

PHUSE 2025

Spot it – A Shiny Application for Visual QC at Scale



Authors:

Pravin Lakkaraju - Servier Pharmaceuticals

Samiar Ashtiany - Phastar

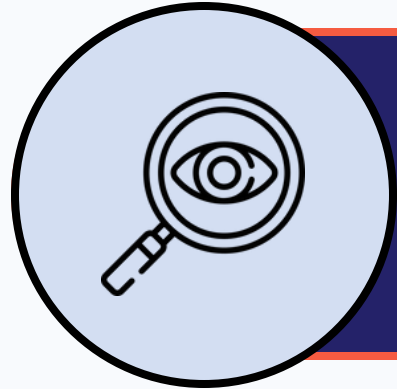
SERVIER 
moved by you

Agenda

1. Introduction
2. Challenges in Oversight QC
3. Spot-It at a Glance
4. Performance and Metrics
5. Workflow Overview
6. Key Modules
7. Future Enhancements
8. Summary
9. Q&A



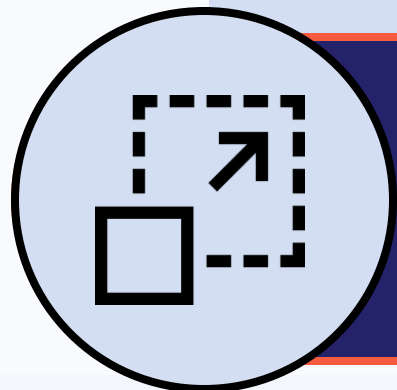
Challenges in Visual Oversight QC



Visual QC: High labor effort required for reviewing outputs.



Regulatory Compliance: Precision required to meet strict standards.



Scalability: Increasing document volumes challenge manual QC.

Spot it – At a Glance

- ▶ **Primary Stakeholder:** Statistical Programming.
- ▶ **Why?:** To address the common QC challenges by reducing manual effort and improve accuracy.
- ▶ **What does it do:**
 1. Compare large number of RTF files: A 1-1 comparison of RTF files. Differences are reported in an Excel.
 2. Extracts and compare: Titles and footnotes from Table Shell documents are extracted and automatically compared against analysis outputs.

Performance and Metrics

Visual QC of Outputs for a Standard CSR

3 days

Time without App

1-3 minutes

Time using App

**\$196K/
year
per study**

Cost Savings

Extract and compare titles/footnotes for standard CSR

4-5 days

Time without App

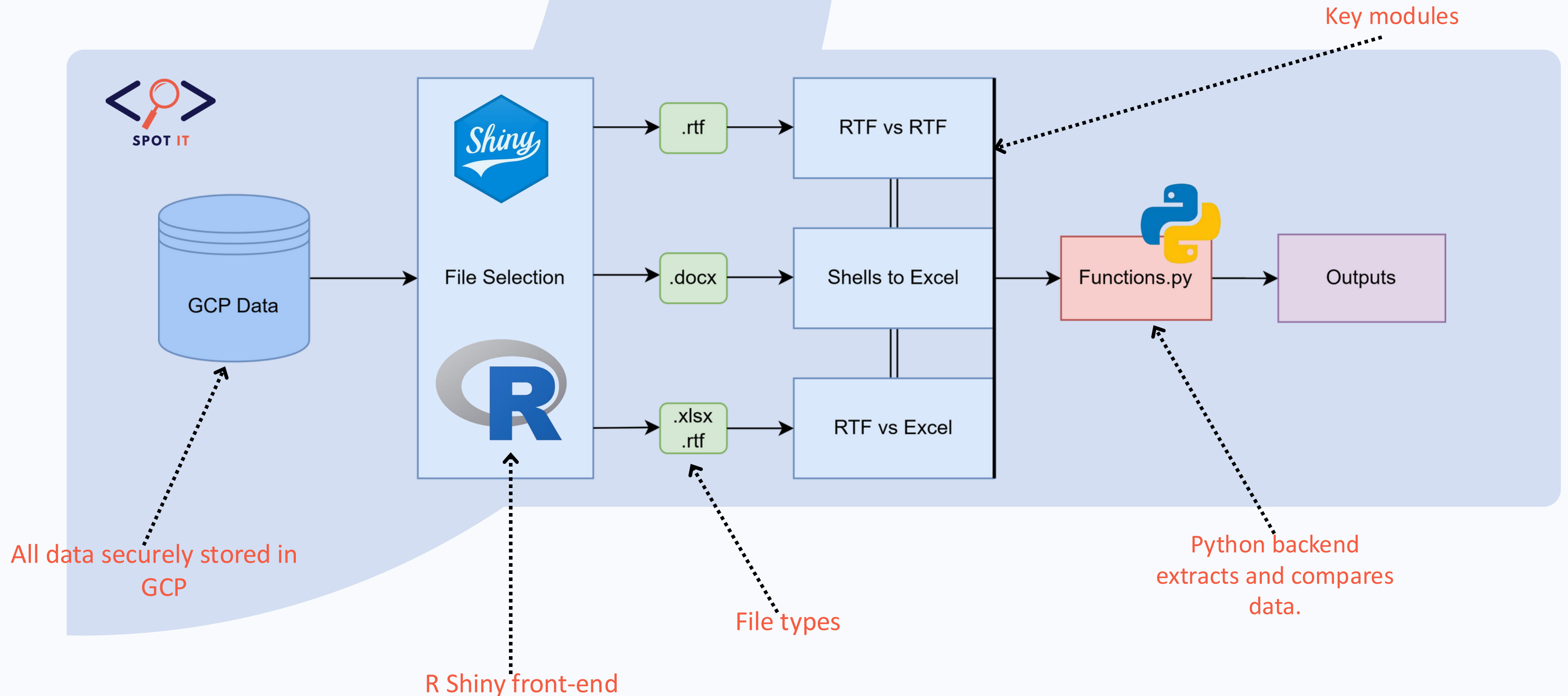
3-5 minutes

Time using App

**\$196K/
year
per study**

Cost Savings

Workflow Overview – Hybrid Programming



Workflow Overview – Hybrid Programming

Why Python Backend Instead of R

Challenges with R:

- Limited document processing capabilities for DOCX, RTF formats. Packages in Python are far superior.
- Performance bottlenecks with complex data pipelines. Especially since Spot-It handles large files and many at a time.

Python Advantages:

- Superior document processing libraries (docx2txt, RTF processors).
- Better memory management for large documents. Example: When Comparing a hundred different RTF files at a time.
- More flexible data manipulation through Pandas (Pandas is essentially Python's equivalent to R's data frames, but with powerful features for data analysis).

Module 1: RTF vs. RTF Comparison

SERVIER

RTF vs RTF

Comparison

Report

Shells to Excel

RTF vs Excel

User Guide



RTF Comparisons

Select RTF pair

T1-1.rtf vs T1-1b.rtf

Difference in line 12:
T1-1.rtf: n (%)|Col4
T1-1b.rtf: n ()|Col4

Difference in line 17:
T1-1.rtf: On treatment|13 (7.7)|16 (83.4)|29 (7.6)| 0 |
T1-1b.rtf: On treatment|14 (7.7)|18 (83.4)|29 (7.6)| 0 |

Difference in line 18:
T1-1.rtf: Discontinued treatment| 5 (9.3)| 5 (16.6)| 76 (2.4)|12 (100) |
T1-1b.rtf: Discontinued treatment| 5 (9.3)| 5 (16.6)| 76 (2.4)|12 (100) |

Difference in line 20:
T1-1.rtf: Adverse event| 1 (57) | 3 (2.0) | 1 (4.0) | 0 (1.2) |
T1-1b.rtf: Adverse event| 1 (57) | 3 (2.0) | 1 (4.0) | 0 (1.2) |

Difference in line 39:
T1-1.rtf: Data cutoff date: 2024-07-31|
T1-1b.rtf: Data cutoff date: 2024-07-30

Difference in line 40:
T1-1.rtf: Summary of Subject Disposition
T1-1b.rtf: Overview and Analysis of Final Disposition Outcomes

Extra line in T1-1b.rtf: Data cutoff date: 2024-07-31|

1) Select RTF's 2) Select RTF's Generate

Set1	Set2	Difference	Extra Lines	Difference_1	Difference_2
T1-1.rtf	T1-1b.rtf	Yes	1 lines in total	Set1: n (%) Col4 Set2: n () Col4	Set1: On treatment 13 (7.7) 16 (83.4) 29 (7.6) 0 Set2: On treatment 14 (7.7) 18 (83.4) 29 (7.6) 0
T1-2.rtf	T1-2b.rtf	Yes	None	Set1: On treatment 13 (7.7) 16 (83.4) 29 (7.6) 0 Set2: On treatment 13 (7.7) 16 (83.4) 29 (7.6) 0	Set1: Adverse event 18 (57) 3 (2.0) 1 (4.0) 0 (1.2) Set2: Adverse event 18 (57) 1 (2.0) 1 (4.0) 0 (1.2)
T1-3.rtf	T1-3b.rtf	No	None		
T1-4.rtf	T1-4b.rtf	Yes	None	Set1: On treatment 13 (7.7) 16 (83.4) 29 (7.6) 0 Set2: On treatment 13 (7.7) 16 (83.4) 29 (7.6) 0	Set1: Discontinued treatment 5 (9.3) 65 (1.6) 76 (2.4) 12 (100) Set2: Discontinued treatment 5 (9.3) 65 (1.6) 76 (2.4) 12 (100.0)

Compare RTF Files for Text, Formatting, and Structural Discrepancies

Uses a sophisticated color -coding system to highlight discrepancies

Generates Detailed Excel Reports Highlighting Discrepancies

Module 2: Shells to Excel

SERVIER

RTF vs RTF

Shells to Excel

Report

RTF vs Excel

User Guide

SPOT IT

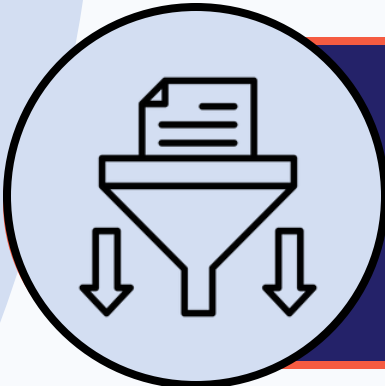
Title and Footnotes Table

Search:

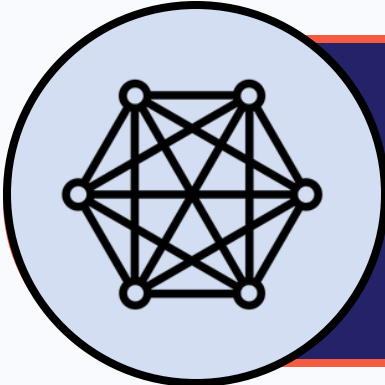
Type	Table No	Title	Population Flag	foot1
Table	1.1	Summary of Subject Disposition	Analysis Set	Completed cases constitute the large
Table	1.2	Summary of Subject Disposition 2	Analysis Set	Completed cases constitute the large
Table	1.3	Summary of Subject Disposition 3	Analysis Set	Completed cases constitute the large
Table	1.4	Summary of Subject Disposition 4	Analysis Set	Completed cases constitute the large

Showing 1 to 4 of 4 entries

Select shells .docGenerateDownload



Extracts Titles and Footnotes for all TLF's in a Table Shells Document



Uses complex algorithms to account for subtle differences in document format



Generates Detailed Excel Report

Type	Table No	Title	Population Flag	foot1	foot2
Table	1.1	Summary of Subject Disposition	Analysis Set	Completed cases constitute the largest category, ac	[1] Includes all cases that met the inclusion criter
Table	1.2	Summary of Subject Disposition 2	Analysis Set	Completed cases constitute the largest category, ac	[1] Includes all cases that met the inclusion criter
Table	1.3	Summary of Subject Disposition 3	Analysis Set	Completed cases constitute the largest category, ac	[1] Includes all cases that met the inclusion criter
Table	1.4	Summary of Subject Disposition 4	Analysis Set	Completed cases constitute the largest category, ac	[1] Includes all cases that met the inclusion criter

Module 3: RTF vs Excel

SERVIER

RTF vs RTF

Shells to Excel

RTF vs Excel

Report

User Guide

< >

SPOT IT

Title and Footnote Comparisons Table

Search:

Type	Table No	Title	Title_rtf
Table	1.1	Summary of Subject Disposition	Overview and Analysis of Final Disposition Outcom
Table	1.2	Summary of Subject Disposition 2	Summary of Subject Disposition 2
Table	1.3	Summary of Subject Disposition 3	Summary of Subject Disposition 3
Table	1.4	Summary of Subject Disposition 4	Summary of Subject Disposition 4

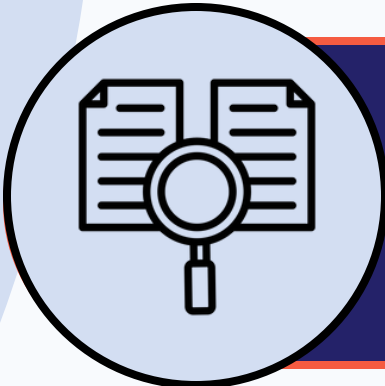
Showing 1 to 4 of 4 entries

1) Select Excel

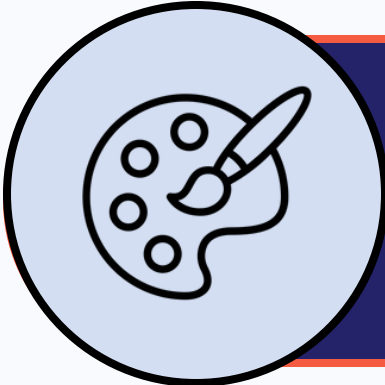
2) Select rtf's

Generate

Download



Compares titles and footnotes from RTF files with Excel output from table shells



Highlights discrepancies in text using colour coding system



Generates Detailed Excel Report

Type	Table No	Population Flag	Title	Title_rtf	Difference in title	Difference in footnotes	foot1
Table	1.1	Analysis Set	Summary of Subject Disposition	Overview and Analysis of Final Disposition Outcomes	Difference found	foot1; foot8; foot9	Completed cases constitute the largest category,
Table	1.2	Analysis Set	Summary of Subject Disposition 2	Summary of Subject Disposition 2	No difference found	foot1; foot4; foot7; foot8; foot9	Completed cases constitute the largest category,
Table	1.3	Analysis Set	Summary of Subject Disposition 3	Summary of Subject Disposition 3	No difference found	foot8; foot9	Completed cases constitute the largest category,
Table	1.4	Analysis Set	Summary of Subject Disposition 4	Summary of Subject Disposition 4	No difference found	foot1; foot8; foot9	Completed cases constitute the largest category,

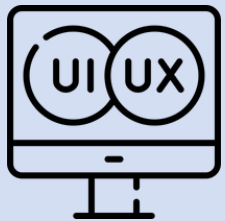
Future Enhancements



AI-powered parsing for improved accuracy in text extraction.



Expanded file support for additional document formats.



UI shift when implementing AI: Streamlit/Java more compatible with Python

Summary

Spot-It automates oversight QC processes with precision and scalability.

②

Significant performance improvements in efficiency and error reduction.

④

Key capabilities: RTF comparison, text extraction, and report generation.

③

Future-ready with planned AI enhancements.

①

Meet the Team



PRAVIN LAKKARAJU

Director - Statistical Programming

Servier Pharmaceuticals



SAMIAR ASHTIANY

Sr. Data Scientist

PHASTAR

Questions?