

BioCelerate White Paper

Unlocking the Value of Shared Toxicology Data: Case Studies from  
DataCelerate and the Toxicology Data Sharing Initiative

**Authors:**

Jean Gerard Sathish<sup>1</sup>, Benjamin Elser<sup>2</sup>, Yukie Murata<sup>3</sup>, Michael Pritt<sup>4</sup>, Michael Giovanelli<sup>1</sup>,  
Mikael Egebjerg<sup>5</sup>

<sup>1</sup>Pfizer Inc., Groton, CT & Pearl River, NY

<sup>2</sup>Bristol Myers Squibb, New Brunswick, NJ

<sup>3</sup>Shionogi & Co., Ltd., Osaka, Japan

<sup>4</sup>Eli Lilly & Company, Indianapolis, IN

<sup>5</sup>Novo Nordisk, Copenhagen, Denmark



## Contents

Executive Summary.....	3
Introduction .....	4
Case Examples & Experiences .....	6
1.1 Using an Internal Comparative Database with Visualization Tools to Increase Toxicology Data Utility.....	6
1.2 Using Historical Toxicology Data to Train New Study Directors .....	8
1.3 Analyzing Body Weight Kinetics for Selected Targets Using Shared Data .....	9
1.4 Exploring AI Opportunities to Enhance Search and Insights Generation.....	12
1.5 Assessing Correlations Between Adverse Events and Compound Vehicles .....	12
Conclusion.....	14
Acknowledgement .....	15
References.....	15



## Executive Summary

BioCelerate, a subsidiary of TransCelerate BioPharma Inc., identifies and develops pragmatic solutions to improve preclinical research efficiency, with the end goal of streamlining drug development and bringing new medicines to patients faster.<sup>1</sup> In 2016, BioCelerate launched the Toxicology Data Sharing (TDS) initiative supported by DataCelerate, a validated global data sharing platform that now includes over 640 studies contributed by member companies. The participating member companies share data with one another,<sup>2</sup> and each company conducts its own independent research analyses.<sup>3</sup>

BioCelerate recognizes the potential of historical toxicology data to support data-driven decision-making across the drug development lifecycle – for example, investigating the underlying causes of toxicity to support proactive risk management, identifying trends in adverse events across similar mechanisms of action to inform go/no-go development decisions, and conducting comparative analysis between compounds within target classes to better understand compound toxicity profiles.

This paper presents case studies demonstrating how historical toxicology data has been applied by member companies to support comparative analysis, training of new study directors, exploration of AI tools, and standardization of vehicle terminology. While differences in study design and terminology and non-standardized data fields create challenges for aggregating and analyzing the data, the use of SEND-formatted data and modern visualization tools has enabled meaningful insights. The paper highlights how collaborative data sharing can reduce animal use, improve decision-making, and accelerate the development of safer therapies.

---


<sup>1</sup> [Homepage - BioCelerate](#)

<sup>2</sup> Each member company determines in its sole discretion which studies to donate, and there is no mechanism for users to request particular studies. Participating member companies do not need to submit research proposals to use data, but they do need to comply with the terms and conditions of the Data Sharing Agreement (“DSA”). Any information relating to a natural person, like a trial investigator, is stripped from the data before it is contributed.

<sup>3</sup> Member companies conduct their own analyses without consulting with other member companies. The details of analytic methods used in individual analyses is not shared between companies per the DSA.


## Introduction

The Toxicology Data Sharing Initiative (TDS) is designed to accelerate *in vivo* toxicology research through data sharing among participating companies. Motivated in part by the Food and Drug Administration's (FDA) 2011 Strategic Plan for Regulatory Science, which includes objectives to modernize toxicology to enhance product safety, TDS is focused on enabling access to a broader cross-company set of toxicology and background control data. Background control data refers to toxicology data collected from untreated or vehicle-treated control groups across studies. This data provides a baseline for interpreting findings and helps distinguish treatment-related effects from incidental or species-specific observations. Unlike toxicology study data, which includes test article-related findings, background control data focuses on normal biological variability and is essential for contextualizing rare or unexpected events. The knowledge gathered through TDS has the potential to help companies reduce unnecessary animal use by leveraging existing animal data (following the principles of Replacement, Reduction and Refinement, or the 3Rs).

An orange arrow icon pointing to the right, positioned to the left of the text block.

Sharing both toxicology study data and background control data is beneficial in several ways. Access to and analysis of toxicology studies submitted across multiple sources may help to increase drug development efficiency (and therefore reduce cost) by identifying more viable – or, conversely, fatally flawed – compounds earlier in the process. Using a data-driven approach, an organization may approach early decision-making with greater confidence through improved understanding of on-target and off-target toxicity. Leveraging this information, member companies can make more informed decisions about where to invest development resources, enhance study designs, and guide the selection of vehicles for new compounds – all while independently conducting their own analyses.

Furthermore, a larger set of background control data consolidated in one platform may allow companies to better determine the significance of rare and incidental findings, improving the ability to respond to regulators. The convenience of having control data from a variety of sources consolidated in one venue will improve speed of decision-making on the relevance of such findings. Finally, an additional value proposition for TDS is enabled by the industry's implementation of the Standard for the Exchange of Nonclinical Data (SEND), which aligns with the FDA's new nonclinical submission requirements. The structured data format represents a significant opportunity to apply analytics and modelling to data across studies and sources. Usage of SEND for TDS enables direct comparability among studies contributed to the initiative while also offering a side benefit

A large, light blue arrow icon pointing to the right, positioned at the bottom right of the page.



of amassing lessons learned among participating companies, who are becoming more familiar with the implementation of SEND.

As of January 2026, seven companies have contributed 642 studies to DataCelerate, including 416 background control studies and 226 toxicology studies.

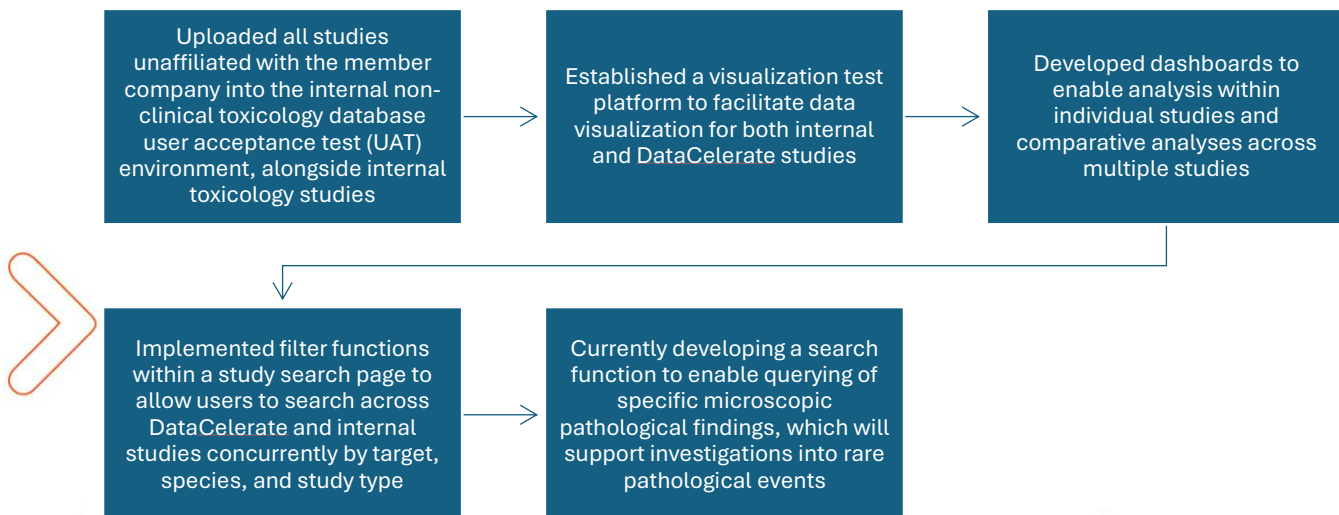


## Case Examples & Experiences

### 1.1 Using an Internal Comparative Database with Visualization Tools to Increase Toxicology Data Utility

Historical toxicology data can help determine whether observed toxicities in new drug candidates are on-target or off-target. On-target drug toxicity refers to side effects resulting from a drug interacting with its intended biological target, while off-target toxicity occurs when a drug or metabolite affects unintended targets. Distinguishing between these mechanisms is critical for guiding compound design and can inform decisions to pursue backup compounds that avoid off-target effects. Recognizing this opportunity, a BioCelerate member company leveraged DataCelerate to help distinguish between on-target and off-target toxicities observed in their internal drug candidates.

To improve the utility of the historical toxicology data, the company developed an internal database that allows scientists to query and visualize both internal toxicology studies and DataCelerate studies concurrently. Figure 1 demonstrates the approach that the company undertook to develop this database.




**Figure 1: Approach to creating an internal database for querying DataCelerate and internal toxicology studies.** Figure 1 shows the approach to develop an internal database that integrates DataCelerate data and internal toxicology study data. The figure outlines key steps in data sourcing, system design, and visualization setup to support comparative analysis.

Once the database and dashboard setups were complete, the company initiated a testing phase to evaluate its functionality. Experts from across the non-clinical safety organization, including pathologists, project toxicologists, and other subject matter specialists, were selected to evaluate the database's searchability and analytical capabilities. Next, the database was tested for its ability to help determine whether unexpected toxicological findings for internal compounds were on-target or off-target. A list of pharmacological targets from external DataCelerate studies was compared to the company's historical portfolio to identify overlapping targets. Two cases were identified where studies existed for both internal and external compounds with the same target.

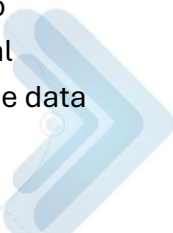
In the first case, an internal drug was observed to affect hematology parameters, which was suspected to be secondary to a metabolite rather than a direct pharmacological effect of the parent drug. To investigate this hypothesis, rat and dog studies from an external compound with the same target (available in the toxicology module of DataCelerate) were examined for similar findings. Comparable hematology changes were not observed in the external studies, supporting the conclusion that internal effects were likely due to the metabolite and not an on-target effect.

In the second case, an internal compound with suspected on-target effects on lymphoid tissue and the gastrointestinal tract was compared to a DataCelerate compound with the same target. DataCelerate studies demonstrated similar effects, supporting the conclusion that the observed toxicity was likely on-target.

An orange outline of a right-pointing arrow is positioned to the left of the text.

These use cases illustrate the value of the DataCelerate data in enabling comparative analyses between internal and external compounds to distinguish between on-target and off-target toxicity. In future cases where the toxicity of a drug candidate prevents further development, determining the mechanism of toxicity can guide critical decisions on discontinuation and/or prioritization of backup molecules. Overall, comparative analyses using DataCelerate can promote more efficient resource prioritization by providing insights into suspected mechanisms of toxicity.

To support broader adoption of comparative analyses using DataCelerate's toxicology data within the company, the company held a training session to demonstrate its capabilities and show how shared data can enhance toxicological interpretation and decision-making. Attendees from across the non-clinical safety group learned how to query and visualize studies within the database, as well as how to distinguish external DataCelerate studies from internal ones. Potential use cases were discussed, and the data was subsequently made available to all members of the non-clinical safety group, providing direct access to DataCelerate data. Moving forward, the team will conduct

A large, light blue, stylized arrow pointing to the right is located in the bottom right corner of the page.

periodic surveys to gather feedback on new use cases and identify opportunities to further refine the dataset's utilization.

## 1.2 Using Historical Toxicology Data to Train New Study Directors

Due to a limited number of internal toxicology studies, one company sought to expose new study directors to historical studies from the DataCelerate platform to deepen understanding of study design, data interpretation, and regulatory expectations.

The training program paired senior and junior researchers to foster mentorship and collaborative learning. The dataset included PDF data tables and SEND-formatted files. Each pair reviewed study data, including pharmacological target, animal species, and administration period (Table 1), and presented their findings to the group. Presentations focused on test article-related changes, toxicity observations, and proposed mechanisms of action. These discussions helped participants learn how to interpret toxicity data, assess adverse effects, and determine the No-Observed-Adverse-Effect Level (NOAEL).

**Table 1: List of data elements used in the training**

Pharmacological target	Animal species	Administration period
Target A	rat	6 weeks
Target B	rat	4 weeks
Target B	monkey	4 weeks
Target C	rat	4 weeks
Target C	pig	4 weeks
Target D	dog	4 weeks
Target E	rat	2 weeks
Target E	dog	2 weeks
Target F	dog	4 weeks

Following the training, a survey was conducted among 10 participants currently working in toxicology-related fields. Nine responded, and 100% of respondents (including both junior and senior participants) indicated that the training was useful. Table 2 summarizes specific situations where participants applied the knowledge that they gained from the training. The results of this survey suggest that training with actual test data is highly effective in improving study directors' knowledge and skills without needing to conduct real toxicology studies.

**Table 2:** Summary of real-world applications of training insights, highlighting how both junior and senior participants benefited from exposure to historical toxicology data.

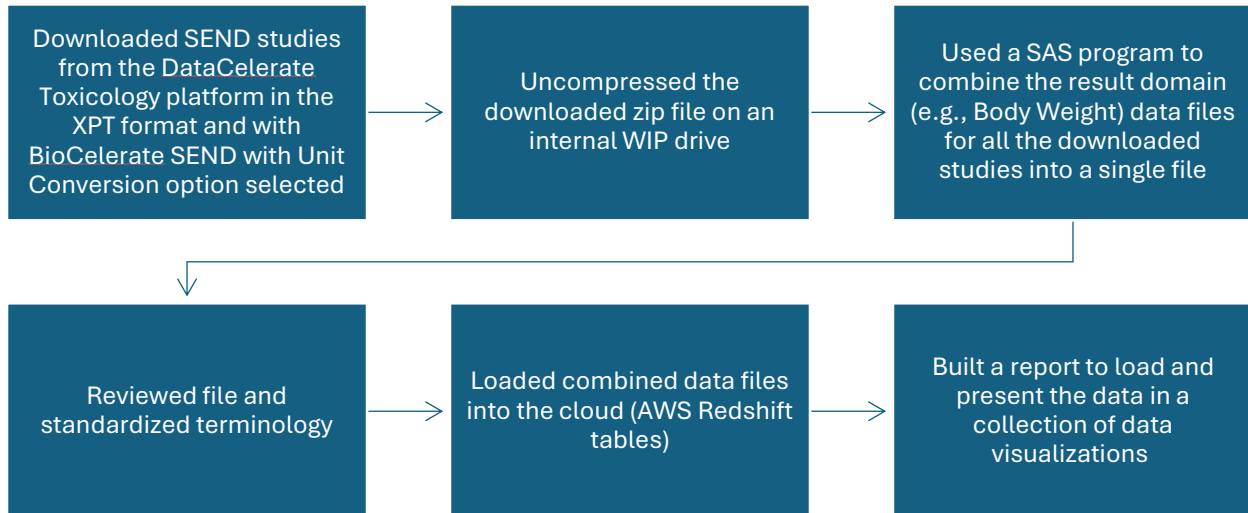
junior/senior	Situation
junior	The training was my first time to consider toxicity study data in non-rodents, so I was able to learn general things, such as the importance of comparing with pre-values.
junior (2 answers)	Since there are few opportunities to oversee study directors for in-house toxicity studies, I felt that the training using data was useful. When I was a study director of in-house compounds, I found this training useful for learning how to deal with the data.
junior and senior	It was possible to have discussions based on actual data, which was useful for gaining experience.
senior	When I come across a study that shows similar changes to the data used in training.
senior	I was able to learn where the members were unsure about their decisions. It was also an opportunity to learn the importance of training to be able to explain adverse changes based on experience or conceptual understanding of toxicology.
senior	This was an opportunity to discuss various on-target toxicities.

This innovative training model allowed study directors to gain hands-on experience in a low-risk environment using real toxicology data. The use of historical data accelerated exposure to diverse pharmacological targets without requiring additional animal studies – a use case that is particularly valuable for smaller companies that may have less internal toxicology data. Based on the positive feedback, the company is now exploring opportunities to expand the program to a wider audience seeking to deepen their understanding of toxicology studies.

### 1.3 Analyzing Body Weight Kinetics for Selected Targets Using Shared Data

Understanding the kinetics and degree of body weight changes in repeat-dose toxicity studies is useful for designing future toxicity studies, particularly when evaluating compounds targeting specific mechanisms. To explore this, the company investigated the use of historical toxicology data from the DataCelerate platform to gather insights on body weight changes in cynomolgus monkeys for selected targets of interest.

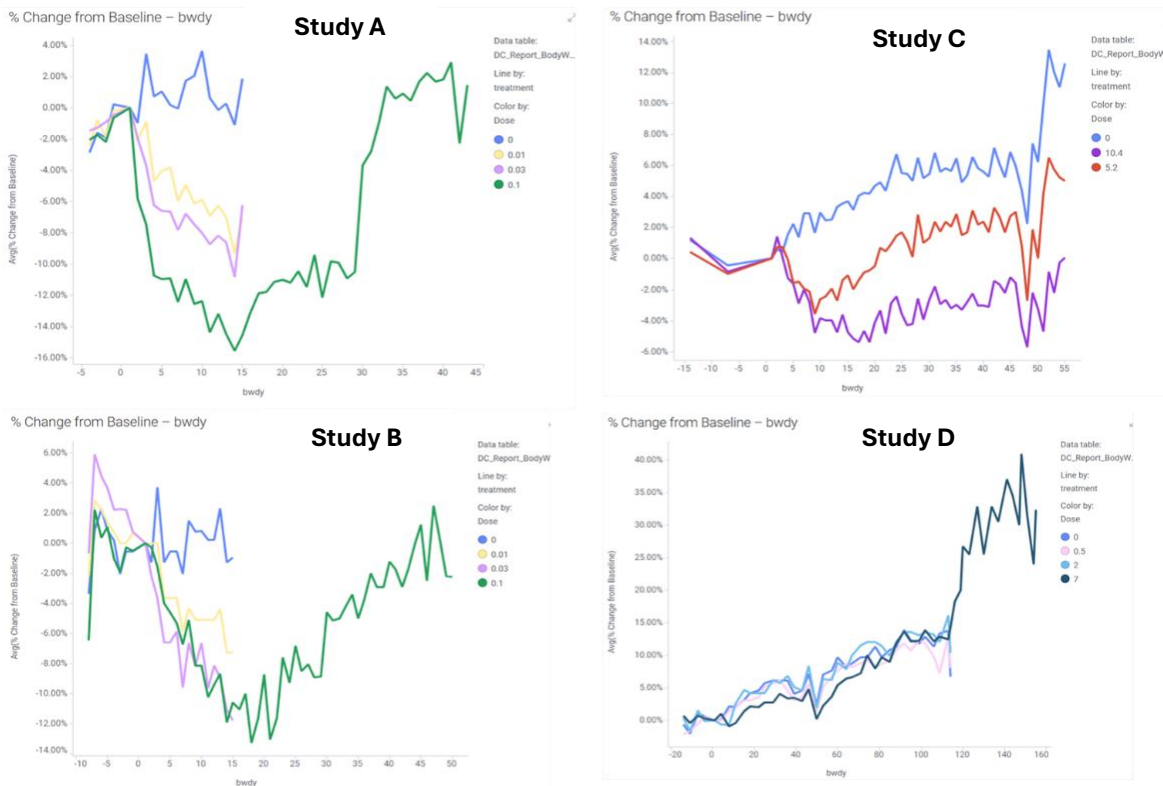
Figure 2 outlines the approach used to convert SEND-formatted studies from DataCelerate into a report that visualizes body weight trends across studies.



**Figure 2 illustrates the workflow for converting SEND-formatted studies into data visualization reports for body weight trends.**

A preliminary search showed variations in the nomenclature for the target of interest. After curating and standardizing the search terms, the team filtered results to display only studies involving cynomolgus monkeys. This search yielded 10 relevant studies, four of which are highlighted in Figure 3. In 3 of the 4 studies, there was a marked reduction of bodyweight over the dosing period, with a ~15% bodyweight change from baseline in Studies A and B and a modest ~6% change in Study C.





**Figure 3: Body weight changes across four studies.**

Figure 3 presents body weight trends across four de-identified studies in cynomolgus monkeys, highlighting dose-dependent effects and variability in pharmacological response.



Interestingly, in Study D, bodyweight increased over the duration of the study, suggesting a potential lack of pharmacology with this candidate (or at least in the monkey) at the doses evaluated. The dose relationship is also clear in Studies A and C, while there is no separation in bodyweight change with increasing dose in Study B.

This exercise enhanced understanding of potential body weight effects by leveraging a larger dataset to assess target(s) of interest. A deeper understanding of potential target findings/effects can support faster, better-informed decision-making regarding targets while also reducing animal use, ultimately accelerating the delivery of new compounds to patients.<sup>4</sup>

<sup>4</sup>Please note that all procedures performed on animals were in accordance with regulations and established guidelines and were reviewed and approved by an Institutional Animal Care and Use Committee or through an ethical review process.

#### 1.4 Exploring AI Opportunities to Enhance Search and Insights Generation

Many pharmaceutical companies are currently exploring AI applications in drug safety, and developing robust AI models requires access to large, well-standardized datasets. One company evaluated the feasibility of developing an AI-powered chatbot large language model (LLM) to improve study searchability. For example, such a tool could allow users to quickly retrieve all studies involving monkey subjects to support the creation of virtual control arms or synthetic comparator groups constructed from historical control data.

The company also considered the possibility that the chatbot LLM could identify patterns and generate insights from the data. The aspiration was to determine if pre-clinical findings could be linked to clinical trial data to assess how pre-clinical findings translate to clinical settings. For example, if the chatbot LLM identified a link between an adverse event and a particular biomarker, clinical teams could leverage that biomarker to determine the risk associated with the adverse event for a particular patient.

However, the creation of an AI-powered chatbot LLM was not feasible at this stage. AI capabilities rely on having available interpretations, but the platform – which relies on SEND – does not include conclusion or contextual interpretation. To address this limitation, one potential solution is to upload redacted PDF study reports to the platform to enable machine learning-powered searches across datasets. While further exploration is needed, these approaches offer exciting potential to improve search precision and relevance, uncover hidden patterns across studies, support translational research by linking preclinical and clinical data, and drive innovation in toxicological analysis and decision-making.



#### 1.5 Assessing Correlations Between Adverse Events and Compound Vehicles

The goal of this analysis was to identify potential associations between study adverse events and the vehicles administered. The dataset included SEND-formatted data from 109 toxicology studies downloaded from DataCelerate and was limited to vehicle control animals to remove the variable of treatment-related effects.

For an initial review of vehicles used in these 109 studies, a summary table was created that lists the vehicles associated with each treatment ID and the specific study IDs associated with each treatment ID/vehicle combination. The summary began with treatment ID, assuming that studies for a given treatment ID were conducted by the same member company and that vehicle values would be most consistent across studies for the



same treatment ID. This approach was expected to produce a single summary row per treatment ID/vehicle with a list of studies in which that treatment/vehicle were tested.

However, significant variability in terminology and formats across vehicle values – for example, the use of abbreviations or spacing differences in vehicle names – led to multiple rows for many of the same treatment ID/vehicle combinations. After reviewing the vehicle values, the focus of the analysis shifted to standardizing what appears to be the same vehicle values with minor differences in terminology and/or formats.

To this end, an R-Shiny tool was used to parse vehicle values into separate vehicle components and to standardize vehicle terms based on the UNII display names for vehicle components. While the tool successfully parsed many vehicle values into separate components with standardized terminology, some vehicle values were parsed incorrectly due to formatting differences – for example, the use of “in” and “and” between components, or the use of parentheses to enclose vehicle components.

To improve the usability and value of vehicle data, member companies could:

- Use standardized UNII display names for vehicle components
- Apply SEND Controlled Terminology for units
- Provide concentrations of components when feasible, as recommended in the Houser & Sato poster

An incremental step to improve the consistency in SEND EXTRT vehicle values could be the development of industry guidelines for standard formatting (e.g., punctuation and sequence of vehicle components and associated properties, such as concentration and pH).

An orange outline of a right-pointing arrow icon is positioned to the left of the text.

Ultimately, the most impactful approach may be to add new SEND fields for populating more granular vehicle-related information, such as individual vehicle components, concentrations, and properties such as pH – the building blocks for the vehicle values currently populated in the EXTRT field. Values in these fields would be “pre-parsed” for direct use in data analyses or could be concatenated in a consistent manner to derive more standardized vehicle names.

## Conclusion

DataCelerate and the Toxicology Data Sharing initiative continue to demonstrate value as a collaborative space for sharing data to enhance preclinical decision-making. Since its inception, member companies have made meaningful strides in:

- Integrating historical data into internal databases for further utilization
- Applying visualization tools to uncover trends and insights
- Using the data to support study training and design

These applications not only improve operational efficiency but also contribute to more informed drug development practices.

Continued investment in study contribution efforts, data harmonization, analytics, and cross-company engagement will be key to unlocking the full potential of DataCelerate's TDS module. As the platform evolves and technology advances, DataCelerate holds promise for further reducing animal use and accelerating the development of safer, more effective therapies.



## Acknowledgement

The 2026 Toxicology Data Sharing Core Team wishes to thank the following companies who have contributed data to the DataCelerate platform, which facilitates the sharing of toxicology and background control data: Boehringer Ingelheim, Bristol-Myers Squibb, Eli Lilly, Novo Nordisk, Pfizer, Johnson & Johnson, and Shionogi. Neither TransCelerate Biopharma, BioCelerate, nor any of the above-referenced companies have contributed to, approved, or are in any way responsible for the authors' research results.

## References

BioCelerate. Toxicology Data Sharing Awareness Manuscript. October 2019. Available from: [BioCelerate-TDS-Awareness-Manuscript\\_October2019.pdf](#)

Houser W, Sato G. Recommendations for Exchanging Vehicle Details Using SENDIG v3.1. Presented at: ACT 2021 Conference, BioCelerate Virtual Poster Session; November 15, 2021. Available from: <https://www.transceleratebiopharmainc.com/events/act-2021-conference/>

U.S. FDA. UNII, Preferred Substance Names, and Synonyms. Available from: <https://precision.fda.gov/uniisearch>

National Cancer Institute EVS. SEND Controlled Terminology. Available from: <https://evs.nci.nih.gov/ftp1/CDISC/SEND/>

