HL7 Version 3—An object-oriented methodology for collaborative standards development[1]

George W. Beeler

Mayo Clinic, 200 First St. S.W., Rochester, MN 55905, USA

Abstract

In January of 1997, Health Level Seven (HL7) began developing Version 3.0 of its standard. The Version 3 effort represents a transformation of the way that HL7 and its Technical Committees will develop future HL7 information interchange standards. This transformation involves applying object-oriented modelling to the development and specification of information interchange standards. This paper discusses the rationale that led HL7 to undertake this change and provides an overview of the Version 3 Message Development Framework which is HL7's new methodology. It also considers the features of the Version 3 methodology that can facilitate the development of international collaboration and consensus in health informatics standards. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

1. What is HL7 today?

In order to understand the changes that HL7 is undertaking, it is necessary to look at what HL7 has accomplished and what it is today. HL7 was formed eleven years ago. It began as a consortium founded at the instigation of a group of health care providers, and set out to develop a protocol for the exchange of healthcare information in clinical settings. The organization adopted a 'just do it' approach, i.e. they followed a pragmatic path to a solution with no particular forethought to the message development methodology they would be using. With this approach, HL7 produced a prototype specification in the first year, and the first formal version of the standard, Version 2.0, followed a year later.

The core concept behind the HL7 Version 2 specifications is the notion that an external event, a trigger event, occurs and is recog-

```
Segment                       Comment
MSH                           Message header
[{NTE}]                       Notes and comments
[    PID                      Patient identification
    [{NTE}                    Notes and comments about the patient
    [{AL1}]                   Allergy data
    [PV1]             ]       Patient Visit
{ORC                          Common Order
    [    Order Detail         chosen from OBR, RXO, RQD [RQ1], {ODS},
                                  {ODT}
            [{NTE}]           Notes and comments about the order
            [{OBX             Observational results
            [{NTE}]    }]     Notes and comments about results
    ]
    [BLG]                     Billing
}
Notation:    [...] is 0 to 1,          {...} is 1 to many,        [{...}] is 0 to many
```

Fig. 1. Segment diagram for one of the HL7 Version 2.3 messages.

nized by a healthcare computer application. After recognizing the event, this application sends a specific message based on that trigger event through the network to one or more receiving applications. It is important to note that HL7 does not specify the communication protocol, only the trigger event and the message.

The Version 2 messages are delimited ASCII strings divided into segments, and into fields within the segments. Generally, the information content of each segment relates to a particular concept or entity in the healthcare domain. Fig. 1 is a sample segment diagram for one of the Version 2.3 messages. This diagram provides the segment structure, segment optionality and cardinality, and a comment indicating the information content of each segment in the context being used.

HL7 has grown significantly using this technical foundation. It has published three more releases of the standard. In 1994, it was accredited by the American National Standards Institute resulting in Versions 2.2 and 2.3 ranking as American National Standards. HL7 has maintained upward compatibility of the message structure in the Version 2 series, and has greatly increased the scope of clinical

functions that its messages support. The current HL7 standard, Version 2.3, specifies over 300 distinct messages and trigger events that use 113 segment types, 50 datatypes, and 1250 defined data elements. HL7 standards are widely used both in inpatient and outpatient clinical care settings.

At present, HL7 has 1700 members including individual members and representatives from 450 member organizations. HL7 has six, active international affiliate organizations that are working to create HL7 implementations for use in their own countries. Over 400 participants registered for the most recent HL7 meeting. The HL7 Working Group is made up of Technical Committees and Special Interest Groups (SIGs) that provide the healthcare domain and technology expertise needed to develop the standards.

The current Technical Committees and SIGs provide domain expertise in: patient administration, financial management, medical orders, inter-enterprise referrals and scheduling, clinical results reporting, information and medical record management, vocabularies and medical terminology, patient care goals and guidelines, decision support, and home health services. Other committees

provide technical expertise in the areas of: message control and queries, interface implementation, modelling and methodology, HL7 architecture and quality assurance, automated data, interface conformance, image management, object broker technologies, secure transactions, and SGML-based messaging.

2. Why adopt a Version 3 methodology?

Given the current strength and scope of HL7, one might ask why HL7 is changing to a new methodology. The answer is that the technical structure of Version 2 has carried the organization about as far as it can.

The principal limitation is that the current standard is not easily implemented. Any particular interface may require many weeks of analyst time to implement. The communicating systems must agree on whether to use the optional fields, and must be certain that they have the same semantic interpretation of the data elements. Without this analysis, an interface will suffer from the inconsistencies that are either inherent in the standard or that arise because of the way various groups interpret the standard.

Secondly, the messages in Version 2 have a large number of optional segments and fields. These preclude rigorous conformance testing. Finally, the amorphous development process makes collaboration between HL7 technical committees difficult, and the standard does not lend itself to implementation in alternate communication protocols.

3. Overview of HL7 Version 3

To address these shortcomings, HL7 undertook development of the Version 3 methodology. This development began formally in 1994. The Version 3 Task Force benefited from prior collaboration between its members and members of other standards groups in such efforts as the IEEE P1157 (MEDIX) Committee, and the activities of CEN TC-251 (especially Project Team 25 for Working Group 3).

In January of 1997, the Version 3 Task Force published the HL7 Version 3 Message Development Framework (MDF) and introduced it to the Working Group for its use. The MDF is a complete, fully documented, model-based methodology for developing message specifications. It specifies four models to be developed in the course of producing a message standard. Fig. 2 is a diagram of these models. This figure includes a box for each of the four models, and symbols representing the documentation for the model. It also shows annotation balloons that indicate the development steps for each model. Although the arrows between models indicate the primary sequence of development, the actual development process will be cyclical. Development teams will return to refine earlier models after they have undertaken definition of one of the later stages, and have learned of additional requirements. (The MDF document can be reviewed at http://www.mcis.duke.edu/standards/hl7/hl7.htm)

The methodology of the MDF is based on object-oriented methodologies for the use case, information and interaction models. These methodologies are constrained in order to ensure that the resulting process meets the needs of message standards development rather than of application development.

HL7 message development is a distributed process. Each Technical Committee contributes to the standard in its domain of expertise. A modelling facilitator from the Modelling and Methodology Committee assists each committee in doing its modelling and helps with the tasks of model harmoniza-
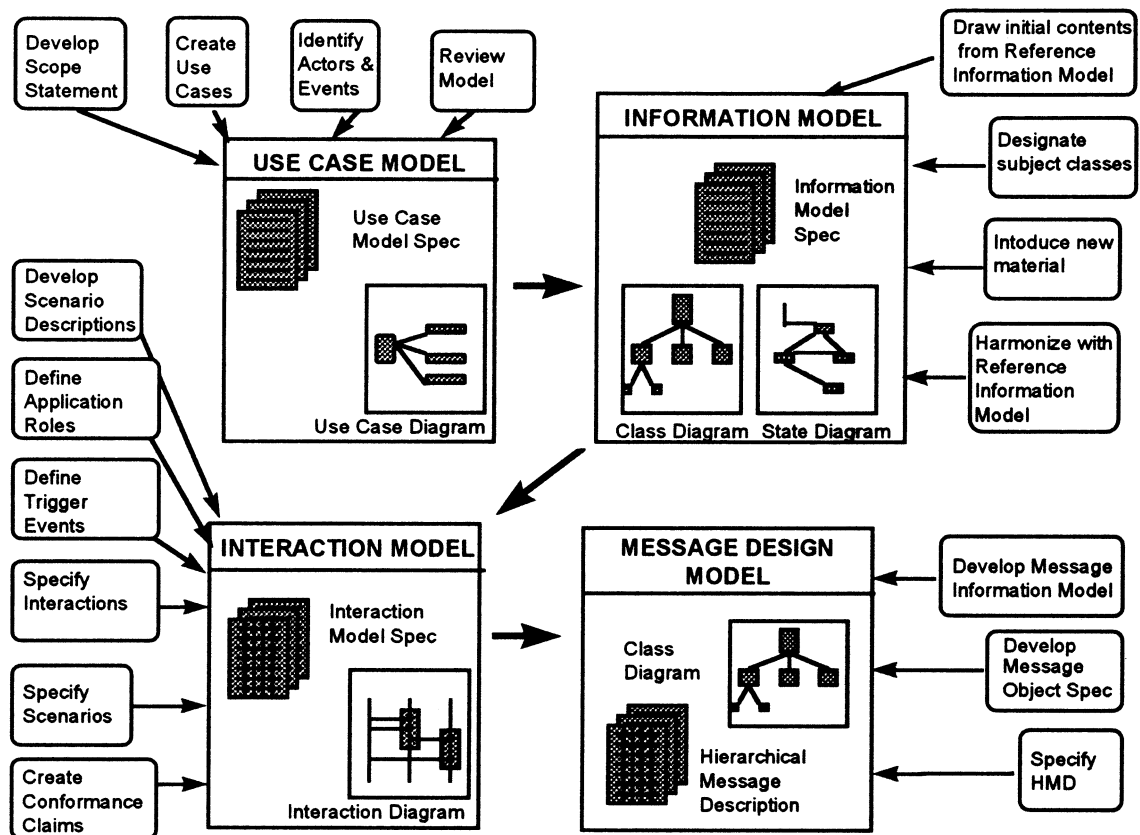
Fig. 2. Diagram of the primary models, development steps and documented deliverables specified by the HL7 Version 3 MDF. Figs. 2–5 are reproduced with permission of Health Level Seven, Inc.

tion. Technical specialists and HL7 support the committees in the specification of the Implementable Message Specifications and in publishing the standard.

3.1. Use Case Model

The Use Case Model is the first in the series of models to be developed. It defines the circumstances in which information must flow between applications, and be managed by them. It documents the expectations for the behavioral relationships between communicating systems that use HL7, and forms the basis for identifying and defining the key information concepts.

3.2. Information Model

The Information Model is perhaps the most critical element in the process, because every HL7 message must draw its information content from a single shared model. The shared model is known as the HL7 Reference Information Model (RIM). It is a complete class model that includes subject areas, attributes for each of the classes, inheritance structures and instance connections or relationships between the classes. In addition, it includes a state transition diagram to express the life cycles for each of the subject classes.

3.2.1. Information model development process

Due to RIM being used by technical committees, and because each committee is expected to provide model content in its area of domain expertise, HL7 develops the RIM using a process that provides for distributed development within each committee coupled with periodic harmonization reviews to combine the changes recommended by the various committees. Each class in the model has a steward committee responsible for coordination of changes for the class. This stewardship provides critical review and continuity during distributed development.

At the beginning of the process, each committee extracts from the RIM the classes of interest to their task. It uses these classes as a working model. Subsequently, the committee develops proposals to extend or change this working model in order to meet the needs of their domain. This usually occurs during the Working Group Meetings. In between each pair of Working Group meetings, the Methodology and Modelling Committee collects the proposed RIM changes from each committee. It posts these on the HL7 web site for review. Finally, before the next Working Group meeting, it conducts a harmonization meeting. During this meeting, each committee presents and defends its changes to a harmonization committee made up of a steward and a modelling facilitator from each technical committee. After hearing the discussion, including the recommendation of the class steward, the harmonization committee votes to accept or deny each proposed change. Regardless of the outcome, the technical committees are expected to use the harmonized version of the RIM as the basis for their message development.

The goal of the harmonization process is to develop the RIM as a coherent, shared information model which will provide all of the content of HL7 messages. The process assures that the RIM is a joint work product of the Working Group as a whole, and is built using the functional knowledge of the Technical Committees.

3.2.2. Subject classes

Subject classes in the information model provide a direct link to the trigger events specified in the interaction model. Subject classes are those classes whose information must be actively managed by clinical healthcare computer systems. For example, in the HL7 domain, classes like patient, orders and results are commonly the management responsibility of one or more computer applications. On the other hand, classes like 'insurance contact' may provide content for messages, but are rarely actively managed by these applications. Because of their central role in HL7 domains, the subject classes are the foci for trigger events. Therefore, a complete life-cycle is specified for them in the form of a state diagram. Each state transition in the diagram is a potential trigger event.

3.2.3. Starter version of HL7 RIM

In order to launch Version 3, HL7 commissioned and funded a project to develop a starter, draft RIM. The project used models from a variety of sources. These include models from HL7 technical committees, from other standards developing bodies, and from several organizations that are members of HL7. The concepts embodied in these models were combined in a single model, giving precedence to the HL7 Technical Committee models. The draft starter model received a preliminary review by representatives of several technical committees and then was published in January 1997 for use by the HL7 Working Group.

HL7 used this starter RIM to initiate the iterative harmonization process described above. As of this writing, there has been one

harmonization cycle, and a second will complete in November 1997. The complete RIM is fully documented in a variety of modalities and can be downloaded from the HL7 web site at http://www.mcis.duke.edu/standards/HL7/hl7.htm.

3.3. Interaction Model

The Interaction Model defines the behaviors of systems that communicate using HL7 messages. In the Interaction Model, the committees define trigger events, interactions and application roles. Trigger events are derived from the use case model and from the state transitions of the subject classes in the RIM.

Interactions are single, one-way information transfers that associate a message, a trigger event, and two application roles. One application role sends the message and one receives it. The interactions are equivalent to the trigger event/message combinations defined in versions 2.x of the standard.

Application roles are new in Version 3. They are sets of responsibilities for messaging that a particular application may take on. Each role is defined in terms of the interactions it must send or receive in order to support a particular function.

Typically, an application will play several roles. For example, an orders system might take on the roles and responsibilities of a 'patient tracker,' an 'order manager' and a 'results tracker.'

Application roles are central to the definition of application conformance. Message formats in Version 3 will not have optional fields or segments. This is possible because the message design process (next section) simplifies the creation of multiple formats or profiles for a given message structure. Because of this simplification, the standard developers will be able to create a specific message format for each interaction. Thus,

each application role will be defined as a specific set of interactions where the role has the responsibility to send or receive a particular, unambiguously defined message. This specificity will allow both exact definition of conformance claims and rigorous testing of these claims.

3.4. Message Design Model

The final design step undertaken by a technical committee is the Message Design Model. It is here that a message format is defined from the information content of the RIM to meet the message requirements of each interaction. The committee extracts from the RIM those classes, attributes and connections needed for a particular set of messages. This extract is the message information model (MIM). Not only is the MIM a proper subset of the RIM, but it may also further constrain elements of the RIM such as object cardinality attributes and allowable specialization. Fig. 3 is an example MIM.

The MIM provides the basis for constructing a message object diagram. This diagram is a structured set of objects that instantiate the classes of the MIM. The process of converting the MIM to a data structure is too detailed to present in full in this paper. It involves a 'walk' along the graph of the MIM, guided by the cardinality of the connections and the known requirements of the task. In the course of the walk, class instances are selected and deposited in the message object diagram. This process was first developed in CEN TC-251 WG-3 Project Team 25 [1], and used in several CEN message standards.

The next message design step expands the objects in the message object diagram by specifying for each object in the diagram the attributes that are candidate data fields for the message. The result is the structural
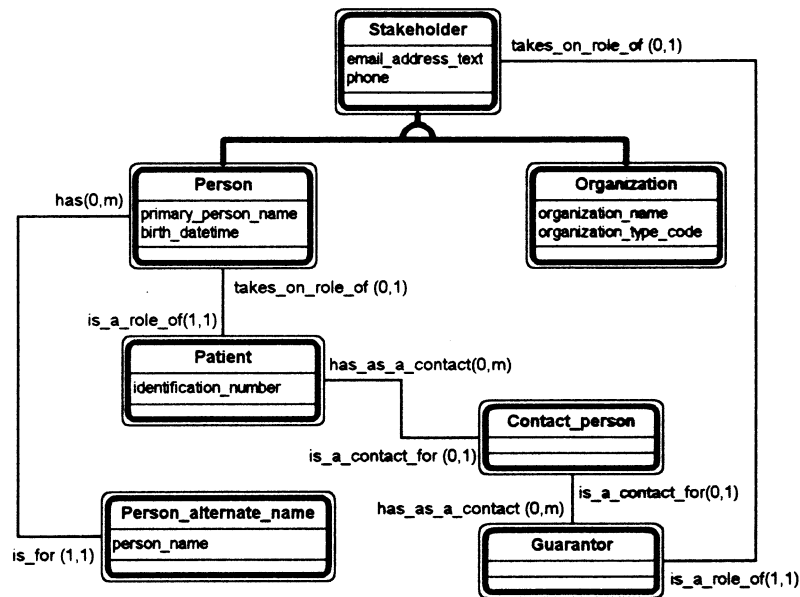
Fig. 3. Hypothetical Message Information Model for exemplar HMD in Figs. 4 and 5.

(right-hand) portion of a Hierarchical Message Description (HMD). This structure is the abstract representation from which each message format is drawn. Fig. 4 contains an example right-hand side of an HMD. The MIM from which the HMD is drawn is shown in Fig. 3. This HMD structures the object views and the attributes, expresses the relationships that led to the inclusion of these elements in the diagram, and provides identifiers, labels and an additional description for each row. Note that the bold-faced rows, those with an assigned 'Seg ID,' are directly analogous to the message segment structure diagrams of the Version 2.x standards. (See Fig. 1)

The final step in the process starts with the core structure contained in the right-hand portion of the HMD. This step defines one or more message formats that are recorded in the left-hand portion of the HMD. These message formats are profiles defined for particular interactions. An example is shown in Fig. 5. Each of the message formats show whether a particular row is included in that format, whether it is conditionally included and what its optionality and cardinality will be in the message. The series of message formats or profiles for a particular HMD are represented in the left hand sets of columns, as in Fig. 5 which shows all columns of the format for message A55, and the first few columns of the format for message E31. A typical HMD will have several such formats. The example includes two different profiles defined to meet the requirements of two different interactions.

To a certain degree, HL7 Version 2.x stopped with the specification of the message structure (right-hand side of the HMD) and left the specification of profiles as an implementation activity. In contrast, Version 3 will include particular profiles or formats for each interaction.

Relationship	Seg ID	Label	Object Views and Attributes	Description	
root	PT		Patient		1
			CD ambulatory_status_code		2
	`'`		ST identification_number		3
is_a_role_of	PER		Person[Pt]	Patient	4
			PN primary_person_name		5
			DT birth_datetime		6
has	ALN		Person_alternate_name		7
			PN person_name		8
has_as_a_contact	GUA		Contact_person		9
is_a_contact_for			Guarantor		10
is_a_role_of			Stakeholder		11
			TN phone		12
			ST email_address_text		13
specialization			*choice*		14
choice	GUAP		Person[Guarantor]	Guarantor	15
			PN primary_person_name		16
choice	GUAO		Organization		17
			ST organization_name		18
			CD organization_type_code		19

Fig. 4. Details of the right side of a Hierarchical Message Description.

3.5. Implementable message specification

HL7 will use the Hierarchical Message Description as the primary focus for its ballots. Once the HMD is approved by ballot, it will be used in combination with one or more Implementable Message Specifications (IMS) to define actual messages and communications. An IMS is specific to a particular communication protocol. It provides the algorithm for using an HMD to generate or interpret message structures appropriate for that communication protocol.

The communication protocols for which IMS's will be defined may be printable character streams, similar to HL7 Version 2, alternate character syntaxes such as EDIFACT, object interfaces suitable for communication with object middleware (such as CORBA ORBS and Microsoft's OLE automation) or alternate representations such as SGML.

4. Features of Version 3 that favor international collaboration

The preceding section of this paper outlines the methodology HL7 will be using in developing Version 3 of its standard. HL7 plans to begin publishing V3 specifications during 1998, and will make its 'work-in-progress' available at the web site throughout this period. The next question to consider is how might the Version 3 process assist in bringing about cooperation and consensus for international health informatics standards. Four particular topics are worthy of consideration—communication adaptability, collaborative standards development, the ability to use a variety of codes and vocabularies; and the option to specialize portions of the standards in order to address region- or nation-specific requirements.

E31		Message Format A55								
Non-ality	Card	Conditional Presence	Required Value	Option-ality	Card	Relationship	Seg ID	Label	Object Views and Attributes	Description
R	1,1			R	1,1	root	PT		Patient	1
R	1,1		IPT	R	1,1				CD ambulatory_status_code	2
M	1,1			M	1,1				ST identification_number	3
R	1,1			R	1,1	is_a_role_of	PER	Patient	Person[Pt]	4
M	1,1			M	1,1				PN primary_person_name	5
R	1,1			R	1,1				DT birth_datetime	6
R	0,m			NP		has	ALN		Person_alternate_name	7
M	1,1			M					PN person_name	8
R	0,1			R	0,m	has_as_a_contact	GUA		Contact_person	9
R	0,1			R	0,1	is_a_contact_for			Guarantor	10
R	1,1			R	1,1	is_a_role_of			Stakeholder	11
R	1,m			R	1,m				TN phone	12
P	1,m		SS	R	1,m				ST email_address_text	13
						specialization			choice	14
R				R		choice	GUAP	Guarantor	Person[Guarantor]	15
R	1,1			R	1,1				PN primary_person_name	16
R				R		choice	GUAO		Organization	17
R	1,1			R	1,1				ST organization_name	18
R	1,1			R	1,1				CD organization_type_code	19

Fig. 5. Hierarchical Message Description HMD of Fig. 4 with a message format added.

4.1. Communication adaptability

Communication adaptability is the ability to realize messaging in a variety of forms in order to meet the specific protocol requirements of a given country or area. This adaptability exists because the entire domain-specific content of messages is captured in the abstract message definitions of the HMD, and these definitions can be readily mapped to multiple protocols.

Thus the methodology provides equally well for communications with ASCII strings or via ORB interfaces; for protocols that are standard protocols or those that are not; for broadcast or point-to-point communications; for file, message or transaction-based messaging; and for the definition of a variety of application roles which aggregate communication responsibilities.

4.2. Collaborative standards development

The second feature of the methodology that lends it to international collaboration is that the methodology is designed to support collaboration between disparate interests. After all, the HL7 technical committees and SIGs collaborate in the development of HL7. The prospect of international collaboration arises because the process is model-based, and thus there is the opportunity to harmonize the models on a broader scale.

Indeed, in August 1997, HL7 agreed to accept recommendations for RIM changes from any accredited standards developing organization. This was undertaken as a first step towards opening the process to a broader range of participants. The HL7 process offers the opportunity for collaboration; for discourse on national and regional differences and similarities; and for specialization of the model to meet local needs. Thus, it may serve as a prototype for harmonizing disparate standards in the future.

4.3. Adoption of various codes and vocabularies

The third advantage of Version 3 lies in options that it provides for using various codes and vocabularies in its messages. In 1996, HL7 formed a Special Interest Group on vocabularies. This SIG is co-chaired by Drs Stan Huff, James Cimino and W. Ed Hammond. According to its charter, the SIG

seeks to provide an organization and a repository for maintaining a coded vocabulary that can be used in conjunction with HL7 and related standards and that will enable the exchange of clinical data and information in such a way so that sending and receiving systems have a shared, well defined, and unambiguous knowledge of the meaning of the data being exchanged.

To achieve this goal, the group will work cooperatively with other groups having an interest in coded vocabularies used in clinical computing. These groups will include: standards development organizations, creators and maintainers of vocabularies, government agencies and regulatory bodies, clinical professional specialty groups, vocabulary content providers, and vocabulary tool vendors.

In order to accomplish its goals and to minimize redundant effort, the SIG will seek to use existing vocabularies or term lists from authoritative governmental and private groups. During its meeting in August 1997, the SIG began to articulate a series of principles defining the kinds of terminologies that might be used. These preliminary, unballoted principles would accept: terminologies that are compliant with the semantics of the HL7 message structure; terminologies that are contributed as a source to the UMLS meta-thesaurus; terminologies for which there exists an organization committed to the timely maintenance and update of the terminology; terminologies that are free from excessive license fees; and terminologies that are comprehensive for the intended domain of use.

These principles will not exclude country-specific terminologies mandated by regulations, provided that such terminologies are compliant with the semantics of the HL7 message structure. Finally, the group will probably favor terminologies that have oversight by nonprofit professional organizations in the healthcare field. Once the SIG has refined these principles, it will offer them for ballot by the HL7 membership.

4.4. Specialization to meet region- or nation-specific requirements

The models upon which Version 3 is based provide numerous opportunities to define solutions that meet region-specific or nation-specific needs. In particular, the information model can be adapted to meet these needs by adding new classes or by creating specializations of existing classes. Local terminologies and coding structures unique to a particular area can be used with the standard messages. Application roles may be tailored to the specific requirements of a set of users. New trigger events and/or new interactions can be defined, thus creating a region- or nation-specific message interaction. Regional or cultural differences in such things as patient demographics (e.g., name, address, orthographic variety of name, etc.) are readily accommodated via specialization.

5. Conclusion

The strength of HL7 Version 3 comes from the flexibility of the models upon which it is based, coupled with the defined methodology for using these models to develop standards. These same strengths make HL7 Version 3 a worthy candidate as a methodology to use in generating collaboration and consensus between multiple standards developers in an international effort.

Because of its limited length, this paper has provided only a flavor or sense of what HL7 Version 3 entails. The reader is encouraged to visit the HL7 web site at http://www.mcis.duke.edu/standards/HL7/hl7.htm to learn more. Links from that page lead to the complete Version 3 Message Develop-

ment Framework; the current rendition of the HL7 Reference Information Model; and the Work in progress of each of the many technical committees and SIGs of HL7.

References

[1] Segars, D., Methodology for the Development of Health Care Messages, CEN/TC 251/PT3-025 Draft Final Document, Dirk Segars Project Team Leader CEN Central Secretariat, Brussels, 1995.

.