Common Data Elements for Clinical Documentation and Secondary Use: Diabe-DS Proof-of-Concept for "Collect Once, Use Many Times"

Project White Paper: Draft for Comment - September, 2011
Please submit comments to Crystal Kallem by September 30, 2011.

White Paper Authors:
Rachel L. Richesson, PhD, MPH; Crystal Kallem, RHIA, CPHQ; Donna DuLong, RN, BSN; Luigi Sison, MSM; William Goossen, PhD, RN; Wendy Huang; Patricia Van Dyke, RN; Cynthia Barton, RN; Donald T. Mon, PhD

See Acknowledgements for a full list of contributing experts.

Abstract

Common data element (CDE) projects are a response to the healthcare industry's increasing need to develop clinical data content standards that can support both patient care and secondary data uses, such as disease surveillance, population and public health, quality improvement, clinical research, and reimbursement. We describe our experiences with a pilot project to develop a disease-specific Domain Analysis Model (DAM) to represent the data elements important for Type 1 Diabetes (T1D). The purpose of the pilot was to rally interest and define and harmonize data definitions for T1D across multiple data use scenarios, ultimately informing a "collect once, use many times" paradigm and facilitating the development of an official T1D DAM that will later be vetted through multiple interested professional societies and balloted as an HL7 informative standard. This 2-year volunteer effort has produced a set of common data elements and related artifacts that need to be reviewed by domain, standards, and technical experts. We are also seeking a professional group to adopt and maintain these content standards, so they can be of use in future healthcare data collection and secondary data use activities. The authors have deliberately documented Diabe-DS strategy and experience to serve as a resource for other content standards development activities. This white paper serves as a comprehensive summary of the Diabe-DS motivation, activities, experience, impressions and suggestions for future work.

Table of Contents

Introduction

Increasing demands for clinical data representations that support both patient care and data re-use, such as disease surveillance, population health, quality measurement and clinical research, are driving requirements for clinical data content standards that are robust enough for all of these purposes. These clinical data content standards are being developed in disease- or domain-specific contexts, such as those initiated by medical specialty groups, the U.S. Food and Drug Administration's Center for Drug Evaluation and Research (FDA, 2010), and the Dutch Diabetes Quality of Care Project (NDF, 2011). The number of these types of projects is likely to explode, especially in light of the recent report released by the U.S. President's Council of Advisors on Science and Technology, which calls for a "universal exchange language" (PCAST, 2010).

Documented, clinician-friendly methodologies to guide these types of projects are desperately needed and virtually absent. The lack of standard methodology in these endeavors means, at best, that multiple groups are duplicating work. For example, chronic hypertension might be included as an important data element in diabetes and cardiac disease. Even more troublesome, and inevitable, is the likelihood that data elements will be defined in different ways across diseases, creating multiple and contradictory "standards." Clearly, there are many abstract data constructs (e.g., laboratory test results, medications, clinical findings, etc.) that are common across most diseases and settings; however, the application of these broad domain data standards, as well as the harmonization of data elements across data uses and disease domains, are beyond the charge of disease-specific content standards groups, and actually threaten the "get it done now" approach of hard-working, motivated, and focused volunteers. Additionally, skilled informatics experts with a clear understanding of multiple data standards and their proper application, are often missing from content-focused data standards groups, further limiting opportunities for harmonization of cross-disease standards.

The development and availability of methods, best practices, and required tools and resource requirements for common data element projects will enable the specification of data content standards. A method for developing data content standards will also facilitate the efficiency of future projects enormously by allowing them to learn from earlier projects. A method for binding those data content standards to EHR system functional specifications will facilitate harmonization and minimize risk of heterogeneity. And, in order to reuse data specifications from one domain to the other, contextual and meta-information are important and should be included.

We describe our methodology and experiences with a prototype project to develop a set of disease-specific data elements important for Type 1 Diabetes (T1D) clinical documentation and secondary use. Our intent was to specify the data requirements for the EHR based on harmonized definitions and value sets, and to demonstrate how the data elements link to EHR functional and interoperability specifications in order to facilitate both primary (i.e., patient care) and secondary data requirements (specifically research, quality measurement, and public health).

Background

Secondary Use of Clinical Data

The demand for clinical health data is increasing with the prospect of widespread EHR adoption. Electronic capture and storage of clinical data has the potential to increase the precision and comprehensiveness of information, which in turn can foster secondary uses of the data. Many organizations have started to demonstrate the usefulness of clinical data for secondary uses (Kallem, 2011; Safran, 2007; NIH, 2005; CDISC, 2009; Goossen, 2002), but data standards issues are still too great to allow meaningful secondary use on a broad scale. Reports from the National Quality Forum (NQF), for example, describe how difficult, if not impossible, it is to derive clinical data for quality measurement from EHR systems (NQF, 2008; NQF, 2009). Public health has had some significant challenges as well, and still does not generally leverage EHR data (exceptions are laboratory reporting of notifiable conditions and emergency department); however, the importance of these quality monitoring and public health functions is enough to keep emphasis on how to utilize electronic data for various healthcare functions. Implementations that support structured data capture for multiple needs will enable the "collect once, use many" vision, and reduce the extra time and expense clinicians spend on data entry, electronic documentation and other new IT-related administrative tasks (Prokosch, 2009); however, in the long term, when the results of secondary data use become available at the point of care, the investment will prove beneficial for the clinicians too, because it can directly be used to improve care for the individual patient.

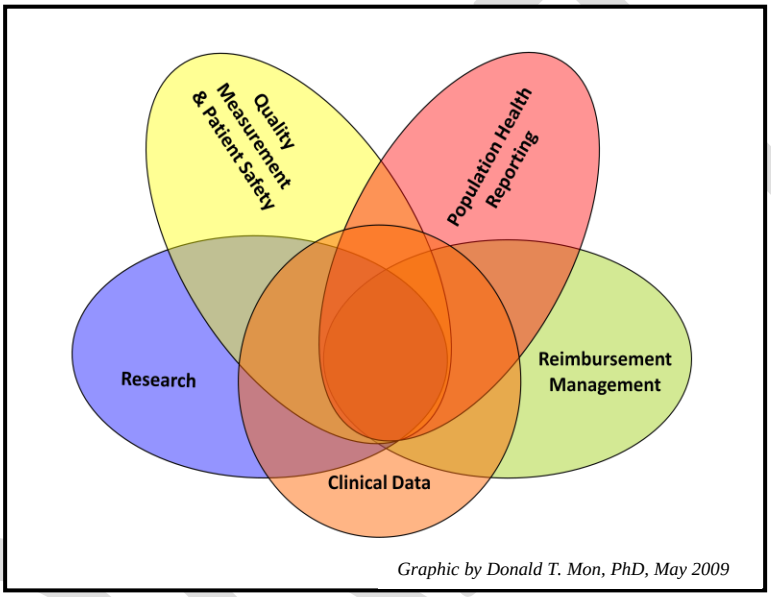
To demonstrate the secondary use of EHR, Kush et al. (2007), in their STARBRITE proof-of-concept study, used data standards from CDISC and HL7 to reuse electronic healthcare data from a clinical care setting for a clinical research study. Their analysis demonstrated that "even in cases where the same data were present in both the clinic note and the case report form, presentation and sometimes even values differed." The authors attribute much of the success of integrating new technology into clinical workflow directly to the high degree of clinician involvement in the design effort.

There is growing interest in the electronic capture and storage of computable clinical data that is consistent, precise, and comprehensive. Unfortunately, there is not enough widespread or consistent use of EHRs to extract valid and reliable data for secondary use. Most healthcare data residing in EHRs today is locked within a proprietary data store and not linked to a standardized vocabulary. Analogous to this scenario is the fact that secondary data uses are as numerous and varied as proprietary EHR data stores. *The ideal scenario is to have structured, computable, semantically interoperable data collected at the source and available for multiple clinical and secondary data uses.* This vision implies the need to harmonize data requirements for various secondary uses; however, these secondary uses all represent 'large' domains and silo-ed communities. While several projects have focused on extracting data for one particular secondary use case (Kush, 2007; NQF, 2009; PHDSC 2007), the diabetes project is innovative in that it addresses *multiple* secondary data use scenarios within a single practice domain. This project is a good practice example for the ISO and HL7 work on standardization of Detailed Clinical Models, in which exactly the requirements of different purposes of data use are included (ISO, 2011).

Common Data Element (CDE) Projects

The notion, then, of *content standards* (i.e., disease-specific data elements that could be used widely across a domain) is a popular and practical approach to identifying data standards that are meaningful, understandable, and implementable in specific settings or professional areas. These content standards are often called Common Data Elements (CDEs) because they are common to a knowledge domain. Generally, CDE projects to date have focused on a given domain or single purpose (e.g., comparative effectiveness research in cardiology or clinical trials in cancer). The Diabe-DS project we describe here; however, strives to harmonize requirements from multiple uses within a domain. In other words, 'common' refers to the single domain of T1D, but also refers to the intersection of data captured for various secondary uses, including population health, quality monitoring, and research, as shown in Figure 1.

Figure 1: Uses of Data Have Significant Overlap: Collect Once, Use Many Times



CDEs have been or are in the process of being developed for oncology, anesthesiology, tuberculosis, cardiovascular diseases, emergency medical services, Alzheimer's disease, Parkinson's disease, polycystic kidney disease, and undoubtedly many others. To date, only a few CDE projects have summarized their process in informatics publications (ACCF/AHA, 2011). Common themes comprise the inclusion of domain experts, an iterative, bottom-up process (such as gathering existing data requirements), and the need for communication tools and long-term collaboration. The many CDE efforts point toward their perceived utility, although there are few documented evaluations of CDE sets or demonstrations of their impact.

Although there are many CDE projects, we have found variation in the definition of the term 'CDE'. After reviewing various definitions and usages of CDE, the Diabe-DS team agreed to define a CDE as a 'data element that is represented uniformly and has value across multiple settings or contexts.' The definition of a data element is also subject to interpretation and some debate. We adopt the ISO -11179 standard's definition of a data element to be a question, answer domain, and definitions (ISO, 2005).

Typically CDEs are common to and identified for a particular domain, which might be research, clinical practice, decision support, or quality monitoring. CDEs are usually developed by convening a group of stakeholders for a defined purpose, identifying important data constructs to collect, and negotiating the best format in which to collect them. These efforts are guided by an explicit sense of purpose, and anecdotally, the more defined and restricted the purpose for CDEs, the more likely for fast agreement. (Case in point: there is still no consensus on broad area standards for primary care or pediatrics, but the American College of Cardiology has defined a set of elements to guide practice evaluation for years.)

Other standardization efforts layer on top of CDE development activities, which define individual data elements. We see in practice that often a set of some data-elements belong together. Examples are the systolic and diastolic blood pressure, or the single scores on a scale that go together with a total score. These small sets of combinations of CDE's, together with the background or context knowledge and meta- information are called Detailed Clinical Models (DCM) (ISO, 2011; Goossen, 2010).

The Diabe-DS project emerged from a desire to replicate earlier domain-specific data standardization efforts conducted within consensus-based standards development organizations, like HL7 and CDISC. These projects rallied disease experts in areas such as tuberculosis and cardiovascular disease to define data elements for each disease area. Both sets of data standards were posted for public comment by CDISC and balloted by HL7 in 2007. The projects contained various artifacts describing the information flow in the treatment of each disease domain. Their operations, scope, methods and results were all open and transparent. The methods are documented well (HL7 2007; Nahm 2010), and offer insight and practical suggestion regarding the naming and definition of data elements and the engagement of multiple and widely distributed stakeholder in the vetting and refinement of the data elements. Similarly the documentation of these projects provides guidance for creating a HL7 domain analysis model (DAM) and related HL7 materials, such as Detailed Clinical Models (DCM). What is lacking from the published methods of both these projects, however, are methods derived from an explicit acknowledgement of other disease-specific data element projects – either completed or in progress – that must be considered to prevent the duplication of effort or the creation of competing and contradictory standards. In particular, an HL7 artifact registry might help with providing an overview of available datasets, projects, models, and message artifacts. (The Diabe-DS project has generated some of that guidance.) There are three communities that were addressed by the Diabe-DS project demonstration, described in more detail below.

Clinical Research Community

The research community has also recognized the need for CDEs and developed many relevant resources, and their experience can and should inform data reuse (and CDE projects). Research is certainly generating lots of "standard" data elements and methods for high-quality and reliable data collection; however, these elements and data collection protocols are focused on clinical research, and might not address the needs (including rapid documentation time) for primary care settings, or other secondary users. The item banks and standards development efforts emerging from various clinical research domains will ease standardization in the sense that they at least move toward a consistency in research data collection. Of course, the new "Clinical Research Chart" emerging from NHII is notable. We extend these ideas and previous demonstrations (e.g.,

STARBRITE) and use cases (e.g., HITSP Clinical Research Use Case) that have been generic to this point, and explore the detailed data collection requirements in a specific domain.

Public Health Community

Public Health has a mission to "assure conditions in which people may be healthy" (IOM, 2002). The scope of public health is both population-based and patient-centric, and in order to fulfill its mission health care providers and public health agencies must be able to exchange pertinent health data about both individuals and communities (PHDSC, 2008). The public health community has been actively pursuing standards that further its mission and that recognize public health as part of a bigger system of interoperable data. As such, the public health data standards community has previously articulated its mission, specific objectives, and activities as they relate to primary data collection (PHDSC, 2005; 2008). This widely vetted, clearly articulated, and carefully documented work of public health data requirements collectively facilitated the Diabe-DS project team's work described here. Many information systems used by public health agencies have been in place since before standards for information exchange existed and are not capable of exchanging data electronically within agency programs, much less with healthcare providers (PHDSC, 2005; 2008). Although limited resources are a continual challenge, public health agencies internationally are continuously improving capabilities to send, receive and exchange data with clinical practices. Public health information needs must be taken into account during development of data standards in order for them to be able to send, receive and exchange relevant data with EHR systems as the public health systems are upgraded. The Diabe-DS project leveraged an existing use case related to the public health reporting and chronic disease management activities in Diabetes (PHDSC, 2008) in order to define specific data requirements for secondary data use in public health.

Quality Measurement Community

Of the primary areas of secondary heath data users, the quality measurement community has made great progress, and their experience can and should inform data reuse (and CDE projects). While not a CDE project per se, the NQF, with support from the U.S. Agency for Healthcare Research and Quality, put a great deal of effort into understanding and navigating the relationships between quality measure data requirements (i.e., the secondary data need) and the ability to automate the extraction of these data from EHR systems (NQF, 2009). They have fully explored EHR data representations, including extensive work to characterize data types and dissected such representations into atomic and reuse (quality) data elements. Similarly, the Dutch Diabetes Federation has defined their set of data elements for Quality indicators (NDF, 2011) for the purpose of automatic extraction of the data for the indicators from EHRs.

The NQF has done a lot of work to assess data requirements for many clinical quality measures across multiple disease areas. They have developed a lot of intellectual work around the methods to understand primary data capture and to promote reuse. In addition, NQF has facilitated a lot of conceptual thinking around scope, strategy, scalability, and future utility. In addition, NQF has actively influenced the development and implementation of EHR data standards. The Diabe-DS project leveraged NQF's findings, which are described in more detail in our description of methods below.

HL7 Standards

Within HL7, which is a dominant organization in EHR data standardization, CDEs are typically documented in the form of a Domain Analysis Model (DAM) and/or a Detailed Clinical Model (DCM). DAMs are conceptual representations of these tasks and information units for a whole domain, such as diabetes, against which software can be defined and evaluated. DCMs help organize clinical concepts by combining data element specifications, the relationships between data elements, and terminology into information models that enable the use of various technical formats (Goossen, 2010).

In HL7, CDEs, DAMs, and DCMs are considered part of "clinical content standards." CDEs should be derived from and inform the development of a model constructed from the information units, communication requirements, and clinical workflows of a domain. In a DAM, CDE's and DCMs are usually identified. Similarly, the existence of CDE's and DCMs for clinical content in another domain can inform new clinical domains, hence informing its development and supporting the reuse of CDE and DCM standards (Goossen, 2010).

In 2007, the HL7 Clinical Interoperability Council (CIC) was formed to help bridge the standards development framework and organizational processes within HL7 by providing a forum for the clinical community to define data content standards and other requirements that are relevant to their particular disease or domain area.

Diabetes Data Strategy (Diabe-DS) Project

The Diabetes Data Strategy (Diabe-DS) project was formed in early 2009 for the purpose of demonstrating the implementation of content standards within an HL7 framework. Diabetes was chosen as a test domain out of interest of the investigators, and because of the lack of domain standards or a diabetes-specific domain analysis model for such a prevalent disease with high morbidity and growing incidence worldwide. Diabe-DS was designed to be a proof-of-concept focused on creating a narrow set of artifacts, processes, and methodologies applied to a similarly narrow context: the collection of pediatric T1D assessment data in an outpatient clinic setting (patient care) that can support related clinical research and quality measurement and population health activities (secondary data use). As an HL7 endorsed activity, the artifacts include HL7 products, such as activity diagrams and story boards. Our process was meticulously documented, as were emerging problems and issues. Participants met weekly during 2009 and 2010 as a group. Additional conference calls of smaller subgroups took place as well, to focus on specific issues (i.e., clinical review of definitions, development of use cases, and data modeling). Most work to date has occurred virtually, with only one face-to-face meeting with available project team members during an HL7 working group meeting.

The multi-disciplinary team included international representatives from academic settings, professional societies, standards development organizations, vendors, and government sector organizations. The names of the group, professional roles, and affiliations are listed on the author list. Part of the work originated in discussions between different HL7 working groups discussing developments in different countries such as US, Canada and the Netherlands. Submitting and comparing data sets from different diabetes groups formed a solid base for the current Diabe-DS project.

The Diabe-DS project has been a voluntary, un-funded effort although some organizations have clearly invested significant human resources and leadership into the project. Participants self-selected into the project. A core team includes diverse perspectives including informatics, EHR design and implementation, information technology (including data modeling), clinical research, epidemiology, endocrinology, and pediatrics. The team operates within HL7 and has had an open membership and wiki. The Diabe-DS wiki, listserve, and conference lines were hosted by HL7 and meetings were announced publicly via these venues.

Methods

The project methods fall into 6 main activities, which are bulleted below, and reflected upon in the results section. The activities to date have included:

1. Selection of data elements
2. Organizing, annotating and harmonizing the data elements and definitions
3. Developing use cases
4. Developing information models that assert relationships between data elements
5. Mapping data elements to EHR functional requirements
6. Documenting relevant findings

In general, these methods are derived from HL7 Healthcare Development Framework (HDF).

The results section describes observations and some illustrative findings from each of these steps. The project deliverables include a set of artifacts produced, which include a preliminary set of common data elements, a set of use case scenarios, corresponding information models, and data mappings.

Project conference calls – especially in the first 6 months of the project – involved lots of discussion about scope, limiting scope, purpose, objectives, and implication. This was a lengthy but important process to ensure that the whole diverse team understood the big picture of secondary use and data content standards. Because the "big picture" (i.e., the mechanisms of data reuse and the harmonization of secondary data uses) is not clear, a significant amount of discussion (approximately 20 hours) occurred within the group to refine the aims, scope, strategy, and end criteria of the Diabe-DS demonstration project. These early calls set the stage for constant reflection and revision of objectives and strategy which continued throughout the project. The entire project was set up to be a reflective process, because our goals were to document our process to support other groups seeking to create disease- or domain-specific data content standards in the future. Because of this, we specifically sought to re-use as much content as possible. The team was challenged with reviewing various CDE projects and not conducting this work in a vacuum but rather realizing that other work existed. A continual point of discussion and decision was the shared concern that the project would produce something useful, and that the project was not duplicating or contradicting any existing efforts or other data standards. At times the search for, exploration of, and attempts to integrate other related projects was very time consuming (at the expense of generating and refining data elements) but it was an important part of the methodology that we used and will recommend to future efforts to ensure that the proliferation of content data standards does not undermine broader standards efforts.

Results

Selection of Data Elements

The first tangible step of this project was to identify data elements that met requirements for T1D clinical and secondary data uses. The initial selection of data elements was performed by a graduate student in nursing informatics from Duke University in spring 2009. The initial selection of data elements involved surveying a broad range of sources, and the sources of data elements differed per data use. For example, the clinical data requirements were solicited from domain experts on our volunteer team, current ADA guidelines, and exploration of two diabetes-specific clinical data collection modules from pediatric outpatient clinics in two academic medical center settings (Duke and University of South Florida). The team did not solicit or include data elements that they knew were relevant to other diseases (e.g., demographics, vital signs, anthropometry, etc.). Also, because the focus of the project was to demonstrate the reuse of data, the team focused only on the set of data elements that had potential *overlap* between primary and secondary uses; they did *not* include many data elements that are clearly valuable in primary data collection. In general, the starting point for identifying the data elements was to focus on secondary uses (as these were smaller in scope than patient care). This was frustrating for clinicians on the team but necessary for managing the project scope. Essentially, the project goals were to identify an important, but not exhaustive, set of data elements for T1D. The set of data elements was refined and expanded over time throughout the course of the project as other secondary data use requirements were gathered (e.g., quality measurement and public health).

The following secondary data requirements were considered as sources for CDEs:

1. The Diagnosis Data Collection Form from The Environmental Determinants of Diabetes in the Young (TEDDY) study, and 4-country observational study of 8,600 infants and children that is exploring a variety of risk factors and potential biomarkers for T1D (TEDDY, 2008)
2. The eligibility screening form from the "Effects of Rituximab on the Progression of Type 1 Diabetes in New Onset Subjects" study (ClinicaTrials.gov ID# NCT00279305)
3. RITUXIMAB research protocol
4. Various research protocols related to T1D from the U.S., Canada, and the Netherlands (convenience sample related to project participants), that had in common both the clinical and the secondary data use specification
5. The National Quality Forum's Quality Data Model (QDM) and related data elements for diabetes care clinical quality measures (NQF, 2009)
6. Public Health Data Standards Consortium's Standards for public health data exchange: Functional requirements standard for diabetes care management and surveillance (PHDSC, 2008).

Most of the initial elements were derived from sampling research data forms from a variety of large T1D studies. This was a convenience sample of studies, but represented the spectrum of important diabetes research and clinical trial data elements, including the protocol eligibility process and observational and epidemiologic research. Quality measurement and public/population health monitoring data elements were incorporated next. The team reviewed NQF-endorsed clinical quality measures and identified two diabetes measures that applied to both type 1 and type 2 diabetic patients: 1) Percentage of pediatric patients with diabetes with a

HbA1c test in a 12-month measurement period (ID# NQF0060); and 2) Diabetes care HbA1c control (<8.0%) (ID# NQF0575).

There were many examples of overlap between similar data elements that were common across one or more secondary uses. In these cases, similar data elements defined different value sets. For example, research might define ketoacidosis using a LOINC® code and/or lab value, whereas a quality measure might define ketoacidosis based on a set of ICD-9-CM, ICD-10-CM, and/or SNOMED CT codes that would be obtained from the problem list. Both use cases seek to identify patients with ketoacidosis, but leverage different data sources and value sets to meet the defined need. These two examples were considered separate data elements as part of the project, but in future studies, mapping could be considered.

An important theme of the Diabe-DS project was the identification and leverage of similar or complementary efforts. There were two very relevant international projects well underway that informed this project. The Netherlands had previously compiled a list of important data elements related to discharge status and planning of diabetes patients in hospital settings, and the use of these data for quality indicators. Though the project focus in the Netherlands was a little different (for example, broader scope of diabetes, different patient populations, hospital context), it provided a tremendous resource for the Diabe-DS project in that it represented over 500 data elements – all with definitions and mapped to HL7 DCMs, the HL7 v3 Care Provision message, and DCMs for diabetes (NDF, 2011).

In Canada, there is a pan-Canadian initiative to define sets of allowable values (also known as value sets or reference sets) used to support standardized primary health care data elements with standardized definitions for both the data elements and the value sets, for the purpose of health system use (secondary use) data extraction. The starter set contains 50 data elements including: diagnosis, allergy and intolerance, patient's height & weight, lab test name, lab test result value & unit of measure, prescribed medication, and dispensed medication. International terminology systems such as HL7, SNOMED CT®, and LOINC® were used to develop the allowable sets of values. Implementation guides are in development on how to use the data elements and implement the value sets into Electronic Medical Record solutions.

The data elements from the Netherlands and Canadian projects were included only when they covered similar data elements already in the set. There was less overlap in data elements than we first imagined or hoped. This is perfectly reasonable and predictable given the different motivations for our data elements. The cases where there was overlap provide excellent examples of an international harmonization process.

Organizing Data Elements

The data elements were compiled into a large spreadsheet, and identified by a unique item number. With over 300 starting elements, the team developed a way to annotate / organize data elements (by content type and characteristics using Clinical Data Interchange Standards Consortium (CDISC) domain names (e.g., subject characteristics, demographics, labs, medication).) Although somewhat tedious, taking more than 6 months to organize, this was a critical step for harmonizing common data elements provided by affiliates. The lack of an existing conceptual model to guide this process of organizing data elements made this a time-

consuming task requiring some understanding of medical knowledge, secondary uses, and intent of the various data types. This seems obvious but is lacking. The CDISC domains (subject characteristics, demographics, labs) were the starting point and were used where possible, but the team had to expand the disease specific areas (diagnosis details, diabetes-specific symptoms) and also added new ones (family history, etc.).

The project was also influenced by the work of the NQF Quality Data Model (QDM). Alternative strategies for organizing CDEs could include by their vocabulary domain (e.g., lab tests, procedures, etc.), by their place in the EHR (e.g., admitting diagnosis, problem list, etc.), or by their place in a Diabetes screening guideline (e.g., diabetes symptoms, glucose tolerance tests). We captured some of these constructs as additional attributes (described in the next section). But for purposes of gross organization of data elements in the spreadsheet, the Diabe-DS project used the following groups:

1) Patient data (subject characteristics, demographics, etc.)
2) Diabetes diagnosis information
3) Target values (~ mostly labs)
4) Lab tests
5) Contact moments
6) Medications (e.g., details medication use; medication delivered from EPR; constrains & allergies)
7) Cardiovascular pathology (co-morbidities)
8) Patient participation and reported data (~ includes quality of life)
9) Blood glucose meter data

It was this organization of data elements (though crudely managed in an Excel spreadsheet) that allowed team members to visualize the broad scope and content of our elements (e.g., lots of symptoms, some labs, mostly disease-specific, some general elements like height and weight, etc.). Also, this process helped to identify numerous duplicate and related data elements that were collected from the multiple sources. At this stage in the project, duplicates and related items could only be captured manually.

<u>Annotating Data Elements</u>

Because of the diverse data sources, there was a clear need to develop conventions/standards for naming data elements. For example, one of the data elements gathered was called "hospitalization during the diagnosis period indicator." The team had to review definitions and query experts (from the source projects) to determine if the intent of the data element was to know if patients were *ADMITTED* at the time of diagnosis, *INPATIENT* at the time of diagnosis, or *IN THE HOSPITAL DUE TO DIABETES KETOACIDOSIS (DKA)* at the time of the diagnosis. There was no clear standard for naming variables and creating clear definitions, nor did we see any guidance or published methodology for this from other CDE projects. The team borrowed heavily from data modeling and statistics literature for naming variables and data objects; first identifying a list of class words (e.g., type, date, name, code, indicator, value, etc.) of which one would be used within the data element name to help clarify the intent of the data element.

Each data element was given a definition, and heuristics were developed to create consistent definitions across the elements. For example, many of the data elements were turned into "indicators." Unfortunately, many of the data elements described "rolled up" clinical concepts, rather than concepts that would be directly charted by a clinician at the point of care. This presented issues later in the project when trying to create a UML representation of the data elements, as the "derived" indicators had to be broken apart into the discrete data elements that might be stored in a patient's EHR. Methodology for exactly deriving an indicator from clinical data elements and clinical data from the EHR is worthy of additional study.

The annotation of data elements on our spreadsheet generated new questions about what exactly constituted a data element and how much data we would need. For example, the data element "ketoacidosis" could be collected and recorded differently, with different representations having different implications for their use and interpretation. In these cases we made the decision that separate data elements should be created for the "ketoacidosis" concept based on the data source. For example, one "ketoacidosis" data element with yes/no for physician interpretation, another "ketoacidosis" data element for diagnosis, another "ketoacidosis" data element based upon pH value, etc., with each different data element having a detailed narrative definition that sufficient to distinguish it from other related data elements.

Another question about defining data elements relates to the granularity. For example, the Oral Glucose Tolerance Testing (OGTT) measure, like many lab tests, can be reported in multiple units. We developed a unique data element for each set of units, and annotated each of the two data elements accordingly (by name, definition, and answer set/units) so that they are clearly distinguishable.

Because the focus and inclusion criteria for the Diabe-DS data elements are ultimately on secondary uses that overlap with clinical documentation (see Figure 1: The Intersection of various data uses is the scope of the Diabe-DS data elements), the team also annotated the Diabe-DS data elements by their likelihood of being captured directly in the EHR or derived from data captured in the EHR. The project team defined the following annotation criteria:

1) Data is in EHR now (secondary data is native to the EHR and format is compatible, or secondary data can be derived from EHR data)
2) Data is not in EHR now, but can be (data has clinical and secondary value)
3) Data has secondary, but no clinical use (deemed out of scope for this project; data must be collected outside the EHR)
4) Data is in EHR now, and has clinical but no secondary use (deemed out of scope for this project; not repurposed)

Ten different attributes were used to describe (annotate) each of the Diabe-DS data elements. The spreadsheet used to manage the data elements is referenced in the appendix, in addition to the final annotated Diabe-DS data elements. In the Excel workbook used for organizing data elements, the team created a separate tab for the value sets. [Note: Value sets were not explored in detail as part of the Diabe-DS pilot project, but are available on the wiki for future discussion. The analysis of value sets represents a separate granular level of analysis for the future.]

Harmonizing Data Elements

The harmonization process consisted of two different activities. First, the organization of the spreadsheet of data elements as described above allowed us to iteratively examine groups of related elements (first by topic, then by keyword in the data element name, then by value types and definitions.) For related elements, if all of these items were the same (i.e., the intent and meaning of data elements from different sources were the same), then we merged this into a common data element. If they were different, we went back to sources (where possible) to determine if the difference was meaningful or arbitrary. If meaningful then the data element was distinct and the team made sure that the data element *name* was distinct as well as the *definitions* to represent any differences.

A second level of harmonization occurred at the modeling stage. In a first pass modeling exercise, the data elements were re-organized into logical groupings of where the data elements might reside within an EHR. This effort served two purposes: 1) help to orient the data elements to where they might match in the EHR-S FM, and 2) identify duplicates and group "related items". This made the data modeling process easier. We organized all elements into the following subject areas:

1) Patient
2) Symptom
3) Medication
4) Test
5) Diagnosis
6) Patient History
7) Physical Exam
8) Disease Management

Examples of these organizational models are referenced in the appendix.

Basically, we organized data elements into groups of 'related' things. Then when we found duplicates, we compared value sets. When value sets differ, the team had two options:

1) Determine if the difference was important. If so, we kept them as separate elements and made sure that the names of the elements differed and were adequately descriptive.
2) If one or more stakeholders were willing to change, we integrated value sets, or choose one. (Often, this was done speculatively by the team level, since we had little contact with many sources.)

Harmonization was supported by a clear definition of value sets, adequate definition of the data element, and an understanding of the source data element (or driving use case). The latter could determine whether the data element could be changed in the Diabe-DS set, and that the changes were tolerable to the source. For example, some of the research elements were improved and it was likely that these changes were reasonable to request of research studies moving forward, whereas changes to a HITEP data element might be less likely to achieve.

Although one of the most critical attributes to support harmonization of related data elements, the definition of the data element was by far the most time consuming to develop. The Diabe-DS project did not attempt to define value sets, but other initiatives, such as the DCMs, have had

similar experience that should inform future efforts in this area. In the HL7 balloted DCM examples, definitions have been dealt with twofold: if they are small value sets, they are enumerated (e.g. 0 = no function, 1 = some function, 2 = optimal function), where in other examples, the value set is only referred to and can be populated from a standardized terminology (for instance all instances of allergy to x,y,z in a SNOMED CT hierarchy).]

Developing Use Cases

As with most data standards projects, the idea of a use case to guide the work is always an important aspect of the project. A suite of use cases was developed to help refine the scope of the project and describe how the data could be used for various purposes (e.g., delivery of patient care, continuity of care, research, quality reporting, etc.). The use case scenarios aimed to demonstrate the same core data within four different scenarios:

a) Patient presents in ambulatory clinic and is diagnosed with T1D, treated, and followed through several office visits.
b) Patient is enrolled in a research study.
c) Patient data is aggregated and reported for quality measurement.
d) Patient data is aggregated and reported for public health.

The use cases created a great deal of discussion around the scope of the project and helped to keep the project focus clear. Development of the use cases was an iterative process. They were influenced by the CDEs in the spreadsheet. In fact, the first iteration use case was a general 'data reuse for EHR data generated from a new T1D diagnosis in outpatient setting.' But in a desire to fully demonstrate the elements, the team elaborated on the use cases to accommodate both the clinical workflow and secondary data use scenarios.

Developing Information Models

A critical component of the project was to demonstrate how the primary data collection elements could be re-used. This was done by developing a representation of data elements in UML format with linkages to other standard representations (e.g., HITSP, HL7 standards, described in previous section). In addition, the modeling was extremely valuable in helping to identify the "atomic data elements" needed to be derived from the patient's EHR. The UML models developed for the HL7 Diabe-DS project are referenced in the appendix.

A key finding of the modeling process, especially in identifying the alignment with other data models such as HITSP, identified significant gaps in the HITSP medication model. Since much of the medication data focused on medication administration, we found the HITSP model to be more focused on the ordering and fulfillment process. This was not a huge surprise considering the use cases that drove the HITSP specification, but did identify a gap in the existing standard.

Mapping Data Elements to EHR System Functional Requirements

Once the data elements were selected and represented in a HL7 compatible format, the data elements were mapped to the EHR-S FM. In this sense we use the content, the modeled version, and the use case for context to demonstrate how those data elements could be defined, represented, and how they relate to current HL7 EHR-S FM standard.

This process relates patient care and secondary data use requirements for the assessment of T1D (via a DAM) to EHR system functions for patient care (via the EHR-S FM) and clinical research and quality measurement, as well as data content standards (via the DCM) and protocol eligibility (via the RCRIM Agreement of Standardized Protocol Inclusion Requirements for Eligibility (ASPIRE) model). It is envisioned that the EHR Interoperability Model and Lifecycle Model will be used as an alternate method to capture clinical and data requirements.

At the technical level, the specific intent of this project is to:

1) Continue the process to allow data exchange via CDA and HL7 v3 messages that applies the clinical statement pattern between EHR systems and systems serving aggregate data purposes.
2) Test the relationship between the more generic but domain specific DAM and the specific and granular DCM.
3) Test the relationship between the DAM, the DCM, the ASPIRE model, the EHR-S FM, and functional profiles (not previously attempted within HL7).

In summary, the domain of diabetes is described on the more abstract level in a DAM and on the most detailed level as CDEs. Where relevant, combinations of CDEs can be grouped into DCM's allowing very specific knowledge to be added, such as contextual information for a specific clinical or secondary use. Once the domain has been described, it becomes possible to map this to the more technical environment. Even so, in order to allow for multiple uses, a logical modeling step is needed between the domain and the technical specifications.

Two specific models serve an important purpose:

1) the EHR-S FM, which describes the functional requirements of the Electronic Health Record System, and
2) the HL7 message models, which define the way electronic data communication can take place between different systems, such as between two or more EHRs, or from EHR to external databases for research or quality measurement.

If this mapping is done consistently, it will reveal opportunities to align and harmonize data requirements within and across domains, align the functional and data requirements, and promote overall consistency and reuse of data.

Discussion

The achievement of interoperable health related data – whether across systems, settings, institutions, or nations - is a an important and unresolved problem, made more difficult because many parts of the solution have yet to be developed. The development of common data elements to support multiple secondary data uses was complicated by the fact that there was no consensus data elements in any of the secondary user communities to start harmonization efforts with. There were no defined data standards for primary or secondary use in T1D at the start of this project, and so it is likely that the elements we have compiled will be subject to debate and revision when vetted in a broader T1D stakeholder community, which should include perspectives from diabetes care, research, population monitoring, and quality measurement.

Despite being an unfunded volunteer effort, with no formal support from diabetes stakeholders, the Diabe-DS project has generated a fair amount of interest. A number of lessons learned should inform the future efforts related to T1D data standards as well as those in other practice or disease areas.

Lessons Learned

There are two outcomes of this project—the production of data content (i.e., a preliminary set of T1D data elements sufficient for clinical and secondary uses), and technical specifications that tie T1D domain-specific data capture needs to EHR and registry functional requirements and to specific messages. The Diabe-DS technical specifications and artifacts include the harmonized data element definitions, use case descriptions, UML information models, and sample mappings of the data elements to the HL7 EHR System Functional Model (EHR-S FM).

The team assembled a collection of over 235 data elements (question, value set, and narrative definition) from various sources including clinical note and EHR specifications, observational and interventional clinical research data forms, and standard (affiliate/jurisdictionally approved) quality measures.

Taking an incremental approach, this project was designed as a proof-of-concept focused on a narrow set of artifacts, processes, and methodologies applied to a similarly narrow use case—T1D assessment in children in an outpatient clinic setting compared to data requirements for various secondary data uses (clinical research, quality measurement, and public health) scenarios. There were many aspects to this issue that the team tried to address, including the need for:

- A standards-based process for defining and harmonizing data elements;
- A methodology to analyze the requirements that satisfy the data needs for both patient care and secondary data use;
- Ascertaining the clinical and data requirements with clinicians, and tying these clinical data requirements to EHR system functionality;
- Assuring that data content standards (e.g., CDEs, DAMs, DCMs) are not recreated across domains;
- A centralized resource to manage and query CDEs over time.

In addition, the project team found it challenging to align the Diabe-DS data requirements with the EHR-S FM, as this standard was currently undergoing a major revision at the time the project took place.

<u>CDEs Relative to Purpose</u>

One operational principle for terminology experts is that standards can not be evaluated (hence, should not be named) without an explicit sense of purpose and context. For example, the specificity required to code diabetes in a quality measure case is different than in screening for an inferential trial on diabetes, which will likely have strict diagnostic eligibility criteria, including the identification of detailed diabetes subtypes that are likely not relevant to clinical management or quality evaluation. Similarly, the notion of 'foot exam' (and the subsequent data collection and value sets that describe it) is notably different for orthopedics versus endocrinology. These issues arose many times and the project and represent fundamental issues in any harmonization intent. We addressed this as much as possible by providing clear and

detailed narrative definitions of each CDE. However, because the CDEs were individually motivated by a certain objective or secondary use context, any evaluation of the collective set of CDEs will need to be informed regarding the spectrum of uses that the Diabe-DS data elements were developed to support. This will be an important but challenging step to implement as the Diabe-DS CDEs are vetted to professional groups and stakeholders in the future.

On a similar note, or work on the Diabe-DS demonstration was complicated by the fact that there are some areas of overlap, but not complete overlap, between the data elements desirable for various uses (clinical, research, quality measurement, public health). The domain experts reviewing CDEs tended to be clinicians, and it was problematic for them to take off their clinician cap entirely, and only focus on a subset of data that we wanted for our limited "collect once, use many" demonstration. It seemed that clinicians are driven to list all the elements wanted in a care sense, and it was a challenge to keep manageable scope but also not exclude data elements that are evidently essential to the clinical management of T1D. The limited scope of the demonstration required us to exclude data elements that other types of experts wanted. For example, a single data element for "Family History of T1D" satisfied clinical experts, but when geneticists or even T1D researchers joined, they wanted additional elements (e.g., which relatives affected, number of primary relatives affected number of relatives unaffected, number of relatives screened.)

Limitations / Recommendation for Future Work

By definition the aim of the Diabe-DS demonstration was to develop an important set of data elements sufficient to describe the domain of T1D and be a starting point for professional vetting and discussion and endorsement. Once endorsed it is hope that these elements would support the development of a standard EHR module for T1D that would support secondary user. Actually, it would also inform EHR and HL7 message specifications so that many different vendors could build such models.

The list of data elements that has emerged from this demonstration work is deliberately small. The data elements represent *overlap* between clinical data collection and one or more secondary uses. This leaves room for this work to be extended. One area in particular is further development of family history (FH) data elements. Exploring a complete set of FH data elements is currently of great interest to diabetes researchers, will becoming increasingly important for personalized medicine, and is relevant to many diseases and CDE efforts. FH could thus be a DCM project, because on top of the CDE, it would include the contextual knowledge on the concept.

The idea behind "collect once, use many times" is that the primary data collection should be at the most granular level. Secondary use needs are different than core documentation for clinical practice. Research and quality secondary uses often involve aggregation, synthesis or summarization of clinical data, which will be the granular date collected. In addition, we are clearly seeing a need that if the CDE prospect is happening in a disease domain (e.g., diabetes), it needs to consider all secondary uses up front to determine the ideal level of granularity required in primary data collection to address the most specific reuse scenario. This principle has been the basis of some preliminary research (Goossen, 2002), but still needs many real world studies to

validate the approach and determine the proper methodology. The issue of evaluating a disease or domain-specific set of CDEs considering a wide range of secondary uses, remains a challenge to the development and utilization of CDEs in other domains.

To move forward, communities that historically have not interacted (e.g., quality measurement, clinical research, public health) will need to cooperate to make CDEs understandable and reusable. These new collaborations and discussions will take time and resources. Developing disease-specific data standards in multi-disciplinary groups will require each stakeholder to consider other uses, new requirements, and – likely – increased scope, likely creating tensions that CDE development advocates should anticipate.

A future challenge for this project might be uncertainty regarding ownership of "clinical knowledge" standards. In the past, efforts related to clinical guidelines have been authored by the specialty organizations—e.g., American College of Cardiologists or Oncology Society, etc. Order sets have been owned by provider organizations or third-party vendors such as Zynx. This is analogous to the challenges of the quality measures/reporting – it sprouted hundreds of similar yet competitive quality measures and increased exponentially the reporting requirements for provider organizations—we do not want history to repeat itself here! Regardless, it is worth noting that the HL7 organization is not claiming the space for DCM, but rather supports the need for a common approach for clinical content specification independent of the technical artifact in which it would be transported (message CDA, or stored (EHR-S-FM). The need for methodology is on the agenda in HL7, in cooperation with ISO work for 13972.

Regardless of who the owner or orchestrator is, there is clearly a need for an owner of the CDEs. This particular Diabe-DS project started off as a demonstration in an area clearly lacking data standards – in either clinical data collection or other data collection like clinical research and public health. Someone needs to own these elements and to maintain them- adding new CDEs as needed to accommodate changes in standard of care, new diagnostics, and new research needs. It would seem logical that these elements would best be maintained by domain experts (rather than multi-disease groups like 'quality measurement') and our future efforts will be to identify a professional medical society or government funder to host Diabe-DS moving forward.

Conclusions

The Diabe-DS is a novel project that is harmonizing data requirements for clinical care and multiple secondary uses, and in turn harmonizing those elements with clinical capture data representations. Much of the process and lessons learned will be applicable to other disease or domain-specific areas. Further, this project provides quality data *definitions*, which are required to thoughtfully apply these standards and to validly reuse technical components or repurpose data. These artifacts and methodologies can be applied to the development of CDEs for other disease domains, as well as to specific clinical research use cases.

Moreover, this work provides the path toward a solution to the *"collect once, use many times"* paradigm which can increase speed and efficiency of evidence-based care, population surveillance, improved quality, and ultimately benefit patients everywhere, because the results of the data for a variety of secondary uses can be more effectively fed back into the practice arena.

References

CDISC. (2009). NIH Roadmap Projects: Therapeutic Area Data Standards–Case Studies in TB &ACS. Presentation to FDA Cardiovascular Standards Committee. White Oak, MD. Retrieved June 20, 2011 from https://www.trialstransformation.org/.

Cimino, J.J. (2007). Collect once, use many. Enabling the reuse of clinical data through controlled terminologies. *Journal of AHIMA / American Health Information Management Association 78 (2), 24-9*. Retrieved June 20, 2011 from http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_033473.hcsp.

Coenen, A., McNeil B., Bakken, S., Bickford, C.,Warren, J.J. (2001). Toward comparable nursing data: American Nurses Association criteria for data sets, classification systems, and nomenclatures. *Computers in Nursing 19(6), 240-6*. Retrieved June 20, 2011 from http://www.ncbi.nlm.nih.gov/pubmed/11764715.

de Lusignan, S., van Weel, C. (2006). The use of routinely collected computer data for research in primary care: opportunities and challenges. *Family Practice 23(2), 253-263*.

Goossen, W.T. (2002). Statistical analysis of the nursing minimum data set for the Netherlands. *International Journal of Medical Informatics 68(1-3), 205-18* .

Goossen, W., Goossen-Baremans, A., van der Zel, M. (2010). Detailed clinical models: A review. *Healthcare Informatics Research 16(4), 201-214*. Retrieved June 20, 2011 from http://pdf.medrang.co.kr/Hir/2010/016/Hir016-04-01.pdf.

Health Level Seven, Inc. (2007). Cardiovascular and tuberculosis data standards (Release 1, package 1). Ann Arbor, MI: Nahm, M., McCourt, B., Walden, A., Diefenbach, J., Kisler, B., Honeycutt, E., et al.

Institute of Medicine (IOM). (2002). The future of the public's health in the 21st century. Washington, DC: National Academies Press. Retrieved June 20, 2011 from http://iom.edu/Reports/2002/The-Future-of-the-Publics-Health-in-the-21st-Century.aspx.

International Organization for Standardization. (2005). *Information technology – Metadata registries (MDR) Part 6: Registration (Reference Number ISO/IEC11179-6:2005(E)).* Geneva, Switzerland.

International Standards Organization. (2011). Health Informatics: Quality criteria for Detailed Clinical Models, Part 1 and Part 2. Geneva, Switzerland: ISO.

Kallem, C., Richesson, R., DuLong, D., Sison, L. Van Dyke, P., Mon, D.T. (2011). Advancing secondary data uses through data standards: HL7 project advances the "collect once, use many times" paradigm. *Journal of AHIMA* 82(4), 38-39.

Kush, R., Alschuler, L., Ruggeri, R., Cassells, S., Gupta, N., Bain, L., et al. (2007). Implementing single source: The STARBRITE proof-of-concept study. *Journal of the American Medical Informatics Association 14(5), 662-673*.

Moyers, S., Richesson, R., Krischer, J. (2008). Trans-atlantic data harmonization in the classification of medicines and dietary supplements: A challenge for epidemiologic study and clinical research. *International Journal of Medical Informatics 77(1), 58-67.*

Netherlands Diabetes Federation (NDF). (2011). Data voor zorgverlening en indicatoren [Dutch: Data for care delivery and indicators]. Retrieved July 28, 2011 from http://www.diabetesfederatie.nl/zorg/data-en-indicatoren.html.

National Institute of Health (NIH). (2005). NIH Roadmap: Accelerating medical discovery to improve health. [cited 2005 March 8]; Retrieved June 20, 2011 from http://nihroadmap.nih.gov/.

National Quality Forum (NQF). (2008). *Recommended common data types and prioritized performance measures for electronic healthcare information systems.* Washington, DC: NQF.

National Quality Forum (NQF). (2009). *Health information technology automation of quality measurement: Quality data set and data flow*. Washington, DC: NQF.

Nahm, M., A. Walden, et al. (2011). Standardising clinical data elements. International Journal of Functional Informatics and Personalised Medicine 3(4): 2010.

Ohmann, C., Kuchinke, W. (2009). Future Developments of Medical Informatics from the Viewpoint of Networked Clinical Research. *Methods of Information in Medicine 48 (1): 45-54.*

President's Council of Advisors on Science and Technology (PCAST). (2010). Realizing the full potential of health information technology to improve healthcare for Americans: The path forward. Washington, DC.

Prokosch H.U., Ganslandt, T. (2009). Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods of Information in Medicine 48(1), 38-44.*

Public Health Data Standards Consortium (PHDSC). (2005). Pediatric electronic health record: public health perspectives. Retrieved June 20, 2011 from http://www.phdsc.org/health_info/ped-electronic-health-record.asp.

Public Health Data Standards Consortium (PHDSC). (2008). Building a roadmap for health information systems interoperability for public health: Public health uses of electronic health record data. [White Paper]. Retrieved June 20, 2011 from http://www.ihe.net/Technical_Framework/index.cfm#quality.

Public Health Data Standards Consortium (PHDSC). (2008). Standards for public health data exchange: Functional requirements standard for diabetes care management and surveillance. Retrieved February 15, 2011 from http://www.phdsc.org/health_info/hrsa_2007_diabetes.asp.

Richesson, R. L., Krischer, J. (2007). Data standards in clinical research: Gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association 14(6), 687-696.*

Safran C., Bloomrosen M., Hammond W.E., Labkoff S., Tang P.C., Detmer D.E. (2007). Toward a national framework for the secondary use of health data: An American Medical Informatics Association white paper. Expert Panel. *Journal of the American Medical Informatics Association* 14(1), 1-9. Retrieved June 20, 2011 from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2329823/.

TEDDY Study Group. (2008). The environmental determinants of diabetes in the young (TEDDY) study. *Annals of the New York Academy of Sciences 1150, 1–13*.

U.S. Food and Drug Administration (FDA) Center for Drug Evaluation and Research. (2010). CDER Data Standards Plan, Version 1.0 (draft). Retrieved January 3, 2011 from http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/FormsSubmissionRequirements/ElectronicSubmissions/UCM214120.pdf.

Acknowledgements

This project was supported by volunteer efforts. The design of the project, analysis and work products were produced over the course of two-years' worth of teleconference sessions using an open and transparent process to ensure broad stakeholder input. The co-authors would like to acknowledge the following individuals for their time and contributions; however, this listing does not imply endorsement by the individual or the organization they represent.

- Cynthia Barton, Graduate Student
- Steve Bentley, National Health Service (NHS)
- Scott Bolte, GE Healthcare
- Joyce Bruno Reitzner, American College of Chest Physicians
- Michael Celeste, Pfizer
- Yong Choi, Duke University
- Gary Dickinson, CentriHealth
- Melanie Dragosljvich, Care Communications
- Donna DuLong, Apelon
- Kristi Eckerson, Emory University/Office of Information Technology
- Davera Gabriel, UC Davis School of Medicine
- Patty Greim, Veterans Health Administration (VHA)
- William Goossen, Results4Care
- Patricia Gunter, Duke University
- Debra Hahn, Association for Healthcare Documentation Integrity
- Monica Harry, Canada Health Infoway, Inc.
- Wendy Huang, Canada Health Infoway, Inc.
- Jeff James, Cerner
- Crystal Kallem, American Health Information Management Association (AHIMA)
- Joy Kuhl, Alliance for Pediatric Quality
- Meredith Nahm, Duke University
- Joyce Niland, City of Hope Comprehensive Cancer Center
- Donald Mon, American Health Information Management Association (AHIMA)
- Craig Parker, Arizona State University
- Maryanne Quinn, Children's Hospital Boston
- Rachel Richesson, University of South Florida
- Mitra Rocca, U.S. Food and Drug Administration (FDA)
- Henry Rodriguez, University of South Florida
- Luigi Sison, LS Associates
- Rachelle Spiro, Spiro Associates
- Patricia Van Dyke, The ODS Companies
- Kendra Vehik, University of South Florida
- Steve Ward, Eli Lilly

We appreciate the support and sponsorship of the HL7 EHR Work Group. We acknowledge the contributions from HL7 committees, especially the Patient Care Work Group, Clinical Interoperability Council, Public Health and Emergency Response (PHER) Work Group, Regulated Clinical Research Information Management (RCRIM) Work Group, and the EHR Interoperability Work Group.

Appendix: List of Diabe-DS Project Artifacts

The following artifacts are available online at
http://wiki.hl7.org/index.php?title=EHR_Diabetes:_Working_Documents

Data Element Spreadsheet

The Diabe-DS data elements represent an important, but not complete, set of data elements for Type 1 diabetes. The spreadsheet contains an inventory of all data elements analyzed as part of the proof-of-concept project. (Last updated 05/11/11.)

Use Cases

The Diabe-DS Use Case includes key activities (represented as a collection of 'mini' use cases) in research, public health, and quality measurement that leverage data collected in a clinical setting. The Diabe-DS Use Case highlights the expected common data elements required for both primary clinical data collection and secondary uses (i.e., research, public health, and quality measurement). (Last updated 08/17/11.)

UML Models

A critical component of the project was to demonstrate the primary data collection elements could be re-used. This was done by developing a representation of data elements in UML format with linkages to other standard representations (e.g., HITSP artifacts, HL7 standards, etc.). In addition, the modeling was extremely valuable in helping to identify the "atomic data elements" needed to be derived from the patient's EHR. (Last updated 08/17/11.)

Modeling Methodology

A description of the modeling methodology used for this project. (Last updated 08/17/11.)

Sample Mapping of Diabe-DS Data Elements to the HL7 EHR System Functional Model

Once the data elements were identified and represented in a HL7 compatible format, the data elements were mapped to the EHR-S FM. The data element content, models, and use case were leveraged for context to demonstrate how those data elements could be better defined and represented, and how they inform the current HL7 EHR-S FM standard. (Last updated 04/27/11.)