Multi-National, Multi-Institutional Analysis of Clinical Decision Support Data Needs to Inform Development of the HL7 Virtual Medical Record Standard

Kensaku Kawamoto, MD, PhD,[1] Guilherme Del Fiol, MD, PhD,[1] Howard R. Strasberg, MD, MS,[2] Nathan Hulse, PhD,[3] Clayton Curtis, MD, PhD,[4] James J. Cimino, MD,[5] Beatriz H. Rocha, MD, PhD,[6] Saverio Maviglia, MD, MSc,[6] Emory Fry, MD,[7] Harm J. Scherpbier, MD,[8] Vojtech Huser, MD, PhD,[9] Patrick K. Redington, PhD,[10] David K. Vawdrey, PhD,[11] Jean-Charles Dufour, MD, PhD,[12] Morgan Price, MD, PhD, CCFP,[13] Jens H. Weber, PhD, PEng,[14] Thomas White, MD, MS, MA,[11] Kevin S. Hughes, MD,[15] James C. McClay, MD, MS,[16] Carla Wood, MS,[17] Karen Eckert, RPh,[2] Scott Bolte, MS,[18] David Shields,[1] Peter R. Tattam,[19] Peter Scott, MBBS,[20] Zhijing Liu, PhD,[21] Andrew K. McIntyre, FRACP, MBBS[20]

[1]Duke University, Durham, NC; [2]Wolters Kluwer Health, Sunnyvale, CA; [3]Intermountain Healthcare, West Valley City, UT; [4]Department of Veterans Affairs, Boston, MA; [5]NIH Clinical Center, Bethesda, MD; [6]Partners Healthcare System, Boston, MA; [7]Uniformed Service University Health Sciences, San Diego, CA; [8]Main Line Health, Berwyn, PA; [9]Marshfield Clinic, Marshfield, WI; [10]VA OIT OED, Salt Lake City, UT; [11]Columbia University, New York, NY; [12]Université Aix-Marseille, Marseille, France; [13]University of British Columbia, Vancouver, BC, Canada; [14]Universityof Victoria, Victoria, BC, Canada; [15]Massachusetts General Hospital, Boston, MA; [16]University of Nebraska Medical Center, Omaha, NE; [17]Altos Solutions, Inc., Los Altos, CA; [18]GE Healthcare, Wauwatosa, WI; [19]Tattam Software Enterprises Pty Ltd, Moonah, Tasmania, Australia; [20]Medical-Objects, Maroochydore, QLD, Australia; [21]Siemens Healthcare, Malvern, PA

ABSTRACT

An important barrier to the widespread dissemination of clinical decision support (CDS) is the heterogeneity of information models and terminologies used across healthcare institutions, health information systems, and CDS resources such as knowledge bases. To address this problem, the Health Level 7 (HL7) Virtual Medical Record project (an open, international standards development effort) is developing community consensus on the clinical information exchanged between CDS engines and clinical information systems. As a part of this effort, the HL7 CDS Work Group embarked on a multi-national, collaborative effort to identify a representative set of clinical data elements required for CDS. Based on an analysis of CDS systems from 20 institutions representing 4 nations, 131 data elements were identified as being currently utilized for CDS. These findings will inform the development of the emerging HL7 Virtual Medical Record standard and will facilitate the achievement of scalable, standards-based CDS.

INTRODUCTION

An important problem facing healthcare systems is the significant gap between optimal, evidence-based medical practice and actual clinical care. For example, in a recent multi-national survey of chronically ill adults living in eight industrialized nations, 14-23% of patients in each country reported at least one medical error in the previous two years.[1] Moreover, a systematic analysis of 439 care quality indicators has found that U.S. adults receive only about 55% of recommended care,[2] and it takes over 15 years for rigorously validated clinical research findings to be routinely implemented in clinical care.[3]

In seeking to address this gap between evidence-based best practice and actual clinical care, a highly promising strategy is the use of clinical decision support (CDS) interventions, which entail providing clinicians, staff, patients or other individuals with knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care.[4] When automatically delivered to clinicians as actionable care recommendations within their routine clinical workflows, computer-based CDS interventions have significantly improved clinical practice in over 90% of randomized controlled trials.[5]

Despite the great potential for CDS interventions to improve care quality and ensure patient safety, robust CDS capabilities beyond basic medication-related CDS is not widely available, especially in the United States.[4] One important reason for the limited deployment of CDS capabilities is the lack of standard clinical information models and associated terminologies that are consistently used across healthcare institutions, health information systems, and CDS resources.[4] Without a common information model, the effort required for cross-system information mapping will become unsustainable.[6] Moreover, different information models may be semantically incompatible and incapable of being mapped to each other.[7] In the context of the HL7

Arden Syntax standard, for example, this problem has long been identified as the "curly braces" problem, due to the implementation-specific nature of data input specifications contained within curly braces in Arden Syntax modules.[8] Thus, the heterogeneity of clinical information models and terminologies in current use represents a significant barrier to the scalable deployment of CDS.

Within the CDS community, it has been recognized for some time that the definition and adoption of a common information model for CDS would be of great value, and this concept of a common CDS information model has been generally referred to as a virtual medical record (vMR).[9-14] To address this need, the HL7 CDS Work Group initiated the vMR project in 2007. The objective of the HL7 vMR project is to support scalable and interoperable CDS by establishing a standard information model for representing clinical information inputs and outputs that can be exchanged between CDS engines and clinical information systems, through mechanisms such as CDS services. Of note, this project intends to leverage existing HL7 information models and to map them to the vMR. The project charter, as well as all other project artifacts, are available on the HL7 wiki.[15]

Following initial work focused on identifying vMR requirements based on four CDS use scenarios (hypertension, diabetes, breast cancer, and cerebral aneurysms), the HL7 vMR project was re-scoped in January 2010 to more formally incorporate a wider range of insights from CDS implementers both within and outside of the Work Group.[15] Accordingly, the vMR project team conducted a multi-institutional analysis of current CDS systems in February and March 2010 to identify a representative set of data elements used for CDS. Here, we describe the results from this analysis, which were used to inform the development of the emerging HL7 vMR standard.

METHODS

Objective. The objective of this analysis was to identify a representative set of data elements and associated terminologies used by current CDS systems, so as to inform the data elements and associated terminologies that need to be included in the vMR standard as potential inputs into a CDS engine. In order to facilitate the gathering and analysis of data from a number of disparate CDS systems, we chose to obtain information on atomic data elements using a flat structure, with the intent to address the structural relationships between the data elements at a later stage. Also, CDS engine outputs were not included within the scope of the analysis, as the HL7 vMR project team felt that development of this aspect of the vMR would be better served through the analysis of specific use cases and existing

HL7 information models for communicating the results of specific CDS inferences (e.g., for vaccination CDS).

Study Participants. Individuals were eligible to participate in the study if they (i) had knowledge of the data used by an operational CDS system or by a CDS system under active design and development, and/or (ii) were active contributors to the HL7 vMR project. A CDS system was defined using the definition provided above.[4] All study participants were invited to be co-authors on this manuscript.

Participant Recruitment. In February 2010, a request for participation was communicated through the HL7 CDS Work Group's list-serv, and this request asked recipients to forward the email to any potentially interested individuals. The HL7 vMR project team also identified relevant experts and proactively reached out to these individuals. All interested contributors were included as long as the inclusion criteria specified above were met.

Data Collection. Each study participant was asked to provide his or her name, degree(s), institutional affiliation, title, and contact information. For each CDS system with which the study participant had familiarity, the participant was asked to provide the following information: (i) description of system, including purpose, deployment scope, operational status, and any references; (ii) the participant's relationship with the system (e.g., co-designer; knowledge engineer); (iii) data elements used by the CDS system for making CDS inferences (e.g., procedure code, encounter date); (iv) a definition and example of the data element; (v) if applicable, value sets and terminologies used; (vi) example(s) of data element usage for CDS; and (vii) any comments. To expedite data collection, an initial data entry template was created by the vMR project team based on the draft vMR previously developed by the team. To minimize misinterpretation by the contributors, each data element in the data entry template included a definition, examples, and clarifying comments. In a second round of data collection, the template was revised to include data elements that were not in the original template, and study participants were asked to explicitly identify their usage of these additional data elements.

Data Analysis. As needed, collected data were consolidated through an open, consensus-based process by members of the HL7 vMR project team. For example, equivalent data elements identified by contributors using different terms were merged, either through project conference calls involving the relevant contributors or through direct phone or email communications between the primary author and the relevant contributors. Following consolidation, the data were summarized in terms of data elements used

by at least one CDS system, instance examples, the proportion of CDS systems reporting the use of each data element, and use case examples. We provide below the salient aspects of this analysis.

RESULTS

Study Participants and CDS Systems. A total of 28 individuals from 22 institutions participated in the study. Together, these individuals contributed data on the data requirements of 20 CDS systems from 4 nations, which included both large-scale home-grown CDS systems (e.g., CDS systems of the Veterans Health Administration, Intermountain Healthcare, and Partners Healthcare) as well as a number of commercial CDS systems (Siemens Soarian, Eclipsys Sunrise, Medical-Objects CDS, Altos OncoEMR, Hughes riskApps, Wolters Kluwer Health Infobutton API, and Medi-Span) (**Table 1**).

Multi-Institutional CDS Data Needs. A total of 131 data elements were identified as being in use by the 20 CDS systems. Of these data elements, 22 (17%) were not in the original data collection template and were identified by the data contributors.

These multi-institutional CDS data needs are summarized in **Table 2**, and the frequency of their use across the systems is shown in **Figure 1**. As shown in the figure, most of the data elements were used by 20-80% of the CDS systems.

With regard to terminologies, the contributors reported using both standard and non-standard terminologies and value sets. Standard terminologies and value sets reported to be used for CDS included
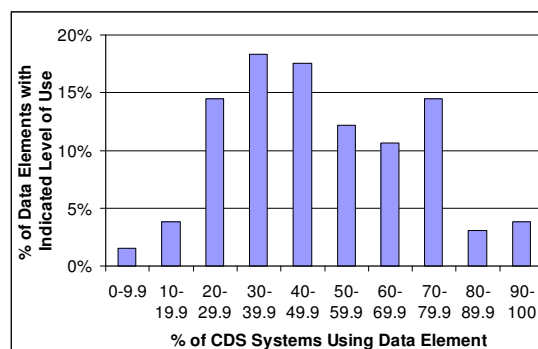

Figure 1. Data element usage pattern across CDS systems.

Table 1. CDS systems included in analysis.

CDS system	Institution	Brief Description
Altos Solutions, Inc. OncoEMR	Altos Solutions, Inc. (U.S.)	Supports oncology and hematology care provided by ~400 medical oncologists and thousands of support staff.
Eclipsys Sunrise Clinical Information System	Columbia University (U.S.)	Deployed at multiple inpatient and outpatient locations. Includes various CDS enabled via Arden Syntax Medical Logic Modules.
DoD Distributed Decision Support & Knowledge Management Repository	Department of Defense (U.S.)	Used by the open source FHA-Connect project to provide a reference implementation for delivering NHIN CDS functionality.
Duke chronic disease management system and enterprise care quality reporting system	Duke University (U.S.)	Supports chronic disease management and health maintenance in 20+ primary care clinics.
HL7 Context-Aware Knowledge Retrieval (Infobutton) Standard	HL7 (International)	Facilitates the context-aware integration of online clinical knowledge resources into EHR systems. Widely adopted.
Intermountain Foresight enterprise-wide decision support infrastructure	Intermountain Healthcare (U.S.)	Responsible for executing the content (rules and protocols) currently available in the HELP2 platform.
Siemens Soarian® Clinical Information System and its Rules and Workflow Engines	Main Line Health (U.S.)	Deployed across 3 acute care facilities and 1 rehab facility to provide point-of-care CDS.
HealthFlow	Marshfield Clinic (U.S.)	Used operationally to support disease management at Marshfield Clinic; integrated with institutional EHR system (Cattails).
Hughes riskApps	Mass. General Hospital (U.S.)	Used at over 35 breast centers to support the early identification of women at increased risk for hereditary cancer.
Medical-Objects GELLO enabled Clinical Decision Support System	Medical-Objects (Australia)	Used in a laboratory information system and multiple specialist clinical systems to provide CDS. Uses HL7 GELLO standard.
Nutritionist's Assistant	Memorial Hermann Hospital (U.S.)	Assists with starting and managing early enteral nutrition for patients in trauma / critical care intensive care unit.
Psychiatric Services Clinical Knowledge Enhancement System (PSYCKES)	New York Office of Mental Health (US)	Deployed across all 26 state hospitals and 300+ mental health clinics to support cost-effective, evidence-based prescribing.
Clinical Research Information System (CRIS) (Eclipsys Sunrise Clinical Manager)	NIH Clinical Center (U.S.)	Supports clinical care for inpatient and outpatient research subjects at the National Institutes of Health (NIH) Clinical Center.
Partners Enterprise Clinical Rules Service (ECRS)	Partners Healthcare (U.S.)	Provides outpatient care alerts and reminders. Part of CDS Consortium work. CDS service will be consumed by Regenstrief.
Automatic Selection of Clinical Trials based on Eligibility Criteria (ASTEC) CDS	Université Aix-Marseille (France)	Automates search for applicable cancer clinical trials. Will be used during oncologic multidisciplinary committee meetings.
UNMC Advanced Clinical Applications Program (ACAP)	Univ. of Nebraska Medical Ctr. (U.S.)	Uses the SAGE CDS model to deliver clinically relevant, problem-based tools to support inpatient physician workflow.
Evidence-based Guideline and Decision Support System (EGADSS)	Univ. of Victoria and BC (Canada)	Supports patient-specific point of care reminders, including preventive care and disease management. Open-source.
VistA Clinical Reminders	Veterans Health Admin. (U.S.)	Used at 153 medical centers and 768 community based clinics to support prevention and chronic disease management.
Wolters Kluwer Health - Infobutton APIs	Wolters Kluwer Health (Internat.)	Provides context-sensitive retrieval of Wolters Kluwer Health content in a manner compliant with the HL7 Infobutton standard.
Medi-Span	Wolters Kluwer Health (Internat.)	Supports real time screening for drug-drug, drug-allergy and other interactions in CPOE and pharmacy systems

Table 2. Multi-institutional CDS data needs (full results online[16]; abridged results shown for data elements used by ≥ 30% of systems).

Data Element (DE)	Example(s)	% Systems Using DE*	Data Element (DE)	Example(s)	% Systems Using DE*
Demographic Data Elements			**Adverse Reaction Observation Data Elements**		
Patient Gender	Male	95%	Causative Agent Type	Medication, Food	70%
Patient Race	Black	55%	Causative Agent Code	SMD code for lisinopril	65%
Patient Birth Date	February 19, 1975	75%	Agent Class	ACE inhibitor	47%
Patient Age	45 years	75%	Reaction Code	SMD code for weal	55%
Patient Age Group	19+ years	55%	Reaction Severity	SMD code for severe	65%
Postal Address	100 Main St., Cary, NC	40%	Reaction Date/Time	Early 1980s	50%
Primary Care Provider	Dr. Jenkins, Clinic X	55%	Reaction Status	Active, Inactive	32%
Encounter Data Elements			**Laboratory Result Observation Data Elements**		
Location Type Code	SMD code for ICU	65%	Test Type	Chemistry, Pathology	75%
Encounter Location	Health System A Clinic X	75%	Test Status	Ordered, Completed	75%
Provider Type Code	HIPAA nephrology code	50%	Test Code	LNC code for HgbA1c	90%
Encounter Status	Completed, Missed	60%	Specimen Location Code	SMD code for lungs	65%
Date/Time Interval	3/1/08 to 3/2/08	65%	Specimen Type Code	SMD code for sputum	65%
Encounter Identifier	Encounter ID ABCDEF	45%	Collection Date/Time	3/15/08 3:15pm	85%
Encounter Note	Discharge summary	55%	Value	135 mg/dL	80%
Procedure Data Elements			Normal Range	50 mg/dL – 150 mg/dL	70%
Procedure Code	SMD code for biopsy	85%	Interpretation	Panic High	95%
Procedure Site Code	SMD code for right breast	40%	Nested Observations	Hgb & HCT in CBC	55%
Procedure Modifier Code	CPT laterality modifier	45%	Observation Note	Report contents	35%
Date/Time Interval	3/5/08 3:15 – 7:40 pm	70%	**Physical Finding Observation Data Elements**		
Procedure Status	Active, Canceled	32%	Finding Type	Vital sign, radiology	75%
Procedure Identifier	EHR Entry # 1234567	35%	Finding Code	SMD code for SBP	75%
Procedure Note	Report contents	35%	Patient Position Code	SMD standing code	40%
Data Elements Common to All Types of "Observations" Below			Finding Location Code	SMD code for lungs	40%
Observation Identifier	EHR Entry # 1234567	Ave. 35%	Value	125 mm Hg	76%
Observation Date/Time	3/5/08 3:15 pm	Ave. 58%	Normal Range	90 – 140 mm Hg	50%
Associated Enc. ID	Encounter ID ABCDEF	Ave. 33%	Interpretation	Normal, High	55%
Problem Observation Data Elements			Nested Observations	SBP & DBP within BP	50%
Observation Type	Problem List Observ.	65%	**Goal Observation Data Elements**		
Problem Code	ICD9 code for diabetes	95%	Goal Focus Code	SMD code for SBP	40%
Problem Modifier	Negative/does not have	32%	Value	135 mg/dL	40%
Problem Status	Active, Resolved	65%	**Other Observation Data Elements**		
Status Time Interval	1995 to present	55%	Observation Type	Social History, Survey	70%
Medication Observation Data Elements			Obs. Focus Code	LNC for survey inst.	65%
Observation Type	Prescription, Usage	70%	Value	5 packs/day, true	73%
Medication Code	SMD code for lisinopril	100%	Interpretation	Normal, abnormal	45%
Medication Class	ACE inhibitor	67%	Nested Observations	Items in survey	45%
Medication Dose	30 mg, 2 puffs	70%	**Patient Affiliation Data Elements**		
Medication Route	PO, IV, IM	70%	Affiliated Entity Type	Insurer, Care Provider	45%
Medication Rate	BID prn, 12mg/hr, qam	75%	Entity Identifier	Medicaid, Clinic X	40%
Coverage Time Interval	11/1/07 to 3/1/08	60%	Obs. Date/Time	3/15/08	35%
Refill Information	On 2nd of 6 total refills	50%	Affiliation Status	Active, Inactive	35%
Medication Status	Active, Inactive	63%	Status Time Interval	3/15/08 – 3/15/09	30%
Family History Observation Data Elements			**CDS Context Data Elements**		
Relationship to Patient	SMD code for aunt	45%	CDS System User Type	Physician, Patient	42%
Relative Demographics	57 year old female	30%	Info Recipient Type	Physician, Patient	37%
Relative Age of Death	N/A, 85 years	30%	Info Recipient Language	English, Spanish	47%
Relative Problem	Problem info as above	35%	Task Context	Order entry, lab review	35%
Data Elements for All Orderable Items (e.g., Meds, Labs)			**CDS Resource Data Elements**		
Orderable Item Status	Ordered, Completed	37%	Concept Taxonomy	ICD9 codes for COPD	58%

*n = 19-20 for percentage calculations. LNC = LOINC; SMD = SNOMED CT; SBP/DBP = systolic/diastolic BP.

SNOMED CT, LOINC, ICD9, ICD10, CPT, MeSH, NDC, RxNorm, and HL7-defined value sets (e.g., for gender and race). Many respondents reported that the non-standard terminologies and value sets in use could be, or have been, mapped to standard terminologies of similar granularity.

The full data set and analysis, including details on the terminologies and value sets used by the contributing CDS systems, are available online.[16]

DISCUSSION

Summary and Interpretation of Findings. In this study, we analyzed the data needs of 20 CDS systems from 4 nations to identify a representative set

of data elements used by CDS systems. Through this analysis, we identified 131 data elements used for CDS, all but two of which were used across multiple systems. Also, while both standard and non-standard terminologies were used, many contributors reported that their non-standard terminologies could be mapped to standard terminologies. Therefore, we believe that this work represents a solid step forward in the HL7 CDS Work Group's efforts to define a common vMR for CDS that can overcome the "curly braces" problem and facilitate highly scalable CDS.

Strengths. As one important strength, this study sampled a highly diverse set of CDS systems, including mature home-grown and commercial CDS systems. This diversity minimizes the chances of false negative findings (i.e., the overlooking of important data elements). Second, this study is based on actual CDS systems and their data needs. Consequently, our methodology minimizes the chances of false positive findings (i.e., the inclusion of data elements not truly useful for CDS). Third, the data element set identified appears to be relatively compact and suitable for standardization and adoption. Finally, this study addresses a well-recognized problem and has the potential to facilitate significant advances in CDS scalability and impact.

Limitations. As one limitation, study participants were self-selected based on interest and were primarily from one country (the United States). Thus, it is possible that this analysis did not capture data elements used by non-participants. However, the large number and significant diversity of CDS systems included in this analysis should minimize the risk of such false negative findings. Second, the use of an initial data entry template may have biased responses. However, as indicated by the fact that close to 20% of the data elements we identified were not included in the original data entry template, individual contributors actively pursued the inclusion of data elements regardless of whether they were included in the original data entry template.

Implications and Future Directions. Based on this multi-national, multi-institutional analysis of CDS data needs, the HL7 vMR project team developed an initial proposal for a vMR standard that incorporated a CDS input model that was the focus of this study, a query model for specifying the data required in a given instance, and a CDS output model.[15] This proposed standard underwent balloting in May 2010, and we are currently addressing the ballot comments to improve the proposed standard. Moving forward, all project artifacts will continue to be posted on the project Wiki,[15] and any interested individuals are invited to participate. Ultimately, we envision that this work will serve as an important foundation for the health informatics community to develop and deploy interoperable CDS solutions that improve population health on a widespread scale.

REFERENCES
1. Schoen C, Osborn R, How SK, Doty MM, Peugh J. In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008. Health Aff (Millwood). 2009 Jan-Feb;28(1):w1-16.
2. McGlynn EA, Asch SM, Adams J, Keesey J, Hicks J, et al. The quality of health care delivered to adults in the United States. N Engl J Med. 2003;348(26):2635-45.
3. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. In: Bemmel J, McCray AT, editors. Yearbook of Medical Informatics 2000: Patient-Centered Systems. Stuttgart: Schattauer; 2000. p. 65-70.
4. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc. 2007;14(2):141-5.
5. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765-8.
6. Parker CG, Rocha RA, Campbell JR, Tu SW, Huff SM. Detailed clinical models for sharable, executable guidelines. Medinfo. 2004;11(Pt 1):145-8.
7. Kawamoto K. Integration of knowledge resources into applications to enable clinical decision support: architectural considerations. In: Greenes RA, editor. Clinical Decision Support: the Road Ahead. Boston: Academic Press; 2007. p. 503-38.
8. Choi J, Lussier YA, Mendoca EA. Adapting current Arden Syntax knowledge for an object oriented event monitor. AMIA Annu Symp Proc. 2003:814.
9. Johnson PD, Tu SW, Musen MA, Purves I. A virtual medical record for guideline-based decision support. AMIA Annu Symp Proc. 2001:294-8.
10. Huang C, Noirot LA, Heard KM, Reichley RM, Dunagan WC, Bailey TC. Implementation of virtual medical record object model for a standards-based clinical decision support rule engine. AMIA Annu Symp Proc. 2006:958.
11. Jenders RA, Dasgupta B. Challenges in implementing a knowledge editor for the Arden Syntax: knowledge base maintenance and standardization of database linkages. Proc AMIA Symp. 2002:355-9.
12. Boaz D, Shahar Y. A framework for distributed mediation of temporal-abstraction queries to clinical databases. Artif Intell Med. 2005 May;34(1):3-24.
13. Kolesa P, Spidlen J, Zvarova J. Obstacles to implementing an execution engine for clinical guidelines formalized in GLIF. Stud Health Technol Inform. 2005;116:563-8.
14. Sonnenberg FA, Hagerty CG. Computer-interpretable clinical practice guidelines. Where are we and where are we going? Yearb Med Inform. 2006:145-58.
15. HL7 vMR Project Wiki. Available from: http://wiki.hl7.org/index.php?title=Virtual_Medical_Record_(vMR).
16. HL7 vMR Project CDS data needs analysis. Available from: http://wiki.hl7.org/index.php?title=Image:HL7vMR_Cross_Inst itutional_CDS_Data_Needs_Analysis_vFinal.zip.