Paper AD07

TLF Treasure Hunt: Making Legacy Data Instantly Discoverable with RAD Finder

Cielo Marie Castañeda, Pfizer Inc., Makati City, Philippines
Alexis Bernard Dalam, Pfizer Inc., Makati City, Philippines

ABSTRACT

In today's fast-paced regulatory and scientific landscapes, timely access to previously generated Tables, Listings, and Figures (TLFs) is vital for efficient analysis and data-driven decisions. As the volume of clinical research outputs grows exponentially - often spanning tens of thousands of TLFs across diverse programs and indications - the ability to rapidly locate relevant information becomes increasingly challenging. This creates a pressing need for intuitive, metadata-driven tools that can streamline data exploration and maximize the utility of existing assets.

This paper will demonstrate how RAD (Repository of Analyzed Data) Finder, a user-friendly RShiny application that empowers statistical programmers, statisticians, and clinical teams to search and retrieve TLFs through rich, metadata-driven search capabilities using intuitive filters and keyword-based queries. This will significantly reduce turnaround time, duplication of effort in locating TLFs through its user-centric design principles and support faster, more informed decision-making in clinical development.

INTRODUCTION

In modern clinical development, the pharmaceutical industry faces an unprecedented surge in data generation. Each clinical trial produces extensive datasets, including Tables, Listings, and Figures (TLFs) that summarize statistical analyses essential for regulatory submissions, clinical study reports (CSRs), and internal decision-making. These outputs are not merely technical artifacts; they serve as the foundation for interpreting trial results, demonstrating safety and efficacy, and ensuring compliance with global regulatory standards.

The scale of this data challenge is staggering: a single clinical trial now generates an average of 3.6 million data points from site visits, labs, procedures, patient-reported outcomes, and more. As organizations expand their research portfolios and conduct increasingly complex, multi-regional trials, the volume and diversity of TLFs grow exponentially. Traditionally, locating these outputs involves manual searches through shared drives or querying multiple disconnected systems. This process is time-consuming, prone to duplication, and often lacks transparency. The absence of centralized, metadata-driven search capabilities further compounds inefficiencies, delaying critical activities such as integrated summaries, cross-study analyses, and regulatory submissions. [forbes.com]

CDARS™ (Clinical Data Analysis and Reporting System) is a web-based interface that helps Pfizer® produce its clinical study reporting and other analysis and reporting artifacts. Data from studies based on CDISC or legacy Pfizer Data Standards (PDS) are presented as TLFs through the CDARS. CDARS resides on a LINUX server and is driven by a library of standard SAS macros. The CDARS interface produces a standard way of describing our CRF data, analysis data, and the final outputs. [Pfizer Programming Master Handbook™]

To address these challenges, RAD Finder was developed as an innovative solution within the CDARS ecosystem. RAD Finder™ short for "Repository of Analyzed Data" provides a centralized, intuitive platform for searching and retrieving TLFs, leveraging rich metadata to enable precise and efficient access. By consolidating outputs across studies and programs, RAD Finder eliminates silos, reduces redundancy, and accelerates workflows—ultimately supporting faster and more informed decision-making.

The strategic impact of RAD Finder extends beyond operational efficiency. In an industry where timelines are critical and regulatory scrutiny is intense, the ability to rapidly locate validated outputs can significantly influence submission readiness and compliance. Moreover, by enabling cross-study accessibility, RAD Finder facilitates integrated analyses that inform portfolio strategy, benefit-risk assessments, and evidence synthesis for health authorities and stakeholders. This paper will explore the design principles, key functionalities, and implementation of RAD Finder, highlighting its role in transforming TLF retrieval processes. We will also discuss its alignment with broader trends in digital transformation, data governance, and automation within clinical development. Through case examples and performance metrics, we aim to demonstrate how RAD Finder contributes to operational excellence and accelerates the path from data to decision.

BACKGROUND

Modern clinical trials involve numerous endpoints, subgroups, and statistical methodologies, resulting in thousands of TLFs per study. This complexity results in the generation of thousands of TLFs per study, which serve as critical components for interpreting trial outcomes and supporting downstream activities. These outputs are frequently reused for:

- Regulatory submissions (e.g., New Drug Applications [NDA], Biologics License Applications [BLA])
- Safety and efficacy summaries
- Cross-program analyses

Despite their importance, the absence of efficient retrieval mechanisms introduces significant operational challenges. Teams often risk duplicating work or overlooking critical insights, which can delay decision-making and regulatory timelines.

CDARS SYSTEM OVERVIEW

CDARS functions as the primary repository for finalized TLFs within many pharmaceutical organizations. While CDARS provides robust storage capabilities, it lacks advanced search functionality for end-users. Consequently, locating specific outputs often requires manual searches across shared drives or multiple systems—a process that is time-consuming, error-prone, and inefficient. Addressing this gap necessitates a solution that combines centralized access with metadata-driven search capabilities.

CDARS serves as the primary repository for finalized TLFs. While robust in storing outputs, CDARS lacks advanced search capabilities for end-users. RAD Finder complements CDARS by leveraging its metadata to enable dynamic filtering and keyword-based searches.

RAD FINDER OVERVIEW

To overcome these limitations, **RAD Finder App** was developed as an innovative, **Shiny-based application** designed to complement CDARS. RAD Finder enables:

- **Centralized access** to finalized TLFs stored in CDARS
- **Metadata-driven** searches using compound, submission, and protocol filters
- **Keyword and exclusion-based queries** for precise retrieval
- **Direct links** to TLF PDFs and CDARS program files

The application emphasizes **usability, speed, and accuracy**, ensuring that users can locate relevant outputs with minimal effort and maximum efficiency.

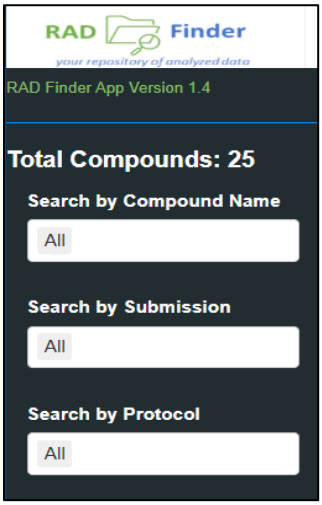
FEATURES

- **Metadata Filters**

Users can filter by:

- Compound Name
- Submission
- Protocol

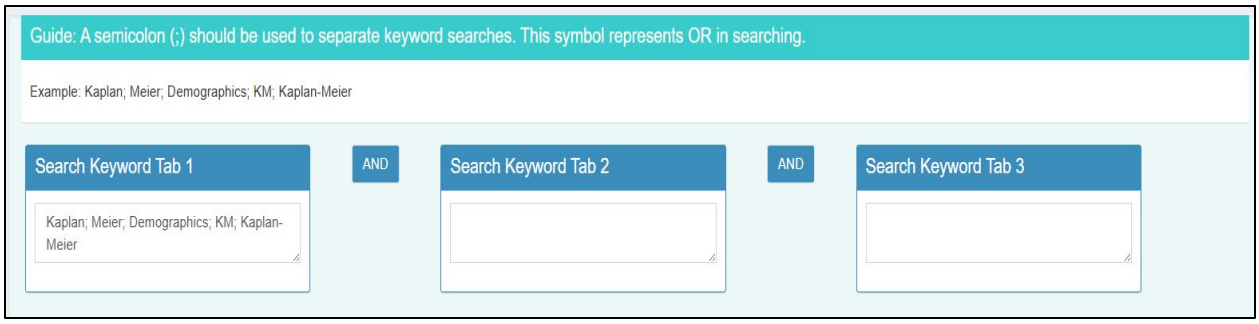
Selections can be made via dropdown menus or manual input, with an option to select "*All*" for broader searches.



- **Keyword Search**

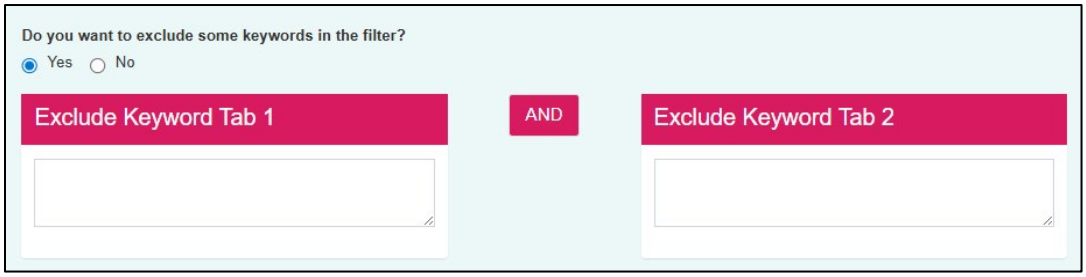
- Supports multiple keywords separated by semicolons (;), representing *OR* logic.
- Combines all keyword tabs using *AND* logic for refined results.

Example: *Kaplan; Meier; Demographics; KM; Kaplan-Meier*



- **Exclusion Tabs**

Users can exclude unwanted keywords to narrow results further.

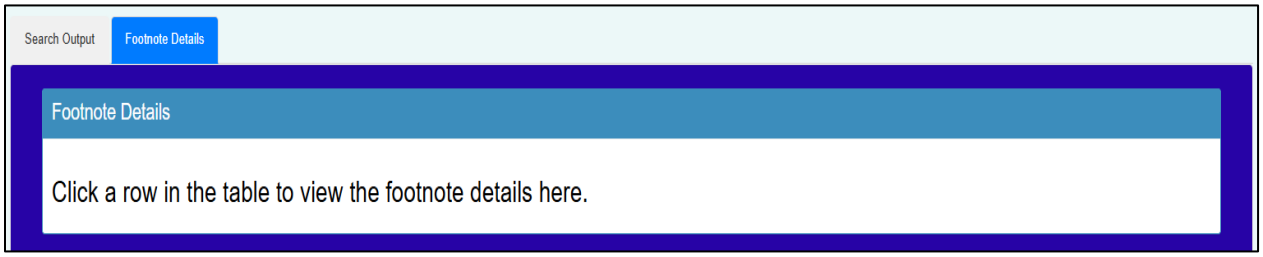


- **Run Date Filtering**

A toggle allows filtering by run dates, with a date range selector for precise control.

- **Footnotes Search**

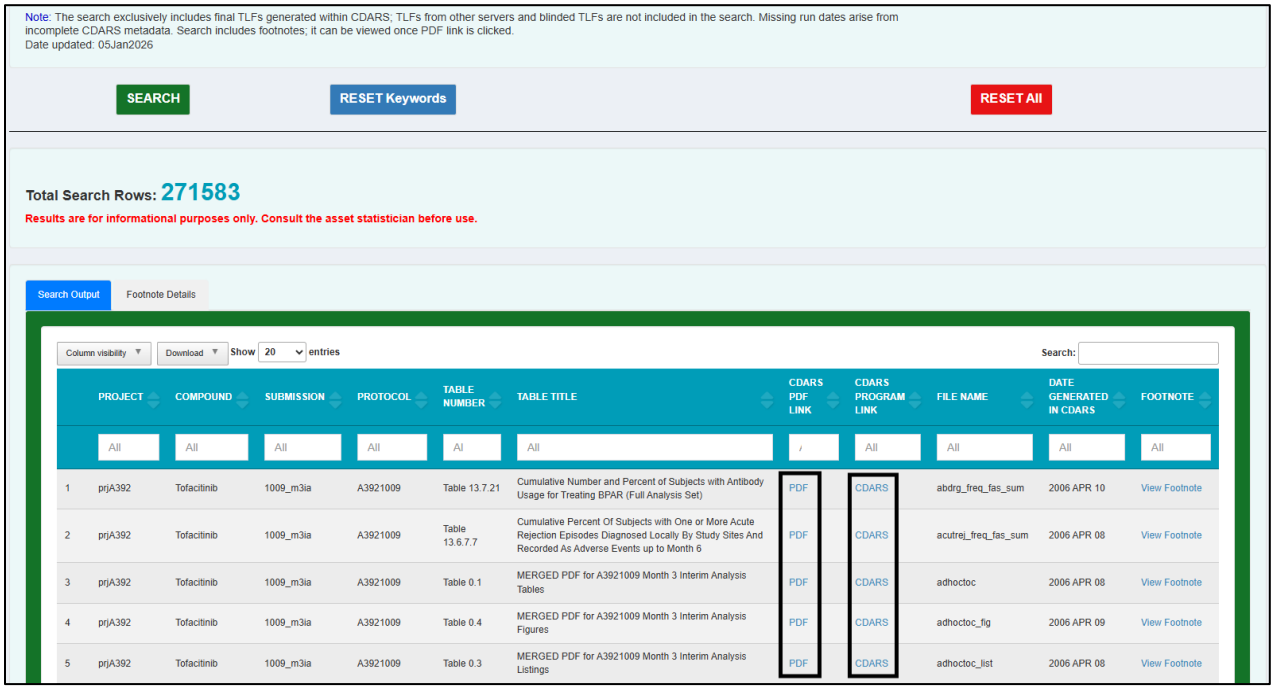
Introduced in version 1.3 and enhanced in 1.4, footnotes can now be searched and viewed in the Details tab.



- **Direct Access Links**

PDF Link: Opens the PDF file of the TLF.
CDARS Link: Redirects to the CDARS TOT file where user can access the parameters and derivations.

Sample Output:



USER WORKFLOW

1. **Launch the App:** Access via RAD Finder Portal.
2. **Select Filters:** Choose compound, submission, and protocol.
3. **Enter Keywords:** Use semicolons for multiple terms.
4. **Apply Exclusions (Optional):** Add keywords to exclude irrelevant outputs.
5. **Set Date Range (Optional):** Toggle run date filter if needed.
6. **Search and Review:** Click SEARCH to generate results.
7. **Access Outputs:** Use PDF or CDARS links for detailed review.
8. **Reset Keywords/All Filters (Optional):** Option to reset all filters or reset keywords.

BENEFITS AND IMPACT

RAD Finder significantly improves operational efficiency by reducing retrieval time from hours to minutes, minimizing redundant analyses, and enabling rapid access to historical data for informed decision-making. Its intuitive interface lowers the learning curve for non-technical users, supporting broader adoption across clinical teams.

LIMITATIONS AND CONSIDERATIONS

Several limitations should be acknowledged in the current implementation of RAD Finder. First, metadata completeness remains a critical dependency; missing run dates or incomplete metadata can adversely affect search accuracy and retrieval performance. Second, the outputs retrieved through RAD Finder are intended for informational purposes only. Users are advised to consult asset statisticians before making any data-driven decisions to ensure proper interpretation and compliance with regulatory standards. Finally, the scope of RAD Finder is currently restricted to finalized TLFs stored within CDARS. Blinded outputs and data housed in other repositories are not yet supported, which may limit applicability for certain workflows.

- **Metadata Completeness**: Missing run dates or incomplete metadata may affect search accuracy.
- **Interpretation Caution**: Outputs are for informational purposes; consult asset statisticians before making decisions.
- **Scope:** Currently limited to finalized TLFs in CDARS; blinded outputs and other servers are excluded.

FUTURE ENHANCEMENTS

Future development efforts will focus on expanding RAD Finder's capabilities to address current limitations and enhance its utility. Planned enhancements include:

- **Integration with Additional Repositories**: Extending beyond CDARS to incorporate other clinical data sources, enabling broader coverage and cross-system accessibility. It is also being explored to be extended to Pfizer's new reporting system – SIGMA™.

- **Advanced Analytics and Visualization**: Introducing interactive dashboards and trend analysis tools to support exploratory data review and portfolio-level insights.
- **AI-Powered Semantic Search:** Leveraging natural language processing (NLP) to enable context-aware queries and improve retrieval precision.
- **User Feedback-Driven Iteration:** Establishing a continuous improvement cycle informed by end-user feedback to refine functionality, usability, and performance.

CONCLUSION

The development and implementation of the RAD Finder application underscore the transformative potential of metadata-driven tools in clinical research and development. Traditional clinical data retrieval processes are often hindered by fragmented data sources, inconsistent metadata standards, and time-intensive manual searches. These inefficiencies not only delay decision-making but also increase the risk of data duplication and missed insights. RAD Finder addresses these challenges by centralizing access to diverse datasets and enabling intuitive, metadata-based search capabilities, thereby streamlining workflows and enhancing operational efficiency.

As the volume and complexity of clinical data continue to grow exponentially, the need for robust, scalable, and user-friendly retrieval systems becomes increasingly critical. RAD Finder exemplifies how leveraging metadata can bridge the gap between raw data and actionable insights, supporting evidence-based decision-making across clinical development stages. Furthermore, its design principles—emphasizing interoperability, usability, and real-time accessibility align with broader industry trends toward digital transformation and data democratization.

Looking ahead, tools like RAD Finder will play a pivotal role in shaping the future of clinical research. By reducing redundancies, improving data discoverability, and fostering collaboration across multidisciplinary teams, such platforms contribute to accelerating innovation and improving patient outcomes. Continued investment in metadata-driven technologies, coupled with advancements in artificial intelligence and machine learning, will further enhance the precision and predictive power of clinical data retrieval systems. Ultimately, RAD Finder represents not only a solution to current bottlenecks but also a foundation for next-generation data ecosystems that prioritize efficiency, transparency, and scientific rigor.

REFERENCES

- https://www.forbes.com/councils/forbestechcouncil/2025/06/25/increasing-productivity-in-pharma-rd-starts-with-data/
- **Shiny for R (Posit) -** https://shiny.posit.co/
- **htmlwidgets for R -** https://www.htmlwidgets.org/
- **Mastering Shiny -** https://mastering-shiny.org/index.html
- **Programming Master Handbook –** Overview of CDARS (2.1)

ACKNOWLEDGMENTS

- **Margaret Gamalo:** Vice President, Head I&I Statistics, Pfizer
- **Gary Chen:** Programming TA Lead - Executive Director, Pfizer
- **Kenneth Kwok:** Senior Director, I&I Statistics Group Lead, Pfizer
- **Darrell Edgley:** Director Stats Program, SDSA Pfizer
- **Clifford Aspinall:** Director Stats Program, SDSA Pfizer
- **Christian Russel Reyes:** Head of DSA Philippines
- **Mary Jane Cadatal:** Senior Manager, Biostatistics, Pfizer
- **Shixue (Melody) Liu:** Associate Director, Statistics, Pfizer

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)
Your comments and questions are valued and encouraged. Contact the author at:

Author Name: **Cielo Marie A. Castañeda**
Company: **Pfizer, Inc.**
Address: **8 Rockwell, Makati City, Philippines**
Work Phone: **+63 9190764251**
Email: **cielomarie.castaneda@pfizer.com**
Website: **https://www.linkedin.com/in/cielomariecastaneda/**