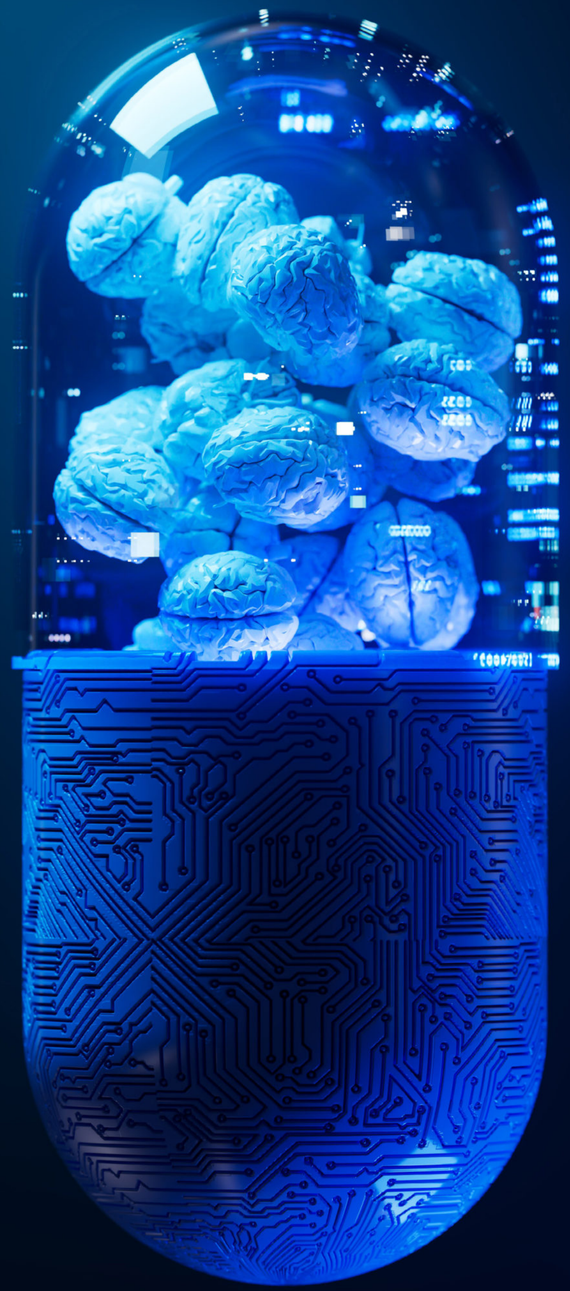


CAS Custom Services℠

From strategic data initiatives to therapeutic breakthroughs

CAS
A division of the
American Chemical Society

Table of Contents

3
Leveraging knowledge management to tackle drug repurposing challenges

4
Mastering data management:
The FAIR Principles as a roadmap

5
Laying your data foundation:
Putting the FAIR Principles into action with knowledge management strategies

11
Building upon your data foundation:
Strengthening AI models for high-impact outcomes

15
Accelerating your knowledge management initiatives with data expert

16
References



Leveraging knowledge management to tackle drug repurposing challenges

Recently, the pharmaceutical sector has been adopting artificial intelligence (AI) more. **Drug discovery processes now represent over 80% of the global AI market share** within the industry.[1] This surge underscores the crucial role of cognitive tools in transforming the landscape of pharmaceutical development, offering unprecedented opportunities to accelerate drug discovery and repurposing efforts.

As AI-assisted drug repurposing successes multiply, **more researchers are embracing new digital technologies** to streamline their discovery pipelines and improve approval success rates. However, companies integrating AI into their drug repurposing efforts without addressing critical data considerations risk compromising their strategy, financial investments, and patient safety.

For pharmaceutical companies that generate massive volumes of information, getting their data landscape AI-ready can quickly become resource-intensive. However, those who understand essential data governance principles and implement knowledge management strategies can **avoid the typical data pitfalls** of cognitive tool use, streamline their AI-driven repurposing frameworks, and pioneer therapeutic breakthroughs.

Three major obstacles impede the effective integration of AI in drug repurposing frameworks:

POOR DATA QUALITY
Fueling AI-driven analytics with incomplete, inaccurate, and inconsistent data hinders the model's ability to generate reliable insights.

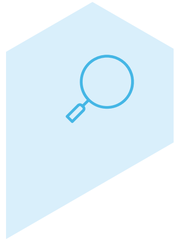
SILOED OR 'DARK DATA'[2]
Unstructured and unharmonized data siloed in disconnected systems limits AI's capabilities by restricting access to critical information.

DATA BIAS
Operating on homogeneous datasets leads to insufficient algorithm validation, biased AI models, and low prediction accuracy.

Mastering data management: The FAIR Principles as a roadmap

Introduced in 2016 by data science experts, the **FAIR Principles** (Findable, Accessible, Interoperable, and Reusable) provide an infrastructure for managing and sharing data to enhance its usage in complex areas like drug repurposing.[3,4] By adhering to the FAIR Principles, organizations can effectively organize unstructured, siloed, and disorganized data into valuable resources that fuel their AI-driven strategies.

FAIR Principles and implications for AI



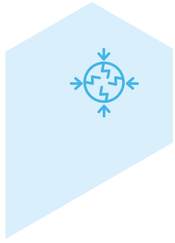
FINDABLE

Assigning unique identifiers to datasets ensures they can be reliably found and referenced, reducing the time spent searching for data sources.

Indexing data in a searchable resource allows researchers to locate relevant datasets, simplifying information retrieval.

Implications for AI

Making information findable ensures AI models can locate the necessary elements for effective predictive analytics, even across massive data volumes.



INTEROPERABLE

Establishing common data standards and formats enables data integration from various sources, mitigating compatibility issues for streamlined AI-driven analytics.

Using shared terminologies and vocabularies ensures consistent data interpretation across systems, enhancing AI's ability to connect disparate datasets.

Implications for AI

Harmonizing data formats and languages empowers AI models to easily combine diverse datasets and generate enhanced insights.



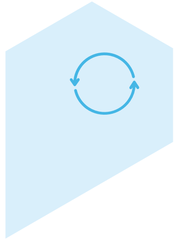
ACCESSIBLE

Establishing clear permissions and licensing terms that consider intellectual property (IP) makes it easier to access data without legal issues.

Storing data in standardized formats enables seamless information retrieval and real-time integration, even when distributed across multiple locations.

Implications for AI

Standardized permission protocols and formats allow AI models to access, retrieve, and use data from different sources without significant accessibility barriers.



REUSABLE

Providing detailed provenance information, including methodologies and usage rights, ensures data can be replicated or combined in different settings.

Adhering to domain-relevant community standards simplifies data reuse by other researchers, enhancing its long-term value and impact.

Implications for AI

Unified and well-curated data contributes to high-quality algorithm training and reusability across AI models without extensive reprocessing.

Introduced in 2016 by data science experts, the **FAIR Principles** (Findable, Accessible, Interoperable, and Reusable) provide an infrastructure for managing and sharing data to enhance its usage in complex areas like drug repurposing.[3,4] By adhering to the FAIR Principles, organizations can effectively organize unstructured, siloed, and disorganized data into valuable resources that fuel their AI-driven strategies.

Laying your data foundation: Putting the FAIR Principles into action with knowledge management strategies

Transforming data into FINDABLE and ACCESSIBLE assets

DIGITIZE

From aborted reprofiling projects to past success stories, legacy data overflows with valuable insights. Yet, locked in physical formats or scattered across unsecured digital platforms, this high-potential dark data often remains unused.

By digitizing legacy documents, companies can convert latent information into accessible and actionable digital assets, making even decades-old documents available for researchers and AI models.

Essential documents to digitize include:

- Lab journals, research reports, and formulation data
- Historical clinical trial and patient records
- Regulatory submissions, approval documents, and compliance reports
- Patents, trademarks, and proprietary research documents

COLLECT

Pharmaceutical companies must integrate legacy documents properly and ensure the integrity of new entries to maintain data quality and support sustainable knowledge management. While most organizations already have a digital acquisition strategy, inconsistencies in protocols and systems can impede information sharing and hinder seamless integration into AI-driven frameworks.

By deploying a unified data collection system with clear entry guidelines and data quality check protocols, pharmaceutical companies can eliminate the need for manual consolidation and significantly reduce the time spent on verification and AI analytics preprocessing. Data governance automation contributes to maintaining data integrity, accuracy, and consistency by minimizing the risk of human errors. This strategy ensures a high-quality data foundation for AI-driven analytics while simplifying cross-departmental dataset merges for faster drug reprofiling pipelines.

SEARCH

Having legacy and new data in digital format significantly improves accessibility. However, easy information retrieval requires properly integrating robust search tools into the company's knowledge management infrastructure.

Appropriate search tools allow pharmaceutical researchers to efficiently search extensive databases of existing drugs, clinical trial outcomes, and patent records across the organization, accelerating their R&D endeavors and reprofiling initiatives.

By implementing search tools in their knowledge management system, pharmaceutical companies can:

- **Optimize data searchability:** Enhance data discoverability and retrievability so users can navigate large datasets or document repositories to locate essential information quickly.
- **Refine data filtering:** Use advanced filtering options to exclude irrelevant data and invest only in the most promising candidates.
- **Customize search parameters:** Tailor searches to meet specific research requirements by setting precise criteria and parameters aligned with drug repurposing projects.
- **Configure to stay up-to-date:** Use real-time updates and alerts on specific drugs or diseases to always work with the most current information.

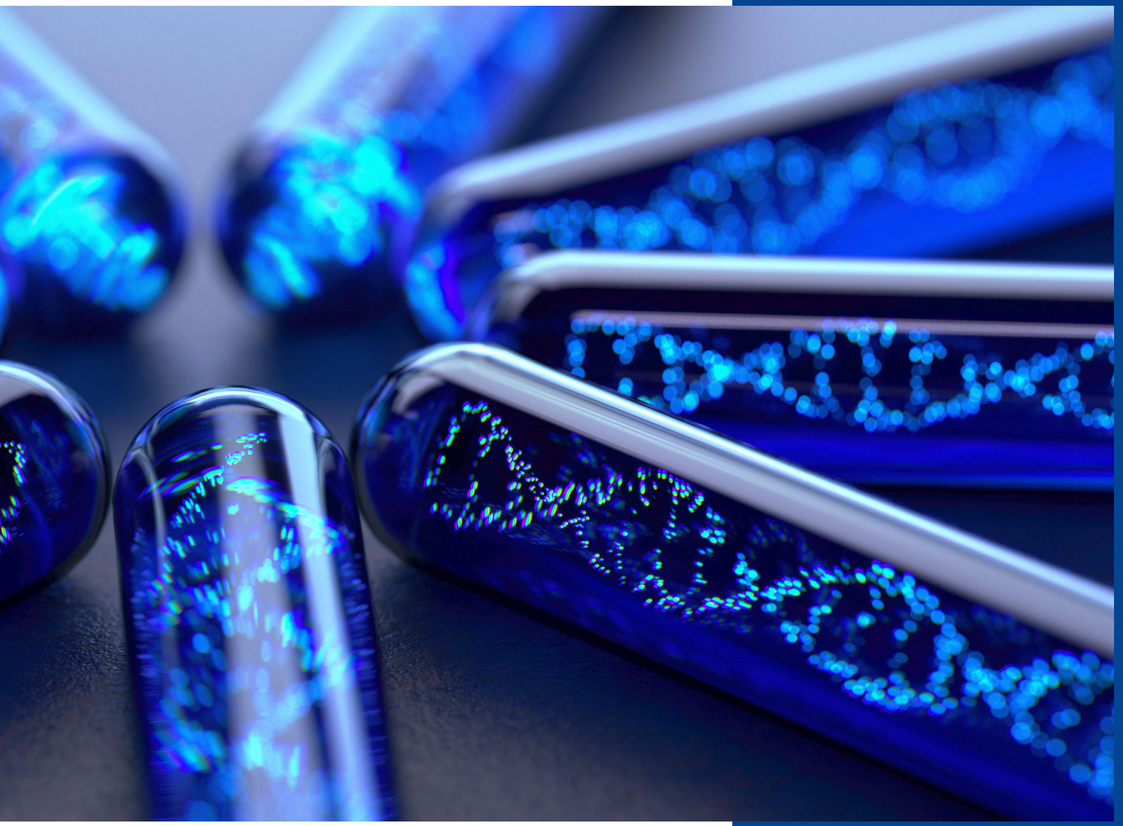
From hours to minutes: Accelerating data retrieval with CAS custom solutions

A large health-tech organization partnered with CAS Custom Services℠ to convert decades of dark data into searchable assets and avoid costly redundant experiments. Thanks to a custom knowledge management system designed with advanced search capabilities, scientists can now quickly tap into years of research documentation, preventing experiment repetitions and accelerating the innovation of new health-tech solutions.[5]

One search that took a senior scientist ~8 hours, now takes minutes with scientifically aware search from CAS.

Company estimate





Transforming data into INTEROPERABLE and REUSABLE assets

HARMONIZE

Successful drug repurposing endeavors involve delving into research from different teams or departments, sometimes on an international scale. However, discrepancies in data entries across your organization can lead to misleading insights, poor decision-making, and costly mistakes.

Through terminology, abbreviation, and format standardization, pharmaceutical companies can:

- **Ensure consistency and compatibility** by facilitating dataset combination and comparison across different sources.
- **Simplify data integration** by reducing the need for data preprocessing for advanced analytics.
- **Promote data reusability** by ensuring all collected information adheres to universal standards.
- **Streamline regulatory compliance** by improving data transparency, traceability, and retrieval.

SHARE

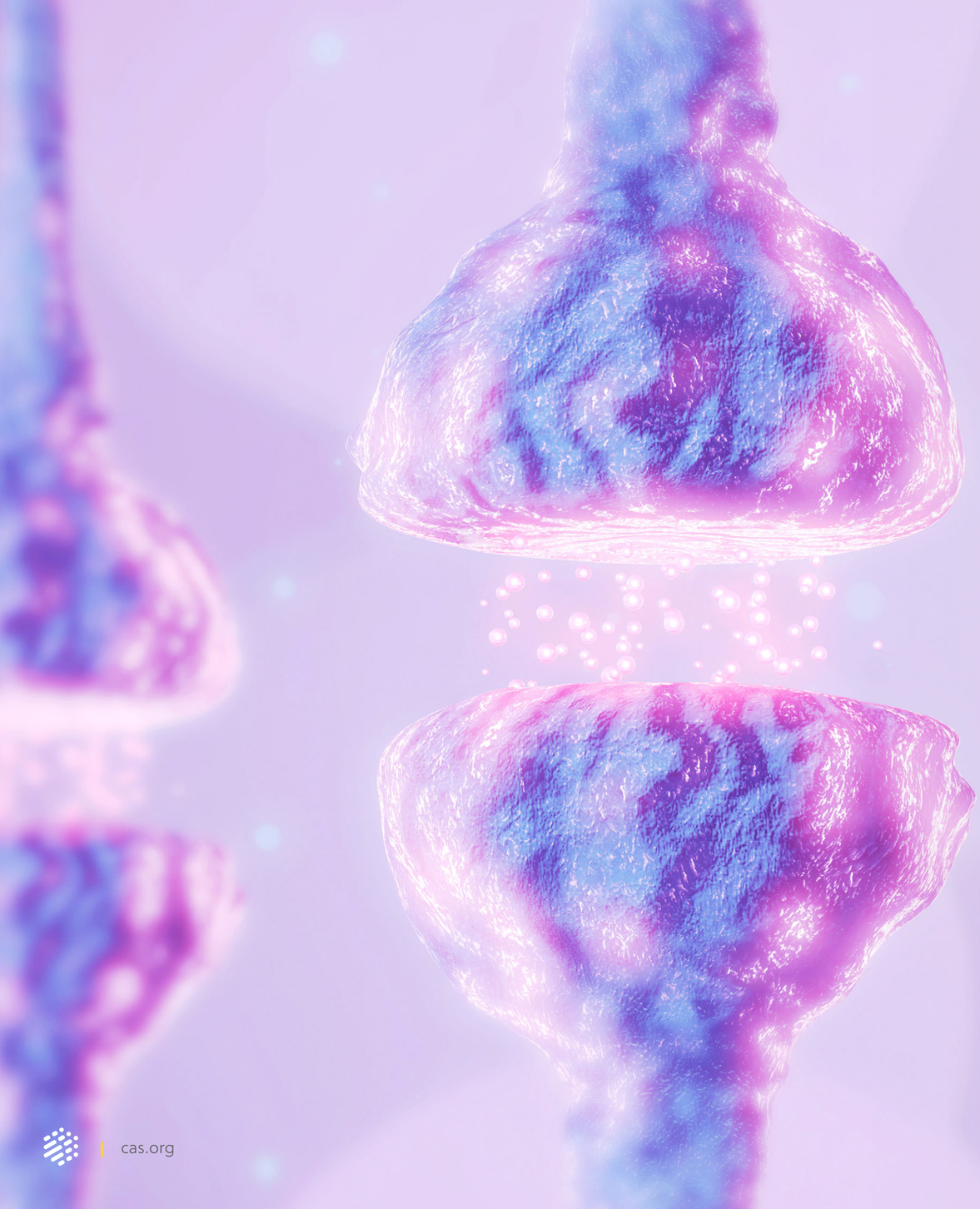
Effective drug repurposing is fueled by cross-team data sharing and multidisciplinary collaboration. To reduce barriers to knowledge exchange and speed up the reprofiling pipeline, pharmaceutical companies can implement a unified knowledge management platform that allows for secure communications, data sharing, and iterations.

Implementing a centralized knowledge management system will:

- **Increase operational efficiency** by reducing the time spent searching for information.
- **Simplify collaboration and data interoperability** by allowing real-time data-sharing and communications.
- **Prevent redundancy and resource waste** by offering a unified view of information.
- **Streamline compliance and auditing processes** by standardizing data practices and simplifying information retrieval.

CONNECT

Interoperable and reusable data enables seamless integration into advanced systems, including knowledge graphs. Knowledge graphs provide a powerful way to organize and explore large, complex datasets, making it easier to discover new insights and make informed decisions. By integrating structured data from multiple sources, knowledge graphs can connect information to uncover hidden relationships in complex data landscapes and guide decision-making for enhanced drug repurposing efforts.



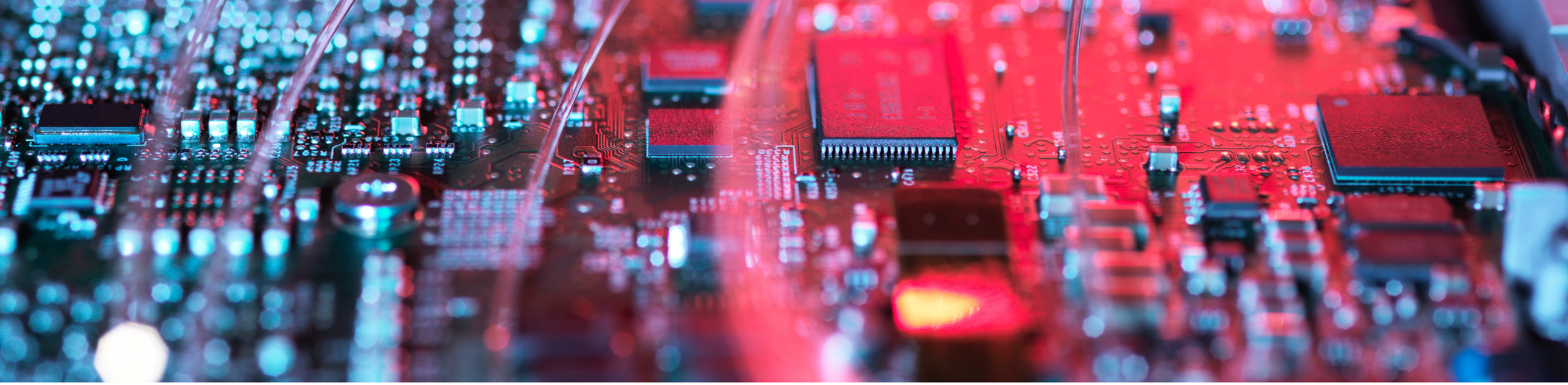
Combining CAS biomedical knowledge graphs and algorithms to maximize drug repurposing efficiency

By integrating over 6 million small molecules, 8,000 biological pathways, and 26,000 human and viral genes, the CAS biomedical knowledge graph created a comprehensive and interconnected dataset that identified 1,350 small molecules as potentially repurposable drugs for COVID-19. By leveraging AI algorithms for automated ranking and prioritization, researchers rapidly identified the top 50 drug repurposing candidates, shaving months off the traditional R&D timeline.[6,7]

Building upon your data foundation: Strengthening AI models for high-impact outcomes

While a solid data infrastructure lays the foundation for effective AI-driven drug repurposing strategies, AI models trained and validated on inadequate or limited data can compromise the quality of insights and jeopardize reprofiling workflows. To ensure reliable predictions for effective decision-making, pharmaceutical companies must follow key data strategies that prevent selective bias and enable comprehensive model development.





Diversify to mitigate bias

Limiting AI to internal data can restrict its ability to recognize patterns, skew predictions, and confine its effectiveness to a narrow set of scenarios. By incorporating additional data from scientific literature and patents into your internal data landscape, you can broaden the scope of your models, reduce the risk of biased outcomes, and significantly enhance the impact of your AI-driven strategies. However, in complex fields like drug repurposing, publicly available datasets often fall short due to transcription errors, legal restrictions, or unsearchable formats.

Unlocking critical insights: Elevating your internal database with the CAS Content Collection™

The CAS Content Collection, meticulously extracted, verified, and indexed, offers access to an unmatched depth and breadth of information.[8] **As the world's largest human-curated repository of scientific data,** this resource provides high-quality data that can be seamlessly integrated into your internal database, enhancing your AI model's training and broadening its focus and simplifying information retrieval.



Customize to reach full potential

Training AI models is crucial, but relying on generic datasets can introduce irrelevant and noisy data, diminishing the accuracy and reliability of predictions. This lack of precision and potential bias undermines the model's effectiveness, particularly in complex fields like drug repurposing.

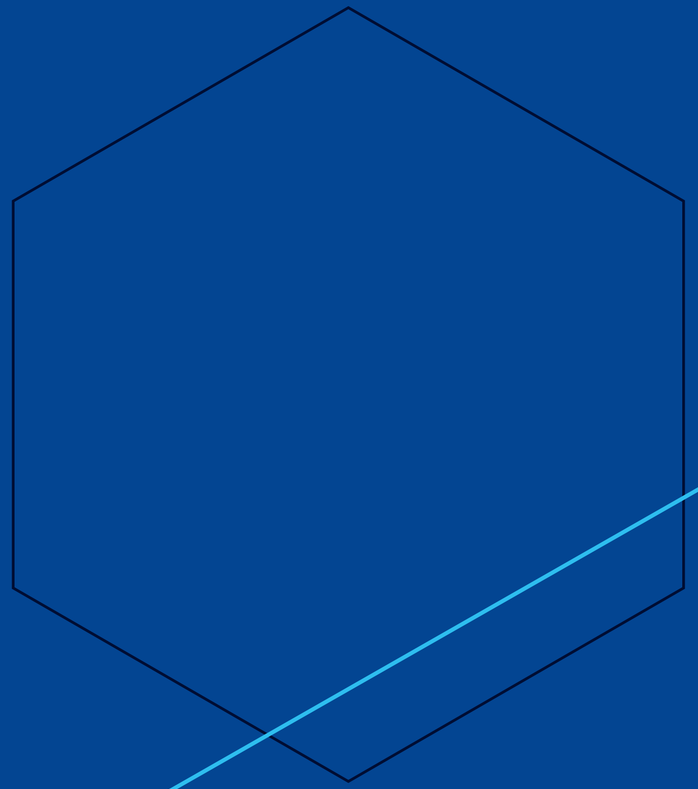
Designed with specific objectives in mind, **custom-curated training datasets can optimize prediction accuracy by fueling your algorithms with relevant information.** For pharmaceutical companies, these tailored datasets significantly enhance AI model precision, leading to faster and more accurate identification of viable drug repurposing candidates.

By training AI on custom-curated datasets, pharmaceutical companies can:

- **Increase analytics relevance** by training on superior data that minimizes the risk of selective bias.
- **Boost recommendation precision** by capturing subtle patterns and nuances specific to your settings.
- **Enhance algorithm transferability** by making your model more adaptable and applicable to real-world scenarios.
- **Optimize resource utilization** by maximizing computational efficiency and minimizing the need for extensive model retraining or fine-tuning.

Cutting discovery time by half with CAS custom-curated datasets

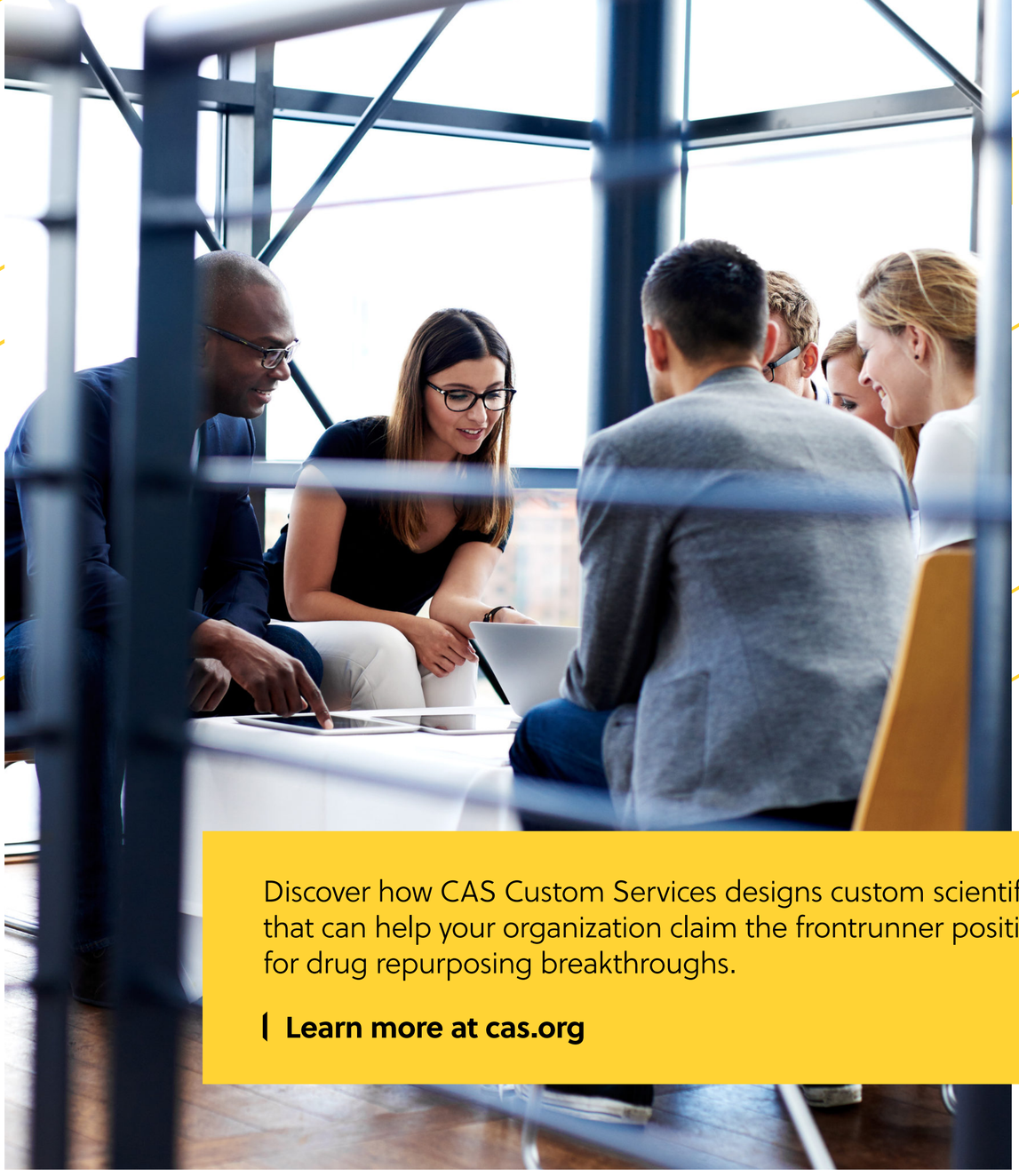
By partnering with **CAS Custom Services** to expand its internal data and train its AI model on highly targeted, custom datasets, a major chemical company accelerated the discovery of new materials and devices by 50%.[9]



Accelerating your knowledge management initiatives with data experts

As AI becomes central to drug repurposing, pharmaceutical companies must refine their data to avoid misguided investments and reduce risks. Strategic knowledge management aligned with the FAIR Principles transforms disorganized data into valuable assets that fuel effective AI models. However, this process demands specialized knowledge management expertise to optimize AI-driven drug repurposing efforts.

Partnering with knowledge management experts empowers pharmaceutical companies to elevate their internal capabilities, streamline AI initiatives, and secure a competitive edge in the fast-paced world of drug repurposing. With data specialists crafting custom solutions tailored to your needs, you can turn unstructured information into powerful insights that propel your AI-driven drug repurposing efforts toward groundbreaking success.



Discover how CAS Custom Services designs custom scientific data solutions that can help your organization claim the frontrunner position in the race for drug repurposing breakthroughs.

Learn more at cas.org



References

1. AI in Pharmaceutical Market Size, Share, Report 2023-2032. Accessed August 16, 2024. https://www.precedenceresearch.com/ai-in-pharmaceutical-market
2. Dark data in R&D: How knowledge management can uncover hidden value. CAS. October 25, 2022. Accessed May 30, 2023. https://www.cas.org/resources/cas-insights/digital/dark-data-knowledge-management
3. FAIR Principles. GO FAIR. Accessed August 16, 2024. https://www.go-fair.org/fair-principles/
4. Wilkinson MD, Dumontier M, Aalbersberg IjJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3(1):160018. doi:10.1038/sdata.2016.18
5. Health-tech unlocks hidden insights. Accessed August 22, 2024. https://www.cas.org/resources/gated-content/cas-unlocks-potential-dark-data
6. Knowledge graph-based approaches to drug repurposing for COVID-19. Accessed August 16, 2024. https://www.cas.org/resources/cas-insights/knowledge-graph-drug-repurposing
7. Al-Saleem J, Granet R, Ramakrishnan S, et al. Knowledge Graph-Based Approaches to Drug Repurposing for COVID-19. J Chem Inf Model. 2021;61(8):4058-4067. doi:10.1021/acs.jcim.1c00642
8. CAS Data. Accessed August 16, 2024. https://www.cas.org/cas-data
9. How Curated Training Sets from CAS Improved the Prediction Accuracy and Transferability of an AI Model. CAS. Accessed June 5, 2023. https://www.cas.org/resources/gated-content/case-studies/ai-predictions

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

© 2024 American Chemical Society. All rights reserved.
SVCGENENGWHP10225241030

CAS
A division of the
American Chemical Society