For example, polygon can be drawn to emphasize the contour of certain areas so that audience's attention can be led to certain areas; legend, made by the annotation of text and symbols, can be added to wherever you want in the map and with format that are not in standard SAS legend templates; images can be overlaid with transparent background so that the map is not obscured by the background of image. **Results:** These SAS macros for mapping VDW data have the advantages of creating highly customized maps by reusable code without extra cost of specialized mapping software or training.

Keywords: Mapping; VDW; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-41

PS2-42:

Developing High-Quality Laboratory Results for the Virtual Data Warehouse: The Importance of Single-site Quality Assessment

Andrew Sterrett¹; Marian Bailey¹; Gwyn Saylor²; Marsha Raebel¹

¹Kaiser Permanente Colorado; ²Kaiser Permanente Northwest

Background/Aims: Assessment of data incorporated into the Virtual Data Warehouse (VDW) is crucially important when building a new content area such as laboratory results. At KPCO, we identified two main challenges while constructing tables of laboratory results for the VDW. The first was ensuring that we constructed clinically meaningful laboratory test types. The second was confirming that we identified and classified the correct set of records from source data repositories. The aim of our work was therefore to ensure that these two challenges were successfully met **Methods:** Initial priority tests were broadly categorized as chemistry, hematology, microbiology, or challenge tests. Many laboratory results could be characterized into a single test type. Other laboratory results required one or more of the following characteristics to differentiate subtypes: assay method; test subtypes such as isoenzymes; result type; fasting status; specimen site; and challenge dose and time since dose. Laboratory results were extracted from two data repositories. The first database includes records from 1999-2004. The second database stores records since 2005 with partial backfill of records from 2004 and earlier. Variable names and values did not follow a common naming convention between the two data sources. Extraction and classification rules were based on Logical Observation Identifiers Names and Codes (LOINC)s, test names, result units, and Current Procedural Terminology (CPT) codes. **Results:** We identified numerous issues while developing the VDW laboratory content area: incomplete clinical information; distinct laboratory types with similar names; missing, incorrect, obsolete, or site-specific component codes; incorrect sample collection times; changes in source data streams; clinical practices that differ from guidelines; and little background knowledge of laboratory tests among those processing the data. Each challenge and the approach to addressing each issue will be illustrated with an example. **Conclusions:** A systematic approach to collection and processing of laboratory results can yield high-quality, clinically useful data. Key elements of this systematic approach include a logical and flexible taxonomy of laboratory types and careful consideration of identifier codes and supplemental information to categorize laboratory results. Developers should expect a variety of deviations in the data from expected patterns.

Keywords: Quality; Laboratory Results; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-42

PS2-43:

VDW Data Sources: Group Health

Roy Pardee¹; Gene Hart¹; William Tolbert¹; Dustin Key¹; Tyler Ross¹

¹Group Health Cooperative

Background: The Virtual Data Warehouse (VDW) was created as a mechanism for producing comparable data across sites for purposes of proposing and conducting research. It is "virtual" in the sense that the data remain at the local sites; there is no multi-site physical database at a centralized data coordinating center. At the core of the VDW are a series of standardized file definitions. Content areas and data elements that are commonly required for research studies are identified, and data dictionaries are created for each of the content areas, specifying a common format for

each of the elements—variable name, label, description, code values, and value labels. Local site programmers have mapped the data elements from their HMO's data systems into this standardized set of variable definitions, names, and codes, as well as onto standardized SAS file formats. This common structure of the VDW files enables a SAS analyst at one site to write one program to extract and/or analyze data at all participating sites.

Methods: This poster demonstrates the wide range of data sources used at Group Health to feed information into our local implementation of the VDW datasets. **Results:** The Group Health local implementation of the VDW contains detailed medical information on Group Health members. These files contain details on 143 million pharmacy dispensings (1977-2011), nearly 98 million unique medical encounters (1993-2011), including .75 million hospitalizations, 68 million ambulatory visits, 141 million diagnoses, and 224 million procedures. We have some 11 million Vital Signs observations, and 82 million lab results. The VDW Enrollment and Demographic files are derived from several historical and current membership files; the VDW Pharmacy and utilization files are derived from internal Group Health systems plus claims files; the VDW tumor data is derived from the Puget Sound SEER Registry. **Conclusions:** The VDW at Group Health provides an easily employed unified central repository of data from all available source files. This resource enables the sharing of compatible data in multi-site studies, and also improves programming efficiency, accuracy, and completeness for local single site studies by expending resources to link these legacy systems only once.

Keywords: VDW; Datasources; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-43

PS2-44:

VDW Data Sources: Kaiser Permanente Northern California

Srivardhan Chimmula¹; Ranjeeta Dhuru¹; Bruce Folck¹; Jamila Gul¹; Mei Lee¹; Daniel Ng¹; Wei Tao¹; Stas Yurkevich¹

¹Kaiser Permanente Northern California

Background: The Virtual Data Warehouse (VDW) was created as a mechanism for producing comparable data across sites for purposes of proposing and conducting research. It is "virtual" in the sense that the data remain at the local sites; there is no shared multi-site physical database at a centralized data coordinating center. At the core of the VDW are a series of standardized file definitions. Content areas and data elements that are commonly required for research studies are identified, and data dictionaries are created for each of the content areas, specifying a common format for each of the elements—variable name, label, description, code values, and value labels. Local site programmers have mapped the data elements from their HMO's data systems into this standardized set of variable definitions, names, and codes, as well as into standardized SAS file formats. This common structure of the VDW files enables a SAS analyst at one site to write one program to extract and/or analyze data at all participating sites.

Methods: This poster demonstrates the wide range of data sources used at Kaiser Permanente Northern California (KPNC) to feed information into our local implementation of the VDW datasets. **Results:** The KPNC local implementation of the VDW contains detailed medical information on KPNC members. These files contain details on 350 million pharmacy dispensings (1997-2011), 310 million unique medical encounters (1996-2011), including 1.7 million hospitalizations, 107 million ambulatory visits, 650 million diagnoses, and 224 million procedures. We have 173 million Vital Signs observations, and 530 million lab results. The VDW Enrollment and Demographic files are derived from several historical and current membership files; the VDW Pharmacy and utilization files are derived from internal KPNC system; the VDW tumor data is derived from the KPNC SEER Registry. **Conclusions:** The VDW at KPNC provides an easily employed unified central repository of data from all available source files. This resource enables the sharing of compatible data in multi-site studies, and also improves programming efficiency, accuracy, and completeness for local single site studies by expending resources to link these legacy systems only once.

Keywords: Clinical Trial; Tracking Database; Virtual Data Warehouse
doi:10.3121/cmr.2012.1100.ps2-44