MEDINFO 2015: eHealth-enabled Health
I.N. Sarkar et al. (Eds.)

doi:10.3233/978-1-61499-564-7-7

A Practical Approach to Governance and Optimization of Structured Data Elements

Sarah A Collins[a,b,c], Emily Gesner[a], Steven Morgan[a], Perry Mar[a,b,c], Saverio Maviglia[a,b,c], Doreen Colburn[d], Diana Tierney[d], Roberto Rocha[a,b,c]

[a] *Clinical Informatics, Partners eCare, Partners Healthcare Systems, Boston, MA*
[b] *Department of General Internal Medicine, Brigham and Women's Hospital, Boston, MA*
[c] *Harvard Medical School, Boston, MA*
[d] *Accenture Health Practice, Accenture*

Abstract

Definition and configuration of clinical content in an enterprise-wide electronic health record (EHR) implementation is highly complex. Sharing of data definitions across applications within an EHR implementation project may be constrained by practical limitations, including time, tools, and expertise. However, maintaining rigor in an approach to data governance is important for sustainability and consistency. With this understanding, we have defined a practical approach for governance of structured data elements to optimize data definitions given limited resources. This approach includes a 10 step process: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, and 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) Monitoring for new evidence requiring revisions to reference model.

Keywords:

Structured data elements; EHR optimization; Data governance.

Introduction

A lack of consistent data capture has profound implications on reporting, e-measures, and clinical decision support and other secondary uses of EHR data [1]. Much work has been performed on the development and use of controlled terminologies to mitigate inconsistent data capture [2–8]. In an ideal EHR implementation project, the initial definition of structured data elements would require mapping to a controlled terminology *with* comparison of each newly defined data element to all other data elements that have been defined previously. Such a data governance process requires extensive resources, particularly: *training* for analysts, clinicians and informaticians in terminologies and clinical content management; *sophisticated data management tools* to continuously search, view and compare all draft and final structured data elements; and *sufficient time* to allow for iterative, analytical, and collaborative content management and validation cycles [9]. Training, tools, and time are limited resources in large scale EHR configuration and implementation projects. While best practice certainly should require tightly controlled data governance of structured data definitions at the beginning of an EHR implementation, in many cases this is not done due to limited resources, a lack of expertise, or competing priorities [9].

Challenges

A lack of consistent shared data definitions across EHR applications and clinical settings, prevent reuse and interoperability of healthcare data. EHR data collection forms defined without reference models, compromise information consistency and completeness. Adding further complexity, many healthcare organization's EHR implementation and optimization projects are quite large with limited interaction between individuals responsible for distinct applications within the EHR system [9]. These limited interactions may lead to decreased sharing of data definitions and an increase in the number of distinct data elements defined to represent similar topics.

In the absence of a pre-defined reference model for a given clinical topic, data governance requires extensive manual effort to verify completeness and consistency of documentation and this effort constantly increases with the creation of new EHR forms. This manual effort trickles down as a significant negative effect on the accuracy, timeliness, and relevance of downstream processes that use the collected data, such as billing, clinical decision support, reporting, population management, and analytics (risk estimation, prediction) [10]. Overall, a lack of well-defined data definitions that are not shared across EHR applications will require additional resources to design, build, deploy, and manage large quantities of data definitions and to remediate consequences to downstream processes [10]. Adding to the challenges, correction of data inconsistencies after data elements and forms are in active use invariably require expensive and error prone data conversions.

Potential consequences

As noted in the previous section, challenges that arise from poor data definitions are well-known, significiant and far-reaching. A solution needs to be tractable; hence, we believe focusing on measureable consequenses to help target specific efforts to improve the definitions of data across applications in an EHR is a practical approach that can be implemented in organizations with limited resources, expertise, and time.

Measureable potential consequences of poor data definitions include: regulatory and compliance requirements, billing, reporting, and clinical decision support.

Specific consequences of inconsistent data definitions that relate to regulatory and compliance requirements include:

incomplete or inadequate reporting, inconsistent evidence during auditing events, and difficulties to implement quality improvement interventions.

Billing consequences include: inconsistent evidence to confirm level of care and decreased coding accuracy and speed due to data variability. Consequences for reporting include highly complex reports that need to account for data being defined in multiple different ways, the design of reporting processes that attempt to "normalize" data to enable consistent interpretation, and the risk of not including relevant data when data elements are missed or incompatible.

Clinicial decision support (CDS) rules have to account for data being defined in multiple different ways which leads to variability and highly complex CDS design. Incomplete or incorrect data triggers can result in false positive or false negative CDS alerts and reminders. Finally, CDS maintainance requires expensive updates and re-testing to accommodate new data definitions or to remediate CDS interventions after data conversions.

Given the “real-world” resource limitations that often lead to a state of inconsistent data definitions, and the subsequent challenges and potential consequences that may result, our goal in this study was to define a practical approach for governance of structured data elements that is sustainable and repeatable. Within our proposed process, specific analytical processes (e.g., formulas, calculations, and weighted scores) require in-depth exploration of assumptions and variables and are likely dependent on an organization’s strategic aims and priorities. Therefore, detailed descriptions of analytics within our proposed process are out of scope of this paper and are planned for dissemination in a future publication.

Materials and Methods

Setting

Our method development was conducted by a workgroup comprised of informaticians, project analysts, and clinical experts. The combination of these three roles enabled a collaboration comprised of multiple expertises. For example, informaticians had expertise in data modeling, terminologies, software development lifecycle stages, and measurement research. Project analysts were highly knowledgeable about the EHR configuration and project management and helped ensure the practicality of our proposed method. Clinicians provided clinical expertise across a variety of specialties and professions to ensure the clinical relevance of our work.

Acceptance Criteria

To meet our aim of defining a practical approach for data governance for structured data elements that is sustainable and repeatable, we defined criterion for an acceptable approach. This criterion was to: 1) make efficient use of existing, though limited, resources, 2) deliver clinically relevant, reusable information, 3) reuse existing data models, 4) leverage reliable metrics for comparative decision-making, 5) manage competing priorities from various stakeholders, and 6) follow a lifecycle stages process model.

Clinical relevance was maintained by focusing on clinical topic categories as the unit of analysis. To increase reliable and practical decision-making, we defined objective measures to analyze downstream data needs and comparisons against current state. The management of competing priorities from various stakeholders was handled by prioritizing clinical topics and revisions based on consensus defined weighted scores and thresholds. Finally, each process in our practical approach was aligned with a stage from the software development lifecycle framework to provide a foundation for a repeatable approach.

Results

After iterations within our team, we identified a 10 step approach that was efficient, clinically relevant, rigorous, and practical for managing competing priorities from various stakeholders, demonstrating evidence to support decisions overtime, and ensuring sustainability for data governance needs identification. The 10 steps of our method for data gov-

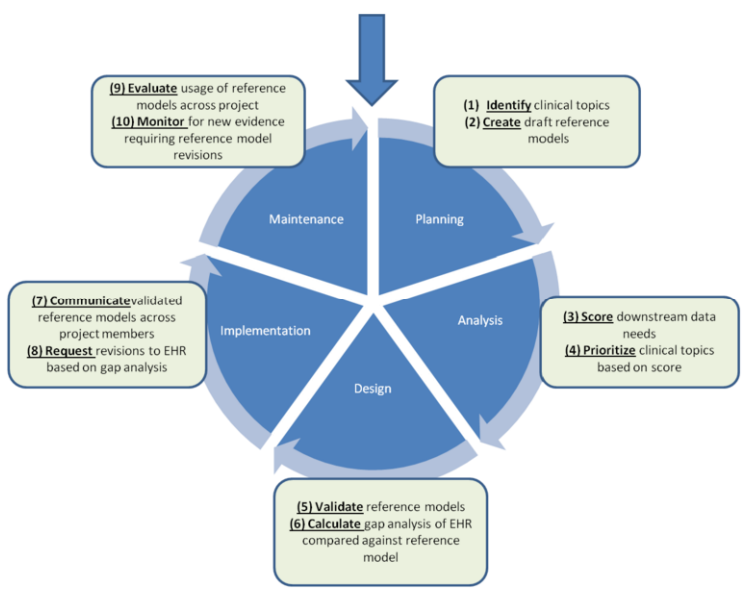

Figure 1 - Practical Approach for Data Governance for Structured Data Elements

ernance needs identification are: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, and 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) monitoring for new evidence requiring revisions to reference model.

Proposed Approach

1) Identification of clinical topics

The aim of identifying clinical topics is to align analytic focus with the strategic aims of an organization and data governance optimization. A key factor in identification of clinical topics is to select topics that are orthogonal to topics previously used to define the clinical content that is being optimized. For efficiency and minimal engagement of resources, certain assumptions may be made that topics focused on previously are a lower priority than topics that have yet to be analyzed. For example, the documentation of intravenous line placement and *urinary catheter management* may have been defined previously, but independent of each other. A focus on *catheter* placement and intravenous line *management* would allow an efficient evalution of discordant and overlapping data definitions among the topics.

2) Creation of draft reference models for clinical topics

The aim of creating a *draft* reference model is to quickly produce an artifact that can be used for initial analyses in the next stage. A reference data model for a clinical topic provides the names of each structured data element that pertains to that topic, data definitions, data types, values sets, abbreviations, synonyms, unique identifiers and original resources. We propose a targeted environmental scan of resources which publish clinical data element reference models such as: 1) HL7 Domain Analysis Models, 2) LOINC, 3) Intermountain Healthcare Clinical Element Model Browser (http://www.clinicalelement.com), 4) salient professional societies, and 5) peer-reviewed literature. This list is not intended to be exhaustive, but rather, is intended to allow for a repeatable and efficient search to identify existing reference models from organizations that are most likely to have produced them.

Existing reference models that result from the search should be compared by mapping synonymous data elements and value sets and forming a comprehensive list of data elements relevant to the clinical topic. This analysis should result in a draft reference model. In the case of retrieval of one extensively validated reference model, that should be used in its entirety.

3) Scoring of downstream data needs for clinical topics

The aim of scoring downstream data needs for clinical topics is to differentiate which clinical topics are most important for an in-depth analysis and development of a validated reference model. Given unlimited resources and time, all identified clinical topics (and in fact all content) within an EHR could be defined consistently. However, we are focused on real-world EHR implementation and optimization projects with competing priorities and limited resources; hence, a requirement of our approach is to focus and prioritize governance efforts for structured data elements.

Consistent data definitions are critical for interoperability, however, there is a lack of measureable benchmarks for interoperability for discrete clinical topics to differentiate which topics are most important to focus on. Yet, there are a set of measurable rationales for increasing the consistency of data definitions: enabling automated reporting, demonstrating adherence to regulatory requirements, maintaining clinical decision support, documenting accurate billing codes, adhering to best clinical practices and terminology described in protocols, guidelines and evidence-based literature and order sets and plans of care.

We propose an approach that identifies measurable downstream data dependencies, calculates a weighted score based on importance of each dependency and results in a total weighted score that can be used to rank the priority of each clinical topic. See Table 1 for an extract of a sample "scorecard" for this downstream data needs analysis.

4) Prioritization of clinical topics

The aim of prioritizing clinical topics is to target the highest priority topics and plan the sequence in which topics will be

Table 1 - Extract of a sample "scorecard" for downstream data needs analysis

[Clinical Topic Name] Scorecard= [total weighted Data Need Score score]

		Data Need Tallies					
Reason Data is Needed	**High Priority Data Needs**	**Data Element 1**	**Data Element 2**	**Data Element 3**	**Data Element 4**	**Data Element 5**	**Data Element 6**
Regulatory	Regulatory requirement A	✓	✓	✓	✓	✓	✓
	Regulatory requirement B	✓	✓				✓
Billing	Billing needs identified by subject matter expert	✓	✓	✓			✓
Reports	Report A		✓				
	Report B		✓				
CDS	CDS Intervention A		✓				
	CDS Intervention B			✓			
Weighted Data Need Score		[score]	[score]	[score]	[score]	[score]	[score]

addressed for data governance optimization overtime. A critical assumption is that this list is dynamic and is constantly being refined, based on scores of downstream data needs for new or re-evaluated clinical topics.

5) Validation of reference models for clinical topics

The aim of validating a reference model for a clinical topic is to produce a standard data model that is clinically accurate and complete and can be disseminated broadly for implementation. The validation of a reference data model may be completed by asking clinical experts to review the draft reference model for missing, superflous, and inaccurate information and asking analysts to review the draft reference model for any feasibility issues or technical constraints to implementation in the EHR system.

Clinical validation of reference models should be based on best practices, is consensus driven, and can be a time intensive activity. Due to the resources required validation of a reference model is done for prioritized clinical topics, and not all clinical topics.

6) Gap analysis of EHR compared against reference model

A validated reference model is useful when used as a reference standard to assert how similar or different EHR data elements are from the defined ideal state (the reference model). We have applied validated metrics for information extraction to this gap analysis [11]. These metrics can be used for a reliable gap analysis measure to identify structured documentation forms within the EHR that do not match the validated reference model. Please see Table 2 for the list of these metrics and their definitions. Documentation forms developed prior to the development of a reference model may not match exactly, but for practical purposes may be "good enough" and not a priority for revisions in the short-term. A pre-defined threshold for a total mismatch score is useful to identify the cut-off score for revisions based on mis-matched data elements. The particular threshold for this cut-off score will dependent on available resources to revise content and push changes into production for the EHR system.

7) Transparent communication of validated reference models across project members

Transparent dissemination and open and ongoing feedback about validated reference models is aimed at increasing the use of the organization's validated reference models across EHR applications and continuing efforts to optimize the models.In our proposed approach, validated reference models are actively used to conduct gap analyses. However, to make efficient use of resources all validated reference models should be communicated across all project members and available for use. Expectations should include that any newly developed content or current revisions will utilize any related validated reference models. Desired exceptions should be communicated for two purposes: 1) to clarify if information is missing or wrong in the reference model that should be added or corrected, and 2) to discover if the exception is appropriate.

8) Request EHR revisions based on gap analysis

The gap analysis provides objective and reliable measure of prioritized areas for revisions in the EHR to match validated reference models. This step to request revisions is included for consistency with change request processes that are common in many organizations. The formality of this step allows for communication about dependencies related to the requested revisions and resources needed for the revision and to manage dependencies of the revision. Additionally, this step offers an opportunity to discuss any contextual information that may impact the appropriateness of the change. For example, a gap analysis may identify 5 documentation forms that vary signficantly from Reference Model A and revisions are requested for all 5 forms. It may be identified at this stage that the 5th documentation form includes structured data elements from a research study that should not be changed during the course of the 12 month study period. A decision could then be make to make revisions for forms 1-4 now and to revise form 5 in 12 months.

9) Evaluation of usage of reference models across project

Use of an electronic collaborative tool to disseminate validated reference models would allow project members to search all validated reference models. An electronic collaboration space could enable the capture of usage metrics with little overhead by providing users the opportunity to document where and how the validated reference model was applied to the EHR application. Usage metrics of reference models can inform: 1) an understanding of the nature of clinical topics that have highly utilized reference models as a method to identify similar important topics, 2) EHR applications that are underutilizing validated reference models and may require training, 3) EHR applications that are proficient in utilizing reference models and could assist teams that are less proficient.

10) Monitoring for new evidence requiring revisions to reference model

An annual clinical subject matter expert review of validated reference models is ideal, and may be extended as limited by resources. As protocols and guidelines change, reference models with related data elements should be reviewed by a clinical subject matter expert to determine if a change is required. Consistent with our goals of a practical, efficient, and reliable approach, the downstream data analysis performed in step 3 should be re-purposed, and not replicated for this step. For example, if a pain documentation ICU protocol was updated to include new evidence, the downstream data analysis that referenced to that pain documentation ICU protocol could be easily identified and the related reference model revised to reflect new evidence.

Table 2 - Proposed Metrics for gap analysis of EHR compared against reference model

Metric	Definition
Match	Data element in the EHR is the same as the data element in Reference model
Partial Match	Data element in the EHR is different but intended to capture the same type of data as the data element in Reference model
Conflicting	Data element in EHR is not equal to the data element from Reference model
Extra	Data element is in the EHR but does not exist in Reference model
Missing	Data element is not in EHR but does exist in Reference model

Discussion

We proposed a 10 step practical approach to governance and optimization of structured data elements. Ten steps may appear to be lengthy, however, we included explicit communication and documentation steps as they are critical to effective governance. It is important to note that organizations may identify shorter paths to prioritizing clinical topics for analyses and identifying gaps in current state for redesign, however, we believe that our proposed approach most effectively combines scientific rigor, collaborative decision-making, and sustainability with efficiency.

Defining data consistently at the outset of an EHR implementation project is the ideal and our research team promotes this pro-active approach. However, we sought to provide organizations who did not or could not effecively govern consistent data definitions initially when configuring and implementing their EHR with a practical approach to remediate inconsistencies and decrease potential downstream consequences. This practical approach is being applied in our organization for a number of clinical topics and is expected to continue as an ongoing activity throughout optimization of our EHR system.

Limitations

The proposed approach is based on the experience of one integrated health system in the Northeast United States during configuration of a vendor based EHR. It is reasonable that the application of this proposed approach to governance of structured data elements would require modifications to fit within the existing infrastructure of other organizations. Our description of the proposed approach is purposefully general to increase the ability for organizations to apply the steps and concepts to meet their needs. As with the design of most best practice and governance approaches, we expect this approach to be iteratively revised and modified to fit different types of health care organizations with varied available resources and expertise.

Conclusion

We defined a 10 step practical and rigorous approach to structured data element governance in EHRs: 1) identification of clinical topics, 2) creation of draft reference models for clinical topics, 3) scoring of downstream data needs for clinical topics, 4) prioritization of clinical topics, 5) validation of reference models for clinical topics, 6) calculation of gap analyses of EHR compared against reference model, 7) communication of validated reference models across project members, 8) requested revisions to EHR based on gap analysis, 9) evaluation of usage of reference models across project, and 10) monitoring for new evidence requiring revisions to reference model. This approach should be evaluated for its applicability across difference types of health care organizations and stages of EHR implementation. We believe that practical and applied approaches that streamline analytical steps while maintaining scientific rigor are highly needed and beneficial to the clinical informatics and health information technology communities amidst the great changes and adoptions that are occurring every day in health care systems throughout the world.

Acknowledgments

We would like to acknowledge all members of the Structured Data Element Workgroup at Partners eCare and all members of Partners eCare project that contributed to this work.

References

[1] (IOM) Institute of Medicine. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC: The National Academies Press; 2013.

[2] Hwang J, Cimino J, and Bakken S. Integrating Nursing Diagnostic Concepts into the Medical Entities Dictionary Using the ISO Reference Terminology Model for Nursing Diagnosis. J Am Med Informatics Assoc 2003: 10: 382–9.

[3] Kawamoto K, Del Fiol G, Strasberg HR, Hulse N, Curtis C, Cimino JJ, et al. Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard. AMIA Annu Symp Proc, 2010; pp. 377–81.

[4] Campbell JR, Xu J, and Fung KW. Can SNOMED CT fulfill the vision of a compositional terminology? Analyzing the use case for problem list. AMIA Annu Symp Proc, 2011; pp. 181–8.

[5] Chen ES, Maloney FL, Shilmayster E, and Goldberg HS. Laying the groundwork for enterprise-wide medical language processing services: architecture and process. AMIA Annu Symp Proc, 2009; pp. 97–101.

[6] Kim H, Dykes P, Mar P, Goldsmith D, Choi J, and Goldberg H. Towards a standardized representation to support data reuse: representing the ICNP semantics using the HL7 RIM. Stud Health Technol Inform 2009: 146: 308–13.

[7] Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, et al. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc n.d.: 16: 81–8.

[8] Fung KW, McDonald C, and Srinivasan S. The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions. J Am Med Inf Assoc 2010: 17: 675–80.

[9] Collins S, Bavuso K, Zuccotti G, and Rocha RA. Lessons Learned for Collaborative Clinical Content Development. Appl Clin Inf 2013: 4.

[10] Cullen T, Bouhaddou O, and Goyal P. Role of Health Informatics in Success of Health IT Projects. HIMSS 2014 Annu Conf Exhib, 2014.

[11] Chinchor N, and Sundheim B. MUC-5 Evalution Metrics. Fifth Messag Underst Conf, 1992; pp. 69–78.

Address for correspondence

Sarah Collins, RN, PhD, Partners Healthcare Systems, Floor 2, 93 Worcester St, Wellesley, MA, 02445, USA