Accelerate Clinical Data Curation with a Low-Code Data Science Platform

Explore Four Use Cases

Clinical teams spend **countless hours** ensuring their study data accurately represents trials and holds water when subjected to the stringent standards set by regulatory authorities. And yet poor data quality still contributes to **41%** of project failures.

Clinical trial data is prepared for submission within a rigid framework. Widespread use of specific programming languages by pharma companies and regulatory agencies to follow standard operating procedures, has led to highly interconnected dependencies.

Though newer technologies provide better functionality for navigating complex data structures, organizations are cautious to replace their systems, fearing loss of data and high costs to reengineer established frameworks and retrain specialized staff.

But organizations don't need to replace their legacy systems to benefit from new technologies. Instead they can extend the utility and functionality of their setups by adding an open data science platform. Open platforms can sit on top of, alongside, or even inside existing infrastructures.

This e-Guide takes four practical examples of the many use cases in clinical data curation to illustrate how, as an open, low-code data science tool, KNIME can fit any clinical data setup.

1. Establishing audit trails with full metadata
2. Transforming raw data into standardized data
3. Assessing quality of standardized data
4. Maximizing re-use of data

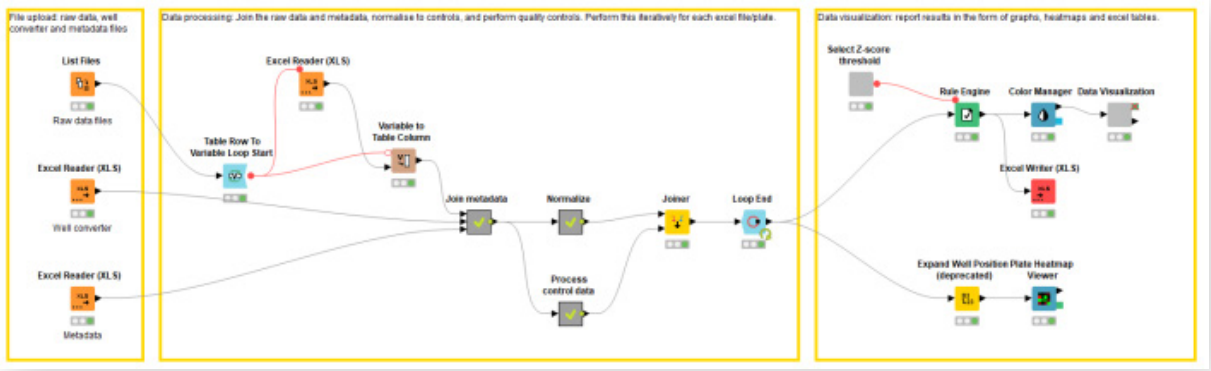
Explore how you can gain access to modern data science techniques with an open data analytics platform without replacing your current system.

Combine Your Existing Setup with KNIME to Boost Efficiencies

KNIME is an open-source data analytics platform. Users can visually create workflows via drag-and-drop. The visual programming environment is intuitive, making it easy for people of different skill sets to explore, process, analyze, and visualize data.

An open platform, KNIME keeps you on the bleeding edge of modern data science. No vendor-developed tool can keep pace with the innovation driven by an active open-source user community.

In addition to data analytics, it offers features for collaboration, reporting, model training, testing, and deployment, making it a versatile tool.



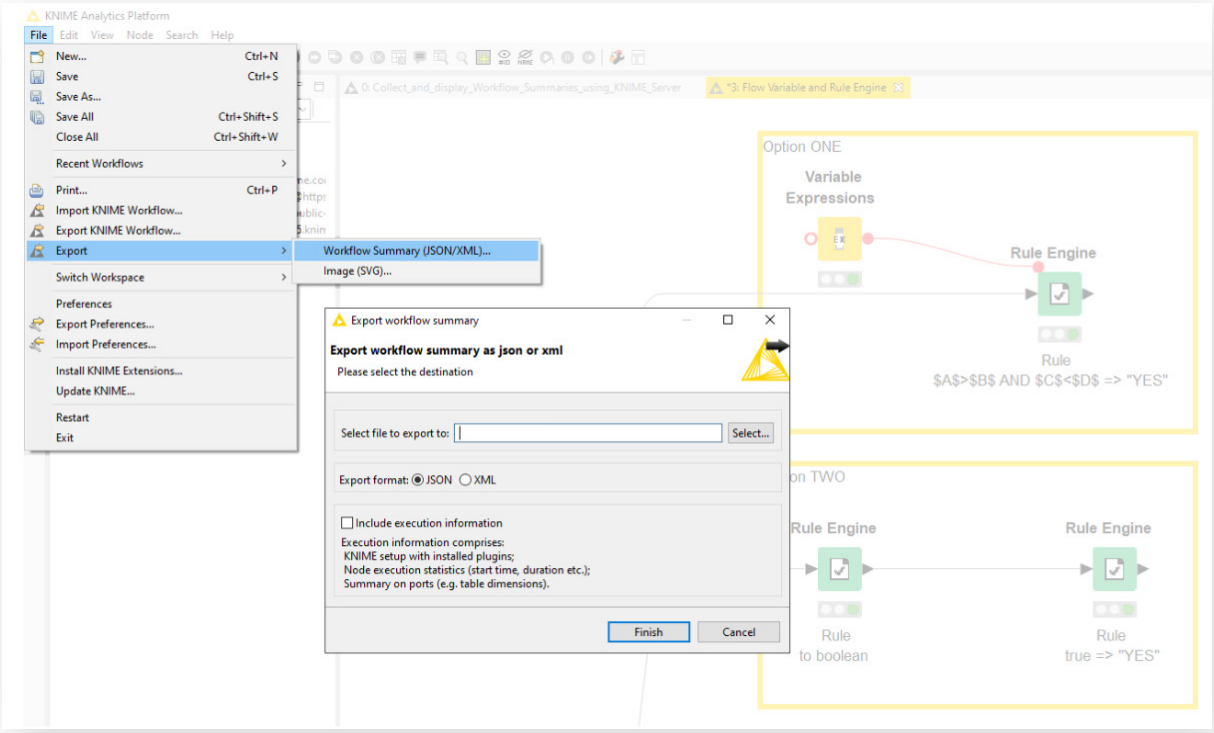
Establish audit trail review procedures with full metadata

Clinical data teams are required to document what happens to all data, where it comes from, and how it is transformed. The first mammoth task is detailing each and every data point, tracing each transformation of the data to produce your audit trail. And then it has to be reviewed.

Being able to capture all this information means capturing metadata. For most data science packages, this is not a straightforward process, because they use scripting languages, which need to be somehow post-processed and converted into the metadata to produce the audit trail.

Once you've got your audit trail, you have to figure out how to load it into a visualization tool for easier review. Often, though, conventional audit trail solutions can't be loaded into these tools directly. This means the data has to be prepared in an additional step before it can be visualized. And to be compliant with audit trail regulations, these data preparation steps also need to be traceable and validatable.

With KNIME, you can export a detailed XML document which traces all the data points and transformations throughout the entire workflow for audit review. You can schedule the XML metadata to be automatically extracted and sent as a report or integrated directly into a visualization tool for review. For review at a future date, you can also archive these files.



Right-click and select Workflow Summary to manually extract the XML document that traces how the data flows through your workflow. Automate this process with KNIME Business Hub.

How Profil, A Full-Service CRO, Improved the Traceability of Data Flow

Profil is a full-service CRO for early clinical trials focused on diabetes and obesity. Based in Germany, they provide clients with scientific and regulatory expertise and services for successful drug and device development.

They have developed a solution that quickly prepares data for review in a reproducible and validatable way. Each data point is logged, with every line in the output data file reflecting any change the data has undergone. These large files can be further analyzed in KNIME or integrated into a third-party visualization tool for review.

If excessive changes have been made to the data, these are easily detected. If a field in a data table has been modified frequently during a short time, possibly to "make a value fit," this will be revealed. Analysis of the metadata also shows whether data changes were performed by the appropriate people.

The ability to capture the full metadata increases **compliance and explainability** while decreasing institutional risk and liability.

"We have created an interactive environment that enables the reviewer to dig into the data and learn what actually happened during a study. With this novel approach, audit trail review not only becomes possible, but meaningful," says Sascha Heckermann, CEO at Profil.

Efficient Transformation of Raw Data into the Study Data Tabulation Model Format

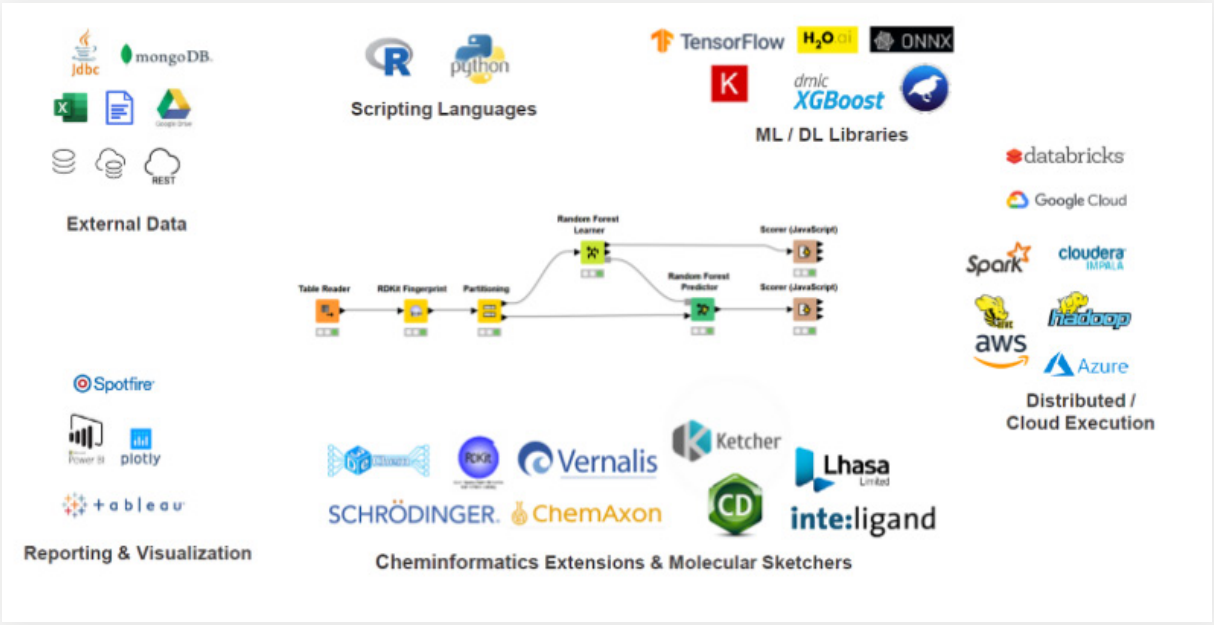
One of the jobs of the clinical programmer is transforming data from its raw format into a standard format (such as the Study Data Tabulation Model format) requested by the FDA for submissions of new compounds. It's a complex and time-consuming process. The raw data comes in from hospitals, labs, or external collaboration partners, and it's messy.

Definitions and standards are in place for the collection of raw data, but many data points might still deviate from these predefined standard protocols. Doctors from a clinic in Arizona note treatment dates in one format, while their UK colleagues use another. These dates need to be transformed into a uniform format, measurements converted into consistent units, terminology standardized, clinical and medical ontologies applied – the list goes on.

While conventional (often proprietary) solutions provide ETL capabilities, these are difficult to maintain. They enable access to any data, but each source is accessed differently. New sources can also mean having to purchase a new connection package.

With KNIME, clinical data teams gain connectivity to 300+ data sources, powerful file handling-capabilities, and support for data types unique to the R&D process, such as sequences, molecules, images, etc. Teams can work within a single consistent environment, with no need to piece together multiple solutions or incur additional costs to access different data sources.

Efficient, automated data integration means R&D teams spend less time collecting, transforming, and loading data, and have more time for analysis.



Some of the integrations with tools and environments required in clinical data curation.

How a Multinational Biotech Company Automated Aggregation of Data from Over 200 Sources

A feasibility study was set up by a multinational pharmaceutical and biotech company to examine **Visual Programming for ADaM Derivation** as an alternative for deriving clinical analysis datasets.

KNIME's complete data science and ETL capabilities make cleaning, mapping, and date derivation easy. At each step of the data transformation process, the actual state of the table can be investigated, which gives KNIME an advantage over classical programming languages.

Accessing and reading data from different sources — e.g. clinical data from the PHUSE Open Data Repository using remote SQL connections, in xml, or .sas7bdat format — is easy with KNIME's multiple connectivity options. Companies can benefit from access to over 200 data sources automatically, without having to pay for extra products or packages to read or write back to those data sources.

The visual programming environment provides immediate clarity of complex processes for anyone, regardless of their programming background. The intuitive interface makes clinical data derivation more accessible and intuitive to learn for new statistical programmers. They don't have to work through complex scripts, but can instead understand the data derivation step by step, following the visual workflow.

The extensive log files and integrated table comparison functionalities in KNIME offer the opportunity for validation by double programming, while a verbose XML document traces all the data points and transformations throughout the entire workflow for audit review.

The study showed that the typical data derivation steps performed using **visual programming with KNIME** not only match the results that statistical programming teams reach with the standard approach, but with the additional benefit of being more efficient, flexible, and intuitively understood.

Streamlined Quality Assessment of Standardized Data

Clinical data managers collect, integrate, and visualize data from multiple systems before handing it over to clinical study teams for ongoing review and decision-making for the safety and efficacy of the particular drug.

Typically this involves pulling the data from multiple source systems, collaborating with different tools and external parties, before the data can be visualized in dashboards or used in reports. Data managers inevitably spend a lot of time chasing the data and troubleshooting between tools. Because all these steps have to be performed within a rigid framework, this has led to widespread use of specific tools and programming languages.

As a result, when organizations look to improve this process, any solution must easily work together with existing legacy systems and tools, like SAS and Python.

Open platforms are all about seamless integration. With KNIME, teams can leverage their existing technology infrastructure. There's no need to rip and replace. Instead KNIME can sit on top of, alongside, or even inside existing infrastructures. If you have to work with SAS programs, you can.

Built-in Python and R integrations enable scientists to work in the tools and environments of choice, and even develop custom functionality as required. If you have to work with Python or R scripts, you can.

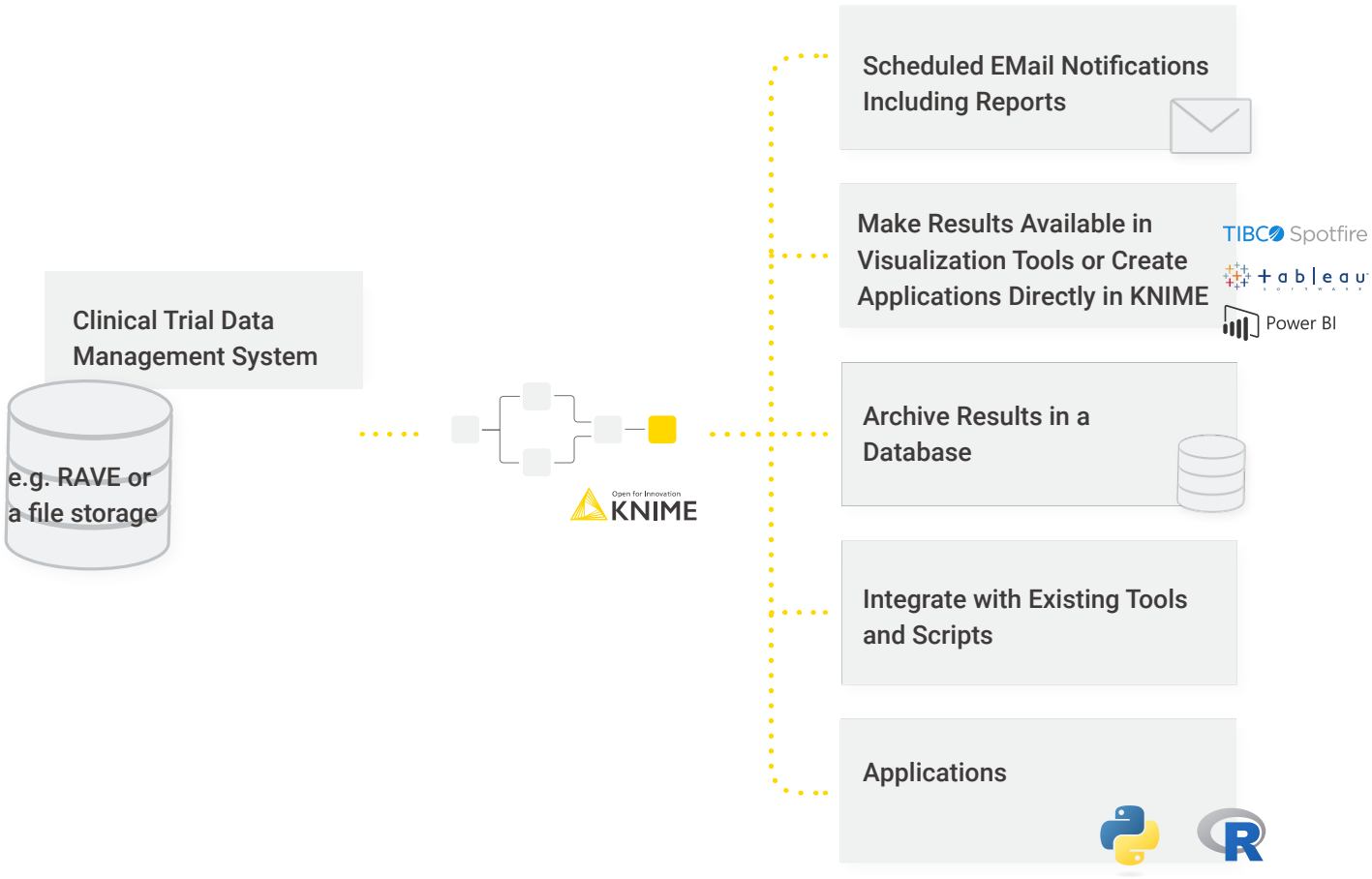
How a Clinical Study Team at a Global Pharmaceutical Company Automated their Data Quality Assessment Process

The clinical study team at a global pharmaceutical company began using KNIME to streamline data standardization and improve their data quality assessment process.

The KNIME solution processes and transforms the data. It runs a completeness check to ensure the data's quality. If the workflow discovers any discrepancies, it automatically sends an email to alert the team. KNIME can be integrated with different visualization tools, like Tableau, Spotfire, Qlik, and more, enabling the team to choose the visualization tool they prefer.

The team appreciates how KNIME coexists with Python and SAS, since these tools are used widely in clinical trial companies. They can develop custom functionality using Python code inside of KNIME nodes as part of the entire workflow. And blueprints to import SAS files or convert SAS dates, for example, can be downloaded from the KNIME Community Hub as needed.

Data managers and clinical trial managers can now focus on executing the trial rather than chasing the data.



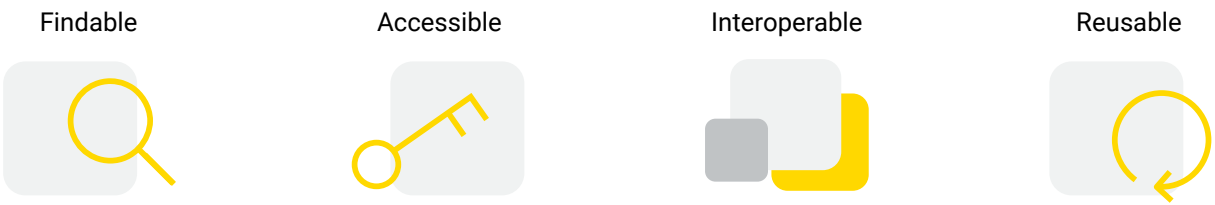
Clinical data operations: KNIME handles data transformation and works alongside different applications and databases.

Maximize Reuse with Findable, Interoperable, and Reusable Data

Pre-competitive data sharing can offer the pharma industry significant benefits in terms of reducing the time and costs involved in getting a new drug to market, via more informed testing strategies and knowledge. However, estimates say that up to 80% of results from publicly funded research are never used again.

When research data cannot be found, or when access, interoperability, and reuse are impaired, the impact can be significant, hampering data-driven innovation and knowledge discovery, jeopardizing collaborations, and tying up budget and resources in areas not contributing to the competitive edge.

The FAIR guiding principles, which were published in 2016, aim at facilitating maximum data reuse, and are now gaining attention in both academic and industry research. Many funding organizations and publishers are now enforcing their implementation to unlock the value of clinical data.



The FAIR Guiding Principles to ensure that research data and objects are findable, accessible, interoperable, and reusable.

Well-structured, machine-readable, and documented data is not only important for people working with big data; it also applies to unstructured data in emails, files, and spreadsheets. KNIME is a useful tool to restructure all kinds of data and bring them into a machine-friendly, interoperable form.

KNIME can read numerous file formats, extract and transform data, concatenate tables, and combine/join on user-defined criteria — all with only a minimal need for coding. It can also be used to extend the metadata with controlled vocabulary, and even automatically write the output to the relevant repositories.

How KNIME Life Sciences Team Dogfooded KNIME to Ensure Data was FAIR

The Life Sciences team at KNIME assembled an academic research use case. In a toxicological testing study, information about compounds and the tested doses were stored in multiple spreadsheets.

To improve the interoperability and reusability of this data, they created a KNIME **workflow to make our data FAIR**.

The workflow combines the data from 48 individual Excel files into one large table. It also adds additional information to make the data more usable for other purposes. For example, the chemical identifiers (CAS numbers) were extended by SMILES, InChI, and InChI keys using the RDKit KNIME community nodes and REST API, enhancing interoperability.

The metadata was extended by domain-specific controlled vocabulary using REST API and programmatic access to the chEMBL and chEBI databases, improving consistency of the data. To further comply with FAIR principles, details about the used databases and ontologies are extracted as well.

The data can now be uploaded to a (project-specific) data repository, meeting the funding requirements of the project. Depending on where the (meta)data is stored, KNIME can also automatically upload the data to this location via a PUT request. Project partners or researchers working on follow-ups could then access a well-structured data table. They can easily recap the published results and integrate findings with newly generated data or data from chEMBL.

Integrate KNIME into Clinical Data Curation

Clinical data curation has undergone significant changes in the last decade, due to the increasing volume and complexity of clinical data. New advances in technology present new opportunities for clinical data curators.

For this reason, organizations are integrating data science platforms like KNIME to gain access to advanced data science techniques to manage and curate clinical data more effectively.

Four Benefits of KNIME for Clinical Data Curation

1. Extensibility and flexibility to integrate with existing systems
2. Detailed traceability of all data points and transformations
3. Intuitive visualization of the data flow within the application
4. Accessibility to powerful ETL capabilities for new learners

Start Building Clinical Data Curation Solutions Immediately

With KNIME's intuitive visual programming environment, teams can easily build and share reusable solutions. In comparison to purely commercial no-code/low-code tools, as an open-source platform, KNIME lowers the barrier to entry. People in your organization can **download KNIME** and start building solutions immediately.

About KNIME

KNIME is a global company that provides data analytics tools for customers across verticals. KNIME software is embraced by life scientists because it enables teams to use a single platform from start to finish, from drug discovery through to manufacturing, sales, and marketing, with all processes verifiable, secure, and easily shared within teams.

Request a Demo with a KNIME Life Sciences Expert