1.
2.
3.
4.
5.
6.
7.

8.

The SEURAT Project and a Discussion of Related Standards:

Response to CORBAmed RFI2, “Clinical Observations,”

OMG Document No. corbamed/97-05-02

Daniel Chudnov, MS, Cushing/Whitney Medical Library

Richard N. Shiffman, MD, MCIS, Center for Medical Informatics,
Department of Pediatrics, Yale School of Medicine

Yale University
New Haven, CT

30 August, 1997

1.Introduction

This submission is a response to CORBAmed RFI 2, "Clinical Observations", OMG Document No. corbamed/97-05-02. As such this is a public document of the OMG and is free to be used by the OMG as seen fit, including making unlimited copies and sharing the submission with other organizations.

The purpose of this submission is to respond directly to CORBAmed's call for information about the various points of contact between systems, applications, instruments, and humans involved in the computerized management of clinical observations. While the experiences described herein by the authors might address only a small subset of the broad scope of the RFI, it is hoped that the information shared is found to be of some utility in informing whatever future actions the CORBAmed DTF chooses to take.

This response draws largely on specific activities undertaken by members of the Yale University medical community which relate to the domain specified in the RFI. A large multidisciplinary project known as SEURAT, a full-production system for capturing clinical observations data using scanning of dynamically generated and completed forms, is described. (Shiffman et. al.) Following this are sections describing specific lessons learned and a more general discussion of high-level issues deemed relevant by the authors, including a discussion of related standards.

2.Affiliations

Daniel Chudnov, MS, is an IAIMS Assistant and Librarian at the Cushing/Whitney Medical Library at the Yale University School of Medicine. He is also a co-chair of the Health Level Seven (HL7) SGML/XML Special Interest Group.

Richard N. Shiffman, MD, MCIS, is a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar at the Department of Pediatrics and an Assistant Professor at the Center for Medical Informatics at theYale University School of Medicine.

3.The SEURAT Experience

3.1.Scope

The SEURAT (Scanning for Evaluation, Utilization Review, Analysis, and Training) Project undertaken at the Yale-New Haven Hospital Primary Care Center leverages several of technologies to improve how clinical data is captured at the point of care. (Shiffman et. al.) A combination of scanning technologies (including mark sense forms as well as optical and intelligent character recognition) is used as the point of digital entry for a discrete set of observation data. This data is initially captured on a variety of age-specific forms, and contains a combination of mark sense data input areas corresponding to direct questions, patient demographics pulled from a remote database and printed on the top of the page, and unstructed handwritten text in several whitespace areas.

The form itself is a product of a complex workflow process. Upon arrival in the clinic, nursing staff retrieves or generates patient registration information, records new measurements, and inputs this data to the system, whereupon the nurse verifies that the form chosen by the system based upon the patient's age at the time of the visit is correct. There are sixteen separate one-page age-specific forms, each with a set of information prompts unique to each age group. Most of these prompts request a choice between three assessment values, e.g. "normal," "abnormal," and "query." Each form is printed, including the up-to-date registration information, from a hardware-based image to improve performance.

The physician completes the form, adding her own provider identifier and writing, by hand, any free-text additional comments into unstructured areas of the form. After each encounter, the form is added to a batch. At the end of each day, the batched encounter forms (generally numbering about 50) are scanned through the recognition system and carefully verified, element by element, by a quality control clerk under supervision by a nursing administrator. Extracted data is parsed through a series of integrity checks before it is stored onto the remote database.

In the eighteen-plus months SEURAT has been used, it has been highly successful both in terms of improved data capture as well as provider satisfaction. All health maintenance data for these encounters is now captured using this system.

3.2.Requirements

SEURAT is designed to accomodate the following requirements relevant to the RFI:

Complex and changing data collection and reporting requirements imposed by various levels of governmental or health maintenance agencies. This aspect of SEURAT was developed, in part, in response to the Connecticut Medicaid managed care program, including the Early and Periodic Screening, Diagnosis, and Treatement Program (EPSDT).

Simplified interface for quality assessment and reporting to identify areas requiring quality improvement.

Active use of guideline-compliant reminders for preventive care.

Ability to capture information not explicitly requested on the form in unstructured handwritten text.

Ability to dynamically combine preformatted forms, remotely available registration information, modularly designed assessment questions organized on an age axis, and measurement information taken during the encounter into a single form (or interface) for physician use.

4.Lessons Learned

4.1.Workflow

The flow of information through the SEURAT system takes advantage of the existing clinical workflow; it is uniquely tuned to offer providers discrete opportunities to record information during the natural flow of a clinical encounter. For instance the form used for a single encounter is scanned and processed as an attested entry approved by a single physician, but the data contained on that form and the form itself has been generated dynamically by a series of individuals in a defined sequence. This sequence includes the output of the form design team in assuring compliance with mandated guidelines, capture or retrieval of registration information by clerks or nursing staff, capture of measurement information by nursing staff, and the capture of assessments and further information by the physician. Thus while the physician attests the validity of the final documents and the data contained therein, as a whole the completed form is the output of a multi-functional service team and the observations recorded implicitly reflect the multi-functionality of the clinical workflow as an accepted business process.

4.2.Variety and dynamics of forms

As described in Section 3.1, the forms used for data capture in SEURAT are created by a team of designers working to assure that complete encounter documentation is captured and that this documentation complies with certain mandated information requirements. For SEURAT, the forms vary along an age axis; this is particularly germane to ambulatory pediatric care, but it is evident that other care units require division of similar forms (or "templates," as called by many current records system implementations) along other axes such as age, diagnosis, or treatment protocols. By modularizing the storage and design of these forms, developers retain the ability to make changes as requirements or guidelines change. Ideally disparate systems across departments might take advantage of standardized descriptions of forms such as these to streamline the creation and management overhead required by this approach. Additionally standardized interfaces to collections of forms might allow their use by disparate systems and implementations.

4.3.Structured data and structured text

The ability to manipulate the data elements recorded using SEURAT is powerful; with the additional capacity to capture free text comments added by the providers within whitespace areas of the forms, we are able to preserve all of the observation information generated at the point of care. We do not, however, have the capability to manipulate the handwritten text as anything more than free text. Ideally a set of mechanisms would exist to enable structuring of the text.

The text recorded on the forms is created in the same context as the structured data, so it implicitly inherits the structure of the form on which it is captured. The text itself, if available in a database, could be connected to the other structured data through the keyed encounter identifiers recorded in the database. These two minimal levels of structure, should they be available online as text (and not a scanned image), could still be used to perform powerful searches given this implicit structure. For example, a query could be written that searches for instances of the word "allergic" in all encounters for all four-year-old patients in the last two years. This feature was outside of the scope of the SEURAT project, and while the technology to enable this might be just a few connections away (for instance, having providers dictate a unique encounter identifier as part of a transcribed report) it represents a significant break from the clinical workflow process embodied (and therefore clinically nonintrusive) by SEURAT.

Parallel to this line of thinking is the current trend in ambulatory record systems toward "template"-based observation capture modules wherein providers are led through a series of online questions (as mentioned in the previous section). Many of these products record the data input by providers using these tools as bits in a database but also extend this kind of implementation with the ability to transform these bits into full-text report streams. The danger faced by institutions implementing these products is the dependence on these proprietary tools for generating full text; as time passes and vendors or support models change the possibility exists that an institution might get stuck without the ability to use the text output of these products.

Ideally over time vendors implementing these modules will grow to conform to a set of standards for the structuring of these reports (see Section 5.2). A trend such as this might imply the need for a set of standards for interfacing internal data storage mechanisms to full-text generation engines.

4.4.Localized knowledge base

Substantial effort was made locally to model the information domain germane to Ambulatory Pediatrics. This effort entailed analysis of a commonly used standard vocabulary, augmentation of terms available from this vocabulary with additional vocabulary to meet the degrees of specificity needed locally, and the creation of a data model and data dictionary to aid in managing the relationships between these terminologies. Generally speaking, it is evident that there is for each clinical domain a gradient of applicability of standard vocabularies. It is also evident that where a particular domain falls on this gradient impacts the extent to which localized vocabularies are integral to the management of clinical observations.

The authors of this response are familiar with the parallel efforts by the CORBAmed DTF to define standards for Lexicon Query Services. It seems likely that many of the services required to manage local knowledge of this kind might be implemented in next-generation systems in accordance with CORBAmed standards to be issued. The need is mentioned here as this effort was a significant activity related directly to the success of the observations capture systems. Ideally lexicon interfaces will work their way into several of the modular pieces of observation systems, such as form development and verification for quality assessment.

4.5.Dynamic involvement of multimedia

An additional aspect of the SEURAT system not described above is the ability to visualize a graphical representation of a particular patient's growth patterns against a range of clinically typical patterns. This visualization is generated by combining the measurement information taken by nursing staff into a predesigned chart of standard growth patterns. The physician may use this chart as an additional tool for assessment which influences the decisions eventually recorded on the SEURAT form.

In this case, the graphical representation of these observations is generated from within the database application. It is possible that as similar systems grow and this pattern of using

multimedia for dynamic presentation and capture of observations data is seen in other domains, specialized applications might come into use for managing only this multimedia functionality. Whether existing OMG or related standards speak to the interfaces required for this kind of distributed processing is beyond the expertise of the authors.

4.6.Human verification and information staging

Critical to the success of SEURAT is the built-in role of the quality assessment clerk. Every form scanned by the imaging and recognition system is verified, field by field, by the clerk, who is trained in using the technologies behind the system. Without this role, the entire system would be subject to the accuracy of recognition applications, which fall short of the quality standards set by the records deparment.
It is hoped that this role might be implemented in conformance with the standards eventually set by CORBAmed. Perhaps this role is part of a larger context of information "staging," wherein automated processing of observations data might move through a series of states. These states or stages could be used to persist information about who created which sets of data and under whose authority and approval it happened. This might be useful both for retrospective case review and for the interim management of data which has not gone through an entire workflow (and approval) process.

It seems likely that significant pieces of this idea exist in naming or directory services such as LDAP and digital signature models with certificate authorities or key-based encryption. The human aspects of this issue are discussed further in Section 5.1. Whether these or related OMG standards speak to this issue is beyond the expertise of the authors.

5.Discussion

5.1.Modeling the humans

Perhaps worthy of consideration by the CORBAmed DTF is the ability to generically represent the roles played by humans in the management of clinical observations. For instance, the role of the quality assessment clerk (described in Section 4.6) is integral to the SEURAT workflow process, as no data is persisted in the database without approval from this authority. A similar example of a human role is the transcriptionist, the "processing unit" involved in transforming dictated reports into digital or paper formats. In a sense, neither of these roles is an essential source of data yet each performs a significant transformation. It is beyond the expertise of the authors to envision how this might affect the development of interface standards.

5.2.Data and documents--Related Standards

There is a useful distinction to be drawn between data and documents in the context of clinical observations. System developers are often interested in the management of "data"--which for this discussion might be best thought of as various ways to map certain values to certain fields and thereafter relate fields and their corresponding values to each other. On the other hand, those providers concerned with the management of clinical observations are more likely used to thinking of "documents"--encounter forms, dictated reports, referral sheets, etc.--as a main frame of reference.

The value of drawing this distinction lies in the added levels of granularity that become available when each is available as a frame of reference for systems operating on clinical observations. A simple example might be to refer to a collection of observation capture forms as a category of "documents" which have varying "data" associated with each. For any particular document, systems might manipulate each data field separately for integrity checks or storage, but the sheet of paper might remain as a permanent part of the medical record. In this case, we can used the stored data in any way we see fit, but as soon as the paper from which any particular data value came disappears the context in which that data was captured (the observation form) is no longer available to relate the data value to another set of values. If the document source for that value is represented with some sort of persistent state, this larger context is preserved. Just as relationships between data elements are often maintained in modular fashion through relational data models, relationships between documents can be managed through classification of document types and relationships between these classifications and appropriate sub-classifications.

This discussion is not intended as an argument for one particular standard over another, nor for any single definition of these concepts. Rather it is included as a suggestion to the DTF that the document metaphor might be of great utility in the managment of clinical observations. There are several efforts in vertical domains to integrate a standardized approach to documents with other horizontal standards. CORBAmed might consider the following vertical domain document architectures (information about each is available from http://www.sil.org/sgml/sgml.html):

The Kona Architecture, currently a draft proposal of the HL7 SGML/XML SIG (a separate response to the CORBAmed RFI is forthcoming from this group)

SAE J2008 (and T2008) Automotive and Truck Standard for mandated documentation

ATA DTDs - Air Transport Association of America architecture for product documentation

Additionally, the following horizontal standards activities might be of similar relevance:

The World Wide Web Consortium Document Object Model, at http://www.w3.org/MarkUp/DOM/

Extensible Markup Language (XML), at http://www.w3.org/XML/

International Organization for Standardization (ISO). ISO/IEC 10744:1992. Information Technology - Hypermedia/Time-based Structuring Language (HyTime). Edited by Charles F. Goldfarb (with assistance from Steven R. Newcomb). Geneva: International Organization for Standardization/International Electrotechnical Commission, 1992. (overview available online at http://www.sil.org/sgml/sgml.html)

An early application for an integrated document approach might be the robust model described in DICOM Supplement 23: Structured Reporting (available at http://www.mcis.duke.edu:80/standards/HL7/committees/image-management/HTML/im-home.html). This model calls for report linking mechanisms currently unavailable in any of the health care data standards. A small informal workgroup (composed of DICOM, HL7, and SGML community representatives) will gather at the 1997 AMIA Fall Symposium in Tennessee in October to examine how the model might be expressed using SGML syntax compliant with the proposed HL7 Kona Architecture. Any report out of this activity will be widely available.

5.3.Integration of local/external, known/unknown

Critical to the success of SEURAT is the ability to recognize that data is missing from the local community of systems. For example, many patients visit Yale-New Haven Hospital for the first time having received extensive preventive care at other instutions. Immunization guidelines impose an extensive series of conditions and results based on both the known and unknown histories of immunizations received. Thus if it is known by a decision support system that certain details about immunization history are missing, this lack of knowledge is often still processable as input sufficient enough to generate a series of suggested interventions.

It is therefore important to be able to record a discrete lack of knowledge in many instances. The corollary need for a system which knows what it does not know might be an ability to discover external resources which might contain information to fill this knowledge gap. It is possible that services provided through the PIDS specifications will enable some or most of the interfaces required for this kind of interaction.

6.References

6.1.Contact Information

The principal point of contact for further communication regarding this response is:

Daniel Chudnov
Cushing/Whitney Medical Library
Yale University
333 Cedar St., P.O. Box 208014
New Haven, CT 06520-8014

Phone: (203) 785-4347
Fax: (203) 785-4369
Email: daniel.chudnov@yale.edu

6.2.Sources

Shiffman RN, Brandt CA, Hoffman M, Wiig W, Fernandes LA. SEURAT: Scanned Entry of Structured Data for a Pediatric Health Maintenance Record System. Accepted for publication in *Proceedings of the 1997 American Medical Informatics Association Fall Symposium*.

The Kona Architecture. Proposal under development by the HL7 SGML SIG, available at http://www.mcis.duke.edu/standards/HL7/committees/sgml/kona.htm.