



From XPT to Dataset-JSON: Enabling Real-Time Validation and Streamlined Submissions

Hugo Signol

Sebastià Barceló

- 19-Nov-2025



Table of Contents

01

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore Lorem ipsum dolor sit amet, consectetur adipiscing

02

Dataset-JSON Viewer

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore Lorem ipsum dolor sit amet, consectetur adipiscing

03

CORE Integration

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore Lorem ipsum dolor sit amet, consectetur adipiscing

04

Conclusion

Introduction



Introduction

Context



For decades, XPT format dictated how data moved and how validation occurred.



XPT works but constrains iteration



Opportunity: Dataset-JSON +
APIs promise faster, traceable
checks



Response: Viewer + CORE delivers
continuous, subset-aware
validation.

Introduction

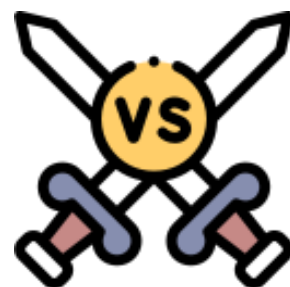
XPT vs Dataset-JSON

XPT format

Proven, widely supported, but binary, 8-char names, shallow metadata

Needs Define-XML for context, SAS dependencies common

File-transfer mindset slows review and iteration



Dataset-JSON

Text-based, human-readable, language-agnostic

Extensible metadata + links to Define-XML

Designed for both files **and** APIs, compact with compression

Currently under evaluation for submissions (v1.1 released Dec 2024)

Dataset-JSON Viewer



Background & Motivation

Dataset-JSON Viewer



Viewer Motivation

CDISC Dataset-JSON-Viewer hackathon (Q4 2024) kicked this off

Goal: SAS-viewer parity plus modern extras

Outcome: prototype viewer with live descriptives and a clear path to rule execution



Viewer features

Built in R Shiny

Landing (file management/summary) + Viewer (grid + tools)

Visual filters (instant, non-destructive) and R-syntax filters (true subsets)

Live descriptive stats (updates with subset; optional group-by)

Conditional row highlighting and clean export on the exact subset

Next, see the viewer...

Data Manipulation

Filter data using R expressions

e.g., AGE > 30 & STATUS == 'Active'

X

Q

You currently have 24 out of 74 observations

ON

AEREL == "NOT RELATED"

X

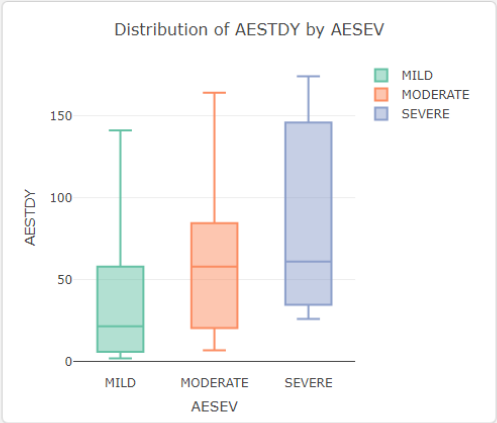
Live descriptives statistics

AESTDY

Add grouping variable

Select grouping variable

AESEV



Select a dataset

AE - Adverse Events



CONG	AESDISAB	AESDTH	AESHOSP	AESLIFE	AESOD	EPOCH	AESTDTC	AEENDTC	AESTDY	AEENDY	AEENRPT	AEENTPT
									2 - 174	0 - 174		
N	N	N	N	N	N	TREATMENT	2012-11-21		7		ONGOING	2013-01-14
N	N	N	N	N	N	TREATMENT	2012-11-21	2012-11-21	7	7		
N	N	N	N	N	N	TREATMENT	2012-11-22		8		ONGOING	2013-01-14
N	N	N	N	N	N	TREATMENT	2012-12-09		25		ONGOING	2013-01-14
N	N	N	N	N	N	TREATMENT	2013-01-11		58		ONGOING	2013-01-14
N	N	N	N	N	N	TREATMENT	2013-01-11		58		ONGOING	2013-01-14
N	N	Y	N	Y	N	TREATMENT	2013-01-14	2013-01-14	61	61		
N	N	N	N	N	N	TREATMENT	2013-09-01		4		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-09-02		5		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-09-02		5		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-09-15		18		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-10-23		56		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-10-23		56		ONGOING	2013-02-13
N	N	N	N	N	N	TREATMENT	2013-06-24	2013-06-24	141	141		
N	N	N	N	N	N	TREATMENT	2013-06-24	2013-06-24	141	141		
N	N	N	N	N	N	TREATMENT	2013-07-17	2013-07-26	164	173		
N	N	N	N	N	N	TREATMENT	2013-02-08	2013-02-08	35	35		
N	N	Y	N	N	N	TREATMENT	2014-10-31	2014-10-31	174	174		
N	N	N	N	N	N	TREATMENT	2013-02-16		118		ONGOING	2013-04-28
N	N	N	N	N	N	TREATMENT	2013-04-27		25		ONGOING	2013-09-18
N	N	N	N	N	N	TREATMENT	2013-04-28	2013-05-11	26	39		
N	N	N	N	N	N	TREATMENT	2013-06-01		60		ONGOING	2013-09-22
N	N	N	N	N	N	TREATMENT	2012-12-18	2012-12-18	2	2		
N	N	N	N	N	N	TREATMENT	2013-01-01	2013-01-19	16	34		

CORE Integration



CORE Integration

What is CORE ?



A governed set of executable **CDISC conformance rules** plus a **reference engine**.



Pulls rules from the **CDISC Library** and runs them consistently **locally or in the cloud**.



Produces **structured outputs** (JSON or XLSX) with rule-by-rule detail for review and traceability.



```
core validate -s sdtmig -v 3-4 \  
-d ./json \  
-ct sdtmct-2024-06-28 \  
-o ./reports -of json
```



COMPLETION

8.8%

TOTAL ISSUES

859

OPEN / RE-OPEN

783

CLOSED / RESOLVED

76

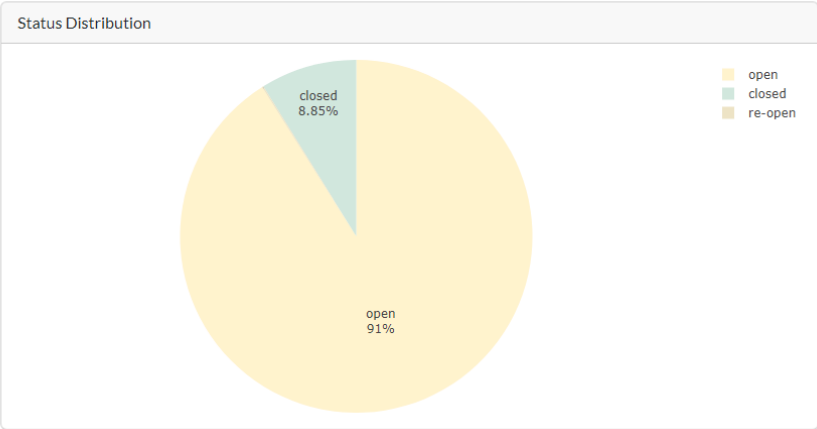
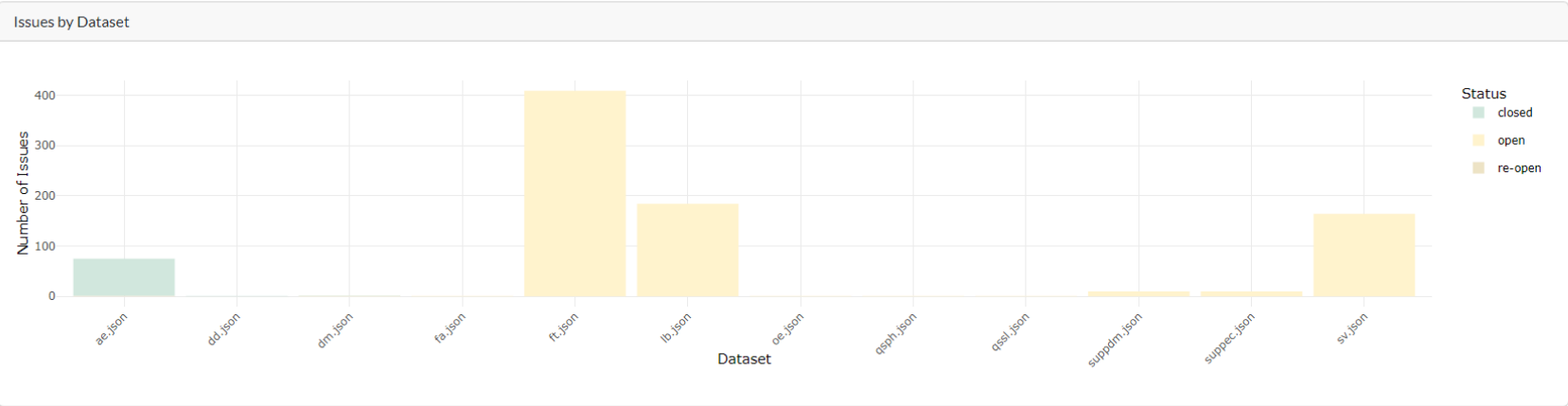
Issues Detail Table

open, re-open, closed, resolved

Copy CSV Excel

Search:

	Dataset	CORE_ID	Issues	Status	Message
1	ae.json	CORE-000266	1	re-open	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".
2	ae.json	CORE-000356	74	closed	At least one Required variable has a null value
3	dd.json	CORE-000594	1	closed	Variable label is not in title case.
4	dm.json	CORE-000238	1	closed	RFXENDTC does not equal the latest value of EX.EXSTDTC or EX.EXENDTC
5	dm.json	CORE-000594	1	open	Variable label is not in title case.
6	fa.json	CORE-000358	1	open	LNKGRP variable is not found in any of the other domains.
7	ft.json	CORE-000594	1	open	Variable label is not in title case.





ISSUE COMPLETION

8.8%

TOTAL ISSUES

859

OPEN / RE-OPEN

783

CLOSED / RESOLVED

76

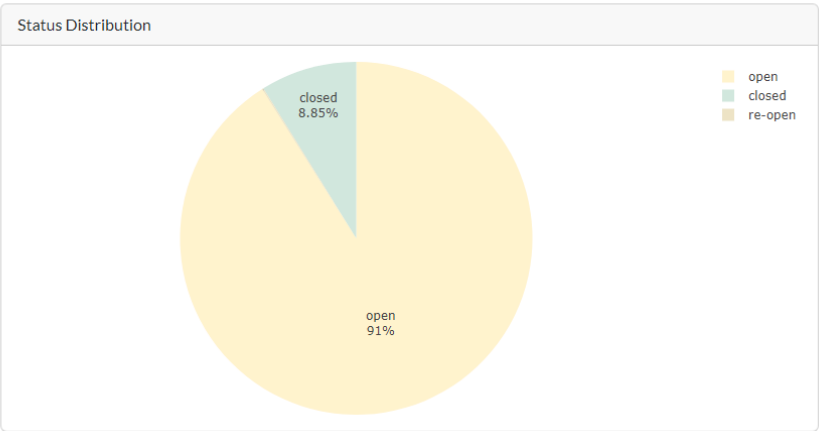
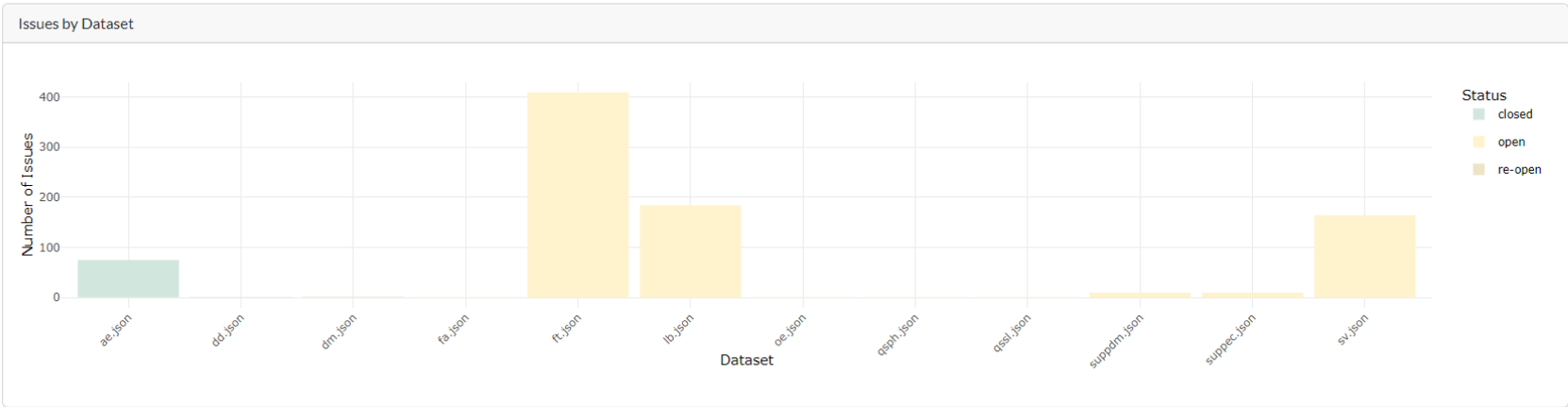
Issues Detail Table

open, re-open, closed, resolved

Copy CSV Excel

Search:

	Dataset	CORE_ID	Issues	Status	Message
1	ae.json	CORE-000266	1	re-open	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".
2	ae.json	CORE-000356	74	closed	At least one Required variable has a null value
3	dd.json	CORE-000594	1	closed	Variable label is not in title case.
4	dm.json	CORE-000238	1	closed	RFXENDTC does not equal the latest value of EX.EXSTDTC or EX.EXENDTC
5	dm.json	CORE-000594	1	open	Variable label is not in title case.
6	fa.json	CORE-000358	1	open	LNKGRP variable is not found in any of the other domains.
7	ft.json	CORE-000594	1	open	Variable label is not in title case.





CORE Validation

Standard

SDTMIG

Version

3.4

Controlled Terminology

2025-03-28

MedDRA Version

28.0

Data path

Please be careful with Windows/Linux f

Run CORE

Windows-only command runner preserved.

Output: Excel report + SQLite cached to .../core/core.sqlite

Dashboard auto-shows when data are present.

COMPLETION

8.8%

TOTAL ISSUES

859

OPEN / RE-OPEN

783

CLOSED / RESOLVED

76

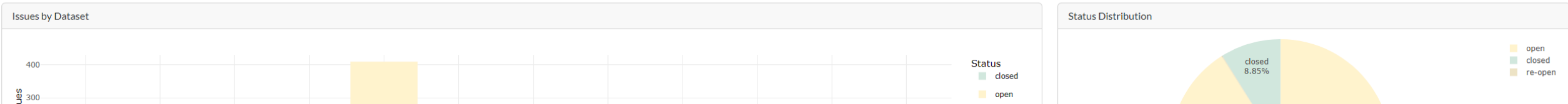
Issues Detail Table

Copy CSV Excel

open, re-open, closed, resolved

Search:

Dataset	CORE_ID	Issues	Status	Message
1 ae.json	CORE-000266	1	re-open	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".
2 ae.json	CORE-000356	74	closed	At least one Required variable has a null value
3 dd.json	CORE-000594	1	closed	Variable label is not in title case.
4 dm.json	CORE-000238	1	closed	RFXENDTC does not equal the latest value of EX.EXSTDTC or EX.EXENDTC
5 dm.json	CORE-000594	1	open	Variable label is not in title case.
6 fa.json	CORE-000358	1	open	LNKGRP variable is not found in any of the other domains.
7 ft.json	CORE-000594	1	open	Variable label is not in title case.





COMPLETION

8.8%

TOTAL ISSUES

859

OPEN / RE-OPEN

783

CLOSED / RESOLVED

76

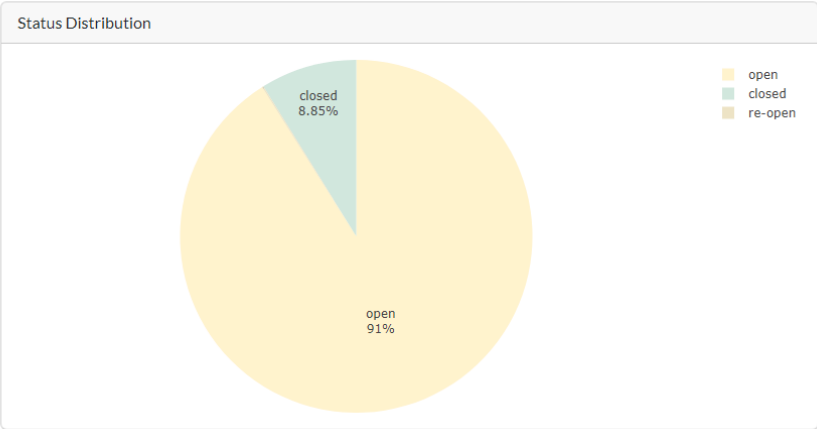
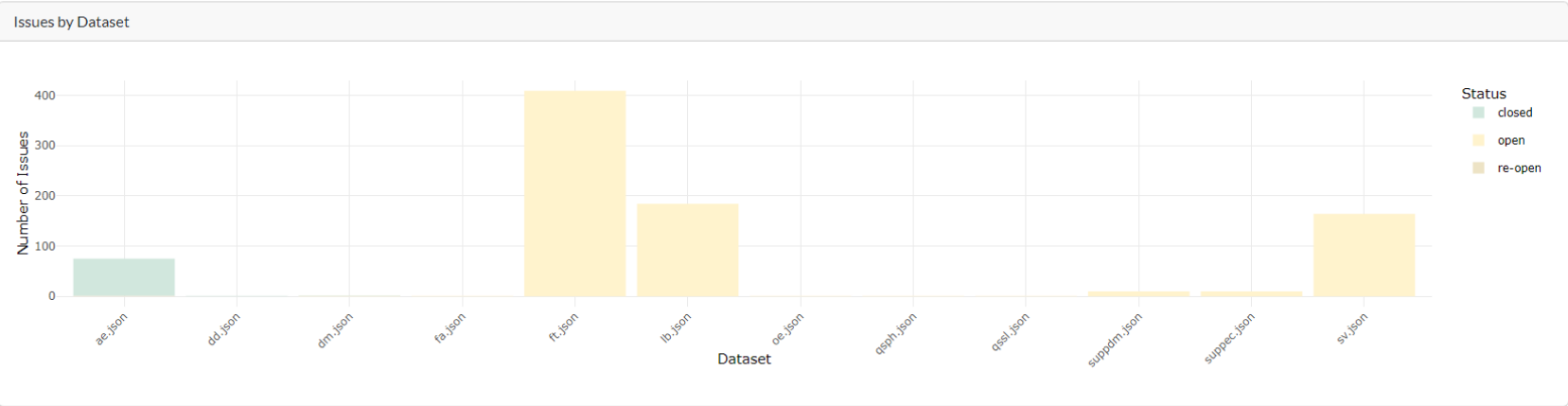
Issues Detail Table

open, re-open, closed, resolved

Copy CSV Excel

Search:

	Dataset	CORE_ID	Issues	Status	Message
1	ae.json	CORE-000266	1	re-open	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".
2	ae.json	CORE-000356	74	closed	At least one Required variable has a null value
3	dd.json	CORE-000594	1	closed	Variable label is not in title case.
4	dm.json	CORE-000238	1	closed	RFXENDTC does not equal the latest value of EX.EXSTDTC or EX.EXENDTC
5	dm.json	CORE-000594	1	open	Variable label is not in title case.
6	fa.json	CORE-000358	1	open	LNKGRP variable is not found in any of the other domains.
7	ft.json	CORE-000594	1	open	Variable label is not in title case.



+

COMPLETION

8.8%

TOTAL ISSUES

859

OPEN / RE-OPEN

783

CLOSED / RESOLVED

76

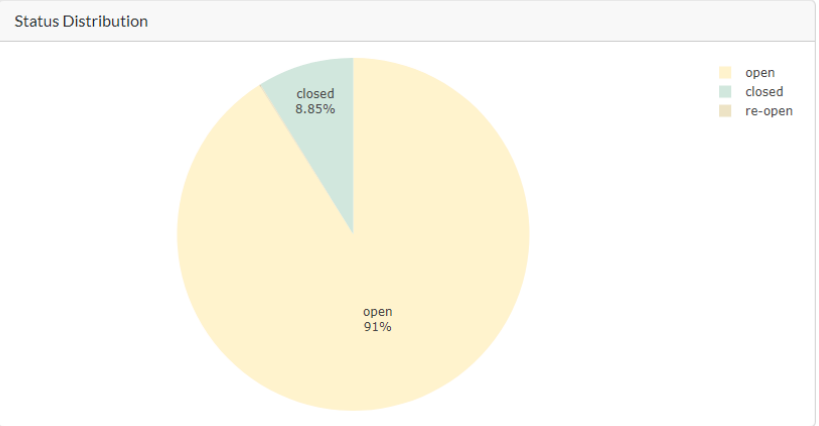
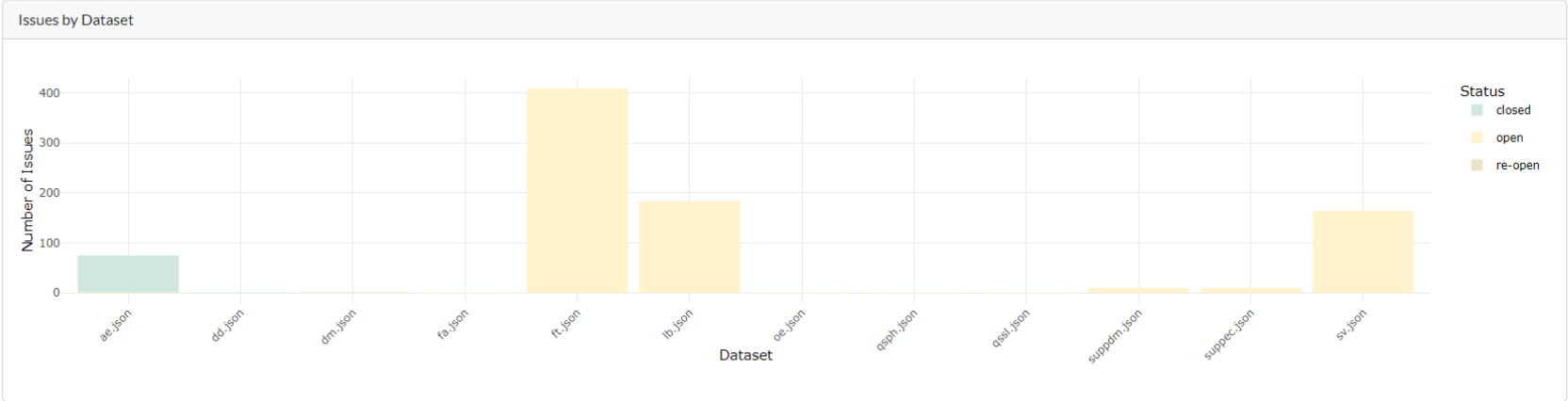
Issues Detail Table

open, re-open, closed, resolved

Copy CSV Excel

Search:

	Dataset	CORE_ID	Issues	Status	Message
1	ae.json	CORE-000266	1	re-open	If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".
2	ae.json	CORE-000356	74	closed	At least one Required variable has a null value
3	dd.json	CORE-000594	1	closed	Variable label is not in title case.
4	dm.json	CORE-000238	1	closed	RFXENDTC does not equal the latest value of EX.EXSTDTC or EX.EXENDTC
5	dm.json	CORE-000594	1	open	Variable label is not in title case.
6	fa.json	CORE-000358	1	open	LNKGRP variable is not found in any of the other domains.
7	ft.json	CORE-000594	1	open	Variable label is not in title case.



Dataset Details

Basic Information

Dataset: ae.json

Rule ID: CORE-000266

Issues: 1

Rule Summary

If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".

Status:

This issue is open and can be commented.

Open

Rule details

AESCAN: N

AESCONG: N

AESDISAB: N

AESDTH: N

AESER: Y

AESHOSP: N

AESLIFE: N

AESMIE: Not in dataset

AESOD: N

Affected Records



STUDYID	DOMAIN	USUBJID	ASEQ	AELNKID	AETERM	AELLT	AELLTCD	AEDECOD	AEPTCD	AEHLT	AEHLTCD	AEHLGT	AEHLGTCD	AEBODSYS	AEBDSYCD	AE
CDISCPLOT01	AE	CDISC003	13	13	EPISTAXIS											

Comments

No comments yet.

Type a message...

Dataset Details

Basic Information

Dataset: ae.json
Rule ID: CORE-000266
Issues: 1

Rule Summary

If AESER = "N" then none of the seriousness criteria (AESCAN, AESCONG, AESDISAB, AESDTH, AESHOSP, AESLIFE, AESOD, AESMIE) could be equal to "Y".

Status:

This issue is open and can be commented.

Open

Rule details

Closing Rule

You want to close this Rule, please see suggestions... or type your own custom answer.

Global Suggestions

- Known issue - Sponsor decision
- Protocol deviation - justified
- Issue due to external data source

CYTEL Suggestions

- To be implemented

AI Suggestions

- To be implemented

Explanation

Please provide an explanation before closing.

This explanation will be logged for audit trail and review.

Reset Explanation X Confirm

Affected Records



STUDYID	DOMAIN	USUBJID	AESEQ	AELNKID	AETERM	AELLT	AELLTCD	AEDECOD	AEPTC
CDISCPLOT01	AE	CDISC003	13	13	EPISTAXIS				

Comments

No comments yet.

Type a message...

Conclusion



Conclusion

To conclude...

Impact: earlier defect discovery; faster “found → explained/fixed”; JSON lowers friction and enables automation

Limits today: Dataset-JSON regulatory adoption is evolving; highlight rules limited to single condition (for now)

Roadmap: multi-condition highlights; broader CLI options; external rule services

Takeaway: keep your familiar table view; get real-time, standards-aligned validation by pairing the viewer with CORE

Q&A / Ask: want access to the app, repo links, or the demo package?

 **GitHub** <https://github.com/sbarcelo777/Dataset-JSON-hackathon>

Cytel

Thank you!