A Decade of Experience in Creating and Maintaining Data Elements for Structured Clinical Documentation in EHRs

Li Zhou, MD, PhD[1,2,3], Sarah Collins, RN, PhD [1,2,3], Stephen J. Morgan, MD[1,2,3], Neelam Zafar, MD, MHA[1], Emily J. Gesner DNP, RN-BC, Martin Fehrenbach[4], Roberto A. Rocha, MD, PhD[1,2,3]
[1]**Clinical Informatics, Partners HealthCare System, Boston, MA;** [2]**Brigham and Women's Hospital, Boston, MA;** [3]**Harvard Medical School, Boston, MA;** [4]**Institute of Medical Biometry and Informatics, Heidelberg University, Germany**

Abstract

Structured clinical documentation is an important component of electronic health records (EHRs) and plays an important role in clinical care, administrative functions, and research activities. Clinical data elements serve as basic building blocks for composing the templates used for generating clinical documents (such as notes and forms). We present our experience in creating and maintaining data elements for three different EHRs (one home-grown and two commercial systems) across different clinical settings, using flowsheet data elements as examples in our case studies. We identified basic but important challenges (including naming convention, links to standard terminologies, and versioning and change management) and possible solutions to address them. We also discussed more complicated challenges regarding governance, documentation vs. structured data capture, pre-coordination vs. post-coordination, reference information models, as well as monitoring, communication and training.

Introduction and Background

Structured clinical documentation is an important component of electronic health records (EHRs) and plays an important role in clinical care (e.g. dependency for clinical decision support [CDS]), administrative functions (e.g. extracted to support billing, quality assessment and reporting), research activities, and other areas.[1, 2] When well implemented in the context of the clinical workflow, structured documents can save clinicians' time, assure professional practice standards and clinical thoroughness, and may reduce potential medical errors through clinical decision support interventions.[2]

Different levels of information aggregation for structured documents are shown in Figure 1. Data elements serve as basic building blocks for composing document templates that are used for generating clinical documents (e.g. provider notes and forms). Appropriate underlying terminologies and information models used to assemble

Figure 1. Constructs for Clinical Documents

these data elements, closely correlate to the quality of the produced documents, and are keys to a successful implementation. The management of these data elements remains a critical and demanding issue for EHR systems.

Researchers have made significant efforts in developing methodologies for specifying the structure and semantics of clinical documents. Standard development organizations, professional associations and other healthcare organizations have developed standards and models in this area. A few known efforts include: 1) HL7 Clinical Document Architecture (CDA),[3] which is a document markup standard and specifies an exchange model for clinical documents; 2) HL7 Reusable Information Constraint Templates,[4] which are used to constrain the structures of a portion of atomic concepts, such as a laboratory report in a CDA document; 3) OpenEHR Archetype Model,[5] which consists of a domain-level definition in the form of archetypes and templates and allows clinical experts to be able to structure their own data in the way they require it; 4) Intermountain Health Care's Detailed Clinical Models[6-9] whose core building block is a "clinical element" represented in a recursive data structure for capturing detailed clinical information; and 5) recent Clinical Information Modeling Imitative by HL7 that aims to create a shared repository of detailed clinical information models and binds the models to standard terminologies.[4] EHR vendors most commonly use proprietary approaches to structured documentation and template management.

In this paper, we present our experience in creating and managing data elements for different EHR systems, focusing mainly on lessons learned from two legacy systems (a homegrown ambulatory EHR system and a commercial inpatient documentation system) and a newly implemented commercial EHR system. In particular, we
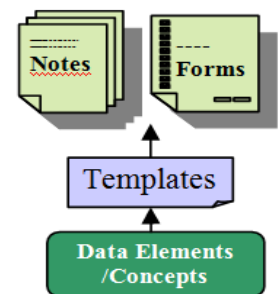
use a basic but nontrivial domain, flowsheet data elements, as an example to illustrate and discuss issues and challenges we encountered.

Case Study Site

Partners Healthcare System is an integrated health care system in the Boston area, founded in 1994 by Brigham and Women's Hospital and Massachusetts General Hospital. It also includes several community and specialty hospitals, community health centers, and other health-related entities. The Longitudinal Medical Record (LMR) is an internally developed, full-featured, and Meaningful Use-Certified EHR, including primary care and subspecialty semi-structured notes, orders, problem lists, medication lists, allergies, laboratory tests, clinical decision support, quality reporting and other functionalities. It was developed and implemented in early 2000s and used across Partners healthcare network. In 2007, Partners carried out the Acute Care Documentation (ACD) project that aimed to develop highly structured clinical content for the inpatient setting using a vendor system.[10] In 2012, Partners cancelled the ACD project and announced a new initiative, known as Partners eCare, to implement a commercially available integrated EHR system at all its sites. On May 30, 2015, the first site, including Brigham and Women's Hospital, Partners Home Care, and Dana Farber Cancer Institute, went live with the new EHR system as part of the Partners eCare initiative. In the following, we describe three case studies based on our experience in creating and maintaining data elements for these EHR systems.

Case Study 1: the Longitudinal Medical Record (LMR) System

A Brief History

LMR allows great flexibility for users to create their own data elements and templates at a local level based on their needs. At the early stages of the systems use, data elements were created on an ad-hoc manner based on the specific needs of individual practices and providers and they were generally not shared among users. Because the granularity (level of detail) and presentation of the data elements change over time and vary by domain, by practice, or by user, their reusability is low. The initial design of these data elements applied the "concept-attribute-value (or section-question-answer)" paradigm. Figure 2 demonstrates an example of a data element for lung symptoms developed by a local oncology team. This type of design mixed data and presentation, compromising reuse and data consistency.

Figure 2. An Example of a Data Element for Composing Document Templates and Structured Notes

In an effort to address these problems, in 2006, the LMR development team and the Partners Knowledge Management (KM) team launched a project that aimed to maximize re-use of data elements and templates, and to support CDS, pay for performance, and quality reporting. A few solutions were proposed and implemented: 1) build an editor that facilitates controlled authoring of data elements and templates; 2) define a shared data element dictionary to manage all elements needed for composing structured clinical documents, where users can create their own data elements or search existing ones; and 3) creation of a limited set of "enterprise" data elements that were flagged and locked (i.e. editable only by authorized KM team members). These solutions did not entirely solve the fundamental limitations of the system. For example, the search function of the editor was relatively simple. LMR Analysts preview existing data elements and templates for new clinical practices and customize as needed. If no existing elements and templates for a specific practice type were found, new ones were promptly created from scratch. In addition, users often expected a high level of customization. Overtime, many similar, or sometimes identical, concepts were added into the data element dictionary. For example, more than 127 same or similar "alcohol use" data elements were created. Compared to a total of 28,400 data elements in 2007, the total number has increased 2.2 times within 3 years to 62,600 data elements in 2010, and continued to grow dramatically in the following years. We coined this phenomenon as the uncontrolled replication of "cancerous" data elements.

Another significant challenge we faced was that these data elements represented concepts from diverse clinical domains, often with duplicate data elements for the same clinical concepts existing and in use in different EHR modules (e.g., problems, medications, procedures, laboratory tests, assessments, and other clinical concepts). For example, a separate "hemoglobin A1c" data element was created specifically for a health maintenance template, but it was not necessarily or directly linked to a patient's laboratory test or result. It was difficult to define an appropriate scope for a separate set of data elements that are only used in clinical notes and forms, what terminologies (if available) should be used to encode them, and how to link these data elements to other concepts within an EHR.

An Example: Flowsheet Data Elements

Here we use flowsheet data elements as an example to demonstrate in more detail the important issues related to the definition, implementation, and management of clinical data elements. EHR flowsheets allow clinicians to track specific elements of a patient's health over time and are helpful in identifying trends within these data. As a common method of documenting clinical observations and physiological measurements, flowsheets are an important component of EHRs and are used as a data source for clinical decision support, safety and quality reporting, clinical trials eligibility criteria, and overall data sharing.[11] When well aligned with the user's clinical workflow, flowsheets can save time and promote consistent documentation, potentially reducing medical errors. However, studies have highlighted the challenges with designing an efficient flowsheet user interface, maintaining the flowsheet ontology, adopting a proper terminology, and measuring compliance in clinician documentation.[12]

Some of the most common data elements of an EHR flowsheet are vital signs, such as height, weight, and blood pressure. In different clinical settings and specialties, flowsheets can be configured to reuse standardized content (such as vital signs and pain scales consistent with quality measures and reportable outcomes), as well as to record more complex data elements. Flowsheets are inherently flexible and analogous to a spreadsheet organized with data element rows and columns that indicate the date and time of recorded observations. This flexibility allows the design and development of flowsheet data elements and templates to meet specific needs of multiple providers, but requires consideration for how data can be standardized and shared among all users of the EHR system. The composite nature of flowsheet data elements and templates requires that the semantic relationships between each components be clearly defined. Figure 3 (left) shows an example of the editor for a flowsheet data element "temperature." In LMR, each data element can be defined with an identifier, name, abbreviation, result type, and at most two attributes. Each attribute can have a list of value options. Figure 3 (right) shows an example of flowsheet data entry form using a template composed by combining a set of flowsheet data elements. Figure 4 shows a patient's flowsheet records in a chronological view.
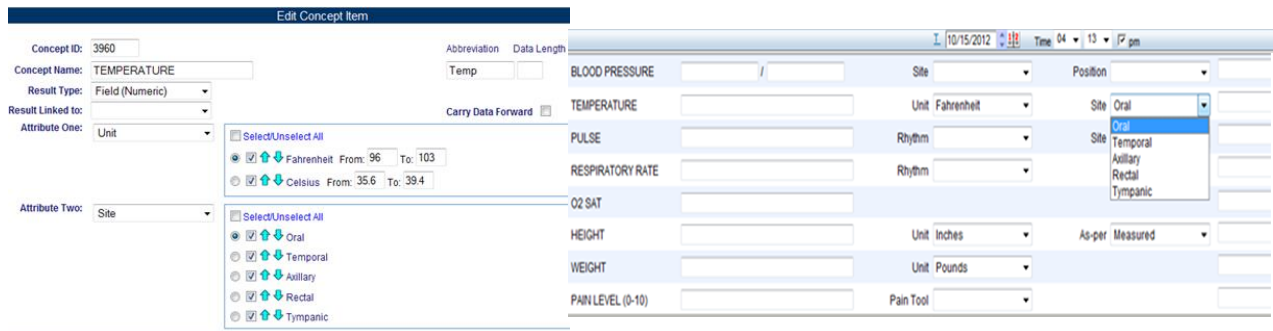

Figure 3(left): An example of creating or editing a flowsheet data element in the LMR's Editor. (Each data element is defined with an identifier, name, abbreviation, result type, and at most two attributes. Each attribute can have a list of value options. The normal range of the attribute value can be defined to validate the user's data entry. The value of an attribute can be carry forwarded in the patient's record.)
Figure 3(right): An example of a flowsheet data entry form: flowsheet template, data elements, attributes, and free-text comment filed.

We performed the following analyses on LMR flowsheet data elements. First, we looked at the total number of flowsheet related data elements and found 3,923 in total. Of these, only 36% had distinct names and the remaining were duplicates (exact same name) of existing data elements. The KM team and the Clinical Quality Reporting team at Partners have identified a set of data elements (by grouping data elements with the same meaning) for the purpose of quality reporting,

Figure 4. Flowsheet data in a chronological view
(clicking on the value itself will allow the user to see more details of the entry)

pay for performance reporting, and CDS. We identified that these important data elements accounted for only less than 3% of the total.

Duplicate or cloned data elements have been frequently created in the LMR dictionary. For example, we found 57 data elements for blood pressure, 55 for weight, 42 for pulse, 21 for height, 15 for temperature, 12 for oxygen

saturation, 6 for respiratory rate, and 5 for pain level (0-10). Using weight as an example, 13 of the 55 data elements are synonyms and 2 are abbreviations. The remaining weight data elements were defined using different combinations of attributes (e.g., unit and measure method) and data type (e.g., string or numeric). Table 1 demonstrates the diversity of attributes or properties and their value sets for some of the flowsheet data elements.

In addition, even with customized structured data elements, we found that free-text comments were often used (sometimes as a workaround) in flowsheet data entries. Collins et al found that nurses use free-text comments in flowsheets as a method to communicate concerning events to physicians, and that these data were associated with survival outcomes of cardiac arrest patients.[12] These phenomena needs to be further studied to provide better understanding of current limitations of EHR flowsheet data elements for automated analyses.

Table 1: Examples of Flowsheet Data Elements and Their Attributes and Value Set Found in LMR

Data Elements	Attributes or Properties and Values
Blood pressure	**Body location**: right arm, left arm, right leg, left leg, both arms, wrist **Method**: doppler, manual, auto **Position**: sitting, standing, lying **Device**: large cuff, small cuff, thigh cuff, pedi cuff, adult cuff, dinamap **As per**: home monitoring, from note, doctor repeated, patient reported, nurse verified **Patient state**: after relaxation, after exam, after medication, orthostatic **Comments**: e.g. patient did not take medication, second reading
Pulse	**Body location**: radial, apical, brachial, carotid, peripheral, pedal **Method**: oximeter, EKG, manual **Position**: standing, sitting, lying **Rhythm**: regular, irregular, abnormal **Abnormal Rhythm**: murmur, pre-mature beats, sinus arrhythmia **As per**: reported by nurse, observed, verified by doctor **Negation**: no
Temperature	**Body location**: axillary, rectal, per ear **Patient condition**: eating or drinking (hot or cold), post medication **As per**: reported by nurse **Comments**: e.g. patient eating chewing gum
Respiratory rate	**Rhythm**: regular, irregular, deep, rapid, shallow **Patient condition**: pre and post nebulizar, coughing, respiratory stress, wheezing **Patient state**: at rest **As per**: verified by nurse

Case Summary

As showed above, the underlying design of data elements was rather simple in LMR (i.e., concept-attribute-value format). One of the original objectives was that by using a relativly simple structure it would allow a high level of flexibility for users to create and customize data elements and templates, which would lead to a wide adoption of structured documentation. However, at the early stage, well-defined data governance strategies were not established. There was no formal model or mechanism for structuring data elements, nor an efficient editing tool. Similarly, linkage and mapping to standard terminologies were not created for these elements. As a result, there is redundancy, inconsistency, minimal reusability, and a lack of interoperability of the existing forms and templates, resulting in continuously increasing maintenance costs.

Case Study 2: the Acute Care Documentation (ACD) Project

A Brief History

Partners Healthcare Systems began the "Acute Care Documentation" (ACD) project in 2007 with the goal of configuring structured clinical documentation using a commercial EHR for implementation at BWH and MGH. The documentation templates were intended for use by nurses, physicians, physical therapists, dieticians and other health professionals that care for patients in the acute and critical care settings. The ACD project was ended just before implementation of the system, but the content creation had produced over 12,000 data elements that were used to create more than 1,200 documentation templates. Based on lessons learned from LMR, two major changes were implemented in the ACD project to control the 'neoplastic' behavior of data elements replication to represent clinical topics: 1) a structured naming convention and best

Table 2. Counts of Vital Sign Data Elements

Content Component (additional Text Search Term)	Count
Temperature	127
Blood Pressure	49
Systolic	14
Diastolic	12
Heart Rate	37
Respiratory Rate	20
Alarm	24
hemodynamic	6
Heart Rhythm	0
Oxygen Saturation (O2)	52
Weight	31
Height	2

practices for data definitions and reuse were defined upfront, and 2) a dedicated team of analysts were the only individuals creating data elements. These analysts were trained in using the defined structured naming conventions leveraging the International Standards Organization/International Electrotechnical Commission (ISO/IEC) 11179 standard.[13] End-users and other clinical stakeholders helped define the documentation templates, but did not have any control in defining or customizing data elements.

An Example –Vital Sign Data Elements

We classified the ACD content in different content categories. A total of 4,188 data elements were classified into 13 categories. Counts for the top 3 categories were: 1,081 wound documentation data elements, 1,440 tube and drain related data elements, and 374 vital sign data elements. Table 2 shows the counts of data elements in the vital signs category. Despite the effort to manage data elements, duplicates and overlap were again present. Some duplicates were unintended and others were "by design" to accommodate system constraints in which a defined data element cannot handle multiple instances of that concept (e.g., multiple arterial blood pressures, multiple wounds). For example, "Blood pressure, diastolic arterial [mmHg]" and "Blood pressure, diastolic arterial 2 [mmHg]" are data fields for measured diastolic arterial blood pressure at a generic anatomical location. These represented essentially the same concept, yet the system requires 2 entries for tracking measurements overtime that take place in two different body locations. This phenomenon repeated for arterial and systolic blood pressure and likely was a factor in the perhaps unintended specification of "Blood pressure, diastolic femoral arterial [mmHg]" and "Blood pressure, diastolic femoral arterial 2 [mmHg]" given that "Blood pressure, diastolic arterial [mmHg]" was already defined. Table 3 shows a set of blood pressure data element names. With this approach a femoral measurement saved in the generic data field, would lead to inconsistent data retrieval.

Case Summary

Based on lessons learned from LMR document template creation, efforts were made to define data elements in the ACD project using a standard naming convention, but also to reuse data elements throughout the system. However, as described above, technical constraints of the system limited the ability to define content according to these identified best practices. This resulted in a data element dictionary with numerous "known" flaws. Such experiences re-confirmed the need to define practical approaches to improve the consistency and reuse of data elements within EHR systems.

Table 3. Examples of ACD Blood Pressure Data Element Names

Atrial pressure left [mmHg]
Atrial pressure right mean [mmHg]
Blood pressure, diastolic arterial [mmHg]
Blood pressure, diastolic arterial 2 [mmHg]
Blood pressure diastolic femoral arterial [mmHg]
Blood pressure diastolic femoral arterial 2 [mmHg]
Blood pressure diastolic noninvasive [mmHg]
Blood pressure diastolic orthostatic lying [mmHg]
Blood pressure diastolic orthostatic sitting [mmHg]
Blood pressure diastolic orthostatic standing [mmHg]
Blood pressure four minute recovery [free text]
Blood pressure highest during intubation [free text]
Blood pressure lowest during intubation [free text]
Blood pressure mean arterial [mmHg] [free text]
Blood pressure mean arterial 2 [mmHg] [free text]
Blood pressure pre-procedure [free text]
Blood pressure Resting [free text]
Blood pressure seated rest [free text]
Blood pressure six minute recovery [free text]
Blood pressure supine rest [free text]
Blood pressure systolic arterial [mmHg]
Blood pressure systolic arterial 2 [mmHg]
Blood pressure systolic noninvasive [mmHg]
Blood pressure systolic orthostatic sitting [mmHg]
Blood pressure systolic orthostatic standing [mmHg]
Blood pressure two minute recovery [free text]

Case Study 3: The Partners eCare Project (PeC)

A Brief History

Partners began the "Partners *e*Care Project" (PeC) in 2012 with the goal of a single patient record across the entire healthcare system using a different proprietary vendor EHR than the ACD project. The PeC project was significantly larger than LMR and ACD projects over a shorter period of time, greatly increasing the complexity of managing data elements. The PeC project went live with its first site on May 31, 2015 and has been in a process of successive implementations since that time. Due to certain functionality, the proprietary system separates structured data elements from flowsheet data elements. At the first went-live site, the total number of structured data elements that have been confirmed to be in use by end users is 15,209. The total number of flowsheet row data elements is 46,575. The project governance was structured based on the EHR system 'modules' that corresponded to a variety

of clinical settings or specialty areas (e.g., inpatient acute care, emergency department, anesthesia, outpatient, and home health). Within each module, reuse and consistent naming of data elements was encouraged; however, reuse across modules was not emphasized as a priority. Problems noted were 1) duplication and overlap of data elements, 2) inconsistent naming convention, and 3) inconsistent data definitions for similar data concepts. During and after initial go-live, efforts were made to implement standardized and consistent data documentation as well as principles and best practices for customized or "local" data elements. We also developed a practical approach to creating reference models for clinical topics that have been determined to be high priority and that should be shared across applications within the EHR (e.g., pain and wound). This effort has been discussed in a previous publication[14]

An Example –Blood Pressure Documentation

We extracted all existing data elements within the structured documentation forms and flowsheet records. We then used keyword search to identify data elements related to "blood pressure". There are 4 data elements used on documentation forms that contained the word "blood pressure", plus 40 others defined for flowsheets. Examples are shown in Table 4. The multiple data elements with overlapping definitions are partially due to the technical constraints of the system as well as the findings that the EHR configuration was based more upon end-user experience rather than consistent data definitions. The definition of data elements and configuration of the user interface was driven by large groups of subject matter experts (SMEs) and prioritized increasing usability and specialty-specific needs. More often than not, it is assumed that there was little to no search for similar data element definitions used in other modules prior to creating a new data element. The result is the overlapping data definitions that currently exist less than a year after going live with the first site.

Table 4. PeC Blood Pressure Data Element Names

R ARTERIAL LINE BLOOD PRESSURE
R ARTERIAL LINE BLOOD PRESSURE 2
R NON-INVASIVE BLOOD PRESSURE
R PHS IP ART LINE BLOOD PRESSURE 3
R IP DOPPLER BLOOD PRESSURE
PHS OB MEAN BLOOD PRESSURE NON-INVASIVE
R AN BLOOD PRESSURE SITE
R AN ARTERIAL BLOOD PRESSURE
R AN MEAN ARTERIAL BLOOD PRESSURE
HOME HEALTH - BLOOD PRESSURE
HOME HEALTH - BLOOD PRESSURE LYING DIASTOLIC
HOME HEALTH - BLOOD PRESSURE LYING SYSTOLIC
ED COURSE - BLOOD PRESSURE CONTROLLED

Case Summary

The lessons learned from our previous experience with LMR and the ACD project did not sufficiently benefit our initial efforts configuring and implementing a vendor-based EHR system. Since those initial efforts, we have established a governance infrastructure and a set of principles and best practices to guide the creation and use of structured data elements and to continuously refine the many inconsistent and duplicate data elements that exist in the system. These inconsistent and duplicate data elements may have been caused by some of the same reasons described in this paper, such as predefined data element structure, the functionality of the editor tool, and constraints for reuse of data elements. We also acknowledge the rapid pace and size of EHR configuration and implementation projects. Our governance infrastructure will continue to remediate and refine structured data elements by focusing on clinical topics that apply across system modules and we have published elsewhere on these processes.[14] These refinement efforts are prioritized based on usage rates of data elements for clinical topics.

Discussion

Structured data elements are important for documenting the care of the patient and for collecting, storing and processing patient care information. In this study, we found that when we develop or implement different EHR systems (either homegrown or commercial products), we encounter various (similar and different) issues and challenges in building clinical data elements and relevant documentation assets used across different clinical settings. In this paper, we mainly used flowsheet data elements as an example, but one can image there are much more data elements needed for building templates and forms to compose clinical documents. Compared to some other coded data in EHRs (e.g., medications or diagnoses), they tend to be more like a "natural" language that contains rich, detailed information with various expressions. Often, they are difficult to be structuralized or standardized across different clinical settings and specialties. Without a comprehensive and efficient strategy, the data elements will grow rapidly to in order cover the diversity.

We faced many challenges to efficiently create and manage clinical data elements in a consistent, reusable, and interoperable fashion. Common themes, requirements, and desirable characteristics for controlled medical

terminologies have been described in previous studies (e.g., Cimino's Desiderata[15, 16]). Goossen et al provided a review of related work in the area of detailed clinical models.[17] Oniki described lessons learned in detailed clinical modeling at Intermountain Healthcare.[9] We have conducted a series of evaluations of the content created for our clinical systems at Partners Healthcare System and reflected on why the desirable characteristics outlined above have not been realized. Importantly, we consider our organization to be process-driven, motivated, and well-resourced; yet still struggle with these challenges outlined above raising the question: "How extensive is this problem at other organizations?". In the following, we summarize and discuss some key challenges and possible solutions that we have learned from our experience. We don't intend to propose data element desiderata or comprehensive solutions in this paper. Instead, we start with basic but important challenges and practical solutions (i.e., challenges 1-3), and then move onto issues and solutions that are important and useful (i.e., challenges 4-6), but they may not achieve an expected impact and need to be combined with other approaches. Finally, we discuss more complicated challenges that need more extensive investigation and work in the future (i.e., challenges 7-8). It is hoped that these discussions may be useful in real-world practice for creating and maintaining clinical data elements.

1. Naming Convention, Attributes and Data Types

Well-defined clinical data elements are important to ensure data accuracy, accessibility, consistency and completeness. The institution should create a guideline for naming convention. When specifying the name for a data element, a preferred description may be chosen, for example, based on standard medical terminologies, and other lexical variations are used as synonyms. The design of data elements and attributes should promote reuse in different templates/forms and sharing among different practices. The value set should be generic (including common items) which allows users to choose the items that they need to create customized templates and forms.

EHR systems often have their inbuilt naming conventions and data types. Due the original design of the EHR system, these conventions may not follow the best industry practices. For example, they may tend to use composite terms for naming their data element (e.g., indicating site, specialty, or purpose within the name) and they may have their own predefined limited data types. Institutions, who implement such an EHR system, may consider developing principles and best practice for naming conventions, including how to name a data element, define its attributes and assign appropriate data types. For example, one may consider what components should be included in the name structure given a certain character limit in the naming field and in which order (e.g., if a data element is designed for a specific purpose and should not be used for other purpose, an indicator for an application or specialty may be added). Metadata may include unique identifier, owner, revision history, type, lifecycle state (e.g., draft, approved, retried), creator, date created, date last modified, sources, long name (i.e., descriptive name with more details about the content and intent), short name, reference terminology, etc.

2. Links to Standard Terminologies

Structured data elements need be encoded in computable forms to be used for clinical decision support, reporting, and other analytical tasks. Linking data elements to standard terminologies also has the potential to help identify duplicate, improve search and reuse. Although this is a basic and important strategy, it frequently gets overlooked. Multiple significant challenges for encoding standardized data elements exist. For example, Kim et al. found that SNOMED terminology provided both complete and partial matches of the Intensive Care Unit (ICU) nursing flowsheet data.[18] The gaps in matching were felt to be mainly due to a lack of appropriate terms used in the original flowsheet and limitations of the concept models.

Manual review and mapping to standard terminologies is tedious and labor intensive. There are many lexical variations (including synonyms, local jargons, abbreviations, misspellings, and other specific symbols) and a lot of data elements have nested structures, thus exact string match or simple algorithms may be inadequate, indicating a need for semi-automated approaches and an efficient tool for mapping (such as the use of natural language processing). Since data elements for structured documentation contains concepts from multiple clinical domains, we need either adopt a comprehensive medical terminology or adopt and combine multiple terminologies at the same time. As mentioned above, similar clinical concept can be defined in both structured data (e.g., problem list) and as a data element for structured documents. This might need to define a scope, i.e., what types of concepts can be created as data elements and what concepts should be reused from other sources. For certain data elements (e.g., flowsheet and health maintenance items), one possible solution might be to keep these separated set of data elements and map them to standard terminologies. For these that are created separately in multiple sources, mapping them to standard terminology may facilitate future data integration and analysis.

3. Versioning and Change Management

Content lifecycle management is a basic and very critical problem. Versioning and change management is important in following the lifecycle of a specific data element. Once a data element is modified (e.g., add an

additional attribute), if the change management is inappropriate, this change may be propagated to every template and structured document using that data element and may potentially break these documents. This requires the system to track state changes (e.g., draft, in review, active, retried) and versions of these knowledge objects. In addition, change management also provides insight for future efforts to see why a particular element was created, the reasoning and thought process into why it was changed, and may provide guidance on when new but contextually related data elements are created. Most editors built within an EHR system often do not have a sophisticate capability for versioning and change management; therefore, additional content lifecycle management and an efficient edit/management tool outside an EHR may be needed.

4. Data governance

Healthcare institutions should establish a set of principles and processes to ensure that clinical data elements and templates are formally managed throughout the enterprise based on applicable hospital accreditation standards, federal, state and institutional regulations, payer requirements and professional practice standards. The authority and accountability of data and knowledge assets should be clearly defined and enforced. Relevant parties within the organization should reach and follow an agreed-upon model, policies, best practices and guidelines which describe, for example, who can take what actions with what data elements, and when (under what circumstances) and how (using what methods and going through what processes). A committee or team that consists of executive leadership, clinical domain experts, informaticians, application analysts, and data stewards should be allocated, who employs certain methodologies and tools for managing, monitoring and improving data elements across the enterprise.[19] For example, the committee may periodically conduct quantitative and qualitative analyses to check the quality and consistency of existing data elements and gather feedback from application teams and end users, and when necessary, refine data governance policies and procedures. Our PeC project has well-defined data governance strategies, but it has not eliminated all the problems, indicating it should be combined with other strategies such like those approaches mentioned above.

5. Monitoring, Communication and Training

We found that continuous monitoring of the quality and quantity of clinical data elements is needed to make sure that they are useful and reusable, and consistently meet users' needs. We also found that the existing data elements in our legacy system (although some of them are duplicate and inconsistent) demonstrate the diversity of users' needs. This resource demonstrating the diversity of users' needs is now a valuable resource for us to use to create higher quality data elements. For example, we were able to analyze usage rates and aggregate multiple duplicate or similar data elements to create reference data elements.

Based on our previous experience, we found that when the users or analysts cannot easily find a data element that exactly meets their workflow, they tended to create or request a new one. Possible reasons for creating a new (duplicate) data elements include, insufficient time to look up existing content, insufficient tools to look up existing content, or insufficient training to look up existing content and meet best practices for data definitions. Effective communication between different stakeholders and training for data stewards and end users is critical to help maximize data element adoption among clinician users, reduce duplicates and facilitate linking to the standards.

6. Documentation vs. Structured Data Capture and Customization vs. Standardization

Each practice or provider may have their unique requirements or preference when creating a same or similar data element. Data elements should preserve details to some extent in order to capture something unique about the patient, the health problem, the practice, and the doctor. However, if each team works only on their own goals and ignore the needs of others, their silo requirements produces redundant information and functions and make integration with other silos very expensive, if not impossible. In contrast, standardization is critical for data reuse, integration and sharing. It is also important for high quality data capture and collection, and has impact on downstream data dependencies, including reporting, regulatory requirements, and CDS.

To alleviate the tension between documentation (e.g., note generation) versus structured data capture, it is important to build data elements in a more uniform form under a well-defined infrastructure. Development of individual information technology solutions at each practice or provider level with little coordination at the corporate level should be avoided. An enterprise-wide strategy may be established to ensure deeper cooperation and formal coordination among different parties (such as clinicians, application analysts and data stewards). For example, instead of allowing each local application team to create their data element directly into the system, a centralized vetting and management process should be in place. A possible workflow may be that the application team at each site send request to create a new data element to the enterprise committee. The committee will review whether or not a same data element already exists. If not, they will check whether the proposed data element follows the best practices. A management methodology is needed for standardizing the clinical data elements, while leaving

room for flexibility and customization. The effort required to implement this workflow during an EHR configuration is large and requires early engagement, prior to any clinician content validation efforts.

7. Pre-coordination vs. Post-coordination

As mentioned by Oniki et al,[9] modeling and creating clinical data elements is not an exact science. A data element builder may face multiple choices when creating a data element, particularly with those that have multiple components. There are pros and cons of each approach and no state-of-the-art rules for deciding when to pre-coordinate and when to post-coordinate. Basically, post-coordination is more structured, so it is easy for computer to process and use for tasks such as information retrieval and clinical decision support. On the other hand, pre-coordinated terms are more like natural language and therefore more user-friendly. Oniki et al has proposed a set of general principles for helping make the decision when to pre-coordinate and when to post-coordinate. Such a decision is often constrained by the functionality that an EHR system provides and the burden created for end-users. For these cases, well-defined naming conventions may be helpful to create useful and consistent clinical data elements.

8. Reference Information Models

There are some existing efforts for structured clinical documentations as mentioned above, such as HL7 Clinical Information Modeling Initiatives, OpenEHR Archetype Model, Intermountain's Detailed Clinical Models, etc., but there are also challenges and gaps in using these standards in real-world practice. Multiple factors affect the standard adoption, mostly due to the complexity of these models and the limited functionality of the EHR system and informatics expertise within local institutions to implement these models.

One of our endeavors is to create enterprise-wide reference data element models based on usage statistics. That is, we identify highly used clinical topics and convene an inter-professional panel of subject matter experts to vote and validate a reference model for that topic. Based on these activities, we create enterprise-wide reference data element models (e.g., with a comprehensive set of attributes) that specialty areas can customize based on their need (e.g., by selecting specific attributes needed for their purposes). [14]

Summary

As discussed in Lean methodology,[20] we cannot only rely on clinicians, practice staff, and site analysts that work on a system to adhere with principles from training and generic advice to not create duplicate or inconsistent data elements. Individual institutions need to develop a comprehensive strategy by considering multiple factors. By designing and deploying a documentation system that fits the workflow of clinicians, it may have the potential to save provider's time for documentation and eventually improve the quality of patient care.

Limitations of This Study

Our study has some limitations. This study was conducted in one integrated healthcare system. Our experiences may not be the same as what others have in other institutions. Our lessons were learned from developing and implementing particular EHR systems.

Conclusion

We presented three case studies of projects within Partners Healthcare to demonstrate the paramount challenges and issues regarding the development and management of data elements for structured clinical documentation in EHR systems. Some of the key challenges faced at Partners were discussed, and we offer some potential solutions to these challenges. It is hoped that our lessons and suggestions will be helpful and useful for institutions who face similar problems when developing strategies for managing their data elements and relevant clinical content.

Acknowledgments

References

1. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of P. Clinical documentation in the 21st century: executive summary of a policy position paper from the american college of physicians. Annals of internal medicine. 2015;162(4):301-3.

2.   Linder JA, Schnipper JL, Middleton B. Method of electronic health record documentation and quality of primary care. Journal of the American Medical Informatics Association : JAMIA. 2012;19(6):1019-24.

3.   Dolin RH, Alschuler L, Beebe C, Biron PV, Boyer SL, Essin D, et al. The HL7 Clinical Document Architecture. Journal of the American Medical Informatics Association : JAMIA. 2001;8(6):552-69.

4.   HL7 Templates Standard: Specification and Use of Reusable Information Constraint Templates, Release 1 (incl Templates Interchange Format), Ballot Cycle: January 2014.

5.   Garde S, Hovenga E, Buck J, Knaup P. Expressing clinical data sets with openEHR archetypes: a solid basis for ubiquitous computing. International journal of medical informatics. 2007;76:S334-S41.

6.   Huff SM, Rocha RA, Coyle JF, Narus SP. Integrating detailed clinical models into application development tools. Studies in health technology and informatics. 2004;107(Pt 2):1058-62.

7.   Coyle JF, Mori AR, Huff SM. Standards for detailed clinical models as the basis for medical data exchange and decision support. International journal of medical informatics. 2003;69(2-3):157-74.

8.   Jiang G, Evans J, Oniki TA, Coyle JF, Bain L, Huff SM, et al. Harmonization of detailed clinical models with clinical study data standards. Methods of information in medicine. 2015;54(1):65-74.

9.   Oniki TA, Coyle JF, Parker CG, Huff SM. Lessons learned in detailed clinical modeling at Intermountain Healthcare. Journal of the American Medical Informatics Association : JAMIA. 2014;21(6):1076-81.

10. Collins SA, Bavuso K, Zuccotti G, Rocha RA. Lessons learned for collaborative clinical content development. Applied clinical informatics. 2013;4(2):304-16.

11. Waitman LR, Warren JJ, Manos EL, Connolly DW. Expressing observations from electronic medical record flowsheets in an i2b2 based clinical data repository to support research and quality improvement. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium. 2011;2011:1454-63.

12. Collins SA, Fred M, Wilcox L, Vawdrey DK. Workarounds used by nurses to overcome design constraints of electronic health records. Nursing informatics : proceedings of the International Congress on Nursing Informatics. 2012;2012:93.

13. ISO/IEC 11179, Information Technology: http://metadata-standards.org/11179/ (last accessed on 3/12/2015).

14. Collins SA, Gesner E, Morgan S, Mar P, Maviglia S, Colburn D, et al. A Practical Approach to Governance and Optimization of Structured Data Elements. Studies in health technology and informatics. 2015;216:7-11.

15. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine. 1998;37(4-5):394-403.

16. Cimino JJ. In defense of the Desiderata. Journal of biomedical informatics. 2006;39(3):299-306.

17. Goossen W, Goossen-Baremans A, van der Zel M. Detailed clinical models: a review. Healthcare informatics research. 2010;16(4):201-14.

18. Kim H, Harris MR, Savova G, Chute CG. Content coverage of SNOMED-CT toward the ICU nursing flowsheets and the acuity indicators. Studies in health technology and informatics. 2005;122:722-6.

19. Collins SA, Alexander D, Moss J. Nursing domain of CI governance: recommendations for health IT adoption and optimization. Journal of the American Medical Informatics Association : JAMIA. 2015.

20. Kimsey DB. Lean methodology in health care. AORN journal. 2010;92(1):53-60.