

# Automated, Harmonized CRO Data Drives Cost-Effective Collaboration

## Who Should Read this Study?

*Externalized research leaders, life sciences start-up founders, pharmacologists, data scientists & engineers, scientists working in research and development, rare disease specialists, CRO leaders and scientists, R&D IT professionals*

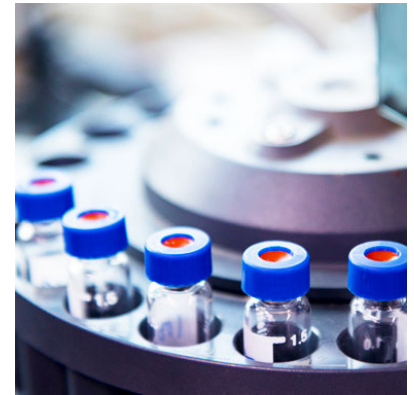
## CASE STUDY

### ENABLING FAIR ACCESS TO PHARMACOKINETIC AND PHARMACODYNAMIC RESULTS

Characterization of absorption, distribution, metabolism, and excretion (ADME) drug parameters inform safety testing and future clinical studies. To augment their current capabilities in a time- and capital-efficient way, pharmaceutical and biotechnology firms often outsource compound synthesis and ADME/DMPK assays to Contract Research Organizations (CROs). Aside from cost savings, the challenges of manually transcribing unformatted or Excel-bound CRO reports dramatically slows the drug discovery and development process. Data availability and quality consumes scientists' cycles, especially when biopharma sponsors work with multiple CROs. Note: most CROs maintain proprietary data formats for standard assays, and these are usually file-based (e.g. Excel or PowerPoint files).

### SUCCESS METRICS

| Saved Time  | Reduced Complexity   |
|---|--|
| <ul style="list-style-type: none"> <li>Save 2+ hours per week, per scientist of manual file transfer work</li> <li>Provide structured and clean data integration for other systems to consume</li> </ul>  | <ul style="list-style-type: none"> <li>Reduce manual errors; Results are directly calculated and derived from raw data</li> <li>Automation allows streamlined downstream analysis via data science applications (e.g. Streamlit apps)</li> </ul> |
| Accelerated Insights  | Improved Transparency and Future-Proofing  |
| <ul style="list-style-type: none"> <li>Improved data-driven decision making (Result is calculated and generated much faster)</li> <li>Harmonized, vendor-agnostic data accelerates ability to run AI/ML and data science projects on top of it</li> </ul> | <ul style="list-style-type: none"> <li>Versatile pipelines are configurable and extensible to handle rich set of technologies and formats</li> </ul>   |



### Customer Profile

- Two Boston-area, public, clinical stage start-up biotechnology companies, who specialize in gene therapy and precision medicine, respectively. Both collaborate with a single key CRO for their ADME/PK needs.

### Product Focus

- Small molecule therapeutics, biomarker development, computational screening

### Client/ Customer

- Biopharmaceuticals

### KPIs / Results

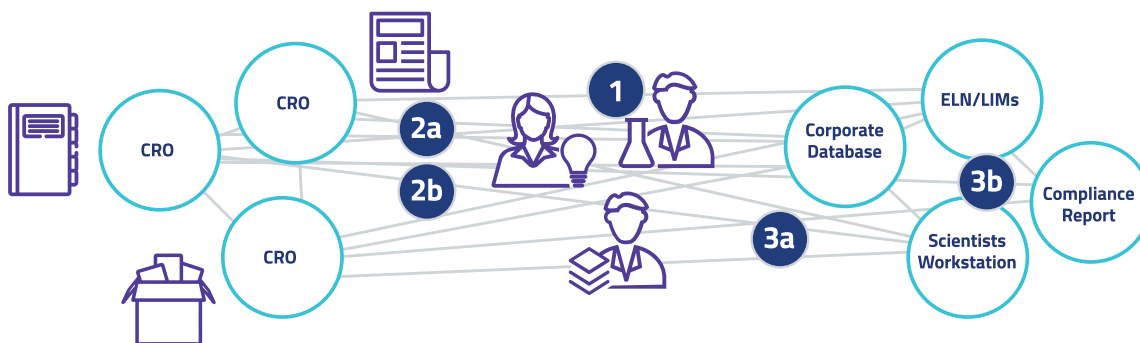
- Gene Therapy, Precision Medicine

To maximally benefit from CRO outsourcing, biopharma companies must automate the returning data flow. Ingestion, validation and curation of the data into internal data stores make it accessible and actionable. Otherwise, the vast volumes of generated data will be stuck in emails, obsolete file versions, or (worst) paper reports. This case study illustrates two recent CRO ADME/ PK reporting workflows from different biotech customers. TetraScience has evolved the ADME /PK report process to be efficient, scalable, and consistent with enforced data quality from the beginning of the pipeline process. Scientists can use their preferred informatics tools (Spotfire, Tableau, Vortex) to construct structure-activity relationship (SAR) tables to better understand how molecular properties influence biological activity.

## ANALYZING TODAY'S MANUAL DATA FLOW

Scheme 1, below, illustrates our customers' current workflow for pharmacokinetics and pharmacodynamics (PK/PD) studies with their CRO partners.

**Scheme 1 - Example ADME and PK Report Workflow (before TetraScience)**



**Step 1:** Scientists receive ADME reports from CROs, perform sanity checks on all required fields for each report - protein binding, kinetics solubility, metabolic stability. Every report submitted must be manually checked, and usually requires ~2 hours of scientists' time.

**Step 2a:** If a required field like "species" - is missing, or a typo in half-life value, scientists contact the CRO for corrections. This step can take multiple emails or phone calls to resolve.

**Step 2b:** CROs resubmit reports to correct mistakes. Scientists repeat steps 1 and 2a (if applicable) and overwrite older versions of this report.

**Step 3a:** Scientists extract results and publish them to the Electronic Lab Notebook ("ELN") and the corporate ADME/Tox database as needed. This step may take several iterations as the scientists may go back and forth between Step 1, 2a, 2b and 3.

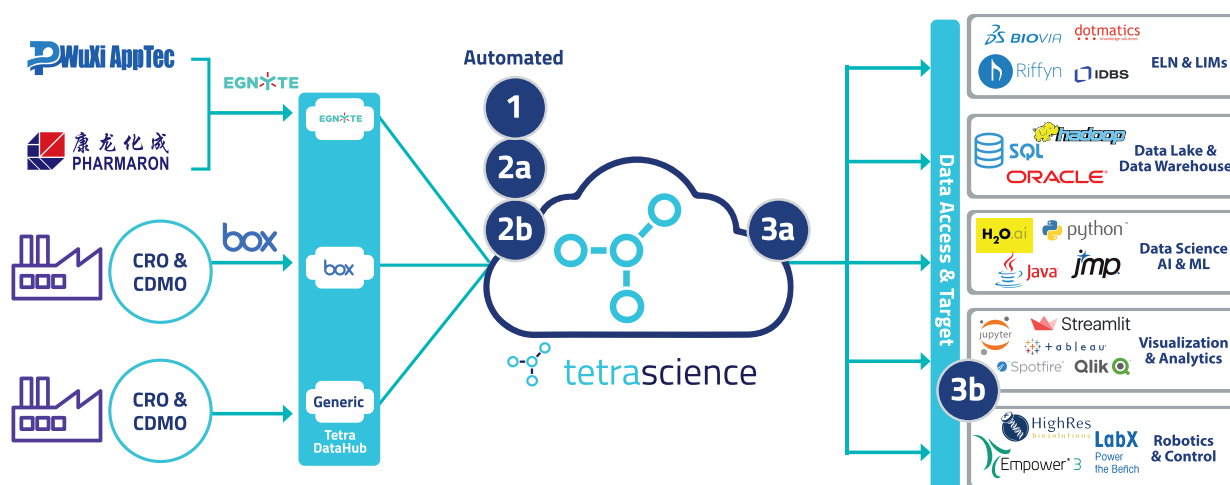
**Step 3b:** Scientists employ SAR tables to interpret pharmacokinetic studies. Reproducible, dependable ADME/PF data is required for good decision-making. If multiple CROs are involved, then scientists must decide how to aggregate results - which metabolic stability, permeability, solubility and protein binding data to keep, reject, or average - for a given compound across different reports. Scientists may repeat the above process until they have all relevant compound data; these manual steps take hours to completely process.

**Key takeaway:** low-cost R&D through externalization may reveal hidden time and data curation expenses. CRO formats hinder this process; protein binding results reported by WuXi AppTec and Pharmaron have different formats and contain non-standard key results fields. How can we connect all of these report formats to a common ontology, consistent results, and ensure FAIR representation of the data?

## OPTIMIZING CRO DATA FLOW

The Tetra Data Platform serves as the common “interpretation engine”. Connecting, parsing and harmonizing ADME/PK reports from various CROs automates manual steps of data entry, processing, and transfer saving time, reducing errors, and increasing throughput. Such harmonized data can directly be used in SAR visualization, data science, AI/ML, and other advanced analytics - read to the end to find out more.

**Scheme 2 - The TDP-enabled ADME/PK workflow today**



**Steps 1 through 3a:** Scientists receive ADME / PK reports via cloud drives (e.g. Egnyte and Box) or emails, as before. However, Box, Egnyte and network drives can now be automatically detected when new or modified reports arrive, which triggers a pipeline to validate and convert to an open data format (JSON). As part of this process, standardized ADME data is also now available to query via RESTful API in the TetraScience Data Lake. Manual data entry is minimized.

When a CRO resubmits a report, a newer version of that report is created automatically in the cloud, while the previous versions are marked as obsolete. Since all queries can retrieve the latest version by default, scientists have peace of mind knowing that their versioning problem has been solved.

**Step 3b:** Scientists can refresh Spotfire or Tableau SAR dashboards to which the cleansed and harmonized ADME / PK reports flow in as new or revised results arrive.

Since all data - across CROs and time periods - is harmonized to a common schema at entry, scientists can search results by compounds and their batches across all CROs, then refine and group the results within the SAR tables to conduct analytics such as correlation analysis.

Let's compare the two processes side-by-side:

| BEFORE  | AFTER            |
|---|------------------|
| <b>Step 1:</b> Manually ADME / PK report review   | <b>Automated</b> |
| <b>Step 2a:</b> Perform data quality checks such as required fields, nonzero results and syntax validation  | <b>Automated</b> |
| <b>Step 2b:</b> Manually delete previously submitted report and make sure the latest version of the report becomes the system of records  | <b>Automated</b> |
| <b>Step 3a:</b> Copy numeric results from the report and enter them in either ELN or the corporate database   | <b>Automated</b> |
| <b>Step 3b:</b> Aggregate ADME results into one table (Structure Activity Relationship "SAR") with compound and batch information. Unifying the results cross CROs is a prerequisite. | <b>Automated</b> |

## BEYOND DATA AUTOMATION: DATA SCIENCE

Now that the data workflow accompanying CRO data exchange is automated, what's next? Disparate pharmacological data are now centralized and harmonized in the Tetra Data Platform. This seems like a prime opportunity to apply some data science! Check out a related blog post about our Intermediate Data Schema (IDS), our open standards method used to seamlessly move data between and across all the different CRO reports, unifying the unique data structure and format from each.

Scientists can now fully utilize their ADME data across different CROs, including querying and visualization of data sets, using the existing data science and analytics tools. For example, scientists can easily query and visualize all active compounds below a certain oral bioavailability threshold in a particular screen, or the behavior of all structurally-related compounds across different screens.

To learn more on how to optimize processes and accelerate discovery, visit [tetrascience.com](https://tetrascience.com)



TetraScience is the Scientific Data Cloud™ company with a mission to accelerate scientific discovery and improve and extend human life.

To learn how you can unlock the power of harmonized, FAIR data in the cloud with the only open, cloud-native platform purpose-built for science, visit [tetrascience.com](https://tetrascience.com)

**Corporate Headquarters** | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

**Atlanta** | 3280 Peachtree Road NE, 7th Floor, Atlanta, GA 30305

© 2022 TetraScience, Inc.



[tetrascience.com](https://tetrascience.com)