

Beyond data silos: Five knowledge management tips to future-proof scientific R&D workflows



Table of Contents

3

Why do R&D workflows
break down?

4

Where does your organization stand?
Five questions to self-assess

5

The four pillars of effective
knowledge management

6

Five actionable knowledge management
tips to future-proof R&D

10

Smarter implementation for faster impact
with knowledge management experts



Why do R&D workflows break down?

R&D innovation depends on trusted data to guide project prioritization, support confident decision-making, and accelerate workflows. Yet scientific information is notoriously difficult to manage. Chemical data, formulation protocols, and experimental results spread across labs and business units make institutional knowledge hard to unify—and even harder to leverage at scale. Such data silos often skew insights and decisions, slowing the pipelines the information is meant to power.

This fragmentation doesn't happen overnight; it builds gradually across teams, tools, and time until momentum stalls. For R&D leaders, identifying the root causes of these knowledge management inefficiencies is the first step toward unlocking more agile, insight-driven workflows. What is slowing teams down?



Fragmented knowledge management systems

Points of friction slowing teams down

- Disconnected or legacy-format internal data (lab journals, siloed databases, etc.)
- Outdated lab software (LIMS, CDMS...)
- Disparate and disconnected data spanning labs and business units

Impact on R&D workflows

- Delays the integration of digital tools critical for data-driven R&D
- Creates barriers between discovery, strategic, and regulatory teams, hindering collaboration
- Slows the identification and validation of viable leads and innovations, impeding products' time to market

Disconnected systems don't just create complexity, they block your ability to harness organizational knowledge and integrate modern data tools to drive faster, smarter decisions.



Errant, low-quality data and poor analytics

Points of friction slowing teams down

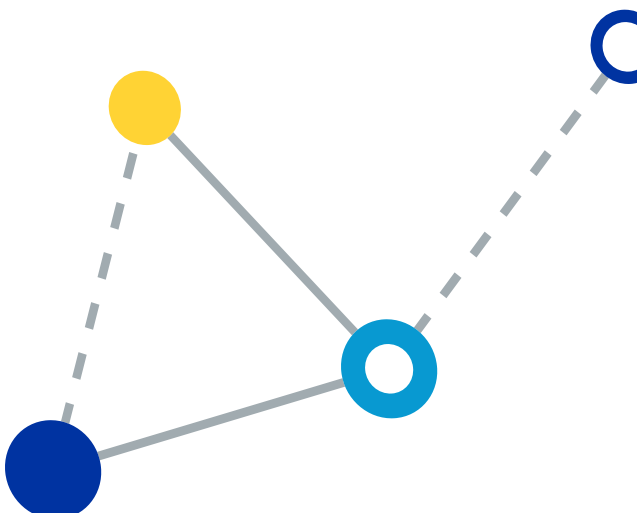
- Unharmonized format, terminologies, and abbreviations
- Inconsistent nomenclatures and chemical structures
- Disparate data collection systems

Impact on R&D workflows

- Leads to low-quality datasets, limiting AI models' power and prediction accuracy
- Increases the risk of redundant experiments, delayed timelines, and missed opportunities
- Skews research prioritization and investment decisions, threatening budgets and pipeline efficiency

Scientific data is diverse and complex, and its rapidly evolving nature presents unique challenges for organizations in consolidating critical information. When data lacks structure and quality, even the best tools and models fail to deliver. R&D teams are left second-guessing outcomes and decisions instead of driving confident innovation.

Fragmented systems, inconsistent data standards, and underpowered analytics quietly drain resources and introduce risk at every stage of the innovation pipeline. How can organizations fix knowledge management gaps? It starts with a clear-eyed assessment of their current state.



Where does your organization stand?

Five questions to self-assess

To optimize R&D workflows, companies must first understand where friction exists within their internal data ecosystems. These five questions will help you diagnose your current knowledge management strategies and identify key barriers to your innovation.

Your self-assessment form	Yes	No
Can teams easily search and retrieve legacy data, including past experiment results and records?	☐	☐
Do your data systems (ELNs, LIMS, registries, knowledge bases) interoperate to enable seamless data flow across teams?	☐	☐
Are your data collection processes standardized to ensure consistency and reusability across teams and AI models?	☐	☐
Is your data structured and harmonized to support analytics and AI initiatives without heavy manual clean-up?	☐	☐
Do you have clear governance protocols for data handling, validation, and ongoing management to maintain integrity and AI readiness?	☐	☐

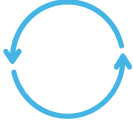
If you answer 'No' to any of these questions, it's a sign that your data may be working against you, inevitably slowing R&D, inflating costs, and weakening decision-making. How can you turn the situation around?

The challenge is that unlocking data's full potential requires more than patching individual gaps; it demands a robust foundation and solutions that connect information, processes, and teams. How can organizations move beyond patchwork fixes and build future-proof knowledge management strategies?



The four pillars of effective knowledge management

As data volumes grow and R&D becomes more digital, building effective knowledge management strategies requires a solid foundation. Successful organizations focus on four critical pillars that improve collaboration, reduce friction, and future-proof their R&D workflows.

Pillar	Why it matters
 Quality	Eliminates extensive pre-analysis cleaning, reduces manual reformatting, and supports seamless integration across systems, modeling tools, and workflows
 Searchability	Improves data discoverability, empowering teams to access and extract key information faster, avoid repeat experiments, and make more efficient use of R&D resources
 Accessibility	Makes data readily available to all contributors in any team or region, supporting smooth collaboration and uninterrupted workflows
 Reusability	Enables teams to build on previous work, streamline analytical workflows, and extend the value of past research and investments to maximize ROI

While understanding the foundational pillars of knowledge management is essential, their true value lies in smart implementation across existing systems and workflows. What practical steps can leaders take today to lay the groundwork for more effective R&D?



Five actionable knowledge management tips to future-proof R&D

1. Digitize and harmonize legacy records

For many organizations, valuable research data remains buried in archives, outdated systems, or inconsistent file formats. Without a structured digitization and data connection strategy to surface these insights, research teams risk duplicating work, overlooking key findings, and delaying high-impact decisions.

Digitizing and harmonizing legacy records turns static archives into searchable, structured assets to enhance visibility and increase the value of past research and investments.

How legacy record digitization helps scientific R&D workflows:

- Preserve critical institutional knowledge that might otherwise be lost or unusable
- Reduce rework and maximize ROI by enabling data reuse and avoiding duplicate efforts
- Simplify the use of modeling tools by converting legacy data into AI-ready formats
- Accelerate decision-making with quick access to historical research and experimental insights

From hours to minutes: How digitization transforms R&D intelligence

A large health-tech organization partnered with CAS Custom ServicesSM to convert decades of unstructured legacy data into searchable assets. With a custom knowledge system, scientists across the organization now uncover insights in record time to accelerate high-impact R&D workflows.¹

"One search took a senior scientist ~8 hours; now it takes minutes with scientifically aware search from CAS."

Company estimate

2. Consolidate and connect existing data sources and knowledge

ELNs, LIMS, and other platforms often operate in isolation across teams and departments. When critical insights only live in personal inboxes or local drives, building on existing knowledge quickly turns into time-consuming backtracking, which slows down decisions and workflows.

Consolidating and connecting scattered systems into a unified knowledge management structure offers teams a clear, integrated view of their data resources. It enables them to search, locate, and leverage organizational intelligence to generate better insights and accelerate operations. When data is centralized and aligned, teams spend less time chasing context and more time advancing innovation.

How unified knowledge management platforms help scientific R&D workflows:

- Break down silos across departments and streamline knowledge sharing
- Accelerate cross-functional collaboration and reduce delays at key handoffs
- Eliminate inconsistencies across systems by aligning formats, terminologies, and metadata
- Provide a unified, trustworthy data source for R&D endeavors and portfolio-level decision-making

3. Unify data collection systems and processes

Companies must verify the integrity of new data entries to ensure quality insights and support sustainable knowledge management. While most organizations already have one or more digital acquisition strategies, inconsistencies in protocols and systems can quickly impede information sharing and hinder seamless integration into multiple R&D workflows.

Establishing a unified data collection system with clear standards, input validation, and governance protocols helps ensure information consistency and integrity from the point of entry. This foundation reduces downstream friction, preventing mismatches between systems while improving the readiness of new data for integration, reuse, and decision support.

How unified data collection systems help scientific R&D workflows:

- Ensure accurate new data input at the source, minimizing variability and human errors
- Capture data consistently across teams, tools, and experiments, improving integrity and interoperability
- Reduce cleaning and validation effort for streamlined analytics and regulatory reporting
- Strengthen trust in data across projects and departments, accelerating decision-making

From inconsistent inputs to trusted insights that boost R&D efficiency

A large diversified chemical company partnered with CAS Custom Services to clean and harmonize its substance data across disconnected systems. By establishing a single source of truth, the organization reduced manual validation, eliminated duplication, and unlocked 3,300 hours in annual R&D efficiency gains.²



4. Curate high-quality datasets and train AI models

Even the most advanced AI and analytics tools are only as good as the data they rely on. Inconsistent, incomplete, or poorly structured datasets can undermine model accuracy, skew insights, and erode trust in outcomes. For organizations looking to use AI and modeling tools to drive faster insights and more scalable R&D workflows, curating data is foundational.

Curated data reduces noise, corrects inconsistencies, and ensures that AI models are trained on relevant, high-quality inputs. By curating datasets with purpose and scientific context, teams enable their models to generate accurate, reliable outputs that reflect real-world R&D complexity.

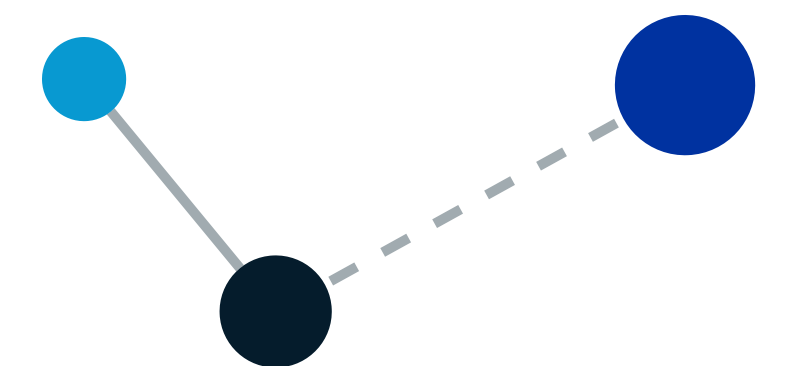
How custom-curated datasets help scientific R&D workflows:

- Increase analytics relevance by reducing the risk of false positives and selective bias
- Boost recommendation precision by capturing subtle patterns specific to areas of interest
- Enhance algorithm transferability and minimize the need for extensive model retraining
- Increase computational efficiency and optimize resource utilization

As models are tested and prove their predictive power, retrieval augmented generation and other ongoing data integration and training are critical to set the stage for longer-term performance and efficacy as new findings and literature are disclosed and the underlying science becomes more complex.

Powering AI success with custom-curated training data

A diversified chemical company partnered with CAS Custom Services to develop high-quality, domain-specific training datasets to improve its model accuracy. With expertly curated content, the organization not only built a scalable foundation for future AI innovation but also reduced experimental trial-and-error, cutting R&D timelines in half.³



5. Train internal users and standardize practices

Inconsistent user behavior can undermine data quality and limit the impact of knowledge management strategies. In organizations where teams span functions and disciplines, aligning on shared data standards is essential for ensuring data can be trusted, scaled, and reused.

Training internal users in data handling practices supports a unified, organization-wide approach to knowledge management. By collaborating with data experts, leaders ensure training is tailored and aligned with workflow needs. With shared standards in place, teams spend less time fixing inconsistencies and more time generating insights.

How knowledge management training helps scientific R&D workflows:

- Promotes data entry consistency by ensuring all contributors follow the same protocols and standards
- Increases trust in data outputs, reducing extensive cleaning and verification
- Minimizes human errors and ensures long-term data maintenance

From scattered habits to scalable systems: How training empowered Toray's R&D teams

Toray Industries partnered with CAS Custom Services to upskill its researchers in data curation and knowledge management best practices. Through customized training and hands-on consultation, Toray reduced inefficiencies, improved data quality, and built a strong, future-proof data foundation for years to come.⁴

Smarter implementation for faster impact with knowledge management experts

Building scalable, insight-driven R&D workflows starts with strong knowledge management. From digitizing and harmonizing legacy records to aligning on data collection and handling best practices, every step drives companies' ability to innovate smarter, collaborate better, and move faster. However, optimizing knowledge management requires deep expertise to find the right strategy that connects the right data to the right people.

CAS Custom Services builds unique data solutions that empower R&D teams with the information they need to make faster progress. With unmatched knowledge management expertise, advanced data technologies, and a century-long legacy of scientific excellence, CAS helps organizations turn information into fuel for impactful scientific R&D workflows.

To find out how we can help you future-proof scientific R&D workflows, contact us at cas.org/contact.

Resources

1. Custom knowledge management system reduces search time from hours to minutes | CAS.
<https://www.cas.org/resources/gated-content/cas-unlocks-potential-dark-data>
2. Unified data systems save a company thousands of hours and dollars | CAS.
<https://www.cas.org/resources/gated-content/cas-data-integrity>
3. How curated training Sets from CAS improved the prediction accuracy and transferability of an AI model | CAS.
<https://www.cas.org/resources/gated-content/case-studies/ai-predictions>
4. Tailored data strategies and training programs empower Toray's scientists | CAS.
<https://www.cas.org/resources/gated-content/toray-knowledge-management>

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