NIST **National Institute of Standards and Technology** • U.S. Department of Commerce

Healthcare Information Technology Standards: General Suitability Analysis

Lantana Consulting Group

NIST **National Institute of Standards and Technology** • U.S. Department of Commerce

Healthcare Information Technology Standards: General Suitability Analysis

Healthcare Information Technology Standards:

General Suitability Analysis

May 2011

Prepared by Lantana Consulting Group

for

National Institute of Standards and Technology

Lantana Consulting Group
PO Box 177

East Thetford, VT 05043
www.lantanagroup.com

Liora Alschuler
Chief Executive Officer
liora.alschuler@lantanagroup.com

Bob Dolin, MD, FACP

President and Chief Medical Officer
bob.dolin@lantanagroup.com

Rick Geimer
Chief Technology Officer
rick.geimer@lantanagroup.com

Lauren Wood, Ph.D.
Senior Project Manager
lauren.wood@lantanagroup.com

Gaye Giannone Dolin, RN, MSN
Senior Nurse Informaticist
gaye.dolin@lantanagroup.com

Jingdong Li, MD
Chief Architect
jingdong.li@lantanagroup.com

Yan Heras
Senior Medical Informaticist
yan.heras@lantanagroup.com

Bob Yencha
Principal Consultant
bob.yencha@lantanagroup.com

Kate Hamilton
Senior Information Analyst
kate.hamilton@lantanagroup.com

Susan Hardy
Senior Technical Editor
susan.hardy@lantangroup.com

Acknowledgments

The methodology described in this document was produced and developed by Lantana Consulting Group for the National Institute of Standards and Technology (NIST). It is part of a series of suitability analsyses on four healthcare standards performed under the following contracts:

Any mention of commercial products or organizations in this report is for information only; it does not imply recommendation or endorsement by Lantana or NIST.

Table of Contents

List of Figures

List of Tables

Executive Summary

This document presents a methodology developed by Lantana Consulting Group for the National Institute of Standards and Technology (NIST) to evaluate the suitability, or "fitness for purpose," of healthcare information technology standards.

We document the process by which we chose the evaluation criteria; this process included a literature review and stakeholder interviews. We defined the criteria such that they may be applied to any given healthcare information technology (IT) standard. We then refined the criteria after applying them to four healthcare IT standards: Health Level Seven (HL7) Continuity of Care Document (CCD), ASTM Continuity of Care Record (CCR), HL7 Quality Reporting Document Architecture (QRDA), and HL7 V2 Biosurveillance. (Analyses of these four standards based on the criteria are covered in separate documents we prepared for NIST.)

We present the criteria as a series of questions. We believe that they provide a useful framework for evaluating the suitability of any healthcare IT standard for its intended purpose.

Introduction

This document presents a general methodology for assessing the suitability of information technology (IT) standards used in healthcare. It was developed for the Healthcare Information Technology (HIT) Standards Analysis Project and used to determine the suitability, or "fitness for purpose," of four standards developed by the healthcare community: Health Level Seven (HL7) Continuity of Care Document (CCD), ASTM Continuity of Care Record (CCR), HL7 Quality Reporting Document Architecture (QRDA), and HL7 V2 Biosurveillance.

The word "standard" as used in information technology has multiple meanings. In this document, "standard" refers to a generally accepted specification created by an organization that includes more than one corporation or company. Ideally, a standards development organization or industry consortium uses a consensus-driven process to create the standard. These organizations may be international, regional, or national governmental, quasi-governmental, or non-governmental entities. Current US Government policy encourages all relevant procurement to be based on standards whenever possible. The National Technology Transfer and Advancement Act of 1995 (NTAA) encourages the use of technical standards that are developed or adopted by voluntary consensus bodies; it does not limit the use of "standard" to international standards development organizations such as the International Organization for Standardization (ISO) or IEEE (originally known as the Institute of Electrical and Electronics Engineers). (See http://standards.gov/ for more details).

Our assessment examines several aspects of a standard including quality (technical errors and inconsistencies within the standard or "parent" standard), testability, compatibility with related standards, and implementability.

Methodology

This project defined a set of suitability criteria and applied them to the standards under review. The Methodology Flow Diagram illustrates this process. As shown in the diagram, the process was iterative; we continued to refine the criteria while we applied them to the standards. We expect the process of adding and refining criteria to continue for as long as this methodology is applied.

The constraint mechanism criterion illustrates a refinement made while applying the criteria. Originally the question was "Does the standard have a well-defined localization mechanism?" After we applied the criteria to the standards, we decided to use the more general question "Does the standard have a well-defined constraint mechanism?" because the more general form also covers extensibility.

To develop the methodology, we researched the four standards to ensure that we knew the requirements for these types of standards. We interviewed industry experts, asking questions about what makes a standard suitable and useful. We also reviewed relevant literature to provide definitions for the suitability components and criteria.

The core of the resulting methodology is a list of criteria that can be applied to each standard. Many of the criteria are qualitative rather than quantitative; in most cases an objective numeric or point-scale answer is not possible. The suitability of any given standard should be considered in terms of how many of these criteria it satisfies rather than any individual component. Not all criteria are applicable to all standards. This general methodology represents the union of suitability criteria.

The Healthcare Information Technology Standards Panel (HITSP) followed a different approach to evaluating and selecting standards. We designed our approach independently of the HITSP work, so any overlap or replication of results shows that different approaches to evaluating standards can produce similar answers.

The Suitability Criteria section describes the criteria and definitions. In the final step of the analysis, we prepared a report for each individual standard, based on these criteria. In the individual reports, we also evaluated whether the standard can support the US Department of Health and Human Services (HHS) Meaningful Use requirements.

Figure 1: Methodology flow diagram
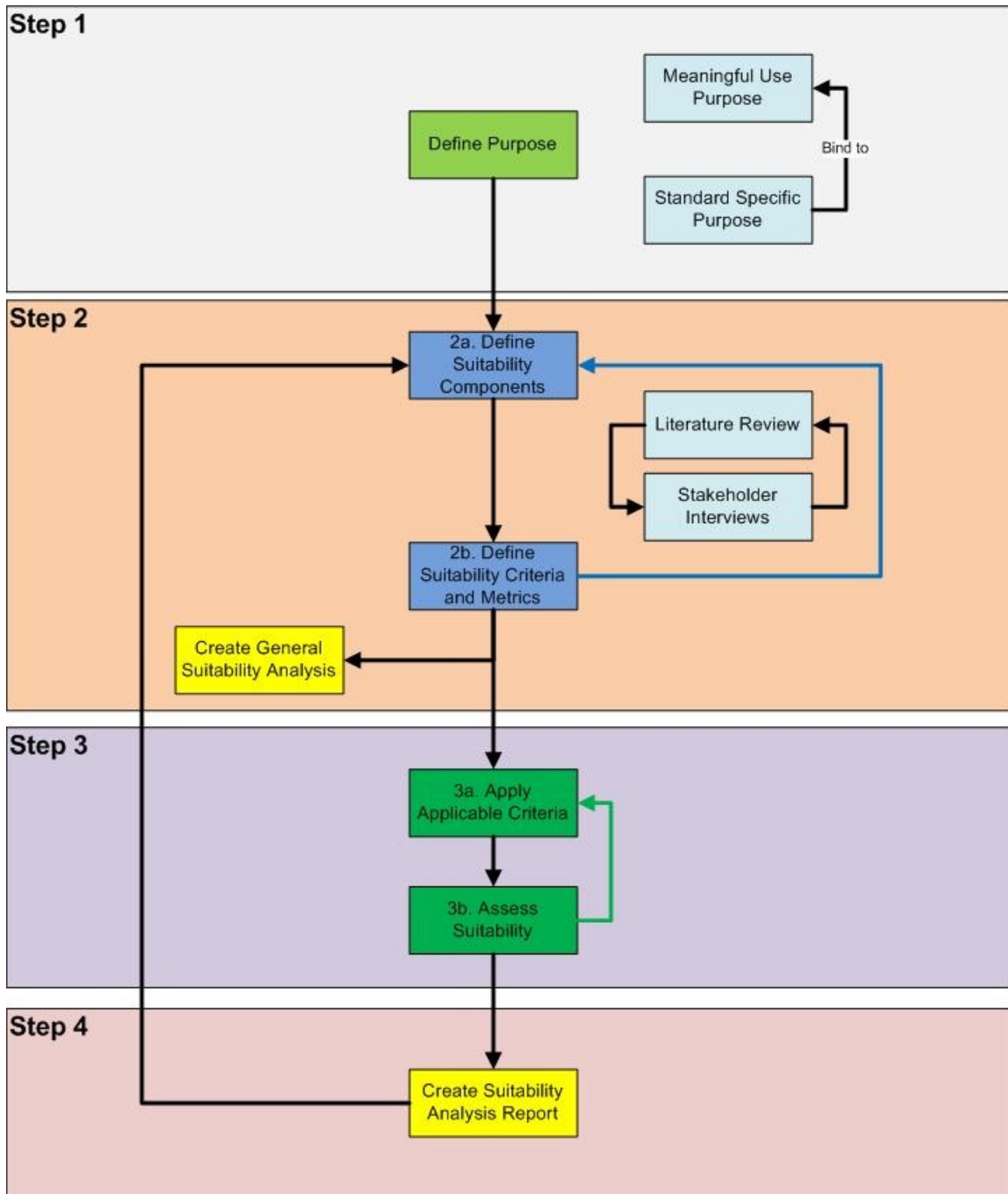
Literature Review

We define the term "literature" to include published journal articles and web sites. We reviewed the sources listed in the References section for potential criteria. We refined the list of criteria based on the stakeholder interviews and internal discussions.

We also reviewed the individual standards to determine their purpose as well as their suitability for Meaningful Use. Each suitability analysis report lists the literature reviewed for the individual standard.

Stakeholder Interviews

We interviewed experts in HIT standards development, quality reporting, and standards implementation to obtain input to the criteria and to evaluate HIT standards in general. We asked these experts questions to evaluate the maturity, robustness, and suitability of the four standards: CCD, CCR, QRDA, and the HL7 V2 Biosurveillance Message Type. The tables on General Interview Questions and Standards-Specific Interview Questions list the questions that guided the discussions.

Table 1: General Interview Questions

Question block ID	Interview Questions
General	- What does "suitability" or "fitness for purpose" mean in an HIT standard"? How would you measure this? - When someone says there is a "gap" in an HIT standard, what does that mean to you? - What makes an HIT standard "testable"? How can you quantify this? - What makes an HIT standard compatible? - With what should an HIT standard be compatible? - Give an example of where it is critical that a standard be compatible. - What makes an HIT standard implementable? - What will it take to make standards easier to implement? - What does it mean for an HIT standard to have "specificity"? - What are the advantages/disadvantages of specificity of purpose? - How would you define "extensibility"? How does extensibility related to quality of a standard? - What does it mean if an HIT standard is "localizable"? How does localizability relate to quality of a standard?

We were fortunate to have experienced people respond to our request for an interview. The discussions included senior developers as well as regional and national program directors who shared their experiences implementing standards and their wide-ranging knowledge about standards quality, informed by their real-world experience. In general, the interviewees were most experienced with CCD, CCR, and/or QRDA.

There was consensus among interviewees that a standard should be complete. In other words, a standard should support all of the rules or requirements needed to implement that standard; the need to read and understand multiple standards for one purpose impedes successful implementations. Interviewees saw implementation guides—along with consistent rules on how

to use standard vocabularies and resources—as critical to successful implementations. Another resonant theme was the concept of a "family" of standards—or the ability to reuse the standard (or a variation thereof) for many use cases without having to learn a completely new standard.

We discussed some of the trade-offs that affect the design of standards. For example, a very specific (or localized) standard provides greater interoperability, but covers fewer use cases. Extensibility helps a standard apply to many use cases, but interoperability may suffer.

The standards-specific questions are listed in Table 2.

Table 2: Standards-Specific Interview Questions

Question block ID	Interview Questions
Government related	• Please briefly describe your experience with <CCD/CCR/QRDA/V2 Biosurveillance Message > to help us understand the basis for your answers to the following questions. • What are the expectations from a government perspective for <CCD/CCR/QRDA/V2 Biosurveillance Message > • What are the highest priorities from a government perspective for <CCD/CCR/QRDA/V2 Biosurveillance Message > • To what extent does fitness rely on clarity of purpose?

CCD	• Please briefly describe your experience with CCD to help us understand the basis for your answers to the following questions.
	• The primary purpose of CCD is to transmit summary data.
	• How well is CCD meeting its primary purpose?
	• Are there errors or ambiguities in CCD that mean it's harder than it should be to implement and use? Do you have examples?
	• Consider the primary purpose of CCD:
	   o What are some measurable criteria that should be assessed to ensure it is meeting this purpose?
	   o In what ways does it not meet its purpose?
	   o What makes CCD testable?
	• Where is "suitability" or "fitness for purpose" lacking or present in CCD? For instance, here are some potential areas:
	   o Narrative Interoperability or immediate, accurate rendering in a receiving system
	   o Data reconciliation
	   o Data reuse
	   o Third party aggregation of data (e.g., ability for disparate systems to send uniform data to a central repository)
	   o Others?
	• Please discuss the one feature about CCD that most supports the Meaningful Use ultimate goal of achieving significant improvements in health care.

Question block ID	Interview Questions

CCR	• Please briefly describe your experience with CCR to help us understand the basis for your answers to the following questions.
	• The primary purpose of CCR is to transmit summary data.
	• How well is CCR meeting its primary purpose?
	• Are there errors or ambiguities in CCR that mean it's harder than it should be to implement and use? Do you have examples?
	• Consider the primary purpose of CCR:
	    o What are some measurable criteria that should be assessed to ensure it is meeting this purpose?
	    o In what ways does it not meet its purpose?
	    o What makes CCR testable?
	• Where is "suitability" or "fitness for purpose" lacking or present in CCR? For instance, here are some potential areas:
	    o Narrative Interoperability or immediate, accurate rendering in a receiving system
	    o Data reconciliation
	    o Data reuse
	    o Third party aggregation of data (e.g., ability for disparate systems to send uniform data to a central repository)
	    o Others?
	• Please discuss the one feature about CCR that most supports the Meaningful Use ultimate goal of achieving significant improvements in health care.

QRDA	• Please briefly describe your experience with QRDA to help us understand the basis for your answers to the following questions.
	• The primary purpose of QRDA is to transmit quality data.
	• How well is QRDA meeting its primary purpose?
	• Are there errors or ambiguities in QRDA that mean it's harder than it should be to implement and use? Do you have examples?
	• Consider the primary purpose of QRDA:
	    o What are some measurable criteria that should be assessed to ensure it is meeting this purpose?
	    o In what ways does it not meet its purpose?
	    o What makes QRDA testable?
	• Where is "suitability" or "fitness for purpose" lacking or present in QRDA? For instance, here are some potential areas:
	    o Conveyance of all data needed by a processing entity to compute a provider's or organization's score on one or more quality measures
	    o Data reuse, whereby data having been originally communicated in a CCD (or other CDA document) is analyzed against a quality criterion, and is repackaged as part of a QRDA report
	    o Conveyance of other aggregation data not directly part of the quality measure (metadata)

Question block ID	Interview Questions
	o Ability to improve the quality feedback loop • Please discuss the one feature about QRDA that most supports the Meaningful Use ultimate goal of achieving significant improvements in health care.

Biosurveillance	• Please briefly describe your experience using V2 messages for biosurveillance data exchange to help us understand the basis for your answers to the following questions.
	 • The primary purpose of using V2 messages for biosurveillance data is to expedite event recognition and response among federal, state, and local public health and healthcare organizations.
	 • How well are V2 messages for biosurveillance data meeting the primary purpose?
	 • Are there errors or ambiguities in V2 biosurveillance messages that mean this messaging is harder than it should be to implement and use? Do you have examples?
	 • Consider the primary purpose of V2 biosurveillance messages:
	     o What are some measurable criteria that should be assessed to ensure it is meeting this purpose?
	     o In what ways does it not meet its purpose?
	     o What makes the V2 biosurveillance messages testable?
	 • Where is "suitability" or "fitness for purpose" lacking or present in V2 biosurveillance messaging? For instance, here are some potential areas:
	     o Event detection of reportable conditions
	     o Outbreak management
	     o Emergency response support
	     o Third-party data aggregation from disparate systems across the country
	 • Please discuss the one feature about V2 biosurveillance messages that most allows you to achieve the Meaningful Use ultimate goal of achieving significant improvements in health care.

Suitability Components

The goal of the Standards Analysis Project is to determine the suitability of HIT standards for their intended uses. Suitability, or "fitness for purpose," is difficult to define precisely, and therefore difficult to measure. It is easiest to look at suitability in terms of components. These components include:

Quality assessment (technical errors, ambiguities, and inconsistencies within the standard or "parent" standard)

Testability of the standard

Compatibility of the standard with related standards

Implementability of the standard

Specificity of the standard

A number of criteria can assess each of these components, and all the criteria support the assessment of at least one of the components. The Criteria and Components Matrix table shows the relationship in more detail.

Suitability Criteria

Many criteria can determine the suitability of a standard. Finding criteria that are both objective and measurable while remaining useful for a range of healthcare IT standards is difficult. These criteria cannot always be assessed with objective numbers or grades, but they are useful nonetheless especially when viewed as a whole. This section presents the general criteria applied to all four standards. Individual assessments also evaluated standard-specific requirements for Meaningful Use.

Is the standard based on a stable, well-vetted data model?

There is a data model underlying the standard that is stable and mature (i.e., not going through rapid changes) and was developed with input from many experts within an established standards body to make sure it meets the needs of the community. The data model can be as simple as an Extensible Markup Language (XML) schema, or it can incorporate a full conceptual model that describes all the data relationships and constraints. Regardless of the form and complexity of the data model, it should be robust, applicable to a large number of use cases within the relevant scope, and implementable. It should also be documented in a clear and understandable way that passes muster with the group of experts that reviewed it as well as subsequent implementers who do not have direct access to the original designers.

Does the standard have a clear, robust vocabulary binding syntax?

A clear, robust vocabulary binding syntax defines links between a specific terminology and the information model and is required to represent clinical information unambiguously.

Does the standard support reusable modules, such as templates or data types?

Reusable modules are tightly defined patterns that can be duplicated for reuse within distinct use cases to which the standard is applied. They are defined in one standard and can be reused in other related standards or specifications. Modules promote implementability and consistency through reuse across specifications. Reuse supports rapid development of new, related standards.

Does the standard have a well-defined constraint mechanism?

A constraint mechanism restricts the original standard to meet business and content needs for specific use cases, geographical locations, or realms. A constrained standard is a subset of the original standard (an *is-a* relationship). Constraints reduce the generality of the standard and focus it on a particular requirement. For example, a standard may allow multiple value sets for a data field, but a realm-specific standard limits usage to only one of those value sets. A well-defined constraint mechanism specifies the constraint process so that a message designer can consistently and repeatedly produce use-case, geographic, or realm-specific standards.

For our purpose, localization is a set of constraints on a base standard; it may also use extensions.

Does the standard have a well-defined extensibility mechanism?

An extensibility mechanism supports unanticipated use cases, where the use cases are in scope. A defined method for allowing extensions indicates the standard's designers recognize that needs may arise that were not originally considered. Existence of such a defined mechanism is a positive indicator of quality for that standard. Extensive use of extensions, however, is a negative indicator of quality as it can result from a poor-quality, poorly-scoped standard or poor documentation.

Are there unambiguous definitions of what is testable?

A testable standard has clear definitions of which data elements are required and which are optional. It provides information on whether these definitions can be tested and which parts, if any, of the specification can't be automatically tested. It should indicate which tests can be automated and which, if any, require natural language testing. Where possible, the definitions should be machine-processable, rather than written in a natural language.

Are there automated test tools and test suites?

The existence of automated test tools, test suites, and test rules for a standard is a positive indicator that documents or messages created according to the standard can be validated and tested. The standard itself is tested for quality (lack of ambiguity, clarity, and completeness) during the development of such tools and test suites. The existence of such test tools and test suites makes it easier for implementers to develop correct implementations of the standard, and may therefore increase the rate of adoption.

Are there reference implementations for the standard?

Evaluating the suitability of a standard requires the existence of two types of implementations: a reference implementation and a real-world implementation (discussed in the next section). A reference implementation is a fully instantiated software solution that is proven to be compliant with the standard; it is often developed concurrently with the standard and serves as a reference to other software developers. It also "helps to clarify intent of the specification where conformance tests are inadequate…[and] serves as Gold Standard against which other implementations can be measured"[1]. In the ideal case, a reference implementation is open source—or at least publicly accessible—and well documented, so it can function as a teaching and reference tool. As a Gold Standard, it can test that the output of one implementation is compatible with that of the reference implementation. Reference implementations of a standard are a strong indicator that the complete standard can be implemented.

Has the standard been implemented by a range of vendors?

If a number of vendors (including government, private, large, and small) have implemented a given standard, it is a strong indicator that the standard is implementable. Real-world

---

[1] Curran, Patrick. *Conformance Testing An Industry Perspective*, Sun Microsystems. 2003.
http://www.nist.gov/itl/vote/upload/6-Curran.pdf

implementations, whether open-source or commercial, prove that the standard, or a subset thereof, can be implemented and support real-world conditions. A high-quality, robust, real-world implementation requires formal testing and thus shows that the standard is not only implementable but also testable.

Is there documented existence of errors, including estimates of the severity?

Errors (including technical errors, ambiguities, and inconsistencies within the standard or "parent" standard) exist in any standard; the question is whether the important ones have been found and fixed. Thus, the existence of lists of errors that contain only minor and no severe known errors is a positive indicator of quality. The lists of errors include change requests and error logs, as well as errors documented in published errata. A lack of known errors may indicate that the standard has indeed few errors, or that it has not been widely enough implemented, or that errors discovered during implementation were not made public.

Is there a defined and effective process for handling errors?

There are several parts to a defined, effective process for handling errors. The most important is to have a process. An effective error-handling process typically means the standard is more likely to be free of significant errors than a standard that does not have one. Ideally, the process is available to all who may find problems and documents a clear path to redress the errors. A thorough review process should eliminate significant errors before a standard becomes normative, but an effective process that is used after the standard becomes normative is also necessary. For example, addendums could be issued that document any discovered errors and that propose solutions.

Do industry associations endorse the standard?

A standard endorsed by industry associations is more likely to be implementable than one that is not. The endorsement process usually requires thorough evaluation of a standard, which includes the implementability of that standard. In some cases, these associations are invited to give input to the standards development process, which should result in a better standard.

Is the standard used in more than one country?

If a standard is in use in more than one country, it is a strong indicator that the standard is implementable (and useful) across a range of use cases and localizations.

Is certification available for developers and architects?

If certification for a standard is available for developers and architects, it indicates that the standard is relatively mature; there is a proven repeatable process for implementers to follow to successfully implement the standard. It also indicates that the certifying body considers the standard important and stable enough to justify the expense of developing the certification criteria.

Criteria and Components

This matrix shows which criteria can assess specific components of suitability. The components are: (Q) quality, (T) testability, (C) compatibility, (I) implementability, and (S) specificity (see the section on Suitability Components).

Table 3: Criteria and Components Matrix

Criteria	Q	T	C	I	S
Is the standard based on a stable, well-vetted data model?	✓		✓	✓	
Does the standard have a clear, robust vocabulary binding syntax?	✓	✓	✓	✓	✓
Does the standard support reusable modules, such as templates or data types?	✓		✓	✓	✓
Does the standard have a well-defined constraint mechanism?	✓		✓	✓	✓
Does the standard have a well-defined extensibility mechanism?	✓			✓	
Are there unambiguous definitions of what is testable?	✓	✓		✓	✓
Are there automated test tools and test suites?		✓		✓	✓
Are there reference implementations for the standard?		✓		✓	✓
Has the standard been implemented by a range of vendors?	✓			✓	
Is there documented existence of errors, including estimates of the severity?	✓	✓		✓	
Is there a defined and effective process for handling errors?	✓			✓	
Do industry associations endorse the standard?	✓		✓	✓	
Is the standard used in more than one country?	✓		✓	✓	
Is certification available for developers and architects?	✓		✓	✓	

Conclusions

We used these suitability criteria to evaluate four standards developed by the healthcare community: CCD, CCR, QRDA, and HL7 V2 Biosurveillance. Individual reports present the analysis for each standard. The suitability criteria determined which of the four standards are ready for widespread use and which would need modification to meet the needs of the healthcare community.

This set of general criteria does not address the HHS Meaningful Use functionality requirements. Those requirements are discussed in the individual standard suitability analyses.

We developed this set of criteria by defining the purpose of each standard, reviewing literature, and interviewing stakeholders. We then applied these criteria to each standard, verified their relevance and refined them as indicated. Appropriate criteria are not always quantifiable, but they are usually measurable in some way. Some of the criteria were not applicable to all of the standards; this is noted in the individual report.

We have evaluated the suitability of four diverse healthcare standards using this methodology. We believe that it is generally applicable and would be useful for evaluating other healthcare standards.

References

- Botella P, Burgues X, Carvallo, JP. *ISO/IEC 9126 in Practice: What do we need to know?*

  http://www.lsi.upc.es/~webgessi/publicacions/SMEF'04-ISO-QualityModels.pdf

- Curran, Patrick. *Conformance Testing An Industry Perspective*, Sun Microsystems. 2003. Available from http://www.nist.gov/itl/vote/upload/6-Curran.pdf

- *HITSP Standards Readiness Criteria* http://publicaa.ansi.org/sites/apdl/Documents/Standards Activities/Healthcare Informatics Technology Standards Panel/HITSP Policy Documents/HITSP 07 N 191 Readiness Criteria Tier2 v1.9.6.pdf

- Moody DL. *Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice.* ECIS. 2003. http://is2.lse.ac.uk/asp/aspecis/20030099.pdf.

- Richesson, R. Viewpoint: Data Standards in Clinical Research: Gaps, Overlaps, Challenges and Future Directions *JAMIA* 2007;14:687-696 doi:10.1197/jamia.M2470

- Wikipedia: Capability Maturity Model: http://en.wikipedia.org/wiki/Capability_Maturity_Model

- Wikipedia: ISO /IEC 9126: http://en.wikipedia.org/wiki/ISO_9126

- Wikipedia: Reference Implementation http://en.wikipedia.org/wiki/Reference_implementation

- Wikipedia: Software Quality: http://en.wikipedia.org/wiki/Software_quality

Acronyms and Abbreviations

AAFP   American Academy of Family Physicians

ASTM   originally called the American Society for Testing and Materials

CCD   Continuity of Care Document

CCR   Continuity of Care Record

CDC   Centers for Disease Control and Prevention

HHS   US Department of Health and Human Services

HIT   Healthcare Information Technology

HITSP Healthcare Information Technology Standards Panel

HL7   Health Level Seven

IEEE   originally known as the Institute of Electrical and Electronics Engineers

ISO   International Organization for Standardization

IT   Information Technology

NCQA   National Committee for Quality Assurance

NIST   National Institute of Standards and Technology

NTAA   National Technology Transfer and Advancement Act

QRDA   Quality Reporting Document Architecture

SAIC   Science Applications International Corporation

XML   Extensible Markup Language