PCORnet CDM Forum

Wednesday, October 5, 2016 from 3–4 PM Eastern
Hosted by Keith Marsolo, PhD
Facilitated by Shelley Rusincovitch and Michelle Smerek

pcornet®

The National Patient-Centered Clinical Research Network

Agenda

- Announcements
- PCORnet at AcademyHealth Concordium
- Considering observation stays
- Data Committee CDM survey
- v3.1 interest group
- v3.1 draft specification
- Wrap up

pcornet®

Announcements

Recap: CDM Forum from August 19

Topics included:

- Challenges in PX_TYPE assignments
- Fixing PX Type issues in OneFlorida
- PCORnet CDM v3.1 development
- Data Committee CDM survey

- Slides: https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-Forum-Materials

- Recording and meeting summary: https://pcornet.imeetcentral.com/p/ZgAAAAAAeax5

pcornet®

Recap: DRNOC Forum on September 19

Topics:

- Implementation Guidance
- Data Characterization Cycle 2

- Slides and meeting recording: https://pcornet.imeetcentral.com/p/ZgAAAAAAeuuH

Recap: DRNOC Forum on October 3

Topics:

- Menu-Driven Query 1 and 2
- Data Characterization Cycle 1 Review
- Review of DataMart refreshes

- Slides and meeting recording: https://pcornet.imeetcentral.com/p/ZgAAAAAAe5fc

(We'll come back to this in a few minutes)

pcornet®

Live component: Review of Issue Tracker

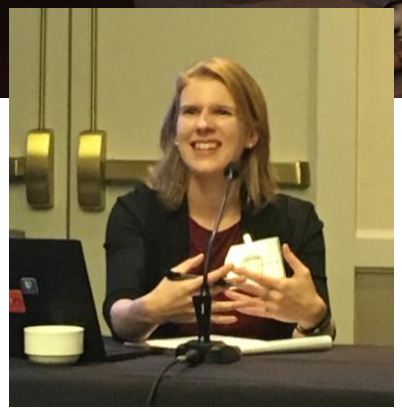
CDM errata issue tracker:
https://github.com/CDMFORUM/CDM-ERRATA/issues

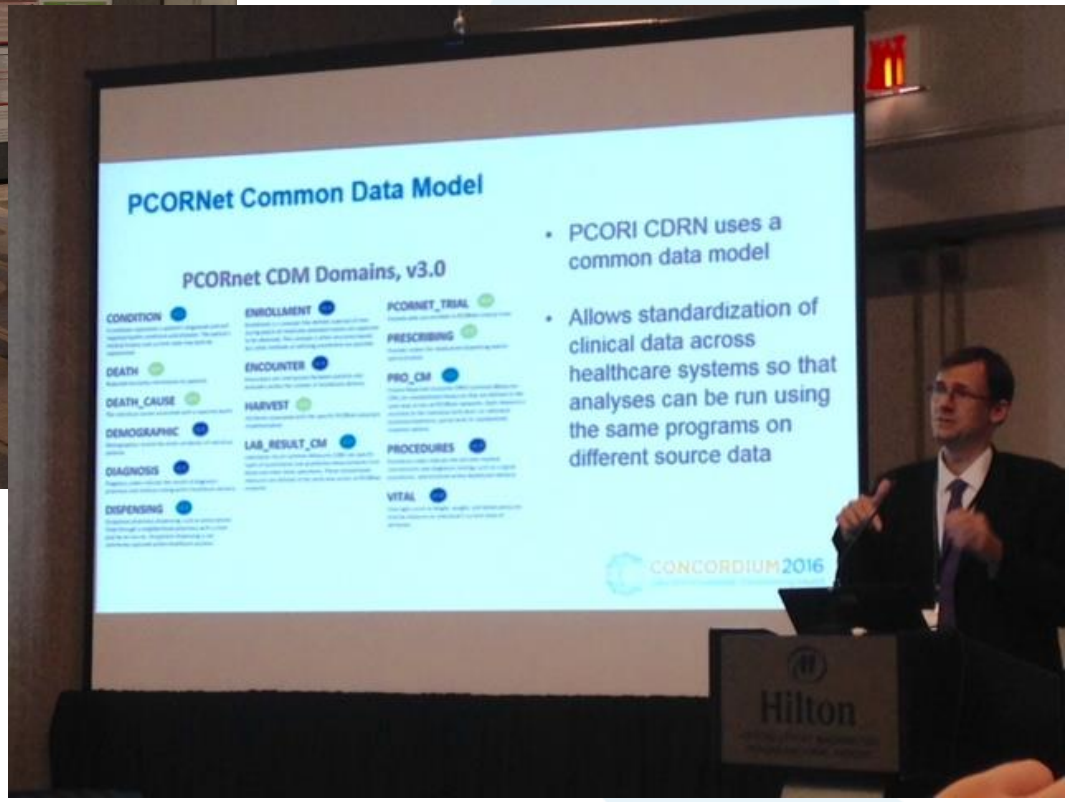
CDM guidance issue tracker:
https://github.com/CDMFORUM/CDM-GUIDANCE/issues

pcornet®

AcademyHealth Concordium

PCORnet at AcademyHealth Concordium, September 12-13, 2016





Representing a Lab Result – Developing a LOINC Reference Table for a Distributed Research Network

Michelle Smerek[1], Shelley Rusincovitch[2], and Keith Marsolo, PhD[3]

[1]Duke Clinical Research Institute, Durham, NC, [2]Duke Translational Research Institute, Durham, NC, [3]Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Introduction

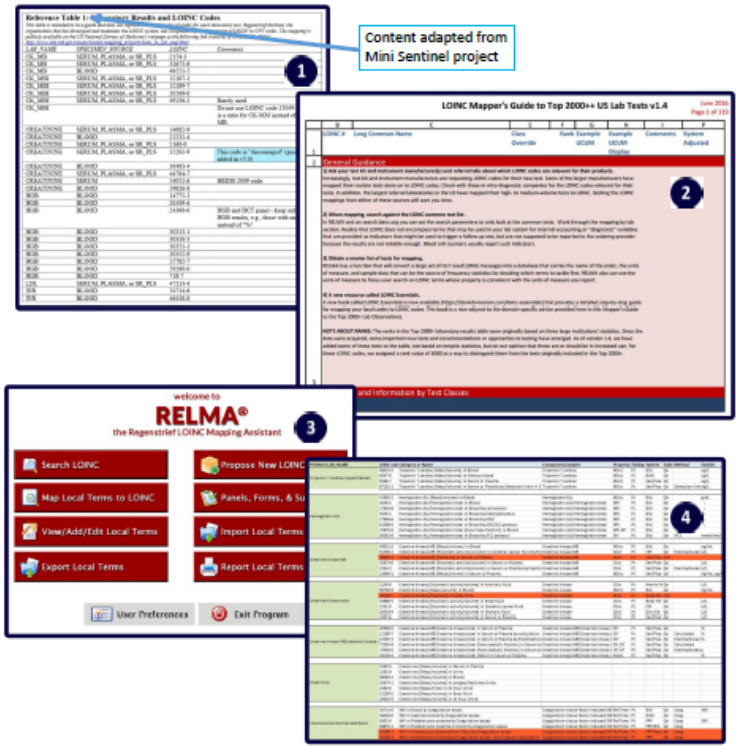
- The PCORnet CDM laboratory results table includes 11 common laboratory tests (CLTs), to which local data are mapped.
- LOINC (Logical Observation Identifiers Names and Codes) values may be included for laboratory results.
- Clinical laboratory results are notorious for their heterogeneity. For many data partners, PCORnet was one of the first large-scale drivers for the assignment of LOINC codes to laboratory results.
- A laboratory result can be represented by several "correct" choices.
- PCORnet data partners asked the PCORnet Distributed Research Network Operations Center (DRNOC) for a reference table that would help them determine whether the LOINC codes they assigned for the 11 CLTs were appropriate.

Methods

The following process was used to expand the lab result reference table:

1. Reviewed content of existing CLT – LOINC reference table (to determine conceptual scope for each CLT in the CDM)
2. Reviewed LOINC Top 2000+ list and LOINC tutorials
3. Attended LOINC workshop to confer with LOINC experts on Regenstrief LOINC Mapping Assistant (RELMA®) best practices
4. Utilized RELMA® to identify LOINC codes for inclusion in updated CLT – LOINC reference table (https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/Lab-Mapping-Resources)

- Received input from PCORnet Lab Mapping Interest Group



Results

- A more complete CLT – LOINC mapping table was created. An additional 59 codes were added, with each CLT having between 2 and 35 codes.

Discussion

- The updated listing of LOINC codes does not address all the nuance associated with the secondary use of laboratory data.
- Much as investigators had to learn which ICD codes to use when identifying patients who were treated for a particular condition, they will need to similarly understand which LOINC code(s) to choose to represent which laboratory test(s).

Lessons Learned

- As new laboratory tests are developed, and existing test methods are clarified, LOINC codes are created and deprecated. Consequently, upstream LOINC assignment (e.g., reference labs/instrument manufacturers) might be more optimal
- While a more complete reference table was created, the LOINC codes that correspond to a categorical name (e.g., "hemoglobin") will depend on the intended analytical purpose of the data, which must be considered on a case-by-case basis

 We are grateful to Daniel Vreeman, PT, DPT, MSC and his colleagues at the Regenstrief Institute for their thoughtful input and engagement in this work.

pcornet® The National Patient-Centered Clinical Research Network

Acknowledgments

The project described was supported by the National Patient-Centered Clinical Research Network (PCORnet). Efforts were led by the Coordinating Center, managed by the Duke Clinical Research Institute, which is partially funded by the Patient-Centered Outcomes Research Institute (PCORI). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the Patient-Centered Outcomes Research Institute.

CONCORDIUM2016
Data and Knowledge Transforming Health
SEPT 12-13,2016 | HILTON CRYSTAL CITY
Sponsored by: AcademyHealth

Contact Information
Michelle Smerek, Research Informaticist
Duke Clinical Research Institute
michelle.smerek@duke.edu

https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-related-Abstracts

PCORnet Lab Reference Table on GitHub

This table is intended to be a dynamic resource containing LOINC codes that may map to PCORnet laboratory result name common measures.
The information contained in this table can be used to search for laboratory results in local data sources that may map to PCORnet laboratory result name common measures.
This table is a living document that will be updated when/if necessary. We welcome feedback from PCORnet data partners regarding questions or suggestions for additional codes that should be considered for inclusion.

Rows highlighted in red indicate LOINC codes that may **NOT** correspond to the PCORnet lab result name common measure listed. They are included for data partners to determine whether or not they are a good fit.

PCORnet LAB_NAME	LOINC code	Category or Name	Component/analyte	Property	Timing	System	Scale	Method	ExUnits	Common_Test_Rank
Troponin T cardiac (quantitative)	48425-3	Troponin T.cardiac [Mass/volume] in Blood	Troponin T.cardiac	MCnc	Pt	Bld	Qn		ug/L	
	6597-9	Troponin T.cardiac [Mass/volume] in Venous blood	Troponin T.cardiac	MCnc	Pt	BldV	Qn		ug/L	
	6598-7	Troponin T.cardiac [Mass/volume] in Serum or Plasma	Troponin T.cardiac	MCnc	Pt	Ser/Plas	Qn		ug/L	291
	67151-1	Troponin T.cardiac [Mass/volume] in Serum or Plasma by Detection limit <= 5 ng/L	Troponin T.cardiac	MCnc	Pt	Ser/Plas	Qn	Detection limit <= 5 ng/L	ng/L	
Hemoglobin A1c	41995-2	Hemoglobin A1c [Mass/volume] in Blood	Hemoglobin A1c	MCnc	Pt	Bld	Qn		g/dL	
	4548-4	Hemoglobin A1c/Hemoglobin.total in Blood	Hemoglobin A1c/Hemoglobin.total	MFr	Pt	Bld	Qn		%	81
	17855-8	Hemoglobin A1c/Hemoglobin.total in Blood by calculation	Hemoglobin A1c/Hemoglobin.total	MFr	Pt	Bld	Qn	Calculated	%	
	4549-2	Hemoglobin A1c/Hemoglobin.total in Blood by Electrophoresis	Hemoglobin A1c/Hemoglobin.total	MFr	Pt	Bld	Qn	Electrophoresis	%	
	17856-6	Hemoglobin A1c/Hemoglobin.total in Blood by HPLC	Hemoglobin A1c/Hemoglobin.total	MFr	Pt	Bld	Qn	HPLC	%	215
	62388-4	Hemoglobin A1c/Hemoglobin.total in Blood by JDS/JSCC protocol	Hemoglobin A1c/Hemoglobin.total	MFr	Pt	Bld	Qn	JDS/JSCC	%	
	71875-9	Hemoglobin A1c/Hemoglobin.total [Pure mass fraction] in Blood	Hemoglobin A1c/Hemoglobin.total	MFr.DF	Pt	Bld	Qn			
	59261-8	Hemoglobin A1c/Hemoglobin.total in Blood by IFCC protocol	Hemoglobin A1c/Hemoglobin.total	SFr	Pt	Bld	Qn	IFCC	mmol/mol	
Creatine kinase MB	49551-5	Creatine kinase.MB [Mass/volume] in Blood	Creatine kinase.MB	MCnc	Pt	Bld	Qn		ng/mL	
	51506-4	Creatine kinase.MB [Enzymatic activity/volume] in Cerebral spinal fluid by Electrophoresis	Creatine kinase.MB	CCnc	Pt	CSF	Qn	Electrophoresis	U/L	
	38482-6	Creatine kinase.MB [Presence] in Serum or Plasma	Creatine kinase.MB	ACnc	Pt	Ser/Plas	Ord			
	32673-6	Creatine kinase.MB [Enzymatic activity/volume] in Serum or Plasma	Creatine kinase.MB	CCnc	Pt	Ser/Plas	Qn		U/L	374
	2154-3	Creatine kinase.MB [Enzymatic activity/volume] in Serum or Plasma by Electrophoresis	Creatine kinase.MB	CCnc	Pt	Ser/Plas	Qn	Electrophoresis	U/L	
	13969-1	Creatine kinase.MB [Mass/volume] in Serum or Plasma	Creatine kinase.MB	MCnc	Pt	Ser/Plas	Qn		ng/mL; ug/L	111
Creatine kinase total	2156-8	Creatine kinase [Enzymatic activity/volume] in Amniotic fluid	Creatine kinase	CCnc	Pt	Amnio fld	Qn		U/L	
	50756-6	Creatine kinase [Mass/volume] in Blood	Creatine kinase	MCnc	Pt	Bld	Qn		ng/mL	
	34160-2	Creatine kinase [Presence] in Body fluid	Creatine kinase	ACnc	Pt	Body fld	Ord			
	16688-4	Creatine kinase [Enzymatic activity/volume] in Body fluid	Creatine kinase	CCnc	Pt	Body fld	Qn		U/L	
	2151-9	Creatine kinase [Enzymatic activity/volume] in Cerebral spinal fluid	Creatine kinase	CCnc	Pt	CSF	Qn		U/L	
	53433-9	Creatine kinase [Enzymatic activity/volume] in Dialysis fluid	Creatine kinase	CCnc	Pt	Dial fld	Qn		U/L	
	2157-6	Creatine kinase [Enzymatic activity/volume] in Serum or Plasma	Creatine kinase	CCnc	Pt	Ser/Plas	Qn		U/L	90
Creatine kinase MB/creatine kinase tota	20569-0	Creatine kinase.MB/Creatine kinase.total in Serum or Plasma	Creatine kinase.MB/Creatine kinase.total	CFr	Pt	Ser/Plas	Qn		%	297
	12189-7	Creatine kinase.MB/Creatine kinase.total in Serum or Plasma by calculation	Creatine kinase.MB/Creatine kinase.total	CFr	Pt	Ser/Plas	Qn	Calculated	%	
	12187-1	Creatine kinase.MB/Creatine kinase.total in Serum or Plasma by Electrophoresis	Creatine kinase.MB/Creatine kinase.total	CFr	Pt	Ser/Plas	Qn	Electrophoresis	%	1391
	72564-8	Creatine kinase.MB/Creatine kinase.total [Pure catalytic fraction] in Serum or Plasma by calculation	Creatine kinase.MB/Creatine kinase.total	CFr.DF	Pt	Ser/Plas	Qn	Calculated		
	72563-0	Creatine kinase.MB/Creatine kinase.total [Pure catalytic fraction] in Serum or Plasma by Electrophoresis	Creatine kinase.MB/Creatine kinase.total	CFr.DF	Pt	Ser/Plas	Qn	Electrophoresis		
	49136-5	Creatine kinase.MB/Creatine kinase.total [Ratio] in Serum or Plasma	Creatine kinase.MB/Creatine kinase.total	Ratio	Pt	Ser/Plas	Qn		%	211

https://github.com/CDMFORUM/CDM-GUIDANCE/blob/master/2016-June1_PCORnet%20Lab%20Common%20Measures_LOINC.xlsx

Practices of Engagement in Developing and Implementing the PCORnet Common Data Model

Michael Park, MMCi[1]; Daniella Meeker, PhD[2]; Keith Marsolo, PhD[3]; Michelle Smerek[1]; Maryan Zirkle, MD, MS, MA[4]; Lesley H. Curtis, PhD[1,5]; Jeffrey S. Brown, PhD[6]; Shelley A. Rusincovitch[7]

[1]Duke Clinical Research Institute, Durham, NC; [2] University of Southern California Keck School of Medicine, Los Angeles CA; [3]Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; [4]Patient-Centered Outcomes Research Institute, Washington, DC; [5]Duke University School of Medicine, Durham, NC; [6]Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; [7]Duke Translational Research Institute, Durham, NC

Introduction

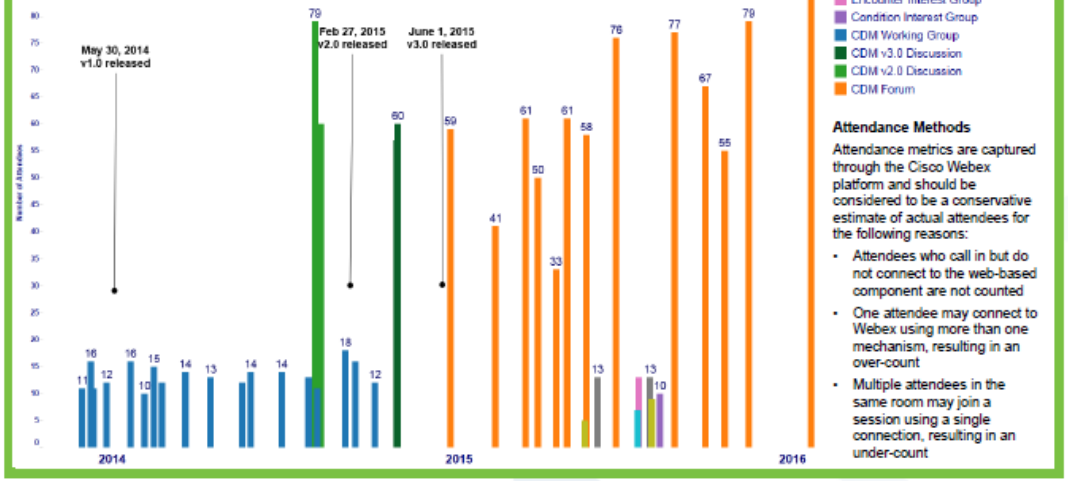
- PCORnet, the National Patient-Centered Clinical Research Network, has developed and implemented an infrastructure to conduct efficient multisite comparative clinical effectiveness research
- The Common Data Model (CDM) is a key part of the PCORnet infrastructure
- We engaged hundreds of stakeholders in CDM development and implementation
- Practices of engagement at this scale are challenging; we have iteratively developed and refined venues and practices over time

Methods

We have established 4 different venues:

CDM Modeling Working Group	Small expert working group instrumental in the early phases of data modeling
CDM Stakeholder Sessions	Targeted sessions focused on sharing themes and responses from CDM feedback cycles with the larger community
CDM Implementation Forum	Large venue intended to regularly update and connect individuals with activity in modeling, implementation, and operations support
CDM Interest Groups	Small, agile groups formed around specific topics; discussion is reported back to the larger Forum

Figure 1. Attendance at sessions related to CDM development and implementation in a 28-month timeframe. This period includes 3 cycles of CDM model development and release, followed by subsequent focus on implementation.



Attendance Methods

Attendance metrics are captured through the Cisco Webex platform and should be considered to be a conservative estimate of actual attendees for the following reasons:

- Attendees who call in but do not connect to the web-based component are not counted
- One attendee may connect to Webex using more than one mechanism, resulting in an over-count
- Multiple attendees in the same room may join a session using a single connection, resulting in an under-count

Results

Over 28 months, we have held **42 sessions with 1,374 attendances** (figure 1). This robust attendance demonstrates committed engagement within the PCORnet community.

QR code for CDM Specification URL:

http://www.pcornet.org/pcornet-common-data-model/

Lessons Learned and Conclusion

- **Complementary venues are valuable**; we have paired small, agile groups with larger forums and stakeholder sessions
- **Early stakeholder engagement** is important to articulate strategy and the process roadmap
- **Large volumes of feedback** can be difficult to articulate back to the full community; it is valuable to distill into themes and larger categories

Pragmatic development of standards intended to harmonize diverse data capture environments requires **engaging both data consumers and hands-on data experts** in **iterative cycles of specification and implementation.**

Contact Information

Shelley Rusincovitch, Project Leader in Applied Informatics & Architecture
Duke Translational Research Institute
shelley.rusincovitch@duke.edu
@rusincovitch

Acknowledgments The PCORnet Common Data Model is a product of shared expertise, feedback, effort, and collaboration from many people. We gratefully acknowledge their work and individual contributions. We thank Ania Berchuck and Luke Aseffa for their assistance in metrics and visualization for this poster. The project described was supported by the National Patient-Centered Clinical Research Network (PCORnet). Efforts were led by the Coordinating Center, managed by the Duke Clinical Research Institute, which is partially funded by the Patient-Centered Outcomes Research Institute (PCORI). The contents are solely the responsibility of the authors and do not not necessarily represent the official views of the Patient-Centered Outcomes Research Institute.

https://github.com/CDMFORUM/CDM-GUIDANCE/wiki/CDM-related-Abstracts

NLM Georgia Biomedical Informatics Course

- Intro to Biomedical Informatics
- Public Health Informatics
- TeleHealth
- Mathematical Modeling
- Funding Trends for Biomedical Informatics Research
- NLM Resources
- Genetics/Genomics
- Practical Bioinformatics for the Clinic
- Electronic Health Records, APIs, Apps
- Data Visualization
- NLP Methods
- Imaging Informatics
- Organizational Issues in Informatics: change management
- HIE and Meaningful Use
- Evidence Based Informatics
- Informatics Standards
- Controlled Vocabularies/Terminologies
- Disaster Health Informatics
- Biomedical Informatics, Data Science, and PMI
- Clinical Research Informatics and Data Management

pcornet®

http://www.augusta.edu/library/greenblatt/informaticscourse/

Considering Observation Stays

Live component: A quick introductory Webex poll on observation stays

1. Are you familiar with the concept of observation stays? (very familiar / moderately familiar / not familiar)

2. Do you think is it possible to identify observation stays - and consistently distinguish them from other encounter types - in your network's data? (very confident / moderately confident / low confident / I'm not sure / not applicable)

3. Do you think observation stays are a challenge in mapping your network's encounter data? (high challenge / moderate challenge / low challenge / I'm not sure / not applicable)

(The usual caveats about Webex polling apply here: this is an informal mechanism, and responses will not be a representative sample across networks. Still, this is a useful way for our discussion to get a sense of the landscape!)

pcornet®

A follow up from Monday's discussion

DRN OC Forum on Monday, October 3

Topics:

- Menu-Driven Query 1 and 2
- Data Characterization Cycle 1 Review
- Review of DataMart refreshes

Link to slides and recording:
https://pcornet.imeetcentral.com/p/ZgAAAAAAe5fc

ETL changes: ENCOUNTER

- Excluding encounters that do not meet the CDM definition
  - Domain description: “Encounters are interactions between patients and providers within the context of healthcare delivery” [1]
  - “Rollback or voided transactions and other adjustments should be processed before populating this table.” [1,2]
- Modifying encounter type for services that occur in an inpatient setting but cannot be combined with the inpatient stay [2]
- Modifying encounter type for observation stays [2]

1. Included in the CDM
2. Included in Implementation Guidance
3. Included in Data Checks
4. Included in ETL ADD

Revised May 2014

Are You a Hospital Inpatient or Outpatient?

If You Have Medicare – Ask!

Did you know that even if you stay in a hospital overnight, you might still be considered an "outpatient?" Your **hospital status** (whether the hospital considers you an "inpatient" or "outpatient") affects how much **you pay** for hospital services (like X-rays, drugs, and lab tests) and may also affect whether Medicare will cover care you get in a skilled nursing facility (SNF) following your hospital stay.

- You're an **inpatient** starting when you're formally admitted to a hospital with a doctor's order. The day **before** you're discharged is your last inpatient day.
- You're an **outpatient** if you're getting emergency department services, observation services, outpatient surgery, lab tests, X-rays, or any other hospital services, and the doctor **hasn't** written an order to admit you to a hospital as an inpatient. In these cases, you're an outpatient even if you spend the night at the hospital.

Note: Observation services are hospital outpatient services given to help the doctor decide if the patient needs to be admitted as an inpatient or can be discharged. Observation services may be given in the emergency department or another area of the hospital.

The decision for inpatient hospital admission is a complex medical decision based on your doctor's judgment and your need for medically necessary hospital care. An inpatient admission is generally appropriate when you're expected to need 2 or more midnights of medically necessary hospital care, but your doctor must order such admission and the hospital must formally admit you in order for you to become an inpatient.

From: https://www.medicare.gov/Pubs/pdf/11435.pdf

Live component: Now, a more detailed Webex poll on observation stays

1. Which encounter type do you currently assign observation stays to? (select all that apply: AV, ED, EI, IP, IS, OA, NI, UN, I'm not sure, Not applicable)
2. What are the potential lengths of stay in an observation unit? (select all that apply: <24 hours, 24-48 hours, >48 hours, Unknown, I'm not sure, Not applicable)
3. How often is Discharge Date populated? (Always / Sometimes / Never / Unknown / I'm not sure / Not applicable)
4. How often is Discharge Disposition populated? (Always / Sometimes / Never / Unknown / I'm not sure / Not applicable)
5. How often is Discharge Status populated? (Always / Sometimes / Never / Unknown / I'm not sure / Not applicable)
6. How often is Admitting Source populated? (Always / Sometimes / Never / Unknown / I'm not sure / Not applicable)

(The usual caveats about Webex polling apply here: this is an informal mechanism, and responses will not be a representative sample across networks. Still, this is a useful way for our discussion to get a sense of the landscape!)

pcornet

PCORnet Data Committee CDM Survey

Slide from August 19, 2016 CDM Forum

The PCORnet Data Committee Charter

I. PURPOSE

The PCORnet Data Committee (PDC) is charged by the PCORnet Council is to oversee PCORnet's data network, which is managed by the PCORnet Coordinating Center. The PDC will provide strategic guidance that ensures the data network is aligned with PCORI's overall goals and that the data network provides the semantic and syntactic interoperability to support PCORnet studies and sustainability. The PDC is also charged with stimulating informatics and research data innovations that advance the goals of PCORnet. (Section 1.1.6 PCORnet Governance Policy).

Learn more about the Data Committee on GitHub:
https://github.com/PCORnet/DataCommittee/wiki

pcornet®

Screenshot from the PCORnet Data Committee charter (page 1)
https://github.com/PCORnet/DataCommittee/blob/master/PCORnet%20Data%20Committee%20Charter%202016.pdf

Slide from August 19, 2016 CDM Forum

The PCORnet Data Committee and the CDM

VI.H: Develop annual updates of the Common Data Model

In collaboration with the Coordinating Center, create the processes and governance that will be used to specify annual updates to information model and domains that are instantiated in the Common Data model, which may include optional modules or new domains. These proposals will be presented to the Council for approval. Included in this process will be the factors used to evaluate whether to include a domain (e.g., effort needed to acquire data, analytic value, ability to standardize, etc.), the updates' cost/sustainability, and methods to gather input and evaluate work from external stakeholders (e.g. NIH/CTSA, FDA, HHS ONC, ODHSI, patients / families, healthcare systems, insurers, medical device, and pharmaceutical companies). The PDC will also be responsible for developing the governance needed to surface and harmonize CDM implementation decisions by individual institutions or networks, with the goal of facilitating high-quality, reproducible research. These efforts will ensure that the structure of the data model continues to serve the purpose of PCORnet and allows the most efficient conduct of research.

pcornet®

Screenshot from the PCORnet Data Committee charter (page 7)
https://github.com/PCORnet/DataCommittee/blob/master/PCORnet%20Data%20Committee%20Charter%202016.pdf

Development of a CDM Survey

- The Data Committee would like to understand experiences by networks who have extended the CDM
- The Data Committee will review results from the survey during our face-to-face meeting on September 22
- The survey results will be shared with the full PCORnet community

PCORnet Data Committee F2F meeting on September 22, 2016

pcornet®

Thank you for contributing!

- Thirty-seven survey responses were analyzed: 31 affiliated with a CDRN and 6 affiliated with a PPRN
- These survey results represent **responses from all CDRNs and 5 of 20 PPRNs**
- Link to original survey questions: https://duke.qualtrics.com/jfe6/preview/SV_3ruGC078iEYUsap

pcornet®

Table 1. Respondent characteristics.

	CDRN	PPRN
Number of respondents who confirmed that they are currently standardizing data elements outside of the CDM v3.0 (% of all respondents)	15 (40.5%)	5 (13.5%)
Number of respondents who offered details of their standardized data elements outside of the CDM v3.0 (% of all respondents)	15 (40.5%)	5 (40.5%)
Number of respondents who offered suggestions on additional data elements for inclusion in the CDM (% of all respondents)	15 (40.5%)	3 (8.1%)
Number of respondents who have structured data sources available that are not captured in the CDM (% of all respondents)	12 (32.4%)	4 (10.8%)

Table 2. Classification of additional data sources reported available by survey respondents.

1. Community metrics	Community Vital Signs data: geocoding of patients, facilities, providers
2. Payer, provider information	Provider NPI Taxonomy Classes, Provider Crosswalk, payer information, type of delivery care facility, risk scores, CESR Benefits, and NPPES NPI database and provider domain
3. Patient-Reported Outcomes (PROs)	Additional harmonized PRO measures (e.g. PHQ2, PHQ9, DAST, AUDIT C, AUDIT)
4. External registries	PINNALCE, AJRR, Cancer Registry, ORIEN project, Asthma, Press Gainey, Stewardship, Trauma Registry, NSQIP, STS Database, VQI, UNOS, ELOS, Robot registry – not otherwise specified, Rectal cancer registry, Breast cancer registry, Thoracic registry, Mayo Clinic Cancer Registry (C/NExt), Genomic test results, HCSRN Tumor Registry and Census data, CNExT for NAACCR tumor registry data, OnCore for Clinical Trials and Tissue Bank registries, Legacy CARRA Registry (full registry dataset) [i2b2]
5. Public health data	HRSA quality metrics (e.g. cervical cancer screening, colorectal cancer screening), Public Health Reporting: Antimicrobial Use, Antimicrobial Resistance
6. Natural Language Processing (NLP)-extracted data	NLP Output, Knowledge extracted form clinical notes
7. Other standard data models for operations or research	OMOP, PR-COIN (full quality improvement dataset) [proprietary format]
8. Other	mHealth data, Standardized list of ICD and Procedure Codes with Code name, CESR Infusion, CESR Pregnancy Outcomes (in development)

Next steps

- Additional insights from survey
- Data Committee CDM Workgroup development of strategic recommendations

pcornet®

CDM Forum v3.1 Interest Group

Slide from August 19, 2016 CDM Forum (updated)

Invitation to participate in the v3.1 interest group

- This group's expert feedback (drawn from the full PCORnet community) will be used to help shape the draft specification, prior to the full feedback cycle
- We will hold 2 sessions in September:
  - Thursday, September 8
  - Monday, September 19
- Please join us! To request a calendar invite, please send an e-mail to Ania Berchuck (ania.berchuck@duke.edu)

We are here

- We will provide updates from the interest group during the full Forum on October 5

pcornet®

CDM Forum v3.1 Interest Group meetings

Many thanks to the people who contributed!

September 8 session

- Link to meeting summary and recording: https://pcornet.imeetcentral.com/p/ZgAAAAAAe0fn

September 19 session

- Link to meeting summary and recording: https://pcornet.imeetcentral.com/p/ZgAAAAAAe0fp

pcornet®

Defining a CDM "minor release"

Here's how we defined it in the v3.0 specification

2.3. History of Releases and Modifications

Note on version conventions: Major releases are denoted with whole number incrementation (eg, v1.0, v2.0, v3.0). Minor releases are denoted with decimal incrementation (eg, v1.1, v1.2) and will be used for bug fixes and minor adjustments.

Reference Table: History of Releases

Version	*Date of Release*	*Description of Release*
v1.0	2014-05-30	The DSSNI Task Force thanks the many individuals who provided thoughtful feedback, comments, and suggestions for this first release of the PCORnet CDM. A special thanks to members of the task force who volunteered to serve on the CDM working group.
v2.0	2015-02-27	The v2.0 release includes: • Four new tables (DISPENSING, CONDITION, PRO_CM, LAB_RESULT_CM) • Four new fields in existing tables (VITAL.TOBACCO, VITAL.TOBACCO_TYPE, PROCEDURE.PX_TYPE, PROCEDURE.PX_SOURCE) • Additional guidance and descriptions
v3.0	2015-06-01	Please note: New and modified fields have been indicated in blue to assist with visually scanning the document (in addition to the descriptive comments). The v3.0 release includes: • Five new tables (PRESCRIBING, PCORNET_TRIAL, DEATH, DEATH_CAUSE, and HARVEST) • Ten new fields in existing tables (DISPENSING.DISPENSINGID, DISPENSING.PRESCRIBINGID, VITAL.VITALID, VITAL.SMOKING, CONDITION.CONDITIONID, CONDITION.ONSET_DATE, PRO_CM.PRO_CM_ID, DIAGNOSIS.DIAGNOSISD, PROCEDURES.PROCEDURESID, LAB_RESULT_CM.LAB_RESULT_CM_ID) • Modification to relational integrity specifications • Modification to date formatting practices • New specifications specific to SAS data types • Additional guidance, clarifications, and descriptions
v3.0	2015-07-29	Document updated with licensing information and new PCORnet.org URL. **No technical specifications have been modified.**

http://www.pcornet.org/pcornet-common-data-model/

Defining a CDM "minor release"

pcornet

Common Data Model (CDM) Specification, Version 3.1 – DRAFT

Draft v3.1 specification feedback cycle October 3–14, 2016

We welcome feedback and comments from the PCORnet Community for the draft v3.1 specification during the open feedback cycle, October 3 – 14, 2016. Please see the announcement on iMeetCentral for more details. Following the processes established in prior versions[1], all comments will be collated, categorized, and reviewed.

More specific scope for the v3.1 draft

The PCORnet Common Data Model v3.1 will be a minor release

The distinction between minor and major releases is especially important to datamart custodians because it allows an estimation of impact, dependencies, and change management processes for technical processes such as database administration, SAS file administration, and related programming. It also affects the PCORnet core operational and analytic infrastructure such as data characterization, query fulfillment, and analytic tools.

The concept of major and minor releases is well-established in software development practices, but because it is a broad concept, it is helpful to define a specific scope for PCORnet CDM v3.1 as:

1. Proposed changes in v3.1 will be **parsimonious**, and higher value will be placed on changes that are recognized as a result of current activity and experiences.
2. The v3.1 minor release **may include** modification to existing value sets, table-level constraints, and/or new fields.
3. The v3.1 minor release **will not include** any new tables.
4. Proposed changes in v3.1 will clearly distinguish where a modification would result in a semantic change from the earlier v3.0 (ie, where the meaning of the data measure itself would be changed and should be carefully recognized in analysis).

With thanks to the CDM v3.1 interest group for their feedback on this definition.

Live component: A quick introductory Webex poll on the parseable file

1. Are you familiar with the Common Data Model parseable file? (very familiar / moderately familiar / not familiar)

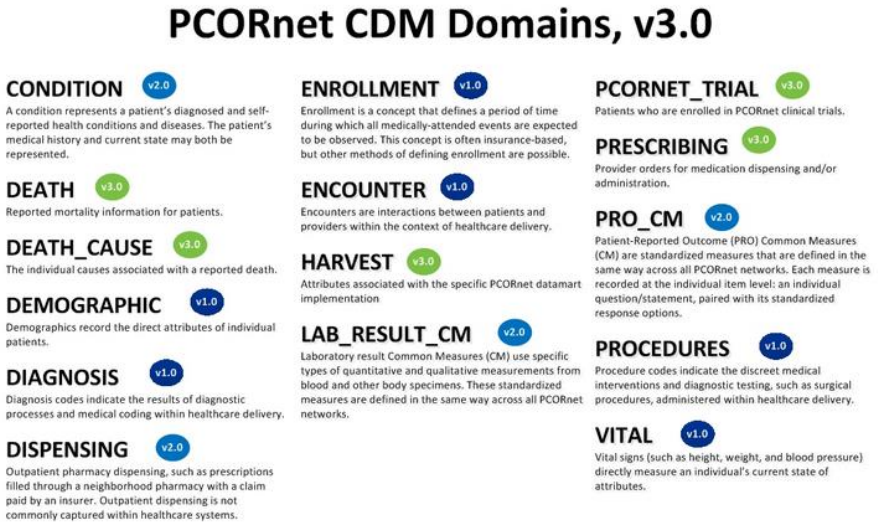
2. Have you (or your technical team) used the Common Data Model parseable file? (yes / no / I'm not sure / not applicable)

The usual caveats about Webex polling apply here: this is an informal mechanism, and responses will not be a representative sample across networks. Still, this is a useful way for our discussion to get a sense of the landscape!

pcornet®

CDM Specifications

- 2015-07-29 PCORnet Common Data Model v3.0
- 2015-06-01 PCORnet Common Data Model v3.0 – parseable
- View useful tools for the CDM, such as the CDM-ERRATA and CDM-GUIDANCE issue trackers, on the PCORnet GitHub CDM Forum.



CDM Lay Guide

The Common Data Model 3.0 Lay Guide is a brief executive summary. It uses less technical, and more everyday language to define the CDM, its purpose, history, and key fields and specifications.

PCORnet CDM Lay Guide [October 2015]

The CDM v3.0 parseable file is part of the CDM specifications at: http://www.pcornet.org/pcornet-common-data-model/

pcornet®

	A	B	C	D	E	F	G	H	I	J	K
1	TABLE_NAME	FIELD_NAME	RDBMS_DATA_TYPE	SAS_DATA_TYPE	DATA_FORMAT	REPLICATED	UNIT_OF_MEASURE	VALUE_SET	VALUE_DESCRIPTION	DEFINITION	CDM_ORDER
2	DEMOGRAPHIC	PATID	RDBMS Text(x)	SAS Char(x)		NO				Arbitrary person-level identifier used to link across tables. PATID is a pseudoidentifier with a consistent crosswalk to the true identifier retained by the source Data Partner. For analytical data sets requiring patient-level data, only the pseudoidentifier is used to link across all information belonging to a patient. The PATID must be unique within the data source being queried. Creating a unique identifier within a CDRN would be beneficial and acceptable. The PATID is not the basis for linkages across partners.	1
3	DEMOGRAPHIC	BIRTH_DATE	RDBMS Date	SAS Date (Numeric)		NO	DATE			Date of birth.	2
4	DEMOGRAPHIC	BIRTH_TIME	RDBMS Text(5)	SAS Time (Numeric)	HH:MI using 24-hour clock and zero-padding for hour and minute	NO	TIME			Time of birth.	3
5	DEMOGRAPHIC	SEX	RDBMS Text(2)	SAS Char(2)		NO		A;F;M;NI;UN;OT	A = Ambiguous F = Female M = Male NI = No information UN = Unknown OT = Other	Administrative sex. v2.0 guidance added: The "Ambiguous" category may be used for individuals who are physically undifferentiated from birth. The "Other" category may be used for individuals who are undergoing gender re-assignment.	4
6	DEMOGRAPHIC	HISPANIC	RDBMS Text(2)	SAS Char(2)		NO		Y;N;R;NI;UN;OT	Y = Yes N = No R = Refuse to answer NI = No information UN = Unknown OT = Other	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. v2.0 amendment: The new categorical value of "Refuse to answer" has been added.	5
7	DEMOGRAPHIC	RACE	RDBMS Text(2)	SAS Char(2)		NO		01;02;03;04;05;06;07;NI;UN;OT	01 = American Indian or Alaska Native 02 = Asian 03 = Black or African American 04 = Native Hawaiian or Other Pacific Islander 05 = White 06 = Multiple race 07 = Refuse to answer NI = No information UN = Unknown OT = Other	Please use only one race value per patient. Details of categorical definitions: American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. Asian: A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. Black or African American: A person having origins in any of the black racial groups of Africa. Native Hawaiian or Other Pacific Islander: A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.	6
									Y = Yes	Flag to indicate that one or more biobanked specimens are stored and available for research use. Examples of biospecimens could include plasma, urine, or tissue. If biospecimens are available, locally maintained "mapping tables" would be necessary to map between the DEMOGRAPHIC record and the originating biobanking system(s).	

INFO | FIELDS | RELATIONAL | CONSTRAINTS | REF 1 (LAB-LOINC) | REF 2 (LAB-CPT) | REF 3 (LAB ABBREV) | REF 4 (PRO-CM)

Purpose of the CDM parseable file: a machine-readable format

Purpose of the CDM specification PDF: a human-readable format

DEMOGRAPHIC Table Specification

Field Name	*RDBMS Data Type*	*SAS Data Type*	*Predefined Value Sets and Descriptive Text for Categorical Fields*	*Definition / Comments*	*Source*
BIRTH_DATE	RDBMS Date	SAS Date (Numeric)	.	Date of birth.	MSCDM
BIRTH_TIME	RDBMS Text(5): Format as HH:MI using 24-hour clock and zero-padding for hour and minute	SAS Time (Numeric)	.	Time of birth.	PCORnet Source of time format: ISO 8601
SEX	RDBMS Text(2)	SAS Char(2)	A=Ambiguous F=Female M=Male NI=No information UN=Unknown OT=Other	Administrative sex. v2.0 guidance added: The "Ambiguous" category may be used for individuals who are physically undifferentiated from birth. The "Other" category may be used for individuals who are undergoing gender re-assignment.	MSCDM with modified field size and value set Source: Administrative Sex (HL7) http://phinvads.cdc.gov/vads/ViewValueSet.action?id=06D34BBC-617F-DD11-B38D-00188B398520
HISPANIC	RDBMS Text(2)	SAS Char(2)	Y=Yes N=No R=Refuse to answer NI=No information UN=Unknown OT=Other	A person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin, regardless of race. v2.0 amendment: The new categorical value of "Refuse to answer" has been added.	MSCDM with modified field size and value set Compatible with "OMB Hispanic Ethnicity" (Hispanic or Latino, Not Hispanic or Latino)

pcornet®

Planned modifications to parseable file format

- We will add a new sheet for VALUE SETS (one row per value set item

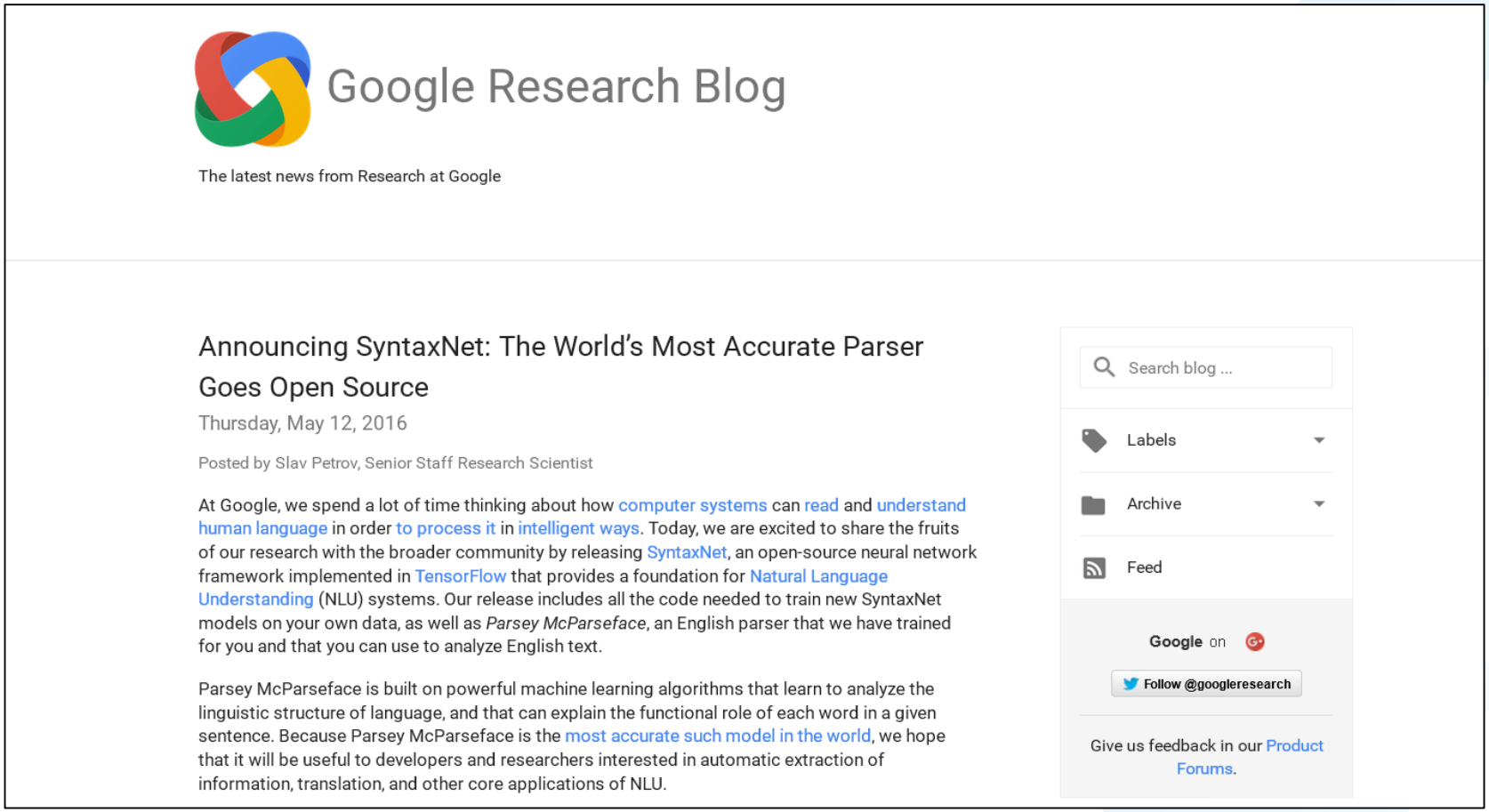
- Lab mapping reference tables will move into the Implementation Guidance documentation

Goals:

- Make the parseable file a valuable resource for creating database scripts for any relational database management system (RDBMS). Each RDBMS (eg, Oracle, SQL Server, Postgres) uses different syntax to create its scripts, so a specification in a common machine-readable format is important.
- Support sharing of database scripts (such as DDL and ETL); for example: https://datahandbook.epic.com/Reports/Details/9000400

pcornet®

Pending question: is “parseable file” the best way to label our resource?



Google Research Blog

The latest news from Research at Google

Announcing SyntaxNet: The World’s Most Accurate Parser Goes Open Source

Thursday, May 12, 2016

Posted by Slav Petrov, Senior Staff Research Scientist

At Google, we spend a lot of time thinking about how computer systems can read and understand human language in order to process it in intelligent ways. Today, we are excited to share the fruits of our research with the broader community by releasing SyntaxNet, an open-source neural network framework implemented in TensorFlow that provides a foundation for Natural Language Understanding (NLU) systems. Our release includes all the code needed to train new SyntaxNet models on your own data, as well as *Parsey McParseface*, an English parser that we have trained for you and that you can use to analyze English text.

Parsey McParseface is built on powerful machine learning algorithms that learn to analyze the linguistic structure of language, and that can explain the functional role of each word in a given sentence. Because Parsey McParseface is the most accurate such model in the world, we hope that it will be useful to developers and researchers interested in automatic extraction of information, translation, and other core applications of NLU.

Search blog ...

Labels

Archive

Feed

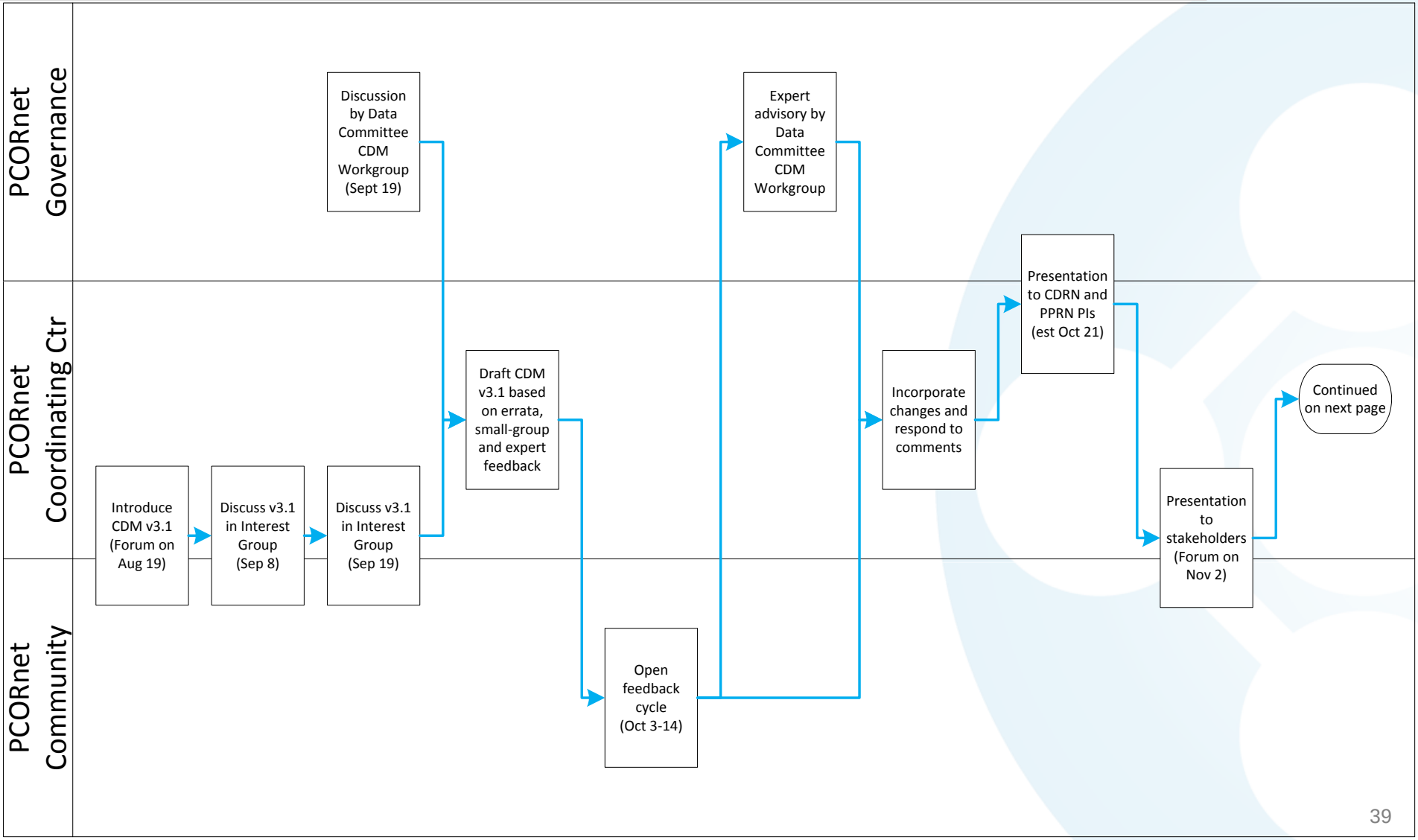
Google on

Follow @googleresearch

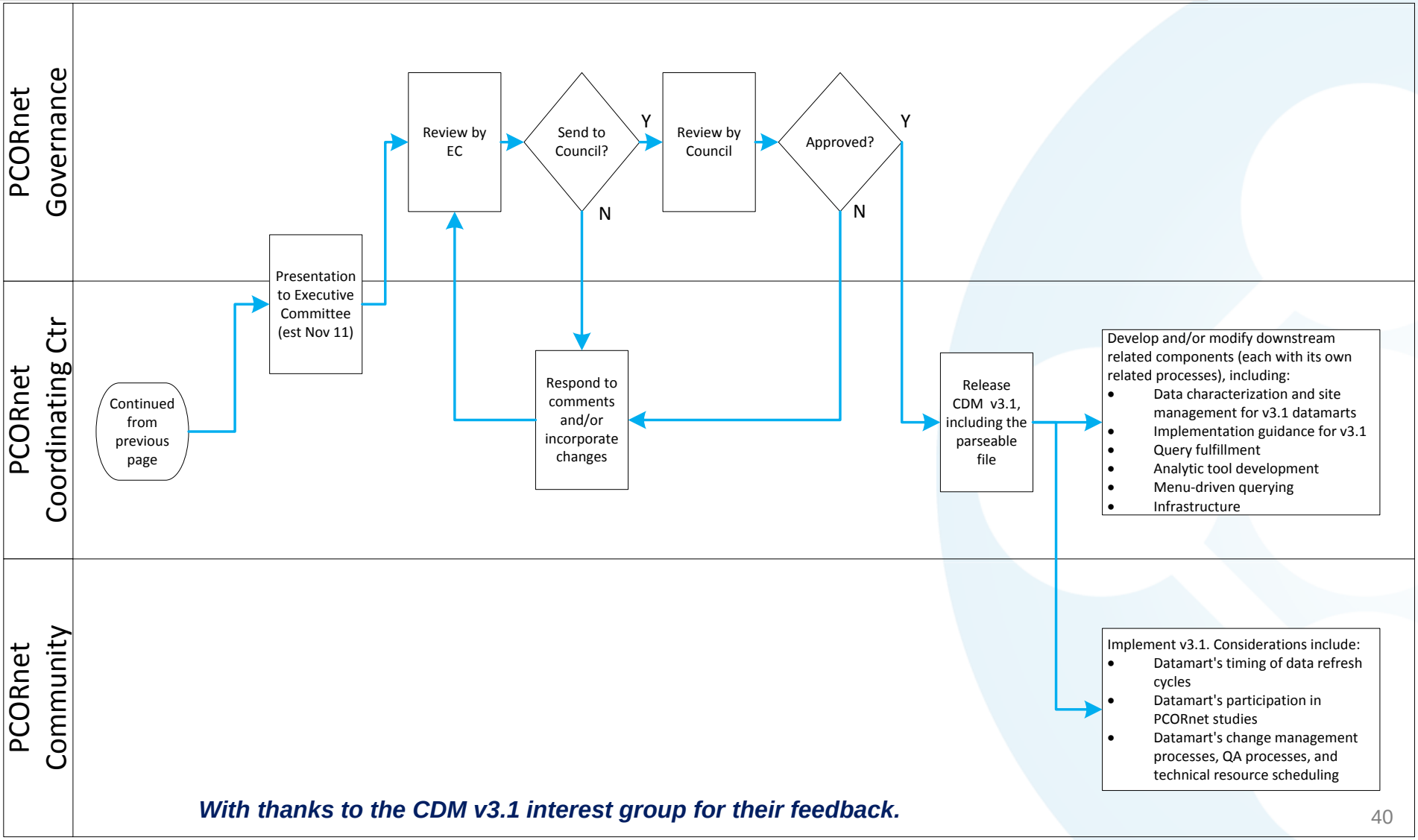
Give us feedback in our Product Forums.

(Yes, “Parsey McParseface” does actually exist!)

Process diagram (part 1 of 2)



Process diagram (part 2 of 2)



With thanks to the CDM v3.1 interest group for their feedback.

PCORnet CDM v3.1 Development

Common Data Model v3.1 development

- A minor release (v3.1) will be developed **in parallel** with development of the larger CDM strategy (more on this in the next section!)
- A **minor** release (by definition) does **not include** major new functionality
- We expect that the current CDM errata will be a primary source (https://github.com/CDMFORUM/CDM-ERRATA)
- We will build on successes in Phase I, and v3.1 offers an opportunity to further refine processes

Important dates for v3.1

Date	*Topic*
August 19	CDM Forum
August 19–September 2	Data Committee CDM survey
September 8	v3.1 interest group meeting (1 of 2)
September 19	v3.1 interest group meeting (2 of 2)
October 3–14	Feedback period on v3.1 draft specification
October 5	CDM Forum: Updates from v.31 interest group
October 21	Discuss with CDRN and PPRN PIs
November 2	CDM Forum: Stakeholder session to review v3.1 feedback, responses, and incorporation of changes
November 11 (tentative)	Present to PCORnet Executive Committee
Mid-November	Review by PCORnet Council

We are here

Goal for CDM v3.1: approval & release by the end of 2016! But…

pcornet®

To learn more about PCORnet governance:
http://pcornet.org/about-pcornet/pcornet-committees-members/

But a CDM release date isn't the same as a network implementation date

- PCORnet does not expect networks datamarts to implement CDM v3.1 in 2016
- Part of v3.1 development will be to assess implementation expectations
- We expect considerations for implementation will include datamart refresh cycles and study participation
- Please keep in mind that v3.1 will be a **minor** release; therefore, it will **not** include new tables

pcornet®

We invite you to contribute to the v3.1 feedback cycle (open now!)

We welcome feedback and comments from the PCORnet Community for the draft v3.1 specification during the open feedback cycle, October 3 – 14, 2016.

Following the processes established in prior versions, all comments will be collated, categorized, and reviewed. Each will be assigned an "action" along with notes for action, and responses will be shared and posted on iMeetCentral.

Please use the feedback spreadsheet template to submit feedback by COB on Friday, October 14 to Ania Berchuck (ania.berchuck@duke.edu).

Thank you in advance for your important feedback and contributions to this feedback cycle!

pcornet

Materials are posted on iMeetCentral:
https://pcornet.imeetcentral.com/drnoc-workgroups/doc/47801510/

Overview of changes in v3.1

Modifications to model include:

- RDBMS numbers have been modified to incorporate explicit references to both precision and scale, and new description added in section 3.1.
- Added new fields in DEMOGRAPHIC for SEXUAL_ORIENTATION and GENDER_IDENTITY.
- ENROLLMENT.ENR_BASIS value set modified to include drug coverage enrollment.
- The domain definition for ENROLLMENT has been updated because the table will now support drug coverage enrollment. The previous definition of "a period of time during which all medically-attended events are expected to be observed" does not apply to drug coverage.
- New DX_ORIGIN field added in DIAGNOSIS table
- PROCEDURES.PX_TYPE and LAB_RESULT_CM.LAB_PX_TYPE: The value set was updated to create one procedure code type for CPT and HCPCS codes. The categories are mutually exclusive, and data partners had significant difficulty in differentiating the appropriate subclassification (for example, the difference between CPT-4 (ie, HCPCS Level I) vs. HCPCS Level II).
- PRO_CM.PRO_ITEM character length increased to accommodate the potential for additional ontologies.
- New definition for CONDITION.ONSET_DATE added.
- The domain definition for PRESCRIBING has been updated to explicitly state that orders may take place in any setting, including the inpatient or outpatient basis.
- PRESCRIBING.RX_BASIS data element definition expanded.
- The RXNORM_CUI field type has been modified to become character, not numeric, so that the CDM harmonizes with the NLM's RxNorm database definition.
- The DEATH table constraint has been modified to remove DEATH_DATE as a required field, given that more than one data partner has reported the situation of having records without a date. Cohort identification is an important use case for incorporating a general mortality status even if a death date is not known.

Modifications to documentation include:

- New and modified content has been indicated in green to assist with visually scanning the document (in addition to the descriptive comments).
- Significant content related to implementation guidance, including the LAB_RESULT_CM reference tables, has been moved into the separate Implementation Guidance document. These changes are not indicated in green because they are extensive.
- Header of "Source" modified to "Data element provenance" to more clearly describe the purpose of this attribute.
- Data element provenance of MSCDM has been qualified as MSCDM v4.0 to more precisely reference the version.
- Added "Important Links and References" in table of contents section.
- The parseable file (machine-readable version of the CDM specification) will be created after the specification is approved. The structure of this file will be updated with feedback from the v3.1 interest group.

pcornet®

Page 2 of the draft v3.1 specification:
https://pcornet.imeetcentral.com/drnoc-workgroups/doc/47801510/

Wrap Up

Recap (the shortlist):

- Please consider contributing to the feedback cycle for the draft v3.1 specification: **October 3–14**
  - https://pcornet.imeetcentral.com/drnoc-workgroups/doc/47801510/
- Next CDM Forum: **November 2**
  - Stakeholder session for CDM v3.1 feedback cycle review

pcornet®

References and links

- DRN OC home page on iMeetCentral: https://pcornet.imeetcentral.com/p/aQAAAAAB6T9b
- CDM errata issue tracker: https://github.com/CDMFORUM/CDM-ERRATA/issues
- CDM guidance issue tracker: https://github.com/CDMFORUM/CDM-GUIDANCE/issues
- PCORnet diagnostic query package: https://github.com/PCORnet-DRN-OC/PCORnet-Diagnostic-Query
- PCORnet data characterization query package: https://github.com/PCORnet-DRN-OC/PCORnet-Data-Characterization
- PCORnet Data Committee on GitHub: https://github.com/PCORnet/DataCommittee

pcornet®

pcornet®

The National Patient-Centered Clinical Research Network