A Knowledge Framework for Representing Family Health History
Jonathan S. Einbinder, MD, MPH1, Wendy F. Cohn, PhD1, James R. Barrett, PhD1, Sandra L. Pelletier, PhD1, Anthony J. Wenzel, MSEE2, Madaline Harrison, MD1, Susan Miesfeldt, MD1, Margaret Reitmeyer, MD1, Amy Tucker, MD1, Bradford Worrall, MD1, William A. Knaus, MD1

1University of Virginia Health System, Charlottesville, VA
2Dominion Digital, Inc., Charlottesville, VA

BACKGROUND
Current methods for collecting and storing family health history (FHx) information are not adequate to exploit the rapidly expanding understanding of the genetic basis for many conditions. In practice, many providers do not even collect or document detailed information on FHx. Also, most electronic patient records and health risk assessment instruments do not support detailed documentation of FHx in a structured form. In part, this reflects limitations of controlled vocabularies, which may allow coded diagnoses for family history of selected conditions (e.g. V17.1 “Family History of Stroke” in ICD-9CM) or append family history as a modifier to other findings or diagnoses (e.g. G-0002 “Family history of” in SNOMED). Neither approach is flexible or detailed enough to permit an assessment of an individual’s risk for a heritable condition. What is needed is a representation that encodes the rationale for why the FHx is “positive,” preserving specific information on family members and enabling an assessment of the importance or severity of the FHx.

METHODS
We have developed a web-based application for collecting FHx directly from patients.1 Underlying this application is a representation that uses detailed information about family members to build a family history profile for the participant. This representation uses a three-step process to collect and organize information. (Figure 1)

1. Detailed information about family members
   "Classification Algorithms"
2. List of conditions in individual’s FHx
   "Risk Logic"
3. Estimation of individual’s risk

Collect detailed FHx for each family member
For each of four disease areas (cardiovascular, neurologic, endocrine, cancer), clinical domain experts constructed disease trees, which were used to guide development of questions about FHx. The questions accommodate participant responses of varying levels of detail, depending on the participant’s knowledge about family members’ diagnoses, e.g. “stroke” (least specific), “intracerebral hemorrhage” (more specific), “amyloid angiopathy” (most specific).

Generate a list of conditions in the participant’s FHx
Responses to questions are used as inputs for classification algorithms, which map information about family members to useful FHx categories for the participant. The resulting list reflects what is typically thought of as an individual’s “family history.” The algorithms attempt to map to the most specific category possible, i.e. maximum depth into the disease tree. However, not all conditions easily fit into a hierarchical tree structure. For example, certain syndromes and states, such as collagen vascular diseases or hypercoaguable states may be expressed with different phenotypes in the four disease areas and/or in different individuals.

Estimate participant’s risk for each condition
The participant’s risk for each condition is classified as high, moderate, or average, using risk logic, developed by the domain experts using epidemiologic data, when available. The risk logic takes into account other potential risk factors, such as alcohol and ethnic background, as well as the number and type (1st degree or 2nd degree) of relatives with given conditions. For each condition/risk level, specific recommendations for further evaluation and management are provided.

CONCLUSION
A method for collecting, storing, and evaluating FHx information will be essential for electronic patient records, health risk assessment tools, and personal health records to realize their potential in detecting heritable conditions. Our approach delineates the requirements for representing FHx. Future efforts will address the use of standard vocabularies for relative and participant conditions, as well as explore new ways to encode the classification algorithms for expressing relationships between conditions.

References
1. Cohn WF, Kinzie M, Barrett JR, et al. The Development of a Web-Based Family