

# Promoting Data Quality in PEDSnet Using GitHub

Aaron N. Browne, Jeffrey W.  
Pennington, L. Charles Bailey, MD,  
PhD

# Disclosures

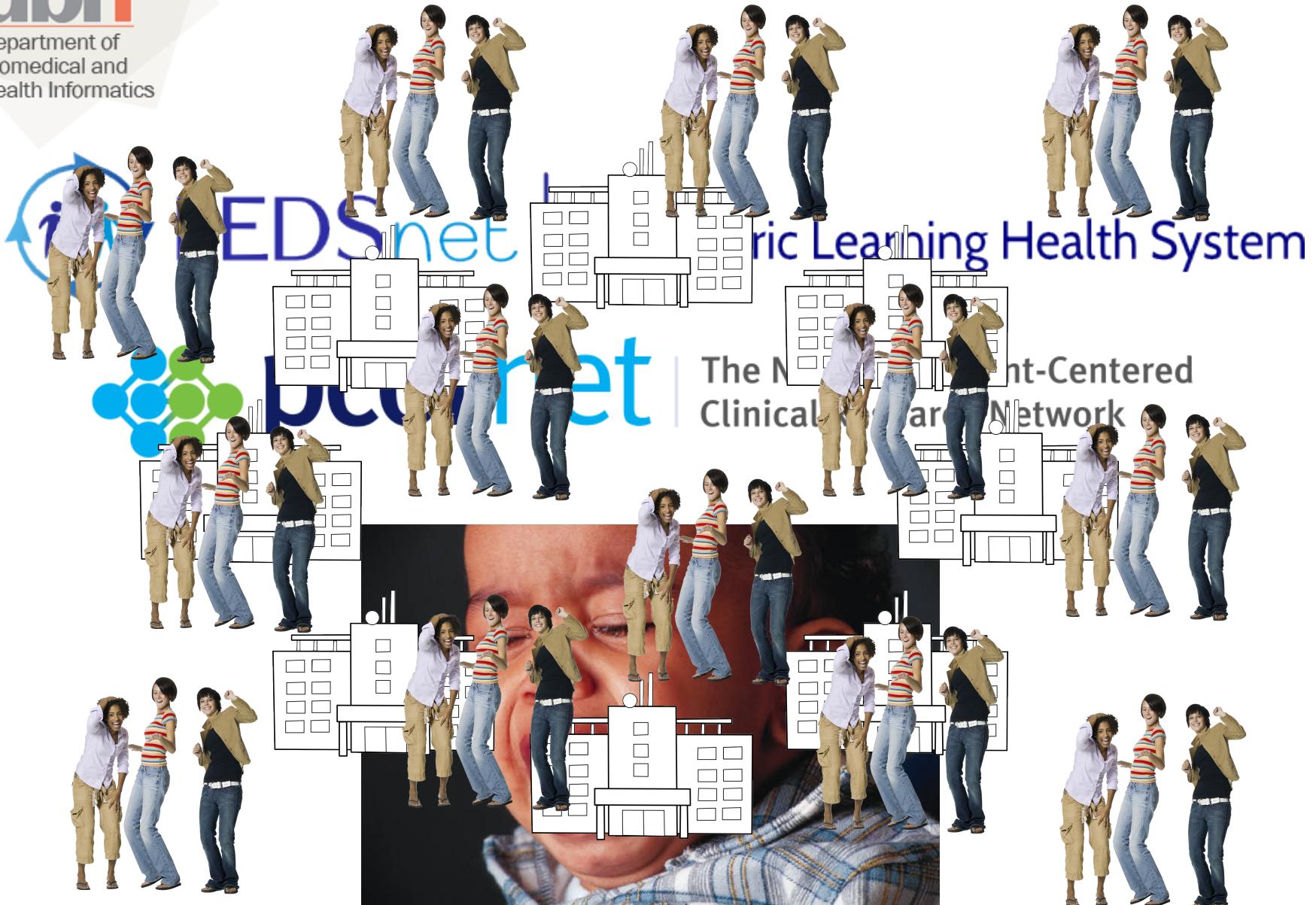
- Nothing to disclose.

# Objectives

- Describe PEDSnet EHR data collaboration
- Highlight potential data quality issues
- Introduce GitHub platform
- Report on GitHub usage in PEDSnet
  - Data Quality improvement?
- Future directions







# PEDSnet

- 2.1 million patients
- 2.8% of US children
- 22 state catchment area

Forrest, Christopher B., et al. "PEDSnet: a National Pediatric Learning Health System." *Journal of the American Medical Informatics Association* (2014): amiajnl-2014.

# Data Quality

- EHR-inherent issues:
  - Accuracy
  - Objectivity
  - Believability
- Extraction and transformation issues:
  - Completeness
  - Mis-representation

Kahn, Michael G., et al. "A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research." Medical care 50 (2012).





Search GitHub

Explore Gist Blog Help

aaron0browne + ➔

## PEDSnet

A pediatric learning health system

[https://wiggio.com/?#tpl=posts\\_0](https://wiggio.com/?#tpl=posts_0)

Filters Find a repository... + New repository

**pedsnetcdm\_to\_pcornetcdm** PRIVATE

SQL scripts and data to convert data in the PEDSNet CDM to the PCORnet CDM.

Updated 13 hours ago

**Data-Quality** PRIVATE

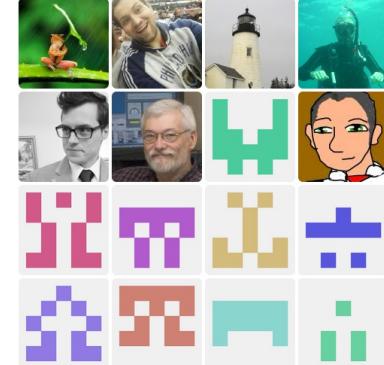
SQL code for Data Quality Queries (OMOP and i2b2)

Updated 14 hours ago

**CCHMC** PRIVATE

Updated 14 hours ago

People 71 ➔



Invite someone

Teams 11 ➔

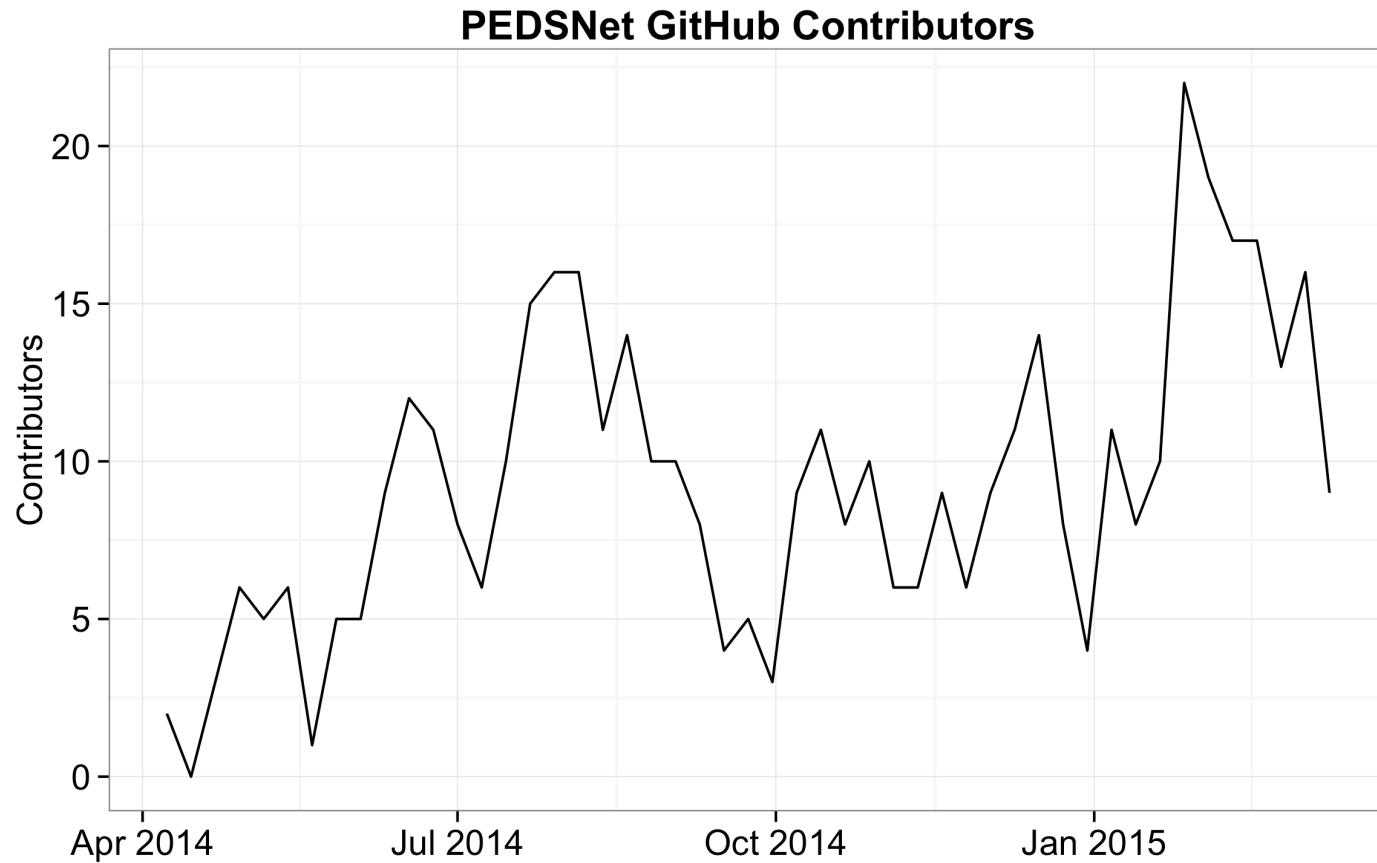
Jump to a team

Owners 7 members · 20 repositories

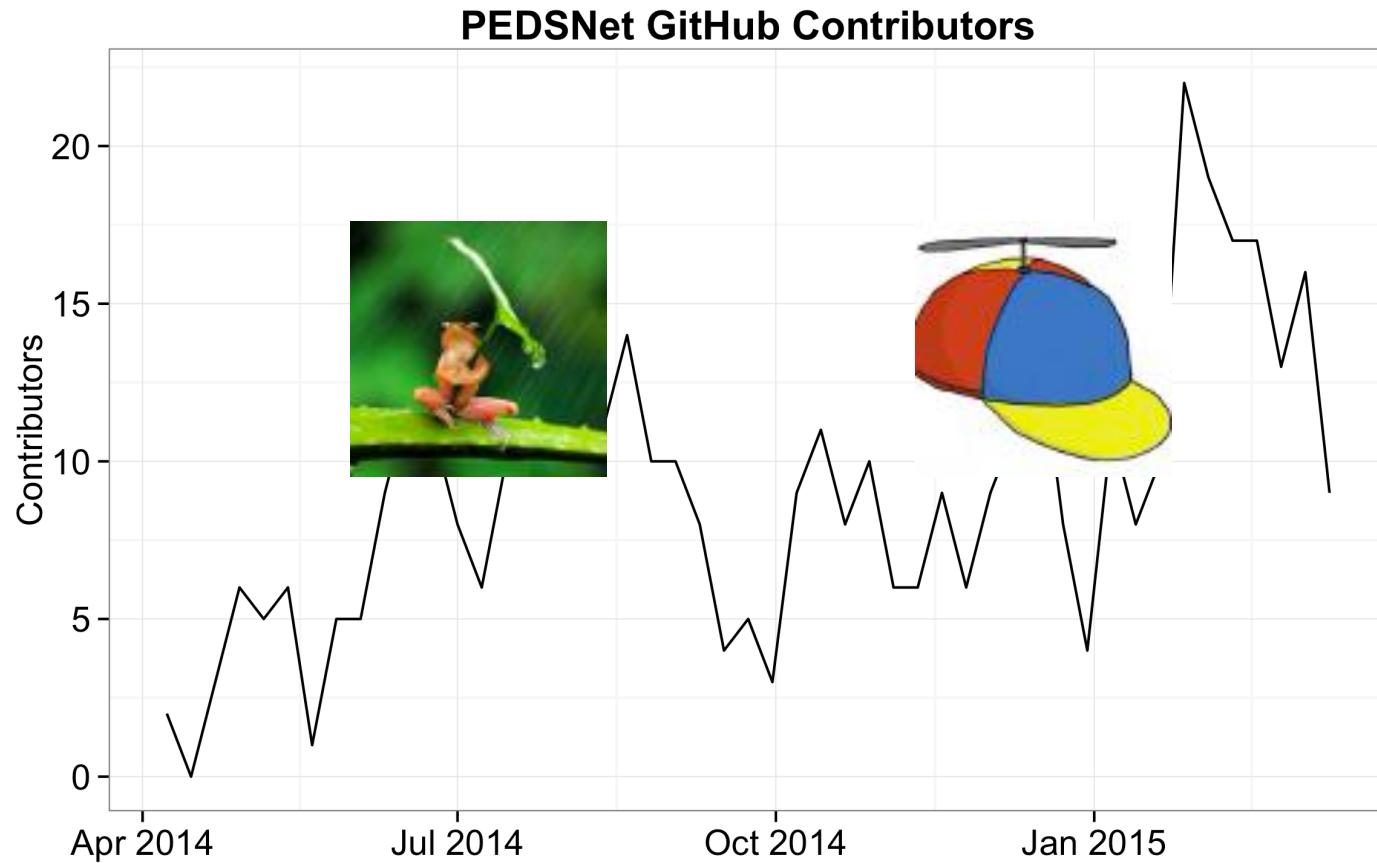
# PEDSnet on GitHub

- 71 members
- 11 teams
- 20 repositories
- Workflow:
  - Issues
  - Comments
  - Commits
  - Pull Requests

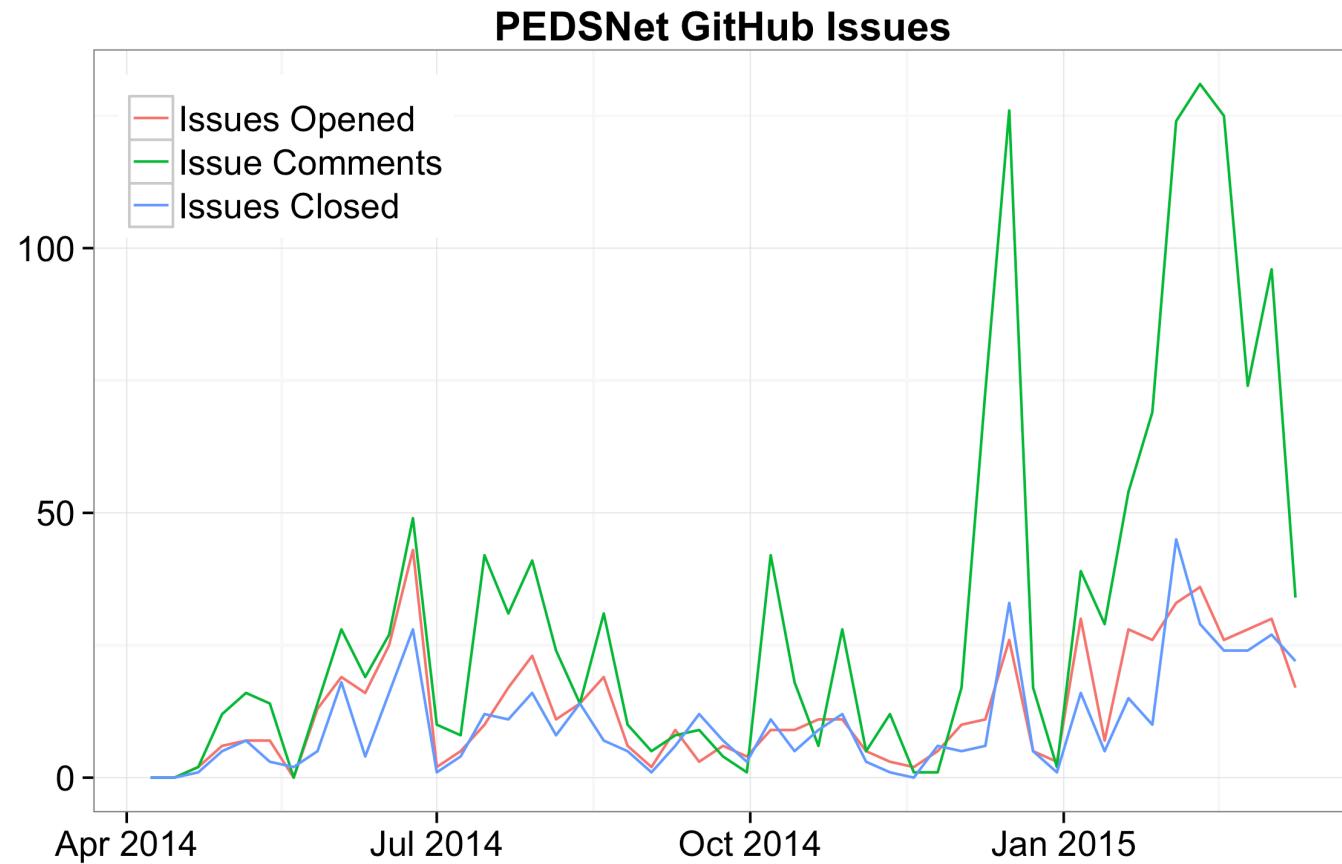
# Contributors



# Contributors



# Issues



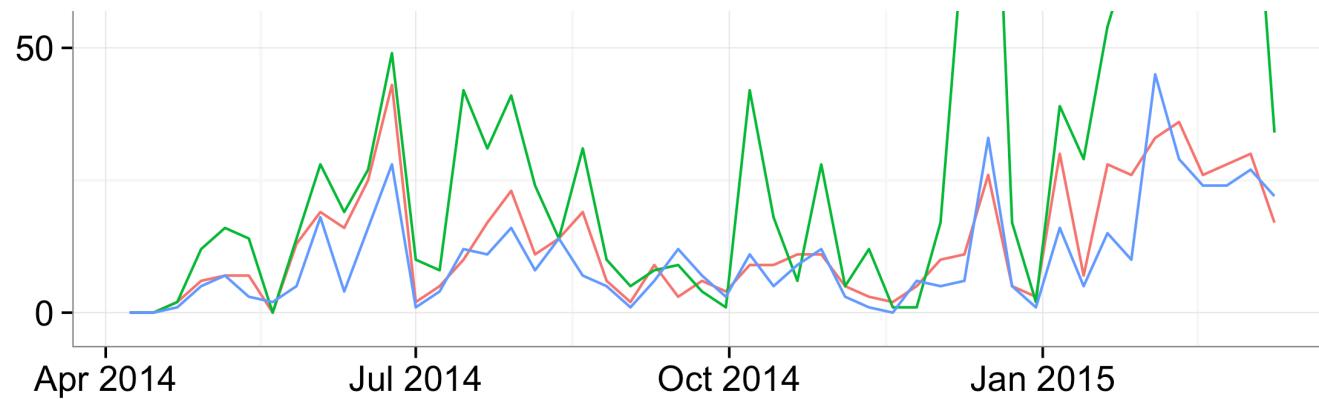
# Issues



## Debugging between Ritu and Greg #7

**Closed**

aaron0browne opened this issue on Dec 8, 2014 · 40 comments



# Issues

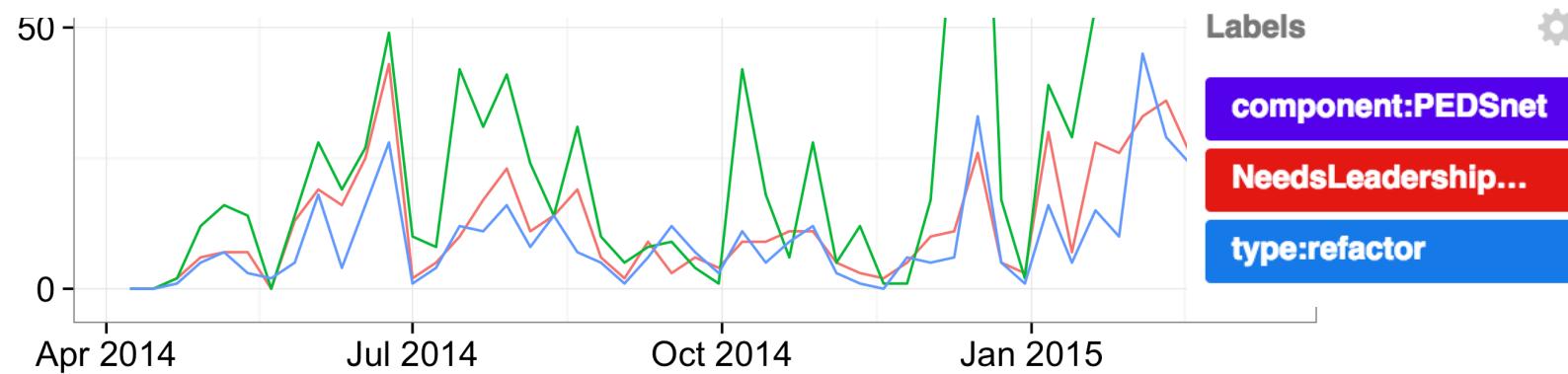
PEDSNet GitHub Issues



Add unique constraints in PEDSnet to materialize DCC natural keys #91

Closed

aaron0browne opened this issue on Jan 24 · 30 comments



# Issues

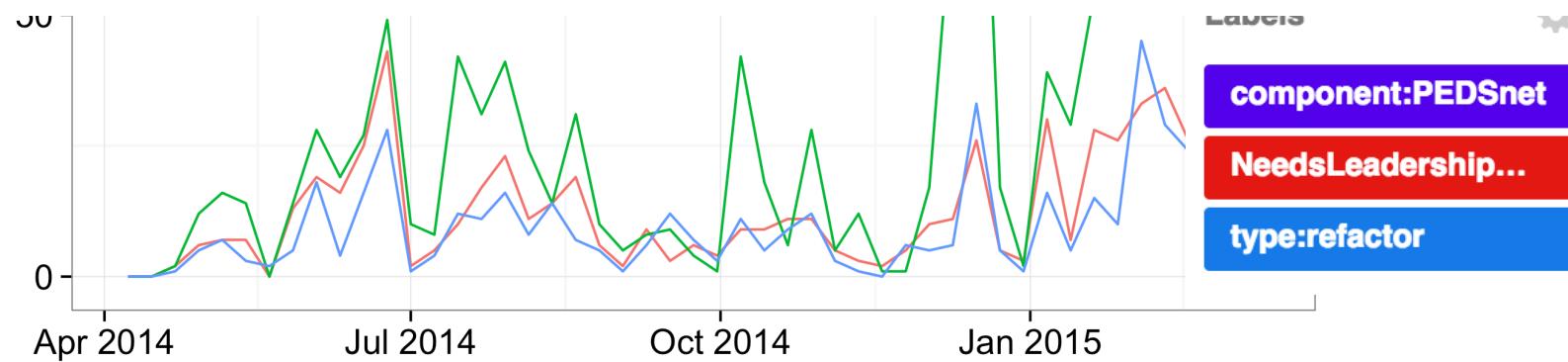
PEDSNet GitHub Issues



## Mapping non-SNOMED diagnoses from EHR through PEDSnet into PCORnet #63

Closed

aaron0browne opened this issue on Dec 8, 2014 · 19 comments



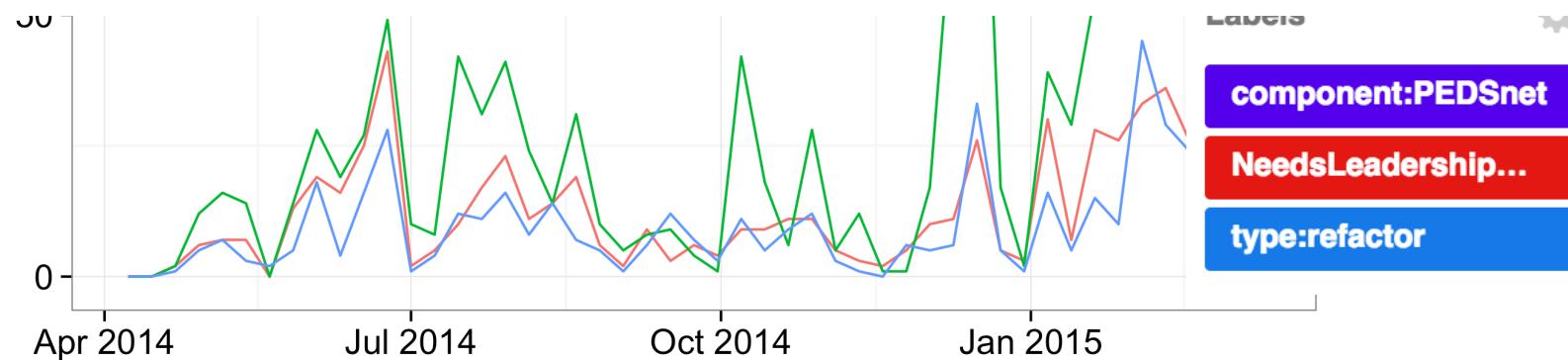
# Issues

PEDSNet GitHub Issues

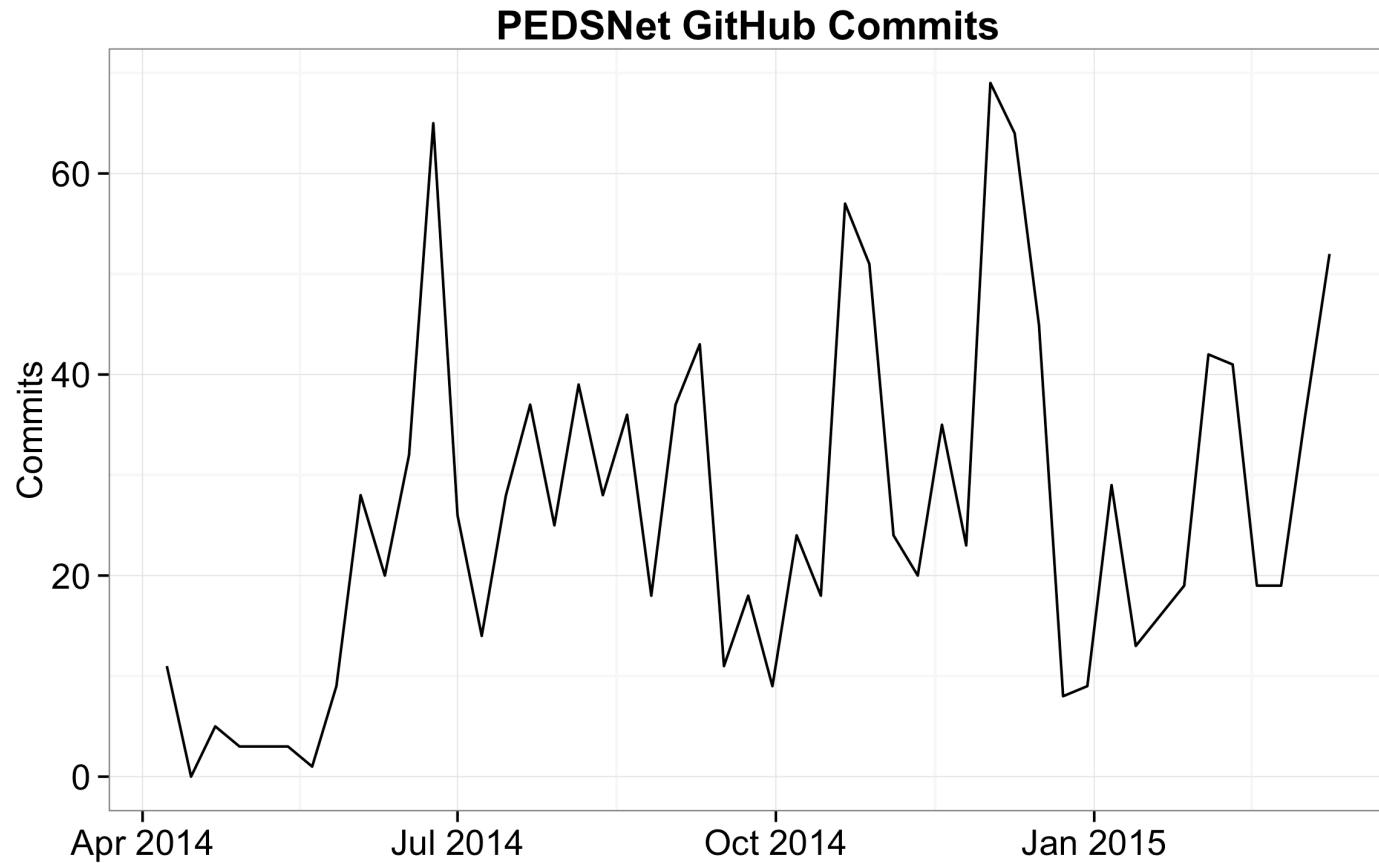


baileych commented 29 days ago

Great that we're having this discussion to flesh things out.



# Commits



# Commits

This repository Search Explore Gist Blog Help aaron0browne + ⌂ ⌐ ⌒ ⌓

 PEDSnet / **pedsnetcdm\_to\_pcornetcdm** PRIVATE Unwatch ▾ 25 ⌐ Star 0 ⌐ Fork 0

branch: master

Commits on Mar 25, 2015

 Merge pull request #8 from PEDSnet/issue-6-colorado-testing ...  
writetoritu authored 23 hours ago

Commits on Mar 24, 2015

 minor edits from testing on CHCO and Seattle  
writetoritu authored 2 days ago

 Merge pull request #5 from PEDSnet/issue-3 ...  
writetoritu authored 2 days ago

 updating admission source codes  
writetoritu authored 2 days ago

Commits on Mar 17, 2015

 addressing issue-3  
writetoritu authored 9 days ago



# Commits

## minor edits from testing on CHCO and Seattle

[Browse files](#)

by master (#8)

 **writetoritu** authored 2 days ago

1 parent e32dbe commit 363aea88ad32a0404a525c13c0066a49f1bf3257

 Showing 1 changed file with 2 additions and 2 deletions.

[Unified](#) [Split](#)

4 ████ ETL Scripts/PEDSnet-PCORnet.sql

[View](#)

		@@ -8,7 +8,7 @@ select distinct
8	8	(case when month_of_birth is null then '-01' else '-'  lpad(cast(month_of_birth as text),2,'0') end)
9	9	(case when day_of_birth is null then '-01' else '-'  lpad(cast(day_of_birth as text),2,'0') end)
10	10	as birth_date,
11	-	pn_time_of_birth as birth_time,
	11	+ date_part('hour',pn_time_of_birth)  ':'  date_part('minute',pn_time_of_birth) as birth_time,
12	12	coalesce (m1.target_concept,'OT') as Sex,
13	13	coalesce (m2.target_concept,'OT') as Hispanic,
14	14	coalesce (m3.target_concept,'OT') as Race,
		@@ -30,7 +30,7 @@ select distinct
30	30	cast(op.person_id as text) as pat_id,
31	31	cast(date_part('year', observation_period_start_date) as text)  '-'  lpad(cast(date_part('month', observation_perio
32	32	cast(date_part('year', observation_period_end_date) as text)  '-'  lpad(cast(date_part('month', observation_perio
33	-	case when o.person_id is null then 'NI' else coalesce(m1.target_concept,'OT') end as chart_availability,
	33	+ case when o.person_id is null then 'N' else coalesce(m1.target_concept,'N') end as chart,
34	34	'E' as ENR_basis
35	35	from
36	36	omop.observation_period op

# Workflow Example

## Manage warehouse schema #96

 **Closed**

aaron0browne opened this issue 14 days ago · 1 comment



aaron0browne commented 14 days ago



The loading user needs to own the warehouse tables in order to turn autovacuum on and off. It also makes more sense to me to create the warehouse tables from the loader than from harvest. Thoughts?



 aaron0browne referenced this issue 14 days ago

**Manage warehouse schema tables #97**

 Merged



bruth commented 14 days ago



Yep.

# Workflow Example

## Manage warehouse schema tables #97

Edit

 Merged bruth merged 1 commit into develop from issue-96 10 days ago

Conversation 3 Commits 1 Files changed 3 +3 -16

aaron0browne commented 14 days ago

Signed-off-by: Aaron Browne aaron0browne@gmail.com

Fixes #96

aaron0browne commented 11 days ago

@bruth @murphyke, does this seem like a reasonable fix for #96? Am I missing anything?

Eventually we may want to have a separate set of models for the warehouse vs the staging dbs, but I didn't think we should overcomplicate before having a specific reason to.

bruth commented 10 days ago

Yes, that should be all that is necessary. It may be worth documenting that the warehouse *should* be created by the loader now.

Labels None yet

Milestone No milestone

Assignee No one—assign yourself

Notifications

Unsubscribe You're receiving notifications because you authored the thread.

2 participants



# Workflow Example

N 58 | 58 |  
59 | 59 | -The PEDSnet warehouse database is assumed to be managed by an external application and therefore is not managed here.  
| 59 | +In addition to the app-specific coordinating database, the PEDSnet warehouse database is also managed by this app.  
60 | 60 |  
61 | 61 |    ### Queue data archives  
62 | 62 |  
    ✖

1 1 pedsnet/conf/settings.py View 💻  
  
@@ -34,7 +34,6 @@  
34 | 34 |  
35 | 35 |    DATABASE\_ROUTERS = (  
36 | 36 |       'pedsnet.routers.CoordinatorRouter',  
37 | - |       'pedsnet.routers.WarehouseRouter'  
38 | 37 |    )  
39 | 38 |  
40 | 39 |    # Alias of the database connection that staging databases will be  
    ✖

12 12 pedsnet/routers.py View 💻  
  
@@ -22,15 +22,3 @@ def allow\_syncdb(self, db, model):  
22 | 22 |       return db == self.database\_name  
23 | 23 |       elif db == self.database\_name:  
24 | 24 |           return False  
25 | - |  
26 | - |  
27 | - |    class WarehouseRouter(object):  
28 | - |       """Router for the PEDSnet warehouse database. This is assumed to be  
29 | - |       managed by an external application and thus is restricted to reads/writes  
    ✖

# Conclusions

- Data quality results:
  - Early distribution and debugging of DQ tools
  - Early and often code reviews
  - Convention discussion and distribution
  - Transparency!
- Future work:
  - Re-use GitHub data for end-user trust

# Questions?



supported by PCORI Contract CDRN-1306-01556