ETL Vocab Mapping

Sprint 4 Meeting

ETL Vocab Mapping Team Updates

Name of the file	Assigned to	Status
intraop_aimsvitals_vitalcode	Liying Pie	Completed
intraop_drugdrug_group1	wong.LeongHui	Completed
intraop_drugmed_group1	wong.LeongHui	Completed
intraop_nurvitals_group1	Lakshmi	Completed
intraop_operation_group1.xlsx (divided into 9 files)	Liying Pie Matsumoto Tanaka Qi Elizabeth	 Dr. Matsumoto - Completed Dr. Tanaka - Completed Dr. Liying Pie - Completed
postop_clindoc_group1.xlsx	Qi	Completed
postop_info_group1.xlsx	wong.LeongHui	Completed
postop_lab_testdesc.xlsx	Lakshmi	Completed
postop_labmicro_antibioticname	Do not Map	
postop_labmicro_microresultedproceduredescription		
postop_labmicro_organismdescription		
postop_labsall_group1	Lakshmi	Completed
postop_pacu_group1	Kosuke Tanaka	Completed
postop_renal_group1	Asif Syed	Will be Completed by 04Oct
preop_char_allergyinformation	Lakshmi	Completed
preop_char_gender	Qi	Completed
preop_char_race	Qi	Completed
preop_lab_prepoplabtestdescription	Lakshmi	Completed
preop_radiology_procedurename	Alexander Miller	In Progress

APAC Community-wide ETL Project

Sprint 4 Review

ETL Team

Sprint Overview

To do	In progress	In review	Done
measurement	condition_occurrence	visit_detail	person
condition_era	observation	note	care_site
drug_era	procedure_occurrence		observation_period
	drug_exposure		device_exposure
			visit_occurrence
			provider
			death
			specimen

***Location table: Lack of source data for mapping

ETL Process Timeline

- **Completed**:
  - ETL Development:
    - Person
    - Care_site
    - Observation_period
    - Death
  - Populated the first batch of the
    - 10 Clinical tables under Done & Review
    - STCM table
  - Merge the 1st draft of the ongoing tables.
- **Update**:
  - Some fields updated using the STCM table provided by the Vocab team.
- **Pending**:
  - Some fields are still to be mapped later.

	Tables	Team	Assigned	Sprint 4		Sprint 5	
				20/9/2024 - 26/9/2024	27/9/2024 - 3/10/2024	4/10/2024 - 10/10/2024	11/10/2024 - 17/10/2024
1	source_to_concept_map	Co-owned	-				
2	person	SQL	Hamcheeze				
3	location	N/A	-				
4	care_site	SQL	Hamcheeze				
5	provider	SQL	Afreen				
6	visit_occurrence	SQL	Hamcheeze				
7	visit_detail	SQL	Chinapat/ Sornchai				
8	observation_period	Python	Sukatat				
9	death	Python	Erwin				
10	condition_occurrence	Python	Hengxian/ Satish				
11	observation	Python	Brandan				
12	procedure_occurrence	SQL	Afreen				
13	drug_exposure	Python	Alicia				
14	measurement	Python	Hengxian/ Satish				
15	device_exposure	Python	Yong Zhe				
16	specimen	SQL	Hamcheeze				
17	note	Python	Mun Chun				
18	condition_era	SQL	-				
19	drug_era	SQL	-				

Sprint Reflections

- **Challenges Faced:**
  - Have some data issues that might need clarifications from the data owner.
- **Data Availability**:
  - Processed data is available within both sub-team schemas (Python and SQL).
    - May consider combining these schemas together later.
- **Next Actions:**
  - Merge existing Pull requests and start 2nd iteration on pending items such as updating standard concept ids.


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank you!

Quality Assurance Team

Sprint Status

- DQD has been set up on both VMs
- R script created for running DQD
- DQD test run on both VMs.

First run – SQL VM

DATA QUALITY ASSESSMENT

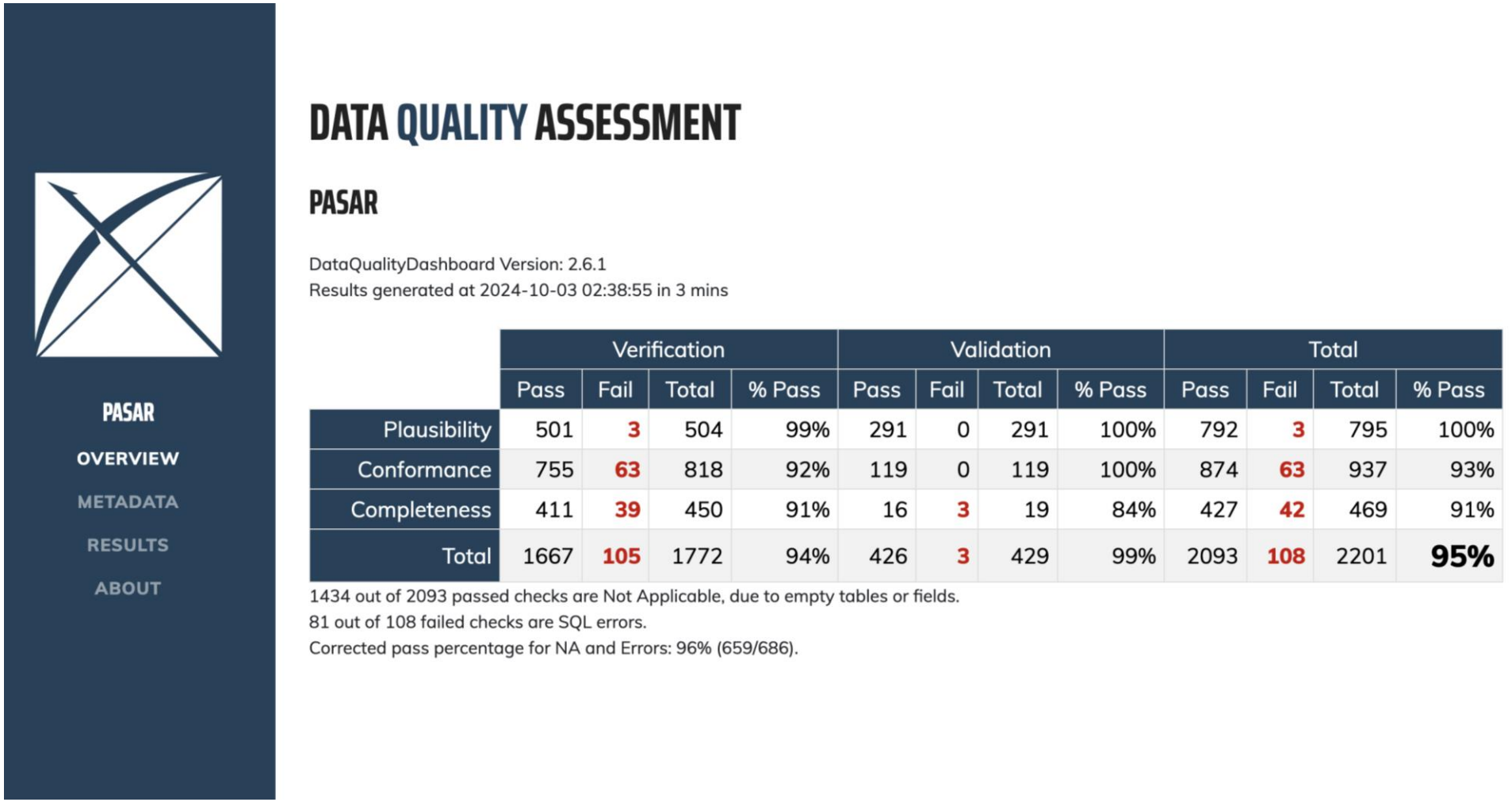
PASAR

DataQualityDashboard Version: 2.6.1
Results generated at 2024-10-01 11:54:25 in 2 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	500	4	504	99%	291	0	291	100%	791	4	795	99%
Conformance	756	62	818	92%	119	0	119	100%	875	62	937	93%
Completeness	408	42	450	91%	19	0	19	100%	427	42	469	91%
Total	1664	108	1772	94%	429	0	429	100%	2093	108	2201	**95%**

1393 out of 2093 passed checks are Not Applicable, due to empty tables or fields.
81 out of 108 failed checks are SQL errors.
Corrected pass percentage for NA and Errors: 96% (700/727).

First run – Python VM



PASAR

OVERVIEW

METADATA

RESULTS

ABOUT

DATA QUALITY ASSESSMENT

PASAR

DataQualityDashboard Version: 2.6.1
Results generated at 2024-10-03 02:38:55 in 3 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	501	3	504	99%	291	0	291	100%	792	3	795	100%
Conformance	755	63	818	92%	119	0	119	100%	874	63	937	93%
Completeness	411	39	450	91%	16	3	19	84%	427	42	469	91%
Total	1667	105	1772	94%	426	3	429	99%	2093	108	2201	**95%**

1434 out of 2093 passed checks are Not Applicable, due to empty tables or fields.
81 out of 108 failed checks are SQL errors.
Corrected pass percentage for NA and Errors: 96% (659/686).

Next steps

- Run DQD on both VMs
- Have a public Github webpage to display results
- Deep dive into results
- Discuss with the other teams on quality results and potential improvements - planning to join next ETL weekly meeting


OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank you!