Automation of ADMET Assays with CDD Vault and KNIME at IOCB Prague

Organization:

Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague

Team:

Anna Janovská, Matúš Drexler, Pavel Kraina

Overview

The **Institute of Organic Chemistry and Biochemistry (IOCB)** in Prague has long been recognized for its pioneering contributions to medicinal chemistry and pharmacology. Within its Biochemical Pharmacology unit, researchers run a suite of in vitro **ADMET** (Absorption, Distribution, Metabolism, Excretion, and Toxicity) assays to characterize compounds at various stages of the discovery pipeline.

As the group's research portfolio expanded, the team began to face an increasingly common challenge in translational research: how to capture, manage, and analyze large, diverse datasets from multiple instruments and assay formats, while maintaining data consistency, traceability, and speed.

The Challenge: Fragmented Data and Manual Workflows

Prior to adopting CDD Vault, IOCB's ADMET group operated through a patchwork of tools: Excel spreadsheets, GraphPad Prism files, and vendor-specific instrument software such as Sciex OS. While each worked well in isolation, together they created a fragmented system with limited traceability. Data were often stored in multiple versions and formats, making it difficult to compare results across assays and timeframes or ensure a single source of truth.

The lack of a central database also meant:

- No barcode-based tracking of compounds or plates
- No systematic backup or ontology structure
- Manual transcription steps prone to human error
- A time-intensive data analysis process that could stretch across several days

The team knew they needed a **digital transformation**, a system that could unify compound registration, data analysis, and reporting into one coherent workflow.

Anna Janovská reflected on the situation before the transition:

> "Before implementing CDD Vault, much of our time was spent manually preparing and transferring data. We had to copy data from Sciex OS into Excel, apply formulas, check each column, and ensure all values were correctly aligned before any analysis could begin. This process was slow, error-prone, and made handling outliers difficult. Now, with KNIME workflows feeding data directly into CDD Vault, everything is aligned, easily accessible, and ready almost instantly. What used to take days now takes only a few hours, with consistent results every time."

The Solution: Integrating KNIME Workflows with CDD

To address these bottlenecks, **the IOCB team** designed a modular and automated workflow that integrates **CDD Vault** with **KNIME**, an open-source data pipelining platform.

This new system seamlessly connects experimental planning, data capture, and analysis for eight core **ADMET assays**, including:

- Microsomal, plasma, S9 fraction (human, mouse, rat), and chemical stability
- Caco-2 and PAMPA permeability
- Kinetic solubility
- Plasma protein binding

The objective of the project was to produce consistent, structured outputs that are automatically evaluated and annotated through CDD Vault. Integration with instrument systems (Echo-650, Echo-MS, LC-MS, Sciex OS) and custom KNIME workflows should allow for the automated transformation and import of data, significantly reducing manual intervention.

From Manual to Semi-Automated Data Handling

The new workflow uses CDD Vault as the **central data repository**, ensuring that all compounds, plate maps, and assay results are consistently stored and accessible. Compounds and source plates are registered in Vault, and **barcodes** now link every physical and digital step of the experiment.

KNIME connects to CDD Vault via the API and serves as the automation backbone, orchestrating data extraction, transformation, and import into CDD Vault through its **API endpoints**. This replaces multi-step manual uploads with semi-automated processes that maintain full metadata integrity.

Custom Tools: MOLD and PADME

Using KNIME and CDD Vault API, the IOCB team developed two workflow tools:

- **MOLD:** A plate layout generator that maps compounds from CDD Vault and designs dilution series, automatically generating the .csv instructions for Echo liquid handlers. It ensures complete reproducibility from source plate to assay plate.

- **PADME**: A data transformation engine built in KNIME that processes raw .txt outputs from instruments (e.g., Sciex OS) and prepares the processed results for upload to CDD Vault, where parameters such as percent remaining, half-life (t1/2), intrinsic clearance (Clint) and permeability are calculated, analyzed and visualized.

The combination of **MOLD, PADME, and CDD Vault** effectively closed the data loop – from experimental setup to data interpretation – in a single digital ecosystem.

Impact: Consistency, Speed, and Reproducibility

The results of the transition were immediate and measurable.

- Data analysis that once required **several days of manual work** now takes only a few minutes.

- All raw data, intermediate files, and evaluated results are automatically **backed up in CDD Vault**, ensuring long-term accessibility and reproducibility.

- **Standardized assay ontologies** now enable seamless comparison across different assay types and species.

- The team can export structured .csv/.xlsx reports directly from Vault, supporting internal decision-making and external collaborations.

"It was a challenge at first," **Pavel Kraina** recalls, "but once the workflows were connected, automation completely changed how we work. Our data are consistent, easy to find, and the process is faster and far more reliable."

This transition also allowed scientists to **focus more on experimentation and interpretation** rather than repetitive data management tasks, a key motivation behind the project.

Broader Benefits Across IOCB

The CDD Vault - KNIME integration has inspired other IOCB researchers to rethink their data practices. Teams across biology and chemistry now see Vault not just as a storage system but as a **collaborative data hub** that connects people, experiments, and analytics.

Matúš Drexler from IOCB noted that this transformation reflects a broader trend:

"Automation and data centralization are not just about saving time; they create a culture of consistency and shared insight that strengthens the entire discovery process."
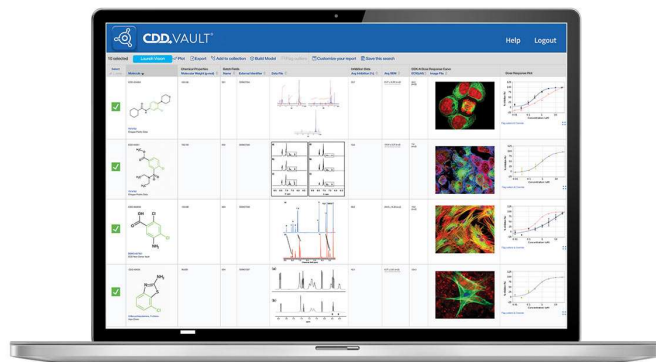
Next Steps

The IOCB Biochemical Pharmacology group continues to refine and expand their digital ecosystem. Future plans include:

- Completing **ontology labeling** for all assays
- Moving to **KNIME server-based deployment** for multi-user access and scalability
- Extending the same automated logic to additional biochemical and pharmacological assays
- Exploring other workflows and **ELN integration** for streamlined documentation of experimental workflows

Conclusion

The integration of **CDD Vault** and **KNIME** at IOCB Prague demonstrates the transformative power of workflow automation in early-stage drug discovery. By replacing fragmented, manual methods with an integrated, ontology-driven, and automated system, the IOCB team has built a scalable foundation for modern pharmacology, one that unites chemistry, biology, and data science under a single, collaborative platform.

This case exemplifies how **CDD Vault** empowers research organizations to modernize their data infrastructure, accelerate experimental cycles, and unlock the full value of their scientific data.

To get your personalized demo or a free trial of CDD Vault, contact **info@collaborativedrug.com**