

Transforming R&D Data (Tables, Listings, and Figures) for a Biotech Company

Clarkston recently partnered with an emerging and rapidly growing biotech company that had been experiencing challenges in quickly accessing key data from their Tables, Listings, and Figures (TLFs) as part of their exploratory research goals.

Teams across Research & Development (R&D) struggled to quickly obtain the data they needed from the TLF documents, often having to go through hundreds of pages, copy-pasting data to glean simple, but impactful, insights about the data.

The client partnered with Clarkston to design and develop an app solution that presented the TLF data in an aggregated view. This enabled R&D users to better understand and interpret critical trial data, helping to advance business goals.

Life Sciences Case Study

PROJECT OVERVIEW

INDUSTRY:



Biotech

PRODUCTS AND SERVICES:



Gene therapies and
RNA-based therapies

PRIMARY OBJECTIVES:

- Understand R&D data needs and challenges and design a proof-of-concept plan to execute against the identified pain points
- Design and develop an application for R&D teams to efficiently visualize critical data from TLFs with ease
- Establish a scalable Python framework to ingest, transform, and surface data in TLF format
- Explore the integration of large language models (LLMs) within the application to enhance capabilities, such as ingesting new studies more effectively or providing instant responses to user queries

RESOLUTION:

- Defined and analyzed challenges related to accessing and visualizing critical TLF data
- Developed and documented detailed user requirements, tailoring the application to the needs of R&D teams in Translational Biology
- Designed a modular architecture with assay-specific views and structured around fixed tabs—Blind ID, Block ID, Subject ID, and Explore—each presenting harmonized outputs from TLFs, enabling users to pivot seamlessly between different levels of patient data aggregation across multiple biomarker assays (e.g., *IHC*, *Western blot*, *ddPCR*)
- Created interactive data filtering and visualization with built-in search and charting tools to allow users to dynamically filter by subject, blind ID, or block, and generate exploratory plots (*line*, *bar*, *scatter*) directly from harmonized TLF outputs, supporting hypothesis generation and signal detection
- Enabled a secure and scalable access model, designed with secure deployment in mind (e.g., *via Posit Server or EC2*), ensuring compliant handling of clinical study data while providing extensibility for future studies and assay types through dictionary-based data ingestion and dynamic transformation pipelines
- Incorporated key study metadata and documentation into an LLM, crafted prompts, and recorded responses to evaluate the potential of LLMs in enhancing the process

KEY BENEFITS:

- Saved significant time for R&D teams by enabling easy access to key TLF data, eliminating the need to manually parse through .rtf files
- Eliminated manual copying and pasting of data tables, reducing risk of inaccuracies due to manual mistakes
- Facilitated more exploratory research, providing deeper insights into the comprehensive patient profile
- Freed R&D teams to focus on higher-value, strategic initiatives



KPIs

- Expedited a process that could take any user days (between requests, execution, and/or manual labor) into minutes
- Developed data governance processes and steps that can be leveraged when developing future proof-of-concept (POC) ideas
- 50+ tables ingested per study and 100+ questions asked of LLM during testing

CONTACT US

