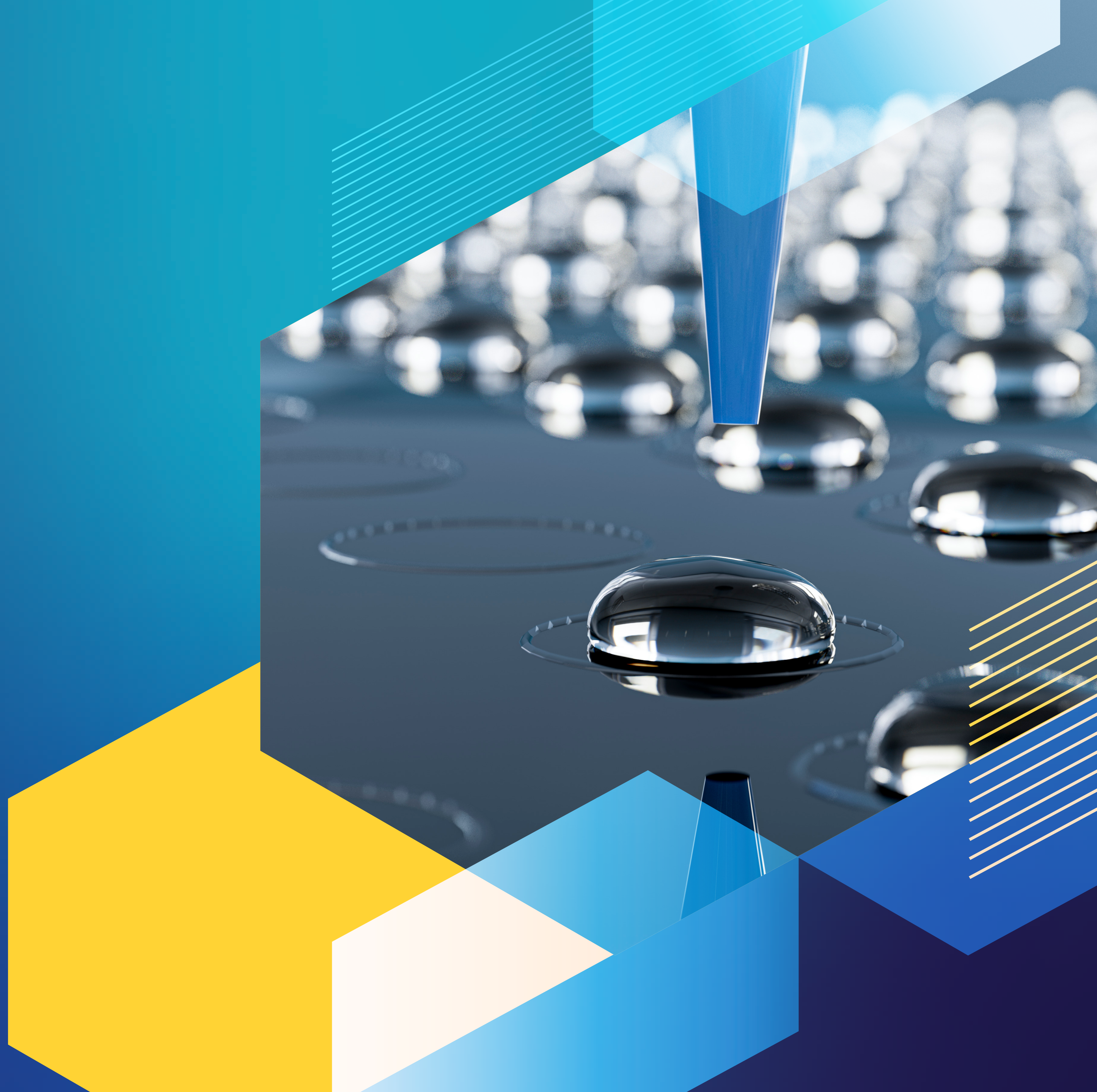


Table of Contents

3

Productivity in drug discovery is declining

4

The R&D time audit: Where do the hours go?

5

Three data challenges limiting
researchers' productivity

8

Drug discovery is drowning in tools that
don't work together

9

Researchers should not be the
integration layer

11

References

Productivity in drug discovery is declining

Global pharmaceutical R&D spending has grown consistently for decades, reaching nearly [\\$288 billion in 2024](#)¹. Yet, the number of new drugs approved per [billion dollars spent](#)² continues to fall. [A 2026 Deloitte report](#)³ puts the 2025 pharmaceutical R&D return at 7.0%, making it the third consecutive year of headline improvement, up from 5.9% in 2024. However, that recovery is almost entirely driven by [GLP-1 obesity and diabetes drugs](#), which now account for 38% of projected sales across the top 20 biopharma companies' late-stage pipelines⁴. Strip them out and projected returns collapse to just 2.9%, painting a bleaker picture for the industry. Not only this, but the average cost to advance a drug from discovery to launch [climbed to \\$2.67 billion in 2025](#), up from \$2.23 billion the year before.

This is Eroom's Law in action: the industry is spending more than ever and getting less in return. Coined by researcher [Jack W. Scannell in 2012](#)⁵, Eroom's law describes the decline in productivity seen in biopharmaceutical R&D despite widespread technological advances. High-throughput screening, computational drug design, next-generation sequencing, and now artificial intelligence (AI) have promised to accelerate discovery and reduce costs, but none have succeeded in changing the trend.

How do drug discovery researchers spend their time? Where do the core inefficiencies lie, and how can organizations implement data curation and harmonized tools to free researcher time for innovation?

The R&D time audit: Where do the hours go?

Scientist time is a precious and expensive resource. Including base salary, benefits, overhead, and laboratory infrastructure, a full-time researcher can cost hundreds of thousands of dollars per year, and the largest biopharma companies employ tens of thousands of them.

There’s no question that researchers work hard to return this investment, with over a third regularly working more than 60 hours a week and more than half working on weekends (Figure 1). However, the distribution of researcher time shows that only 35% is spent on experimental work^{6,7} while most scientific labor is now consumed by activities surrounding R&D rather than R&D itself (Figure 2).

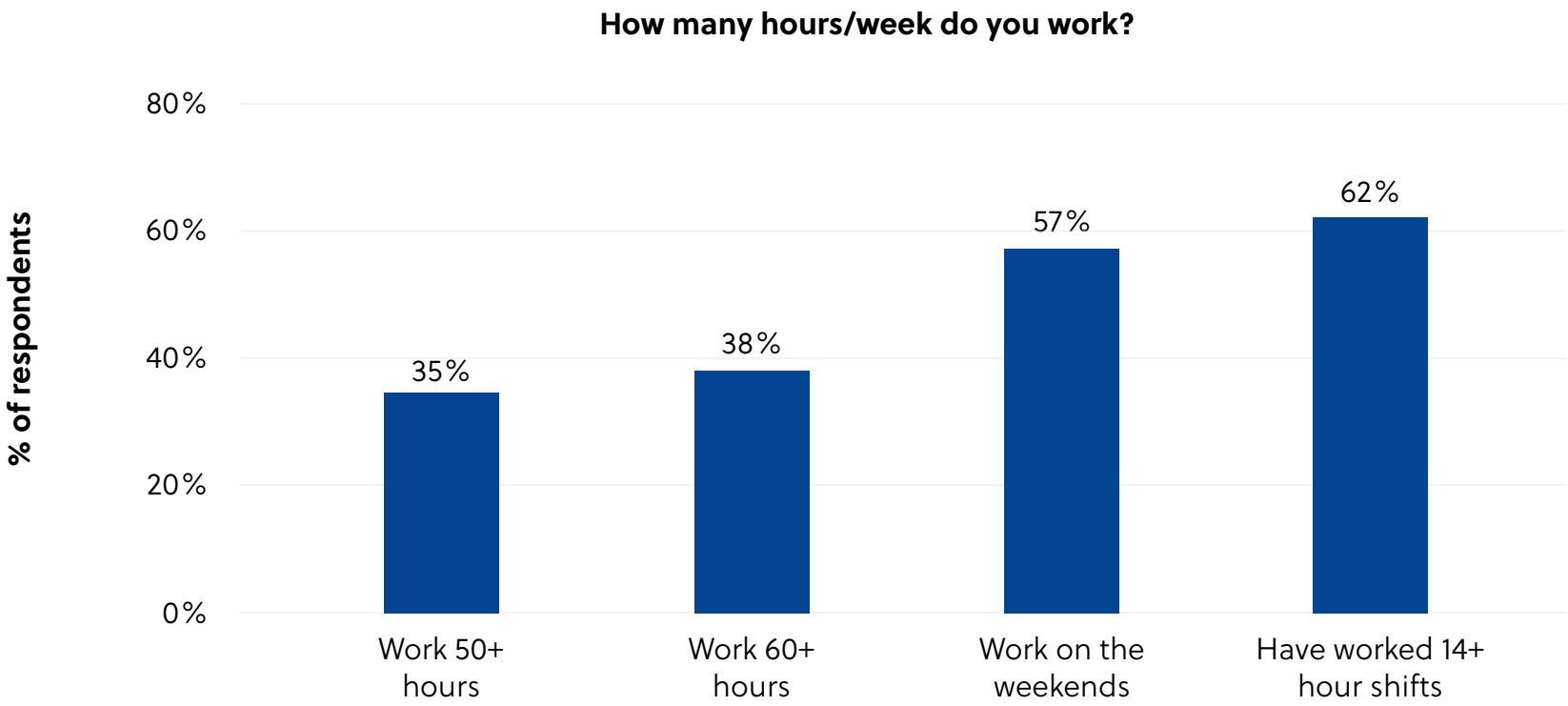
In a research organization of 1,000 scientists, that ratio absorbs the equivalent of 650 full-time scientists into work that does not directly advance experiments. Scaled to the [\\$288 billion of global pharmaceutical R&D spending recorded in 2024](#)¹, the financial weight of this misallocation is tens of billions of dollars each year.

Scientists spend roughly 40% of their time on administrative tasks and reporting⁸, while broader research on knowledge workers finds a similar proportion of time goes to work that doesn’t require specialized expertise and could be delegated to others⁹.

“The hours researchers spend on manual data work are not fixed. Organizations that reduce the reconciliation burden at the data foundation level consistently recover significant scientific capacity, time that goes back to the experiments that matter.”

Dr. Adam Sanford, Sr. Director of Product, Life Sciences

Figure 1. Working hours among life sciences researchers.



Data source:
Powell, Kendall (2016). Hard work, little reward: *Nature* readers reveal working hours and research challenges. *Nature*, <https://doi.org/10.1038/nature.2016.20933>
Stemcell Technologies (Accessed July 06, 2026), The demands of science. Stemcell.com. <https://www.stemcell.com/efficient-research/demands-of-science-report>

While interpreting data requires highly trained researchers and scientific judgment, managing information from literature should not demand the same investment of expert time. The result is a paradox at the heart of modern drug discovery: **organizations recruit some of the most highly trained scientists on the planet, yet their time is absorbed by information management rather than hands-on scientific discovery.**

AI tools can accelerate information extraction, but they often redistribute rather than reduce the burden on researchers. Instead of mining data, researchers spend time standardizing datasets and double-checking AI-driven outputs before they can act on them. Even in the AI era, scientific judgment and data management remain essential: a model that groups structurally similar compounds without accounting for subtle differences in tautomeric forms, solubility, or conformation can introduce errors that require expert correction downstream. Without structured and reliable information, the data burden does not disappear. It shifts from manual reconciliation to extensive dataset pre-processing and output verification, remaining a constant drain on the researcher’s time that should be spent advancing programs.

Three data challenges limiting researchers' productivity

The largest share of researchers' misallocated time is driven by the data itself. Drug discovery is a data-intensive science, spanning chemical structures and synthesis pathways to biological targets, pharmacological responses, and toxicity profiles (Table 1).



Table 1. Key data categories in drug discovery research.

Data category	Subcategory	Description
Chemical data	Compound data	Information about the molecular structures of compounds
	Synthesis data	Details on the synthesis and purification of chemical compounds
	Property data	Physical and chemical properties of compounds, such as solubility, stability, and molecular weight
Biological data	High-throughput screening (HTS) data	Results from screening large libraries of compounds for activity against biological targets
	Biomarker data	Information on biological molecules that indicate pharmacological responses, pathogenic processes, or pharmacokinetic processes
Pharmacological data	Pharmacodynamics	Data on the dose-dependent effects of a drug on its target
	Pharmacokinetics	Data related to drug absorption, distribution, metabolism, and excretion
	Toxicology data	Information on the potential adverse effects of drug candidates
Computational data	Molecular modeling and simulations	Data from computational models predicting how drug molecules interact with biological targets
	AI and machine learning (ML) models	Algorithms and predictive models used to analyze complex datasets and predict drug efficacy and safety

In theory, this abundance of information should accelerate discovery, but in practice, it amplified three scientific data challenges that slow R&D workflows. Understanding where time is lost is the first step toward reclaiming it.

The volume of data is growing faster than researchers can use it

Every major technological advance in drug discovery has increased the amount of data researchers are expected to evaluate before making decisions. Advanced sequencing, screening campaigns, and machine-learning models now routinely produce terabytes to petabytes of data per project, and external literature is growing at a comparable pace. For the [CAS Content Collection™](#) alone, scientists aided by technology process more than 100,000 new publications and patents per day, far exceeding any individual researcher’s capacity to read.

Each new wave of technology has expanded the volume of evidence a researcher is expected to evaluate before making any decision without expanding the working day. As the volume of evidence expands faster than researchers’ capacity to evaluate it, scientific decision-making slows—or worse, pushes researchers into relying on general AI evaluation tools and the likelihood of missing high-value opportunities increases.

“ The exponential growth in experimental data actually makes the discovery process more challenging as there’s more to sift through. It’s not a math problem or technology problem; it’s a content problem.”

Tim Wahlberg, Interim President, CAS

Drug discovery data is inconsistently named and reported

Scientific entities are rarely described in the same way across internal business units and external databases:

- A single biological target may appear under multiple names, identifiers, and formats depending on the source material.
- Chemical compounds may be described using inconsistent structures, naming conventions, or incomplete contextual information.

In addition to different naming conventions, experimental findings are frequently published without standardized metadata, making results difficult to compare, reproduce, or interpret with confidence. [Research suggests that between 80 and 90%](#) of peer-reviewed published biomedical literature cannot be reproduced due to a lack of metadata¹⁰.

These inconsistencies span every stage of the workflow and data type on which researchers rely. Across the chemical, biological, pharmacological, and computational data needed for drug discovery, researchers must continuously reconcile conflicting formats, identifiers, and contextual gaps before they can act on the information with confidence (Table 2).

Table 2. Common sources of variability across drug discovery-relevant data sources.

Data type	Examples	Common inconsistency
Chemical	Compound structures, synthesis pathways, and property data.	Naming conventions, missing structural context.
Biological	HTS results, biomarker data.	Inconsistent identifiers.
Pharmacological	Pharmacodynamics, pharmacokinetics, and toxicology.	Non-standardized metadata, missing experimental detail.
Computational data	Molecular modeling outputs, AI/ML predictions.	Variable formats, lack of traceability to source data.

For example, Janus Kinase 3 (JAK3) is listed across scientific resources as Janus Kinase 3, JAK3, Tyrosine-protein kinase JAK3, HGNC:6193, P52333, LJAK, and JAKL.

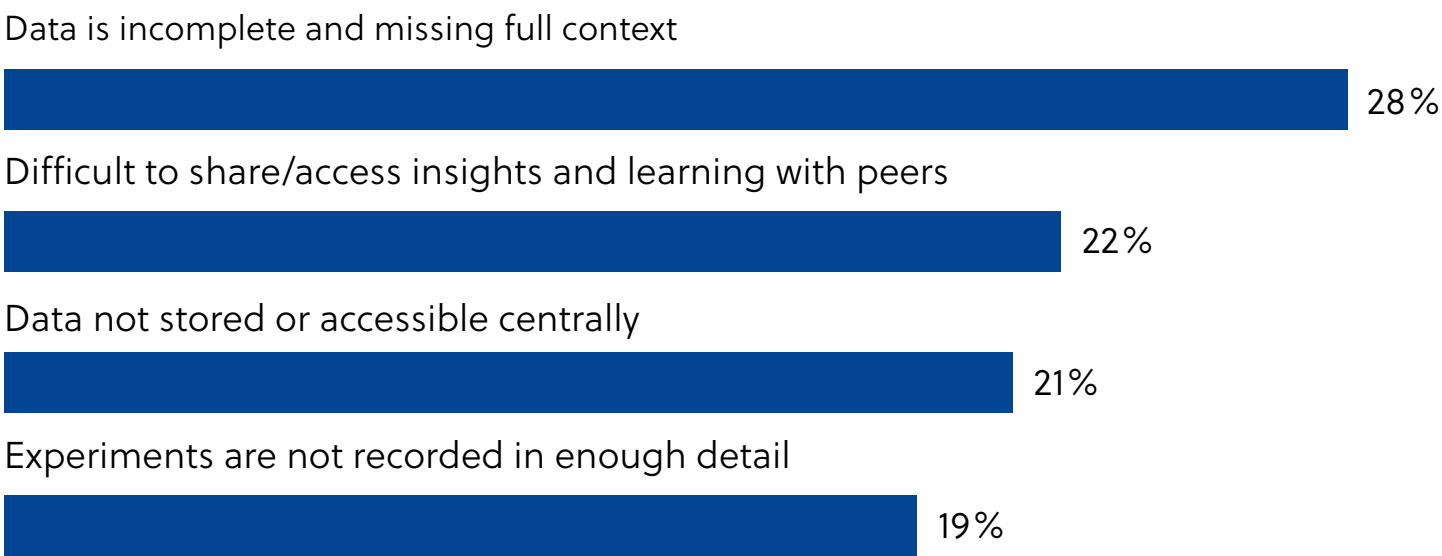


The data is fragmented across disconnected sources

The Nucleic Acids Research journal’s online Molecular Biology Database Collection currently catalogs 2,173 active biological databases, which covers only public, curated resources. Layered on top is each organization’s internal systems: electronic lab notebooks, laboratory information management systems, instrument-specific software, project trackers, and the spreadsheets that fill the gaps between them.

Researchers are conscious of the impact. [Synthace’s 2024 life science R&D trends analysis](#)¹¹ shows that 28% of researchers report that the data they rely on is incomplete, 22% say they struggle to access and share insights with peers, 19% cite missing experimental detail, and 21% identify a lack of centralized access as a persistent barrier to progress (Figure 3).

Figure 3. Top reasons why researchers struggle to build on prior experiments.



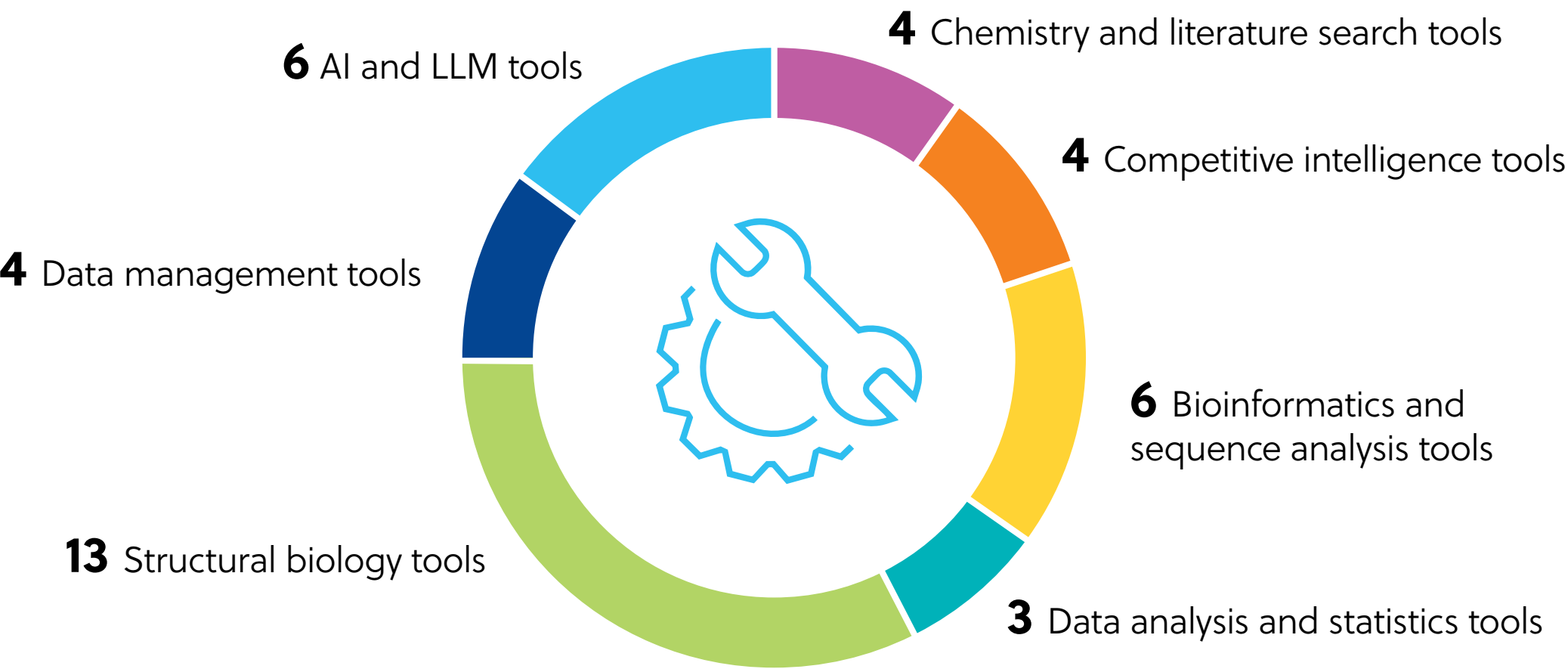
Data source: Synthace (2024). Lab automation & experimentation trends in life science R&D. Synthace.com, 2024, <https://www.synthace.com/life-sciences-trends-automation-experimentation-report>

The pharmaceutical industry has responded to these challenges with significant investment in specialized tools, reflecting a genuine commitment to improving research productivity. The question for R&D leaders is whether that investment is delivering the returns the industry needs, or whether a different approach to the underlying data foundation could unlock more.

Drug discovery is drowning in tools that don't work together

The fragmentation researchers face at the data level is mirrored at the tool level. In a 2023 survey, researchers reported using an average of 22 different tools to source, manage, and visualize discovery data, spanning data sourcing, data management, chemistry and bioscience, and AI/ML applications (Figure 4). In 2026, with newer technologies to mine even more data, this number has nearly doubled.

Figure 4. Tool fragmentation across the drug discovery workflow.



Data source: The Understanding Group. (2025 – 2026). CAS drug discovery research findings.

Each system holds a piece of the full picture, with one database for chemical structures, another for biological targets, and another for pharmacological data. Moving between them requires researchers to manually translate formats, reconcile identifiers, and reconstruct context at every step. This forces them to spend more time navigating systems than generating insights. The cost compounds with every additional layer. Every hour spent switching back and forth between tools to cross-reference information is an hour not spent on discovery.

Beyond the time cost, there is a cognitive one. Researchers working across disconnected systems must hold partial pictures in mind, tracking which database uses which naming convention, mentally bridging gaps between pharmacological and biological datasets, and making judgment calls on data quality before any scientific analysis can begin. This kind of mental overhead disrupts focus, increases the likelihood of errors, and makes it harder to see the connections between datasets that might point toward a promising lead. The insights most likely to be missed are often the ones that require joining information across domains that were never designed to talk to each other.

AI-powered tools are often positioned as the solution to data fragmentation, but they actually amplify the strengths and weaknesses of the underlying data foundation. AI can surface literature, extract relationships, and generate hypotheses faster than ever before, but it cannot eliminate the need for researchers to verify each data point's accuracy and reliability.

A tool built on inconsistent identifiers, missing metadata, and disconnected sources inherits all those problems by default. For high-stakes drug discovery pipelines, faster outputs do not eliminate the need for validation, reconciliation, and scientific judgment.

“What is becoming increasingly clear is that the impact of AI in drug discovery depends fundamentally on the quality and structure of the underlying data.”

Dr. Ulrich Stolz, Flagship Pioneering, former Corporate Vice President, Novo Nordisk

The more fragmented the data foundation becomes, the more researchers are forced to compensate for it. However, fragmentation is an architectural problem, and architectural problems have architectural solutions.

Researchers should not be the integration layer

Organizations investing billions in drug discovery should not depend on researchers to manually assemble and validate fragmented information before advancing programs.

Imagine opening a single system and finding everything needed for target assessment: compound structures linked to biological activity data, pharmacological profiles traceable to source methods, and computational predictions grounded in high-quality experimental evidence. No reconciling identifiers. No switching between disconnected systems. No uncertainty about whether the underlying information is complete enough to trust. This is what becomes possible when data inefficiencies are addressed at the foundation. Researchers spend less time validating information and more time advancing experiments, while connections across chemical, biological, and pharmacological domains become visible without manual integration.

The limitation of current drug discovery-related tools is architectural. Most scientific data platforms were designed to solve isolated workflow problems rather than support the full discovery process.

What research organizations need is a connected scientific data foundation that integrates information across domains, removes reconciliation work from the researcher, and enables teams to move from question to experiment with greater speed and confidence.

For researchers, that means less time spent on data management and more time spent advancing discovery. For R&D leaders, it means more of their investment goes toward R&D itself.

Reclaim discovery time with CAS BioFinder®

CAS BioFinder centralizes the biological, chemical, and pharmacological data that drug discovery teams need into a single connected environment. Powered by the CAS Content Collection, the world's largest human-curated scientific information resource, the platform enables researchers to work from harmonized, contextualized, and traceable data rather than manually reconciling fragmented information across sources. Every day, CAS scientists normalize, review, and confirm measurements, assign authoritative identifiers, and connect every data point back to the underlying methods, assays, and source references so researchers can act on information without second-guessing it.

The result is a workflow in which drug discovery teams can move more efficiently from question to analysis and from analysis to experiment. They can validate targets faster using linked assay data, structural insights, and pathway context. They can compare modalities early and identify safety and feasibility risks before they become costly late-stage development setbacks. They can also narrow candidate lists quickly, reducing the time and resources spent on failed experiments. All of this can be accomplished without wasting time manually piecing together information from disconnected databases.

“By narrowing the candidate list from 66 compounds to four likely candidates in just a few minutes, CAS BioFinder saved at least \$15,000 on external testing for this single project and accelerated the overall project timeline to eight months from initial project design to biological validation. Similar projects often take years without CAS BioFinder.”

Dr. Philipp Natho, Postdoctoral Researcher in Synthetic Organic Chemistry
Università degli Studi di Bari Aldo Moro

The operational impact of a centralized drug discovery platform extends beyond faster access to information. When researchers spend less time locating, validating, and reconciling information across systems, organizations recover valuable researcher time that can be redirected toward experimentation, target validation, and program advancement.

Reclaiming researcher time at scale

CAS BioFinder can help organizations reclaim ~180 hours per discovery researcher per year (around \$25,000) by making data collection and analysis 30% more efficient.

This represents an estimate of:

- \$0.6 million per year for small biotech (with an average of 25 researchers).
- \$5.0 million per year for medium pharma (with an average of 200 researchers).
- \$17.6 million per year for large pharma (with an average of 700 researchers).

Drug discovery has never been more complex or more consequential. Yet across the industry, a significant portion of researcher time and R&D investment continues to flow into data management rather than scientific progress. Organizations that replace fragmented, inconsistent data foundations with a single, curated, and connected environment change that equation. Their R&D teams spend less time questioning information and more time generating evidence that drives pipeline decisions, reduces late-stage risk, and brings viable candidates to patients faster.

See how CAS BioFinder helps drug discovery organizations reclaim R&D time and accelerate pipeline decisions at cas.org/solutions/biofinder.

References

1. BioSpace. Undeterred by Political, Economic Headwinds, Pharma Ups R&D Investment in 2024 and Beyond. BioSpace. May 28, 2025. <https://www.biospace.com/business/undeterred-by-political-economic-headwinds-pharma-ups-r-d-investment-in-2024-and-beyond>
2. Buntz B. Global pharma R&D hits \$276B, triples marketing spend. Research & Development World. 2024. <https://www.rdworldonline.com/how-much-does-the-pharma-industry-spend-on-rd-anyway-probably-more-than-you-thought/>
3. Navigating the GLP-1 boom | Deloitte UK. <https://www.deloitte.com/uk/en/Industries/life-sciences-health-care/research/measuring-return-from-pharmaceutical-innovation.html>
4. Weight loss drugs boom drives pharma R&D returns, but industry could face 'bubble' effect | Deloitte UK. <https://www.deloitte.com/uk/en/about/press-room/weight-loss-drugs-boom-drives-pharma-r-and-d-returns.html>
5. Scannell JW, Blanckley A, Boldon H, Warrington B. Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov. 2012;11(3):191-200. doi:10.1038/nrd3681
6. Effron J. The Changing World of Life Sciences R&D. 2022. <https://vitalsignshealth.substack.com/p/the-changing-world-of-life-sciences>
7. Guzman D. Research Administration: Relieving the Pressure on Researchers. 2020. <https://exlibrisgroup.com/blog/research-administration-relieving-the-pressure-on-researchers/>
8. Schneider, J., et al. (2014). 2012 FDP Faculty Workload Survey Research Report. Federal Demonstration Partnership / National Academies. https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_087667.pdf
9. Bruch, H. & Sander, M. (2014). "Make Time for Work That Matters." Harvard Business Review. <https://hbr.org/2014/01/make-time-for-work-that-matters>
10. Kelm JM, Ferrer M, Bittner MI, Lal-Nag M. Data standards in drug discovery: A long way to go. Drug Discov Today. 2024;29(2):103879. doi:10.1016/j.drudis.2024.103879
11. Synthace. Report: Trends in Life Sciences R&D. 2023. <https://www.synthace.com/life-sciences-trends-automation-experimentation-report>

CAS connects the world's scientific knowledge to accelerate breakthroughs that improve lives. We empower global innovators to efficiently navigate today's complex data landscape and make confident decisions in each phase of the innovation journey. As a specialist in scientific knowledge management, our team builds the largest authoritative collection of human-curated scientific data in the world and provides essential information solutions, services, and expertise. Scientists, patent professionals, and business leaders across industries rely on CAS to help them uncover opportunities, mitigate risks, and unlock shared knowledge so they can get from inspiration to innovation faster. CAS is a division of the American Chemical Society.

Connect with us at cas.org