



# QUESTIONS & ANSWERS ABOUT THE VIRTUAL DATA WAREHOUSE

**Roy Pardee**  
**Group Health Research Institute**  
**HMORN VOC Co-Lead**

# Overview

- **Agenda**

- What VDW Is/Isn't
- Brief Hx
- Use Overview + Benefits
- Survey Existing Implementations
- Documentation & Reference
- Common pitfalls
- Governance & Communications
- Discussion/Q&A

- **Feel free to interrupt with questions.**

- But please mute when not speaking
- And please don't put us on hold!

# The VDW is...

- *a series of dataset standards **and** automated processes in place at each of 15 HMORN sites,*
- *that allow **SAS programs written at one HMORN Site to be run against all the others***
  - *quickly, and*
  - *with a minimum of site-specific customization.*

# It is *Not*...

- **A centralized database.**
  - No secret bunker where you can get at all the data at once.
  - Data stay at the originating sites.
- **A means for fully automating data-based research.**
  - VDW is very human-mediated.
- **A replacement for the local data you already know & love.**

# Brief History

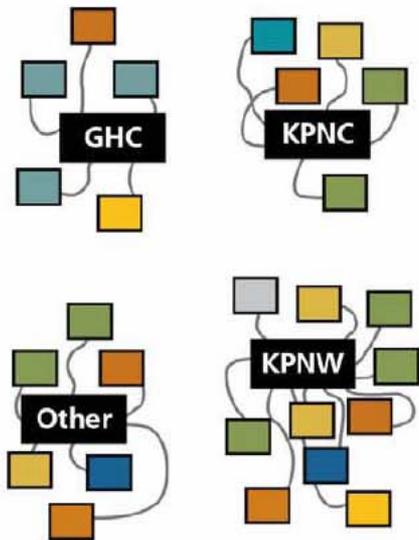
- **Pre-2002: No VDW. Collaborative studies had each site doing its own custom programming.**
  - At project end—toss the data (!)
  - Expensive & Inefficient
- **2002: Cancer Research Network uses infrastructure \$ to**
  - develop standardized dataset specifications,
  - fund the effort of reshaping local data into these standard specifications.

# Brief History (cont)

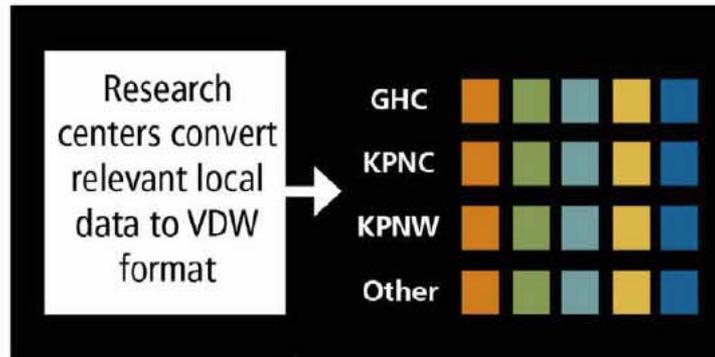
- **2009: Seeing its benefits & general applicability, HMORN took on governance of the VDW.**
- **Site VDW activities funded variously by project infrastructure + Indirect \$.**
- **HMORN VDW Operations Committee (VOC) formed to manage & coordinate:**
  - Specifications
  - Documentation
  - QA Work

# How does it work?

## Individual Health Care Systems Administrative and claims data



## Advance Work Enormous gains in efficiencies and data quality are made through investments in advance work.



Lessons learned

# What are the benefits?

- Eases collaborative data-based research.
- Standard macros & utility code.
- Depending on your existing assets, could be your first unified research data warehouse: one-stop shopping.
- Data vetted through repeated use & the bug reporting/remediation process.

At its best, VDW is *a vehicle for preserving & cumulating the benefits / fruits of project-specific work.*

# How Much Data Is There?

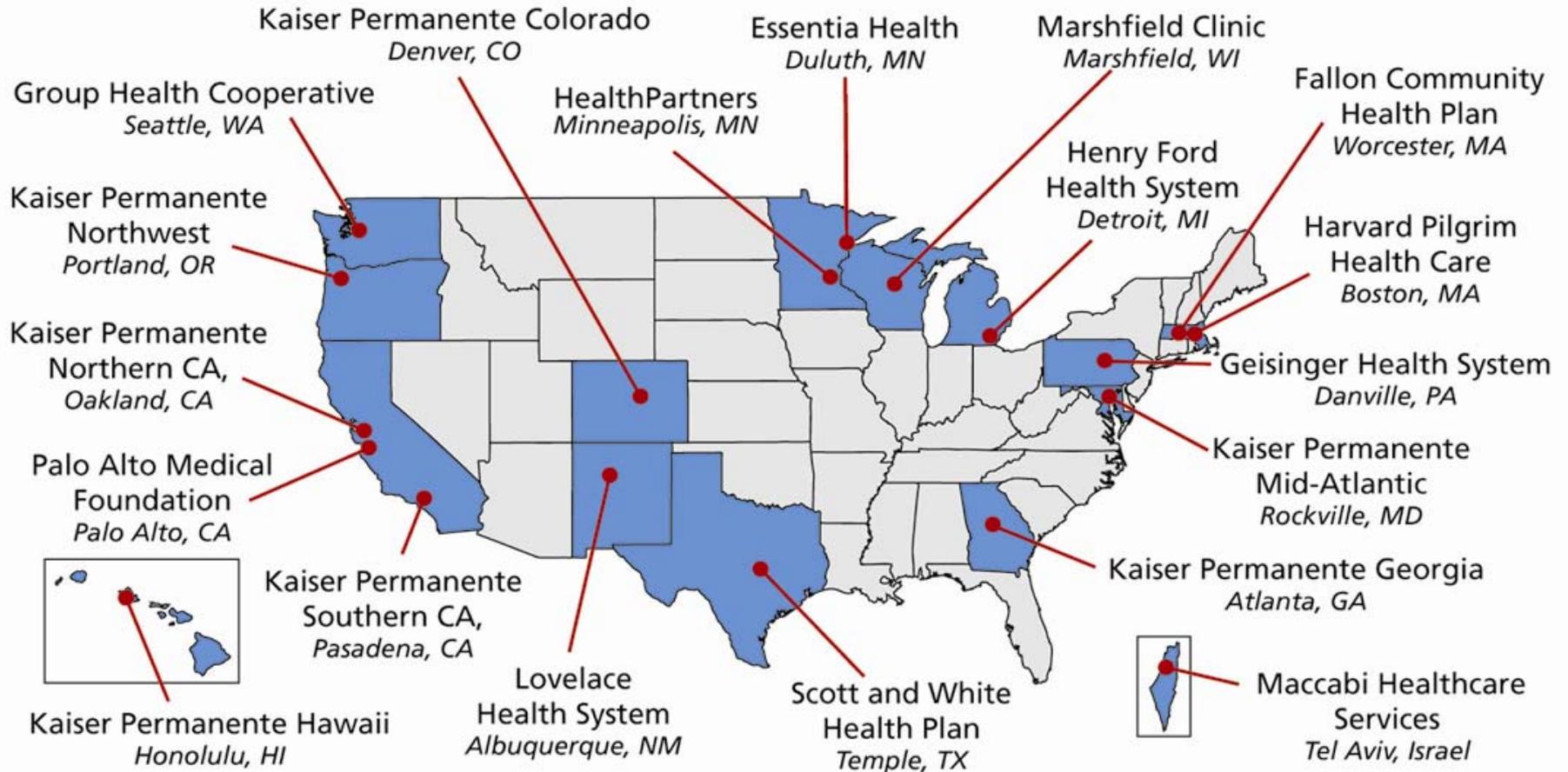
- **15 HMORN sites have implemented one or more VDW datasets.**
- **Currently 8 data areas, comprising 15 datasets.**

1. Tumor
2. Utilization (encounters)
3. Enrollment/Demographics
4. Rx Fills
5. Lab Results
6. Vital Signs
7. Census
8. Death

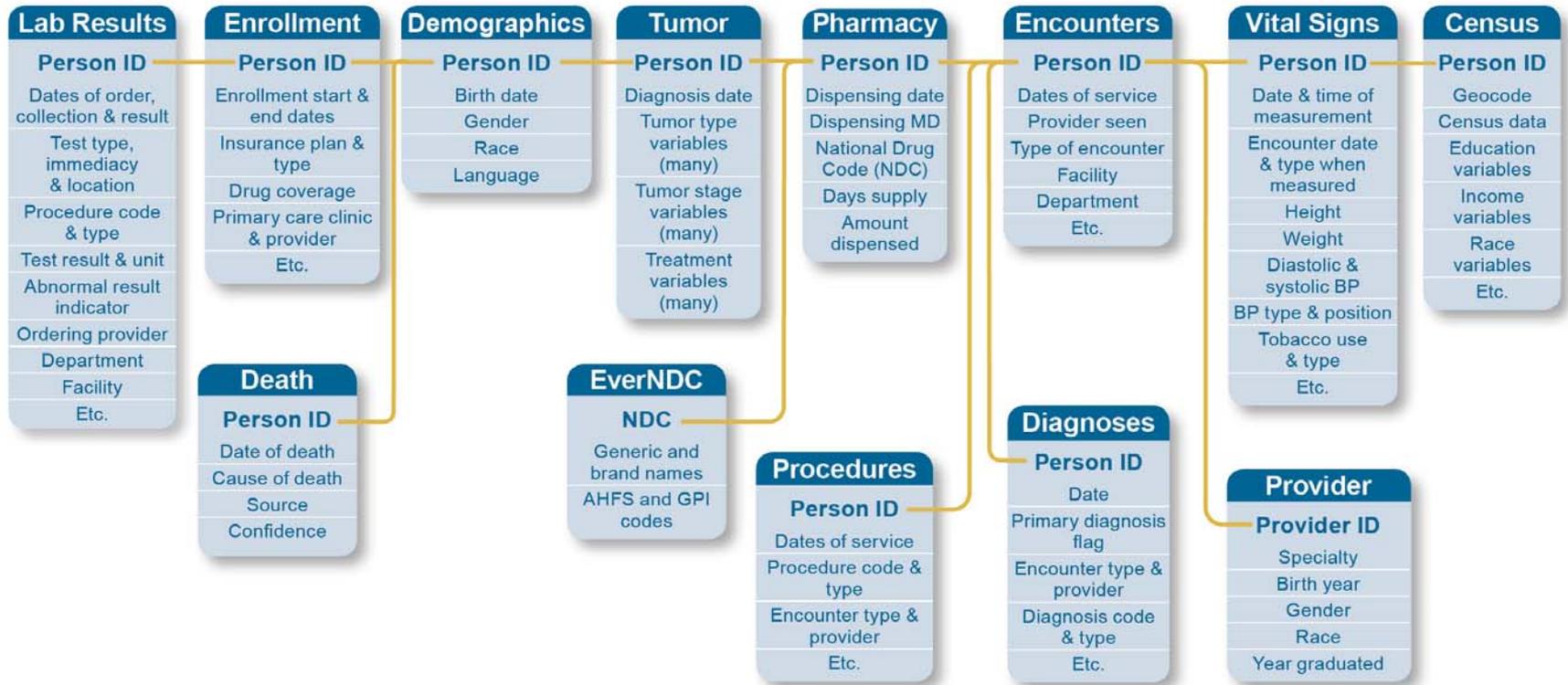
Coming Soon:

1. Infusion (chemo)
2. Orders
3. Dental

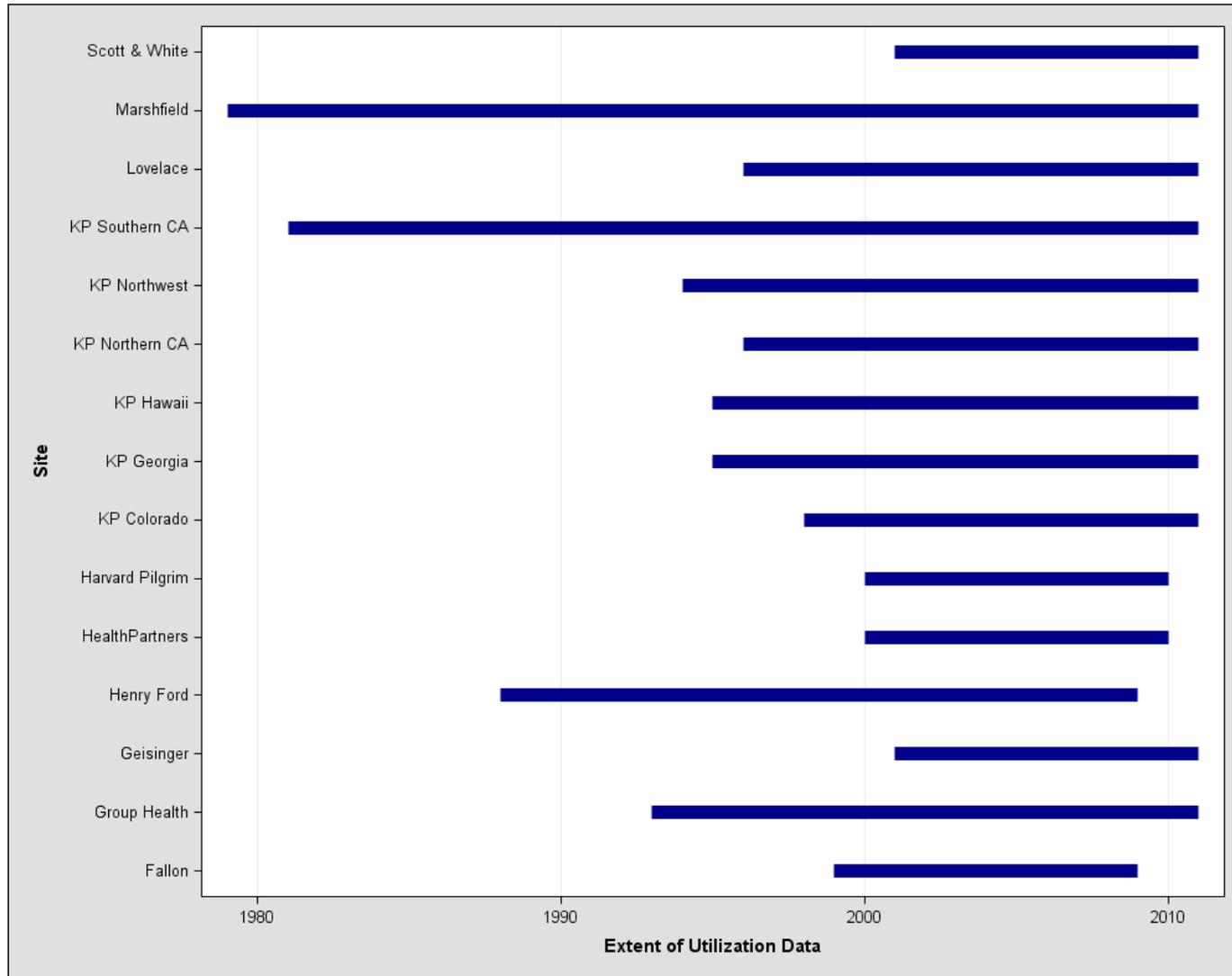
# HMORN Sites



# VDW Data Areas



# Over what span of time?



# What types of people/data?

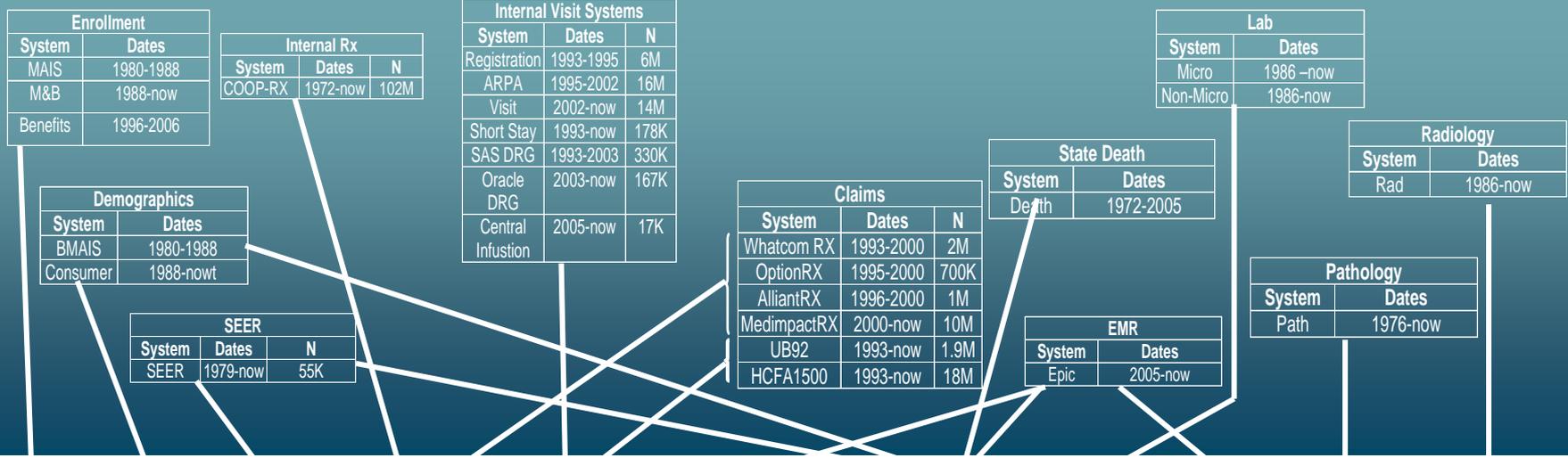
- **Only HMO members? NO.**
- **Only Insured people? NO.**
- **Only (modern integrated) EMR data? NO.**
- **In general, we use data from all compatible sources, including:**
  - Legacy (pre-modern EMR) internal systems
  - Claims (rx, facility, professional)
  - EMR

# Group Health VDW Data Sources

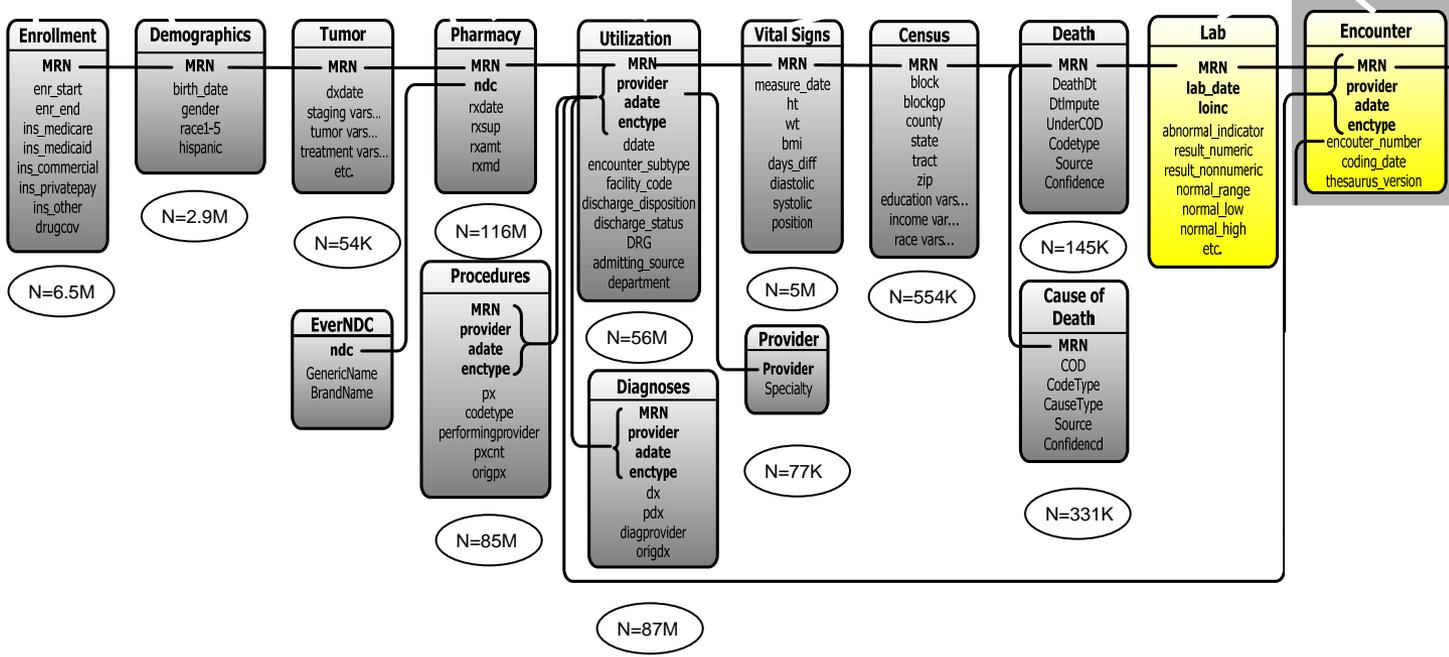
Gene Hart, Roy Pardee, Tyler Ross



LEGACY



VDW



# Where is VDW Documented?

- **Primarily On the CRN Portal**
  - Implementation Overview
    - Dataset Specifications (left column)
    - Site implementation pages (cells)
  - Site Data Managers folder
    - list of
    - File Implementation guidelines
  - Programmer's Guide
  - Issue Tracker

# Additional Material of Interest

- **HMORN.org's Collaboration Toolkit has a wealth of information, including some great presentations:**
  - Getting Your Questions Answered with the VDW
  - Using the VDW
  - VDW Tutorial for Programmers
- **Note:**
  - CRN Portal requires a login, which requires affiliation w/an HMORN member organization.
  - HMORN.org is entirely open to the public.

# Using the VDW

- **Whose permission do I need?**
  - People at the Sites whose data you want.
    - Typically you find Investigators at the sites to collaborate or sponsor your project.
    - Those people navigate whatever IRB/Compliance requirements are in place locally.
- **Who can write a VDW program?**
  - Any reasonably skilled SAS programmer with access to
    - The dataset specifications,
    - (preferably) a set of local VDW files, and
    - [the programmer's guide](#).
  - There is no central programming service or authority.

# Using (cont)

- **Who will run my program?**
  - Again, people at the sites whose data you want.
  - Take-home: budget for some site staff.
  - Some large projects/networks have something like a blanket IRB approval for more trivial requests (counts for feasibility) and standing staff to run things.

# Common Pitfalls Using VDW

- **Biggest is assuming that all HMORN sites are like yours.**
  - Utilization comes completely from claims, or from non-claims.
  - We only treat people we insure.
  - Vast majority of inpatient stays are at external hospitals.
  - We have our own Tumor registry.

# More Pitfalls

- **More “like me” assumptions:**
  - We run the EPIC EMR, and have the Clarity reporting system.
  - Chemo data winds up in Procedures data, not Pharmacy.
- **Assuming all variables will be populated over all time (e.g., Race in Demographics is patchy).**
- **Assuming uniformity of data across sites.**
  - VDW does *some* cleaning/normalization, but not as much as most people assume.

# Still more pitfalls

- **Assuming <<some data element you need>> is available in VDW.**
  - Study the specifications!
- **In general:**
  - Go slow.
  - Explore.
  - Take nothing for granted.
  - Don't hesitate to consult with [Dan](#) & [Roy](#).

# What do I need to implement VDW here?

- **Programmers: intermediate-or-better in SAS or other ETL tools.**
- **SAS Software**
  - Required: BASE and STAT, possibly ACCESS (if RDBMS back-end).
  - Recommended: CONNECT, GRAPH
- **Some sites use an RDBMS (e.g. Oracle, MS SQL Server) for actual data storage—not required.**
- **Hardware**
  - Needs here vary wildly—happy to consult.

# How is VDW Governed?

- **The VDW Operations Committee (VOC) is a consortium of selected Investigators and Programmers from various Research Networks (CRN, CESR, etc.).**
  - Charged with maintaining and setting priorities for the VDW in consultation with HMORN Asset Stewardship.
  - Two co-leads: Dan Ng (KPNC) and Roy Pardee (Group Health) & one Administrative Coordinator: Sarah McDonald (Group Health).
  - One Workgroup per data area (almost) each of which has one Investigator lead and one Technical lead.
- **The VDW Implementation Group (VIG) is the set of Site Data Managers and interested programmers from all sites.**
  - A support group for implementers.
  - Charged with vetting technical/data issues, including especially the feasibility of proposed spec changes.

# How is VDW Governed?

## VOC Organizational Chart



# How is the VDW governed?

- **The VOC Workgroups**

- Steward their specifications

- Correct any errors

- Clarify anything unclear

- Take suggestions for spec improvements & periodically submit to VIG for ratification for spec revision.

- Consult with implementers.

- Generate QA programs and periodically synthesize & report results.

- **Where there is interest, new working groups can be chartered.**

# Communications

- **Conference Calls**

- VIG Conference Calls: Fourth Tuesdays @ 12:00-1:00 pm Pacific Time.
- Workgroups: various—documented on CRN Portal on [this page](#).

- **E-Mail Lists**

- One for VDW Users.
- One for VIG.
- One per Workgroup.
- Subscribe to any or all on [this page](#).

- **We need you!**

- Nearly all of this is volunteer effort.
- Please consider joining one or more Workgroups.

# Thank You!

[pardee.r@ghc.org](mailto:pardee.r@ghc.org)