

Developing a Common Data Model: The PCORnet[®] Experience

Darcy M. Louzao, Ph.D.

March 24, 2018



pcornet[®]

The National Patient-Centered
Clinical Research Network

Questions to Answer Today

- ❁ What is the goal of PCORnet?
- ❁ What is PCORnet's data strategy?
- ❁ How has PCORnet's CDM been developed? And under what key principles?
- ❁ What are the major challenges and how do you navigate them?

PCORnet: the National Patient-Centered Clinical Research Network

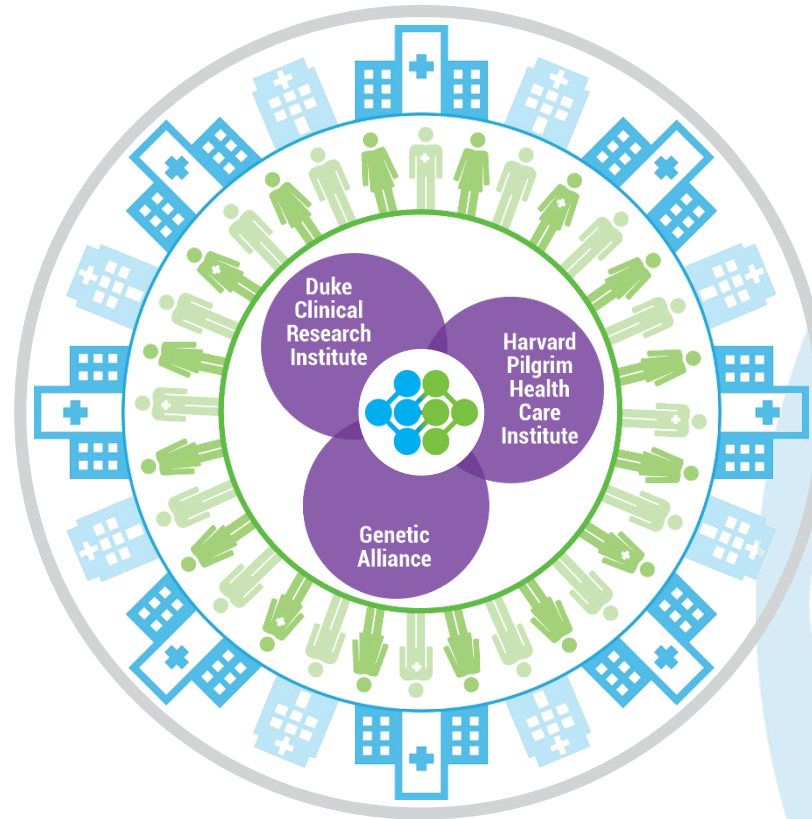
PCORnet is **a large, highly representative, national patient-centered clinical research network.**

Our **vision** is to support a learning U.S. healthcare system and to enable **large-scale clinical research** conducted with **enhanced quality and efficiency.**

Our **mission** is to enable people to make informed healthcare decisions by efficiently conducting clinical research relevant to their needs.



PCORnet[®] embodies a “network of networks” that harnesses the power of partnerships



A national infrastructure for people-centered clinical research

20

Patient-Powered Research Networks (PPRNs)


+

13

Clinical Data Research Networks (CDRNs)

+

2

Health Plan Research Networks (HPRNs)

+

1

Coordinating Center

=

How do we do this? The Data story

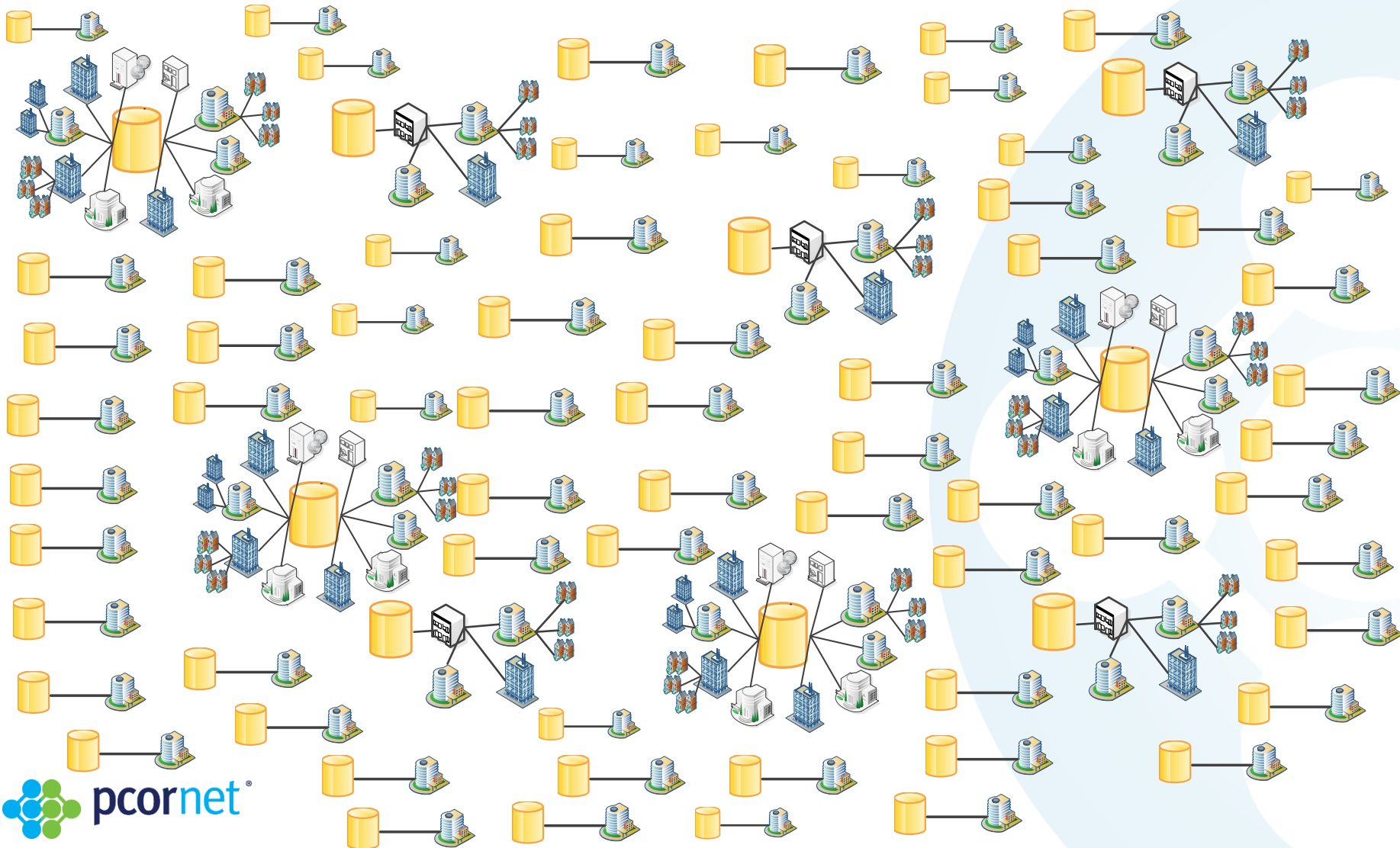
Enabling Research at a National Scale

How do you ask a research question at hundreds of institutions and get back results you can trust?

Option 1 — Write a description and have everyone create a local implementation to run on their data

Option 2 — Create an algorithm that can run against a single, common data model

Density and Heterogeneity of data partners



PCORnet Data Strategy

- Standardize data into a common data model
- Focus on data quality: data curation
- Operate a secure distributed query infrastructure
 - Develop re-usable tools to query the data
 - Send questions to the data and only return required information
- Learn by doing and repeat

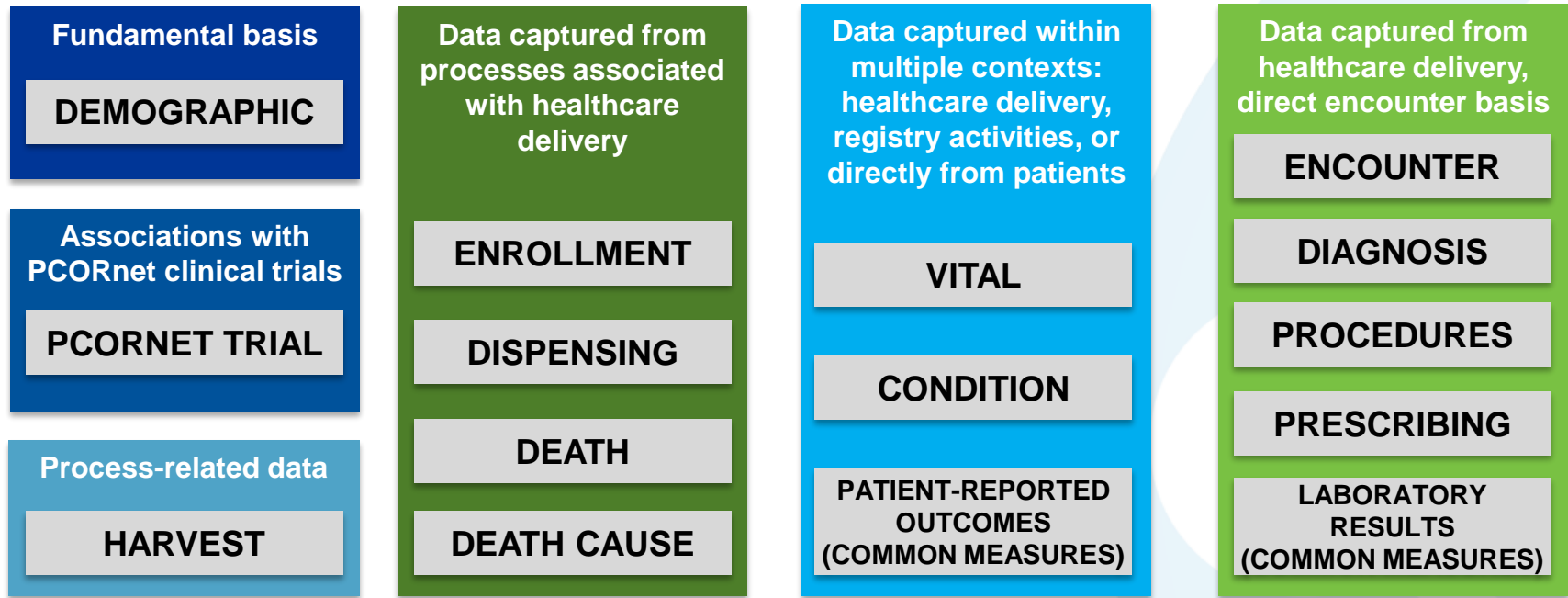
CDM overview

- Strong basis of **secondary data generated by healthcare delivery** processes
- Heavily derived from the **Mini-Sentinel Common Data Model** and **informed by other initiatives**
- Development takes a **pragmatic approach**, supporting the **current state** of these data
- Modeling is **optimized for analytic functionality** and to be **intuitive to investigators**

CDM Guiding Principles (abridged)

- ⊕ CDM will contain data of general interest to PCORnet
- ⊕ Not all networks are expected to populate all parts of CDM
- ⊕ CDM will change over time
- ⊕ CDM will be intuitive & easy to understand
- ⊕ Other CDMs exist – PCORnet will learn from them
- ⊕ CDM will reflect values found in source data
- ⊕ Networks are free to add data/domains to their local CDMs

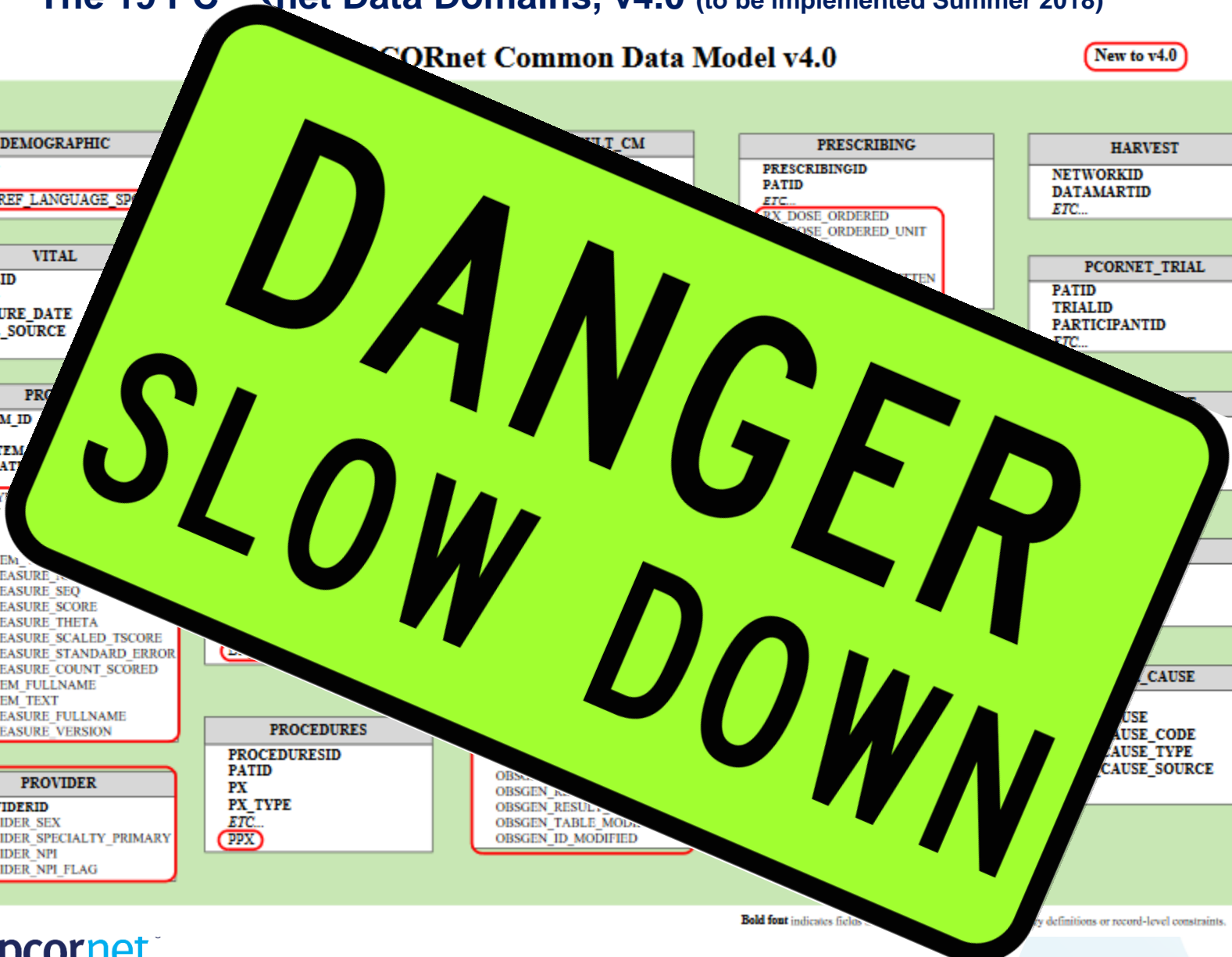
Standardize to a common data model



The 19 PCORnet Data Domains, v4.0 (to be implemented Summer 2018)

PCORnet Common Data Model v4.0

New to v4.0

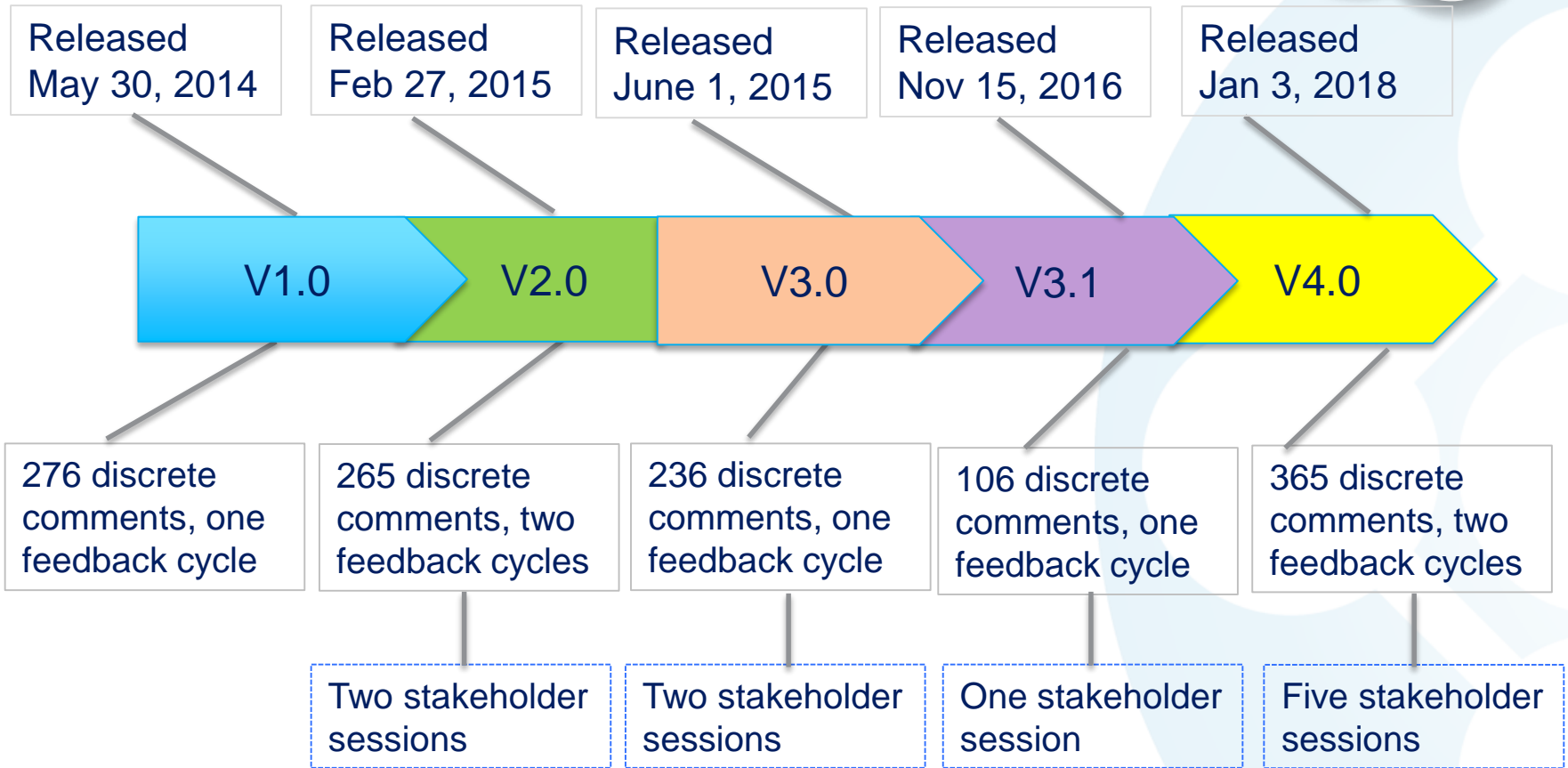


DEMOGRAPHIC	PRESCRIBING	HARVEST
PATID ETC... PAT_PREF_LANGUAGE_SPC	PRESCRIBINGID PATID ETC... DOSE_ORDERED DOSE_ORDERED_UNIT	NETWORKID DATAMARTID ETC...
VITAL		PCORNET_TRIAL
VITALID PATID MEASURE_DATE VITAL_SOURCE ETC...		PATID TRIALID PARTICIPANTID ETC...
PROC		
PRO_CM_ID PATID PRO_ITEM PRO_DATE ETC... PRO_TYPE PRO_ITEM PRO_F PRO_ PRO_ITEM PRO_MEASURE PRO_MEASURE_SEQ PRO_MEASURE_SCORE PRO_MEASURE_THETA PRO_MEASURE_SCALED_TSCORE PRO_MEASURE_STANDARD_ERROR PRO_MEASURE_COUNT_SCORED PRO_ITEM_FULLNAME PRO_ITEM_TEXT PRO_MEASURE_FULLNAME PRO_MEASURE_VERSION		
PROVIDER	PROCEDURES	
PROVIDERID PROVIDER_SEX PROVIDER_SPECIALTY_PRIMARY PROVIDER_NPI PROVIDER_NPI_FLAG	PROCEDURESID PATID PX PX_TYPE ETC... PPX	
	OBSGEN_ID OBSGEN_RESULT OBSGEN_TABLE_MOL OBSGEN_ID_MODIFIED	
		CAUSE
		CAUSE_CODE CAUSE_TYPE CAUSE_SOURCE

Bold font indicates fields that are new to v4.0 or have definitions or record-level constraints.

CDM Development Timeline

5 versions
in 3.5 years



Each release of the CDM has a cascading effect on the Network

- 🌐 Coordinating Center
 - Plan updates to data curation (DC) procedures, Implementation Guidance, analytic tools, etc.
- 🌐 Network Partners
 - Refresh data using current CDM
 - Provide updated DC results on current CDM
 - Respond to queries using current CDM
 - Develop extract-transform-load (ETL) procedures on *new* CDM
 - Populate data into new CDM
- 🌐 First 3 versions of CDM occurred during period of minimal query activity

Evaluate data quality and fitness-for-use



- Data quality checks
- Annotated data dictionaries
- Implementation guidance
- Data curation and CDM implementation forums
- Analyses and reports

Develop reusable tools to query the data



There are many _____ ways to choose a/an _____ to read. First, you could ask for recommendations from your friends and _____ . Just don't ask Aunt _____—she only reads _____ books with _____-ripping goddesses on the cover. If your friends and family are no help, try checking out the _____ Review in *The _____ Times*. If the _____ featured there are too _____ for your taste, try something a little more low-_____. Like _____: *The _____ Magazine*, or _____ *Magazine*. You could also choose a book the _____-fashioned way. Head to your local library or _____ and browse the shelves until something catches your _____. Or, you could save yourself a whole lot of _____ trouble and log on to www.bookish.com, the _____ new website to _____ for books! With all the time you'll save not having to search for _____, you can read _____ more books!

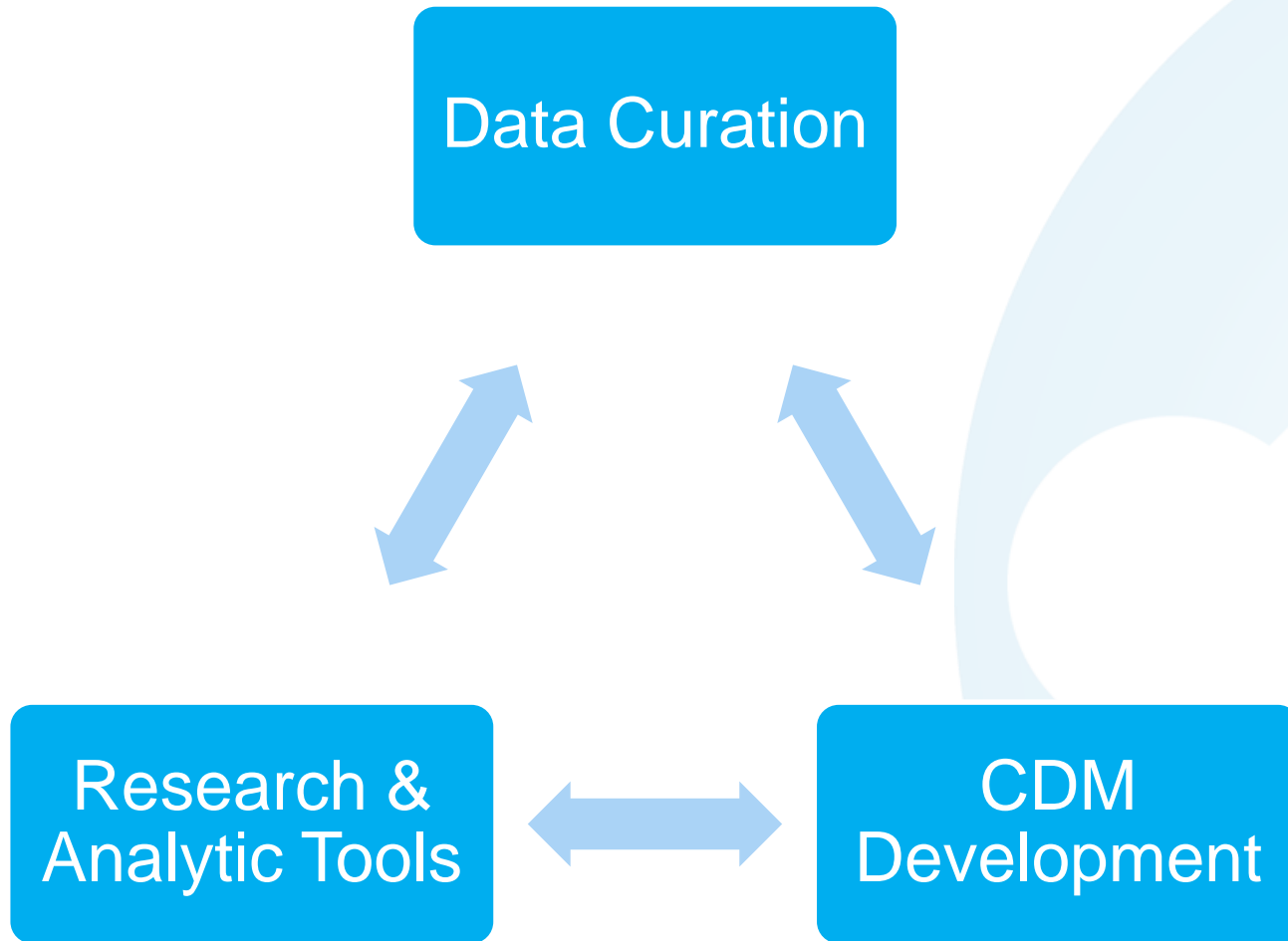
www.madlibs.com • www.madlibsgames.com
© 2014 Penguin Group (USA) Inc. All rights reserved. MAD LIBS and all related trademarks, logos, and devices are trademarks owned and licensed for use by Penguin Group (USA) Inc.

Study Design:

Identify patients <AGE RANGE> with a <NEW / REFILL> dispensing of a <DRUG CLASS>. To be eligible, patients must have met the following criteria in the <LOOKBACK PERIOD> days before the index dispensing: (1) continuous enrollment in <TYPE(S) OF INSURANCE> benefits, (2) no prescription for <DRUG CLASS> or <DRUG CLASS>, and (3) no diagnosis of <DISEASE / CONDITION> in <INPATIENT / OUTPATIENT / ED> care setting.

The primary outcome of interest is <EVENT> identified with <DIAGNOSIS CODE> in <PRIMARY / SECONDARY / ANY> position during an <INPATIENT / OUTPATIENT / ED> encounter.

Learn by doing



What are the challenges and how to navigate?

- ❁ Common data model paralysis
 - Establish guiding principles
 - Don't start from scratch
 - Refine over time, but not too quickly

- ❁ Establishing the rules of the road

Who gets to send queries? How are they sent? Is everyone required to answer every queries? How are results returned? Can query results be shared publicly? Published?

- Begin with a useful, relatively simple use case
- Build trust
- Refine over time

What are the challenges and how to navigate?

- 🌐 Developing analysis code that runs in every environment is challenging
 - Develop reusable programs that are structured to answer common types of questions
 - Engage local experts in the network community
- 🌐 Sustainability of infrastructure

Extract, transform, and load (ETL) of data into a common data model is real work. When source data change, the ETL changes. When the CDM changes, ETL and tools have to change.

 - Be mindful of the cascade and proceed carefully
 - Develop useful products

Thank you!

Questions?