phuse·connect
connect·share·advance

APAC 2026

19–21 February

Novotel Hyderabad International Convention Centre

REIMAGINING CLINICAL DATA EXPLORATION: ANIMATED DATA VISUALIZATIONS FOR INTERACTIVE CLINICAL TRIAL STORYTELLING

Tony Mathew
Email:tony.mathew@pfizer.com
Contact:9496883479
Pfizer, India

DISCLAIMER

I confirm that, the opinions and thoughts discussed in this presentation are subject to my own independent views and are not subject to the opinions of the organization that I represent.

AGENDA

- Introduction
- Overview
- Methodology
- User Interface Walkthrough
- Key Features
- Benefits and Applications
- Future Enhancements
- Summary & Takeaways

INTRODUCTION

- The exploration of clinical trial data has historically depended on fixed tables, listings, and figures, limiting the ability to intuitively understand how datasets change across study milestones

- Recent advances in interactive and animated data visualization provide an opportunity to transform how clinical data is reviewed and conveyed through dynamic and engaging visual stories.

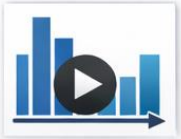
- This tool introduces an animated, interactive dashboard for exploring the progression of clinical datasets across different versions, allowing us to observe changes in data over time.

OVERVIEW

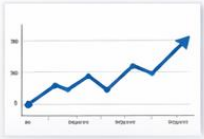
- Reviewing multiple dataset versions can be time-consuming and complex.
- We developed an **interactive Streamlit dashboard** for **animated exploration** of clinical trial data changes.

- *Enhance QC Reviews*
- *Visualize Dataset Changes*

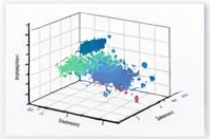
Key Features of the Dashboard



Animated Variable *Distributions*



Metric Trends & Missing Data



PCA Projections

Change Summaries

Metadata Compliance

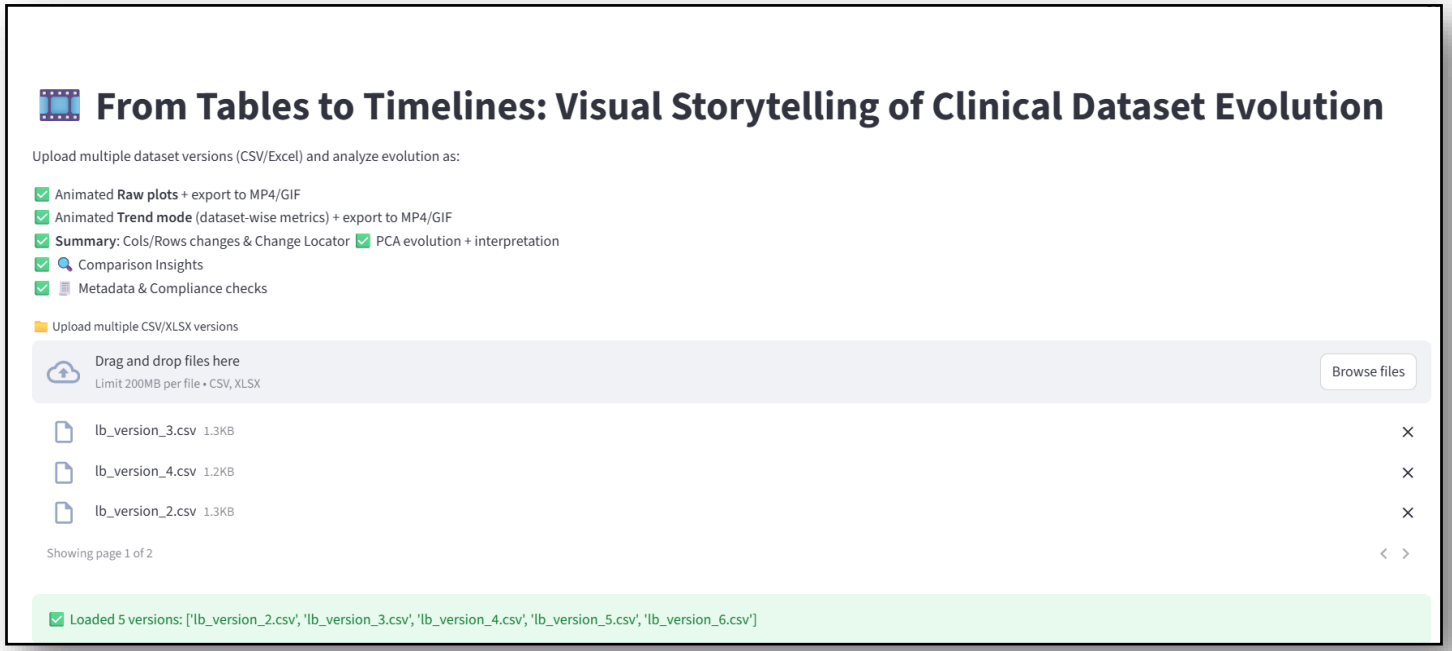
METHODOLOGY

An interactive dashboard was created utilizing a modular framework to facilitate the visualization of the evolution of datasets across various versions.

The system integrates:

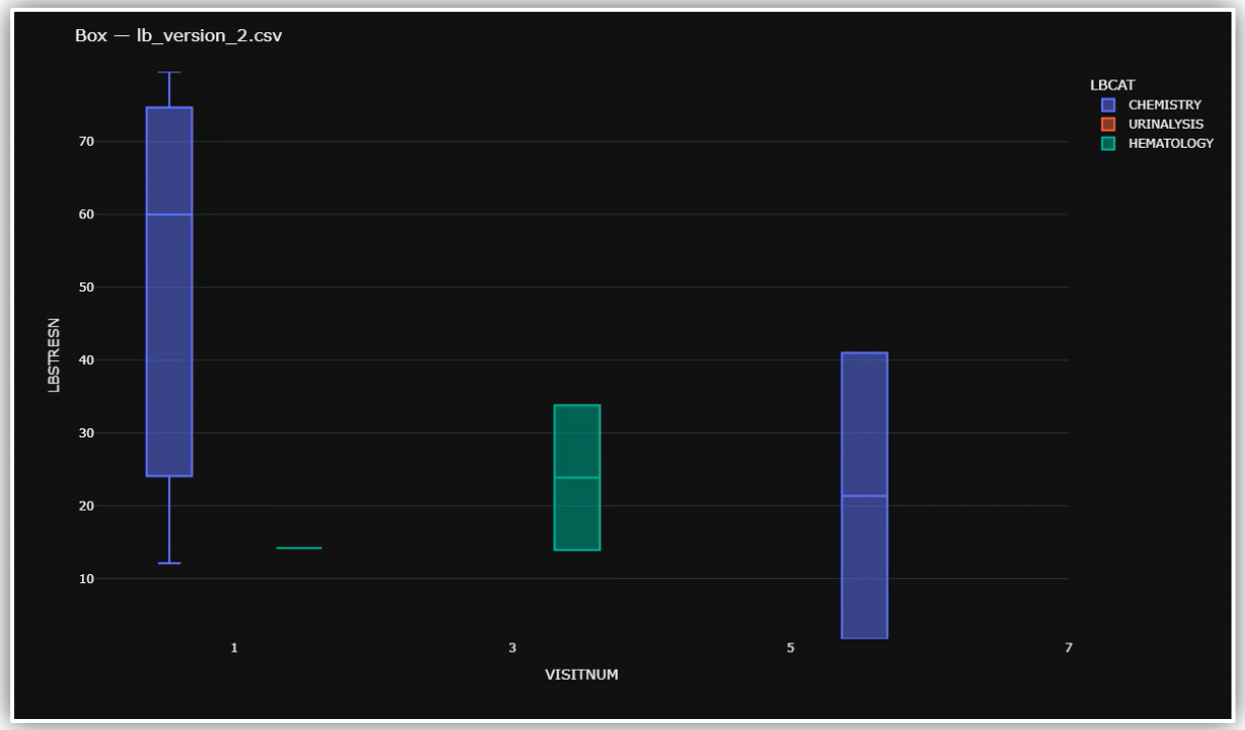
- **Streamlit** for the interactive user interface
- **Plotly** for animated and interactive visualizations
- **Scikit-learn** for multivariate analysis, including Principal Component Analysis (PCA)
- **Image export utilities** to generate GIF and MP4 animations for offline use

USER INTERFACE

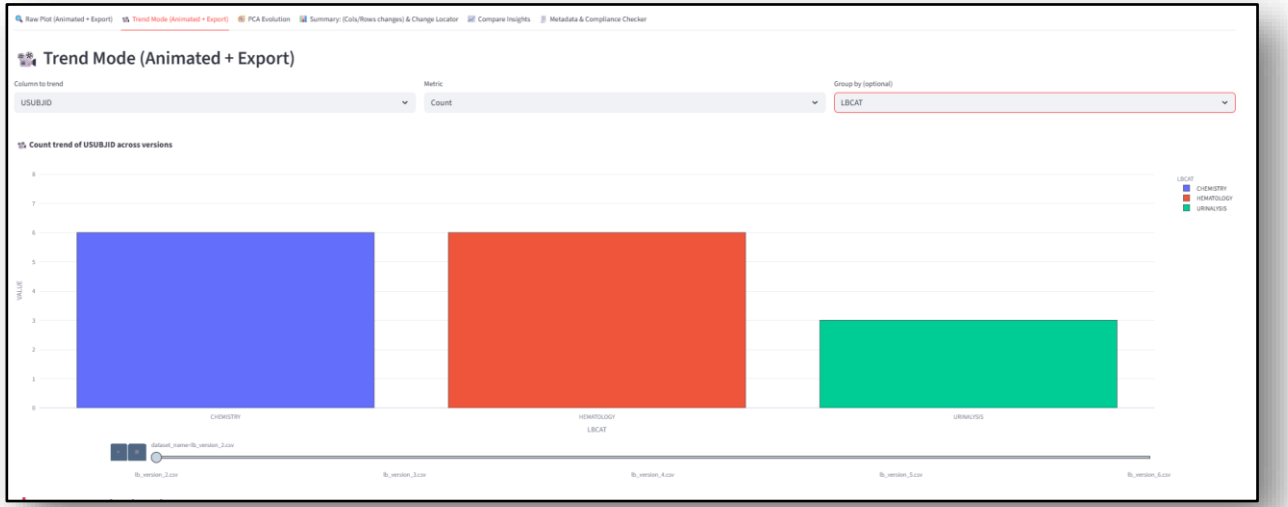


KEY FEATURES

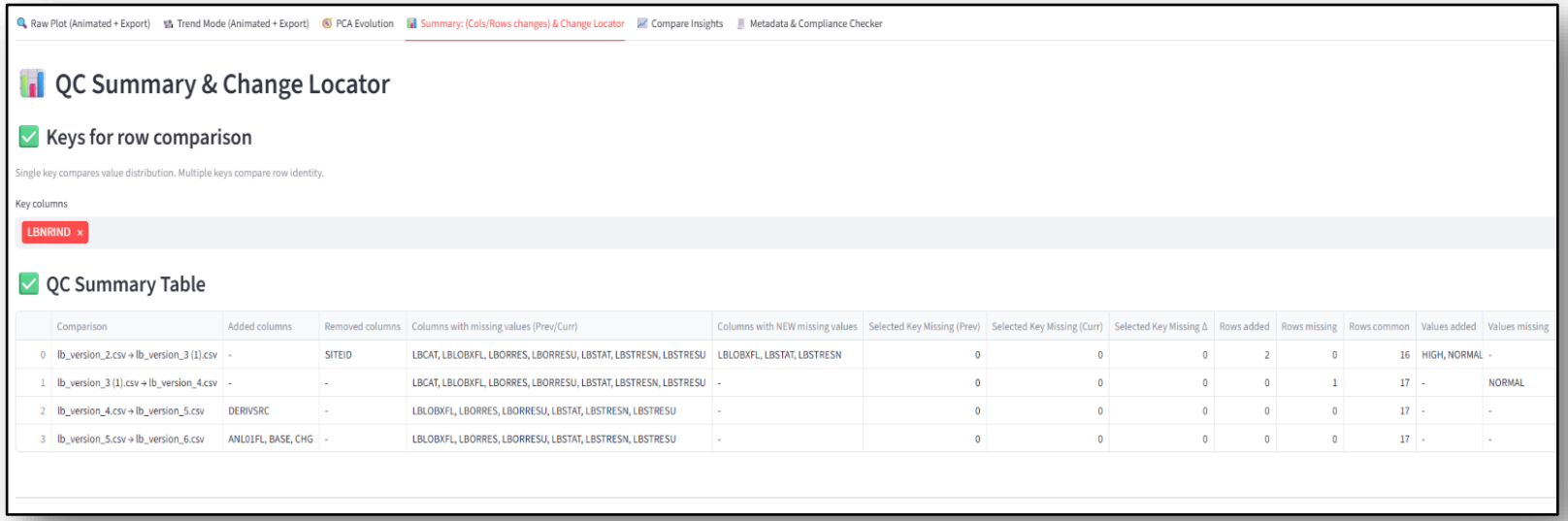
🔍 Raw Plot (Animated + Export)



Trend Mode (Animated + Export)

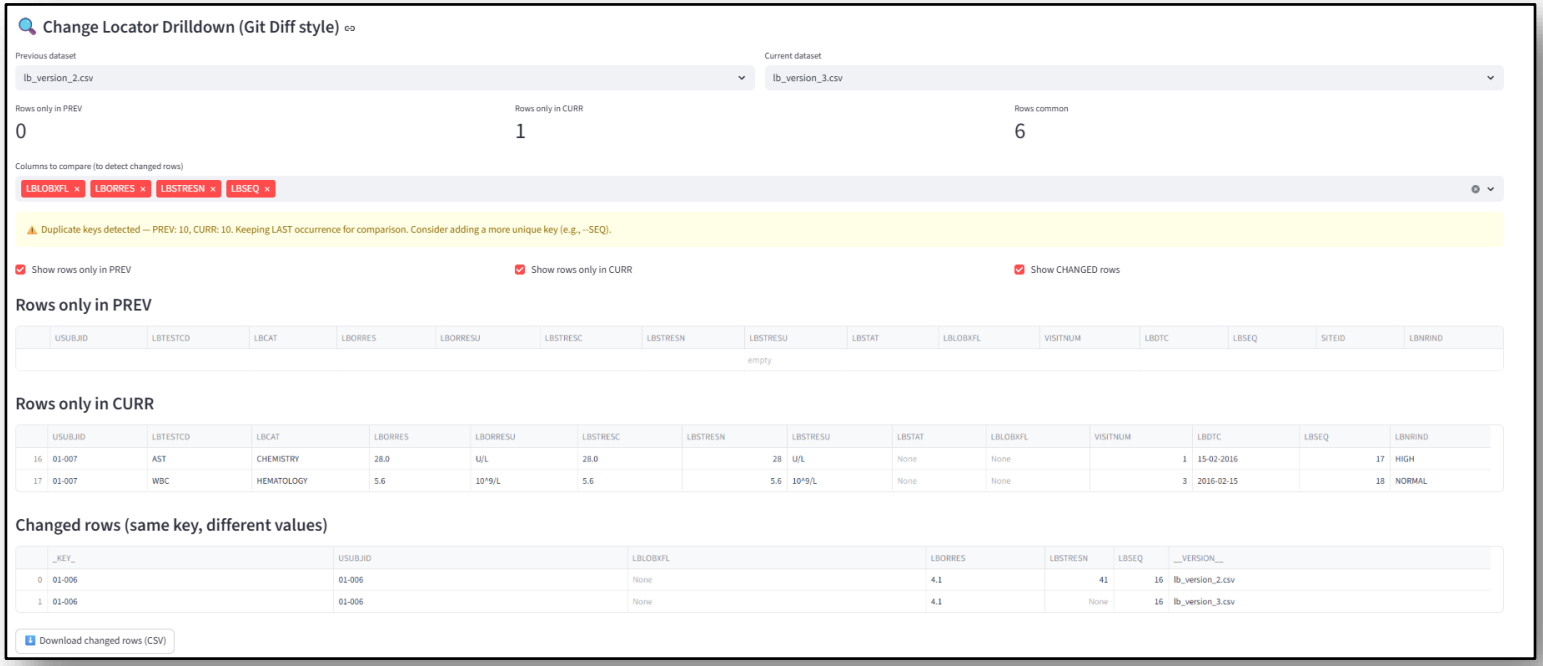


QC Summary & Change Locator



	Comparison	Added columns	Removed columns	Columns with missing values (Prev/Curr)	Columns with NEW missing values	Selected Key Missing (Prev)	Selected Key Missing (Curr)	Selected Key Missing Δ	Rows added	Rows missing	Rows common	Values added	Values missing
0	lb_version_2.csv → lb_version_3 (1).csv	-	SITEID	LBCAT, LBLOBXFL, LBORRES, LBORRESU, LBSTAT, LBSTRESN, LBSTRESU	LBLOBXFL, LBSTAT, LBSTRESN	0	0	0	2	0	16	HIGH, NORMAL	-
1	lb_version_3 (1).csv → lb_version_4.csv	-	-	LBCAT, LBLOBXFL, LBORRES, LBORRESU, LBSTAT, LBSTRESN, LBSTRESU	-	0	0	0	0	1	17	-	NORMAL
2	lb_version_4.csv → lb_version_5.csv	DERIVSRC	-	LBLOBXFL, LBORRES, LBORRESU, LBSTAT, LBSTRESN, LBSTRESU	-	0	0	0	0	0	17	-	-
3	lb_version_5.csv → lb_version_6.csv	ANL01FL, BASE, CHG	-	LBLOBXFL, LBORRES, LBORRESU, LBSTAT, LBSTRESN, LBSTRESU	-	0	0	0	0	0	17	-	-

QC Summary & Change Locator



Change Locator Drilldown (Git Diff style)

Previous dataset: lb_version_2.csv

Current dataset: lb_version_3.csv

Rows only in PREV: 0

Rows only in CURR: 1

Rows common: 6

Columns to compare (to detect changed rows): LBLOBXFL × LBORRES × LBSTRESN × LBSEQ ×

⚠ Duplicate keys detected — PREV: 10, CURR: 10. Keeping LAST occurrence for comparison. Consider adding a more unique key (e.g., -SEQ).

Show rows only in PREV ☑ Show rows only in CURR ☑ Show CHANGED rows ☑

Rows only in PREV

	USUBJID	LBTESTCD	LBCAT	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSTAT	LBLOBXFL	VISITNUM	LBDTC	LBSEQ	SITEID	LBNRIND
	empty														

Rows only in CURR

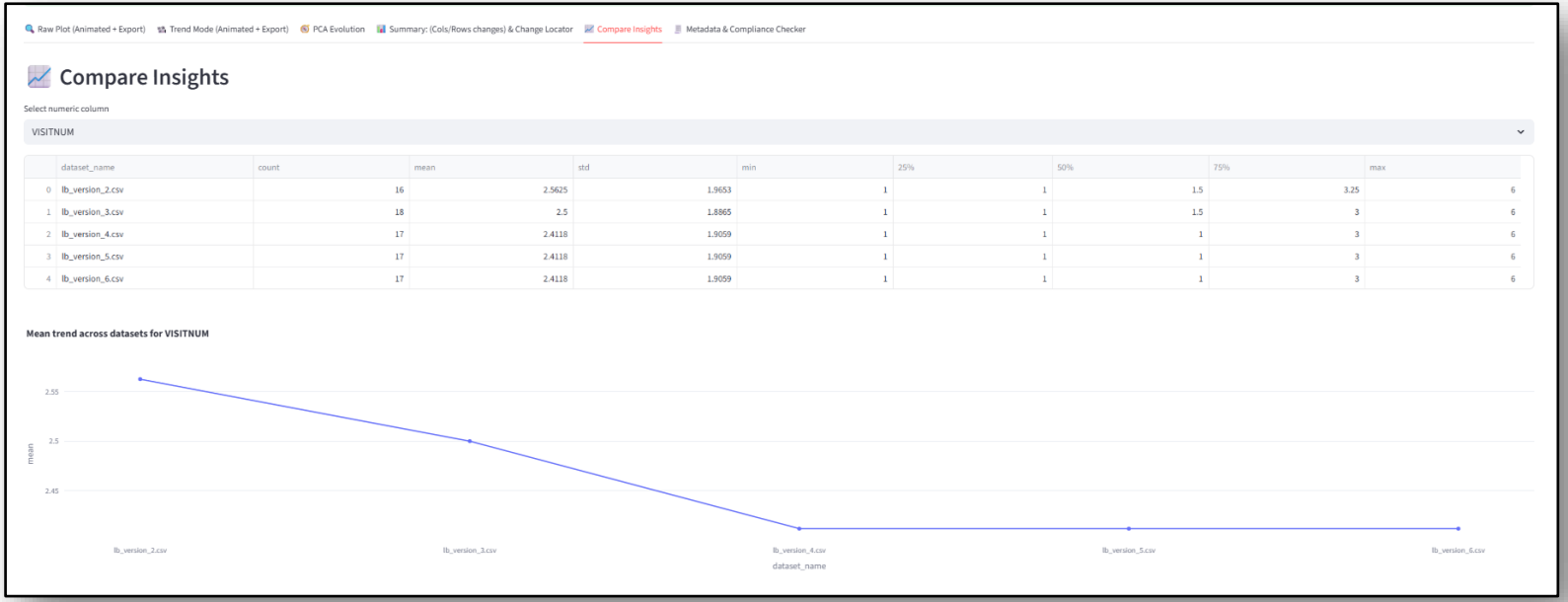
	USUBJID	LBTESTCD	LBCAT	LBORRES	LBORRESU	LBSTRESC	LBSTRESN	LBSTRESU	LBSTAT	LBLOBXFL	VISITNUM	LBDTC	LBSEQ	LBNRIND
16	01-007	AST	CHEMISTRY	28.0	U/L	28.0	28	U/L	None	None	1	15-02-2016	17	HIGH
17	01-007	WBC	HEMATOLOGY	5.6	10^9/L	5.6	5.6	10^9/L	None	None	3	2016-02-15	18	NORMAL

Changed rows (same key, different values)

	KEY	USUBJID	LBLOBXFL	LBORRES	LBSTRESN	LBSEQ	__VERSION__
0	01-006	01-006	None	4.1	41	16	lb_version_2.csv
1	01-006	01-006	None	4.1	None	16	lb_version_3.csv

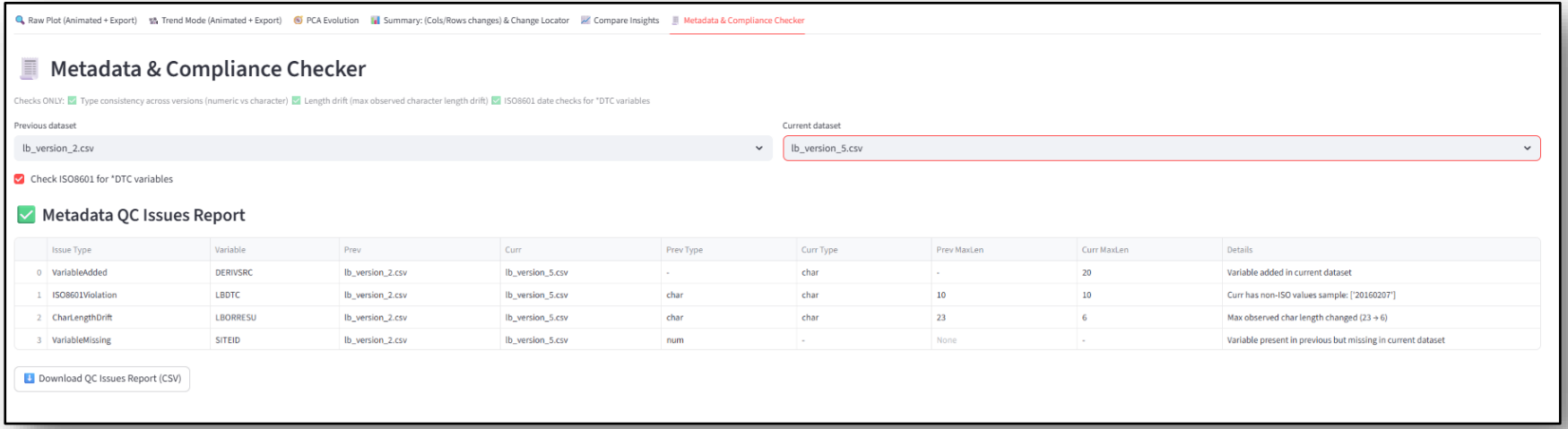
Download changed rows (CSV)

Compare Insights

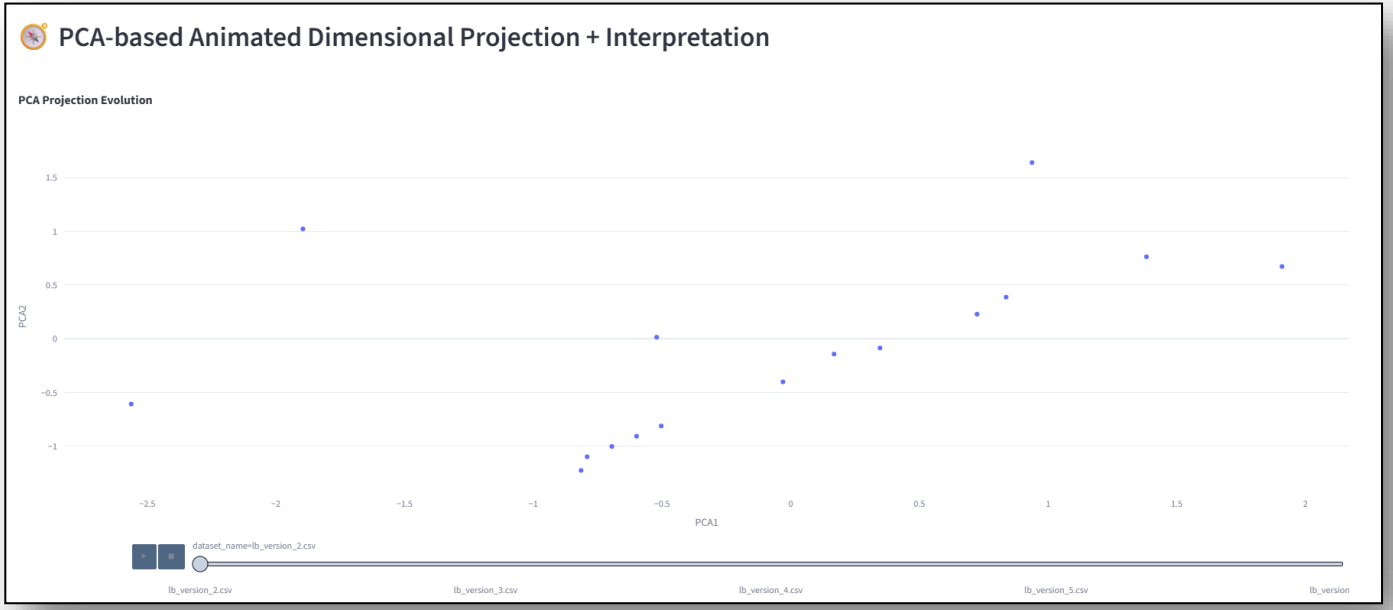


	dataset_name	count	mean	std	min	25%	50%	75%	max
0	lb_version_2.csv	16	2.5625	1.9653	1	1	1.5	3.25	6
1	lb_version_3.csv	18	2.5	1.8865	1	1	1.5	3	6
2	lb_version_4.csv	17	2.4118	1.9059	1	1	1	3	6
3	lb_version_5.csv	17	2.4118	1.9059	1	1	1	3	6
4	lb_version_6.csv	17	2.4118	1.9059	1	1	1	3	6

Metadata & Compliance Checker



PCA- based Animated Dimensional Projection



PCA- based Animated Dimensional Projection

📌 Dataset-level movement in PCA space

	dataset_name	PCA1	PCA2	shift
0	lb_version_2.csv	-0.1315	-0.0969	None
1	lb_version_3.csv	0.1477	-0.1017	0.2793
2	lb_version_4.csv	0.1138	-0.1211	0.0391
3	lb_version_5.csv	0.1138	-0.1211	0
4	lb_version_6.csv	-0.2604	0.441	0.6753

Movement summary

- Total centroid drift across versions: **0.994 PCA-units**
- Largest jump occurred at **lb_version_6.csv** with shift **0.675**

⚠️ Potential outliers (top 1% farthest in PCA space)

Outlier threshold distance: **5.185** | Outlier count: **1**

	dataset_name	USUBJID	PCA1	PCA2	PCA_dist
68	lb_version_6.csv	01-001	-5.7483	3.8909	6.9413

📝 Interpretation Summary

- The PCA projection summarizes how the **overall numeric structure** evolves across dataset versions.
- **PC1 + PC2 explain 58.15%** of numeric variation, so the plot is a compact representation of structural changes.
- **PC1 is mainly driven by**: `VISITNUM, LBSEQ, LBSTRESN, BASE, CHG`
- **PC2 is mainly driven by**: `BASE, LBSEQ, CHG, VISITNUM, LBSTRESN`
- Dataset versions show a **total centroid drift of 0.994 PCA units**, indicating progressive distribution/structure updates.
- The **largest structural change** occurred at **lb_version_6.csv** (shift = **0.675**).
- Top variables showing largest mean change at the biggest jump (**lb_version_5.csv → lb_version_6.csv**): `LBSTRESN, VISITNUM, LBSEQ, SITEID, BASE`
- Potential outliers detected: **1 records (top 1% farthest in PCA space)**. Key fields available for investigation: `USUBJID, LBTESTCD, VISITNUM, LBDTC`

BENEFITS AND APPLICATIONS

Statistical Programmers

Faster quality checks and early issue detection

Statisticians

Clear view of dataset readiness during analysis

Clinical Reviewers

Transparent tracking of data evolution

Study Teams

Shared visual references for team alignment

Practical Use Cases:

- Interim data review & ongoing study monitoring
- Database lock readiness assessments
- Vendor data transfer validation

FUTURE ENHANCEMENTS

Automated Variable Role Identification

- *Identify SDTM & ADaM variable roles*
- *Classify data components automatically*

AI-Generated Narrative Summaries

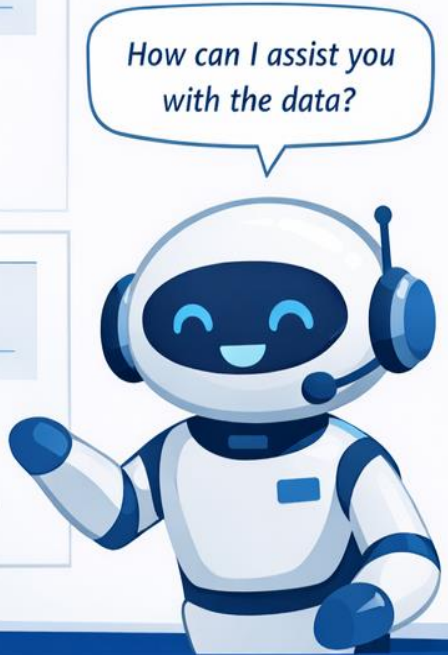
- *Generate key findings & summaries*
- *Contextual insights beyond rule-based text*

Domain-Adaptive Analysis

- *AI- driven contextual data shifts*
- *Analyze semi-structured datasets*

Chat with Data Bot

- *Interactive Q&A with AI Bot*
- *Gain insights through conversation*



SUMMARY

- **Transforms clinical data review from static to temporal** by visualizing dataset version evolution using animated, interactive dashboards.
- **Enables faster and earlier detection of dataset changes** across distributions, missingness, record volumes, and multivariate structure during interim reviews.
- **Analytics views** in one platform, including drift detection, automated change summaries, metadata checks, and comparative statistics.
- **Improves traceability and communication** through intuitive visual narratives (GIF/MP4) while remaining compatible with regulated workflows.

REFERENCES

[1] V. L. West, D. Borland, and W. E. Hammond, “Innovative information visualization of electronic health record data: A systematic review,” Journal of the American Medical Informatics Association, vol. 22, no. 2, pp. 330–339, 2015. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC4394966/

[2] E. Kukafka et al., “Interactive visualization applications in population health and health services research: Systematic scoping review,” Journal of Medical Internet Research, vol. 24, no. 2, 2022. Available: https://www.jmir.org/2022/2/e27534/

[3] M. Scheer et al., “Visualization techniques of time-oriented data for patient and cohort comparisons: Scoping review,” Journal of Medical Internet Research, vol. 24, no. 10, 2022. Available: https://www.jmir.org/2022/10/e38041/

[4] R. Ledesma et al., “Health timeline: An insight-based study of timeline visualization of clinical data,” BMC Medical Informatics and Decision Making, vol. 19, no. 1, 2019. Available: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0885-x

[5] B. C. Kwon et al., “DPVis: Visual analytics with hidden Markov models for disease progression pathways,” IEEE Transactions on Visualization and Computer Graphics, 2019. Available: https://arxiv.org/abs/1904.11652

THANK YOU