ISO/IEC JTC 1/SC 32 WG2 N 1646

Date: 2012-05-09

REPLACES: --

ISO/IEC JTC 1/SC 32

Data Management and Interchange

Secretariat: United States of America (ANSI)
Administered by Farance Inc. on behalf of ANSI

DOCUMENT TYPE	Study Period Report
TITLE	WG2 Final Study Period Report on Metamodel for Forms Registration
SOURCE	Professor Jim Davies and Dr. Steve Harris
PROJECT NUMBER	
STATUS	For discussion at WG2 working group meetings June 1, 2012
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	
DUE DATE	
Number of Pages	7
LANGUAGE USED	English
DISTRIBUTION	P & L Members SC Chair WG Conveners and Secretaries

WG2 Final Study Period Report: forms metamodel

Rationale for the study period

There is an increasing need for the aggregation of data gathered in disparate and distributed systems, often across loosely coupled enterprises. Not only does this requires a clear understanding of the meaning of the data, it also frequently requires the coordination of data capture. Where data input is manual, an important source of data semantics is the design and behaviour of the form used for data entry - indeed if we do not understand the encoding of knowledge in the database schema or we suspect some anomaly in the data captured, we inspect the original form and the context of its use. Furthermore, if we wish to gather interoperable data it is frequently necessary to harmonise aspects of form design and behaviour before data is captured.

This need is recognised and addressed locally in a number of commercial and open source products, particularly in the medical research domain. RedCAP, OpenClinica, Oracle Clinical, Medidata Rave, and the NCI caDSR all provide facilities to design, deploy and share form designs. OpenClinica additionally will output parts of its form design in the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM) which provides some standard facilities for the basic description of forms. However, there is no abstract, universal metamodel for forms that can validate all implementations, support the automatic, accurate registration and exchange of form design instances, and allow for the definition of standard instruments for the coordination of data capture.

Given a standard metamodel for forms, ISO/IEC 19763 Metamodel Framework for Interoperability (MFI) and ISO/IEC 11179 Metamodel for metadata registries provide important facilities for the creation and annotation of form designs. ISO/IEC19763 supports the registration of forms and form sections as models, and provides facilities to record associations between form sections - particularly derivation, specialisation, extension and reuse. ISO/IEC11179 supports the creation and exchange of standard question banks and defines a rich source of question-level metadata with which to explain the meaning of individual data items. Together, both standards can support the rapid design interoperable data capture instruments, wrap and hide the complexity of terminology annotation from subject matter experts, and provide a ready reference of associations and transformations for users seeking interoperable data.

The study period has considered input from the implied forms metamodels of the products described above, together with the experiences of the UK Medical Research Council and its population studies - particularly the Avon Longitudinal Study of Parents and Children (ALSPAC), the Whitehall II study, the Southampton Women's Survey (SWS) and the National Survey of Health and Development - the EU Framework 7 project Union of Light and Ion Centres in Europe (ULICE), the UK National Health Service (NHS), the UK Medical Research Council, the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM), the United States National Cancer Institute (NCI) and to a lesser extent the experiences of the UK Government Cabinet Office and Scottish Executive. From the consultation and prototypical implementation we have been able to

define a draft metamodel which encompasses appropriate considerations from the consultation. We recommend that the study period conclude and development on the standard should begin.

Business relevance for the standard

Forms are everywhere, on paper and online providing a structured means of abstracting information according to a common perspective from surrounding circumstances so that it may be processed in a well understood way. They are used in commerce, science and government to gather data for analysis and management, and are ubiquitous in information systems to communicate configuration and validation parameters for users so that they may adapt software for their purpose.

Requirements for an independent and common means of representation of metadata is enshrined in the ISO/IEC8000 Information Quality standard, as are the requirements for persistence that would be satisfied by a standard metamodel implemented alongside ISO/IEC11179 and ISO/IEC19763.

High level specifications for forms are communicated manually as paper documents in standard business processes such as accountancy, in government through requirements for data capture in legislation or in detailed codes of practice and in science through minimum datasets or standard data capture instruments.

In medicine, key questions about treatment efficacy, safety and quality of life cannot be addressed at a single location, within a single healthcare system/provider, or experiment. Information specifications must be shared and faithfully implemented in appropriate software from a wide range of providers in order to aggregate and understand data from multiple sources. In social science, where opinion and attitude are measured, fine details of presentation can materially affect the outcome of a survey and thus must be communicated if any analysis is to be understood.

In many domains and settings - particularly in government - software suppliers need to be able to demonstrate that specifications for data capture have been faithfully transcribed into an information system so that electronic reports are reliable. In a democratic government and when processing personal data there is an additional duty for transparency - citizens have a right to understand the way data is acquired and processed. Formal, computer readable, semantic metamodels for forms go a considerable way to make system behaviour accessible to wider stakeholders.

Abstraction of the specification of the form from the implementation environment will also lead to tools development for the automatic generation of forms in particular languages and for specific systems reducing system development and customisation costs and increased agility in the development process for governments, scientists and businesses alike. A standard metamodel for forms would also support the development of common, user-oriented tools for forms authoring. and facilitate the migration of forms between applications. In conjunction with automated generation of forms, the design of a significant part of an information system will cease to be a task for a skilled programmer, providing significant cost reductions to medical and social science research.

B. Related Work

A number of areas of related work exist, however there are no standards with the scope or the depth to achieve meet this business case. The CDISC ODM has the ability to exchange form metadata, but the standard does not address how this metadata is linked to domain specific ontologies and question banks, and how a form is derived from other form components. The UML

provides facilities for the description of interfaces that can be interpreted in terms of form interfaces, but this falls short of a proper description of forms. ISO/IEC13606 Health informatics - Electronic health record communication has broadly similar facilities for the description of record templates in terms of *archetypes* but the implementation is specific to ISO/IEC 13606 and healthcare records, and thus would be difficult to establish conformance if the user is outside of the domain. The Health Level 7 (HL7) standards group has a similar (to 13606) standard for 'Clinical Document Architecture' intended to communicate the form used to complete the instance of the clinical document communicated or archived - however, as with all HL7 version 3 specifications it is based upon its unique Reference Information Model semantics making its use outside of an HL7 context difficult.

Several products already provide custom, local solutions to question bank management and form authoring including - in the medical research domain - Oracle Clinical, Medidata Rave, OpenClinica and RedCAP. Similarly many standard accounting and EAP packages provide limited customised slots for the extension of proprietary built in forms. However, it is unlikely that SAP and Oracle will agree on and publish a single metamodel for forms, particularly as this would reduce revenue from system customisation.

The proposed standard would also be closely related to ISO/IEC 11179 Part-6, which provides for the registration of content in a metadata registry. Inheriting the administered item class of ISO/IEC 11179 is would be a natural specialisation of the ISO/IEC19763 standard: facilities within part 10 could be used to persist relationships between forms, sections, questions and data elements, encapsulating bindings to domain specific ontologies and standard data element definitions to allow comparison of data collected across different forms.

C. Technical Status

The technical status of the proposed standard is some way between mature and prospective: early examples of paper forms may be found in 19th Century Law practice in the UK; electronic forms have existed since the first VDUs were devised in the 60s. In this respect, ample examples and expertise exist for the derivation of the metamodel. However, full abstraction of all elements of form design is less clear: the precise way in which standard questions are placed on a form, how their meaning is modified by annotations on the containers within the form and languages for expressing validation and flow amongst questions on the form are less well understood and will require further work informed by early implementations and computer science research.

D. Conformity assessment and interoperability

The project requires normative references with ISO/IEC11179 and ISO/IEC19763, and must subsume elements of standards such as ISO/IEC13606, the CDISC ODM and HL7 CDA, although these standards are designed to operate within tight domain constraints and it may be that clean encapsulation of their capabilities is not possible. However, the standard should be able to support the notion of forms within the NHS implementation of ISO/IEC13606 as the NHS Logical Records program has sought to restrict the application of the standard so that metamodels for record types are better supported.

Another closely related standard is the Data Documentation Initiative (DDI) version 3, which aims to provide a metamodel for the exchange of survey data in social science and population studies. The

standard developed here would provide a natural way to describe the intended semantics of the data capture events and relate them to columns in tabular data. In that the DDI is a portable document format for data and metadata exchange, some mechanism for serialising metadata from the supporting ISO/IEC11179 and ISO/IEC19763 structures will be required.

E. Cultural and Linguistic Adaptability

N.A.

F. Other Justification

N.A.

Scope

This proposed standard will establish a metamodel for the specification of the design of an electronic or paper form. The specification will reuse appropriate elements of the ISO/IEC11179 and ISO/IEC19763 standards so that the form design may be appropriately linked to corporate and enterprise data dictionaries and domain specific ontologies and instance data models, as well as other related form designs. The specification may be communicated to others to describe the meaning of data collected according to the design, to provide a template for interoperable data collection or to exchange aspects of system design between implementations in support of upgrade and migration. The metamodel will be capable of automatic transformation into an implementation language to guarantee the fidelity of transcription into a working system; it will also support the development of form design tools that will enable domain experts to formulate and deploy designs without specific coding skills.

References

Normative

- ISO/IEC DIS 19505-2: Information technology -- OMG Unified Modeling Language (OMG UML) Version 2.1.2 - Part 2: Superstructure
- ISO/IEC 11179-3:2012: Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes
- ISO/IEC 11179-6: Information technology – Metadata registries (MDR) – Part 6: Registration
- ISO/IEC 19763-1: Information technology – Metamodel framework for interoperability (MFI) – Part 1: Reference model
- ISO/IEC 19763-10: Information technology – Metamodel framework for interoperability (MFI) – Part 2: Core model and basic mapping
- ISO/IEC 19763-3: Information technology – Metamodel framework for interoperability (MFI) – Part 3: Metamodel for Ontology registration

Informative

- ISO/IEC 3535-1977: lays down the basic principles for the design of forms, whether discrete forms or continuous forms, and establishes a forms design sheet and a layout chart based on these principles. applies to the design of forms for administrative, commercial and technical use, whether for completion in handwriting or by mechanical means such as typewriters and automatic printers
- BS5537-1991: guide to forms design sheet and layout chart

- CDISC Operational Data Model (ODM) and Study Data Tabulation Model (SDTM): these standards should generalise relevant capabilities
- HL7 Clinical Document Architecture (CDA)
- ISO13606: rigorous and stable information architecture for communicating part or all of the Electronic Health Record (EHR) of a single subject of care (patient).

In addition a number of implementations will be considered as informative including: Google forms; Survey monkey; Adobe forms central; Open Clinica; RedCAP to the extent that open documentation is available, although more detailed information may be sought if it is felt necessary.

Liaisons

The developers hope to liaise with ISO/TC215 for Health Informatics and the International Health Terminology Standards Development Organisation to ensure the standard is technically capable of representing forms in healthcare workflows, Clinical Document Architecture (CDA) in particular, although support for structures specified within ISO13606 and ISO21090 is out of scope of the standard. ISO/TC154 have worked on standards for the layout of forms although liaison may not be required.

Plan

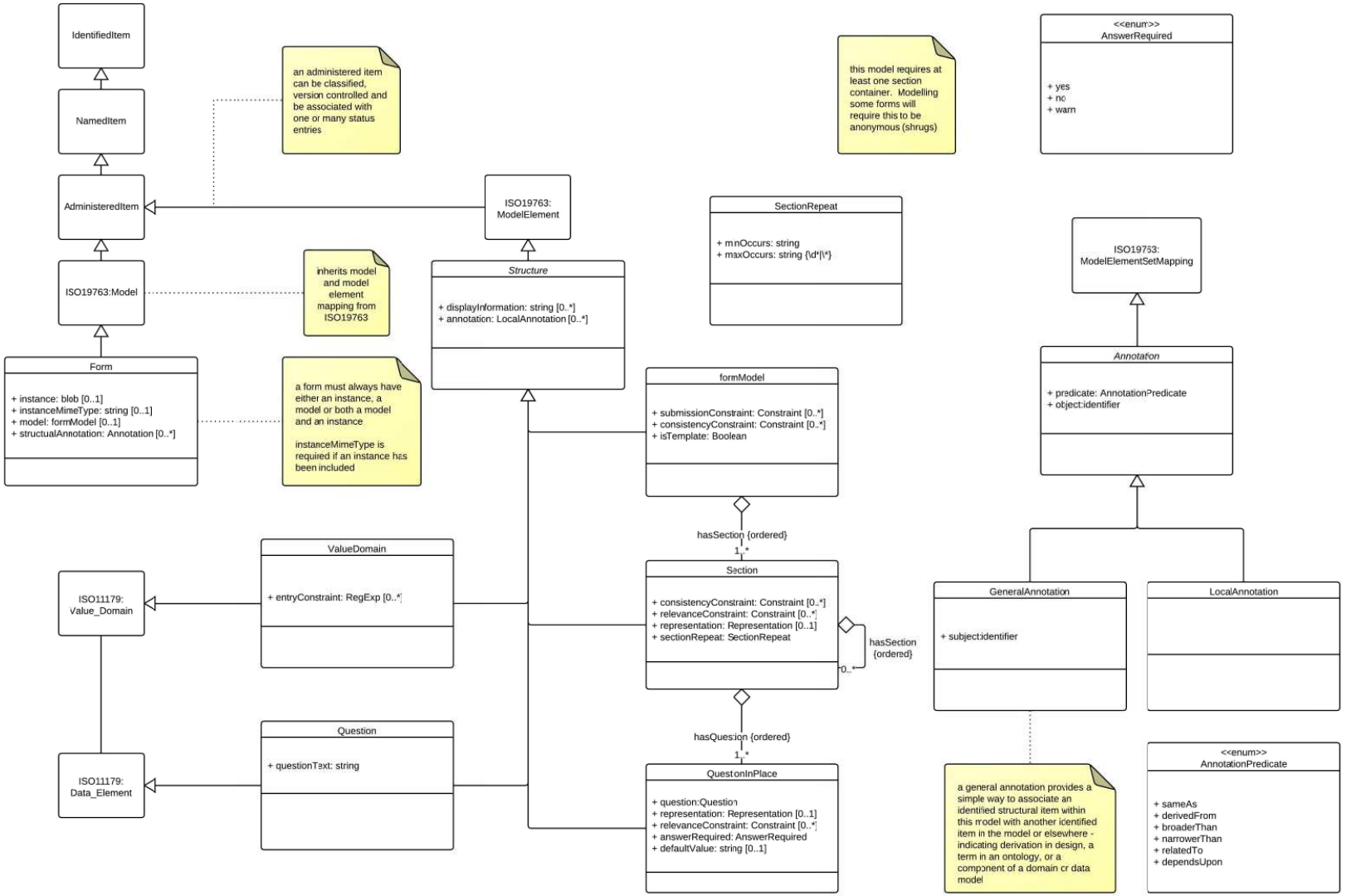
We aim to develop the standard according to the default timeframe with three committee draft stages at successive WG2 plenary and interim meetings leading to publication in June 2015:

- Working draft: interim 2012
- Committee draft 1: plenary 2013
- Committee draft 2: interim 2013
- Committee draft 3: plenary 2014
- Final Draft International Standard (FDIS): interim 2014
- Published standard: plenary 2015

Given the extensive work done by the US NCI and the Oxford Metadata Group, together with the normative and informative references we hope that it will not prove necessary to have three committee drafts and to be able to bring forward the timeline by 6 months. However, given our background in health informatics, we will additionally consult with those working on standards, forms, and interoperability in electronic governance, considering national and international interoperability standards, although it is recognised that the majority of work in this area has been done in health.

The standard will be developed by a core set of principal contributors to include: Jim Davies, Professor of Software Engineering, University of Oxford; Denise Warzel, Associate Director, Semantics Operations Team; Steve Harris, Research Officer and James Martin Research Fellow, University of Oxford, who will coordinate activities and ensure the timeline is met. We hope to obtain further support from within WG2 from Keith Gordon - particularly for conformance to ISO19763 - and Jurgen Stausberg. Harold Solbrig of the Mayo Clinic will help us liaise with TC215, IHTSDO.

Other contributions will be sought from: Dan Gillman, US Chair of DM32.8; Nicholas Oughtibridge, Acting Director - Data Standards and Products for the UK National Health Service (NHS);


IdentifiedItem
NamedItem
AdministeredItem
ISO19763:Model
Form
+ instance: blob [0..1]
+ instanceMimeType: string [0..1]
+ model: formModel [0..1]
+ structualAnnotation: Annotation [0..*]
an administered item can be classified, version controlled and be associated with one or many status entries
inherits model and model element mapping from ISO19763
a form must always have either an instance, a model or both a model and an instance
instanceMimeType is required if an instance has been included
ISO19763: ModelElement
Structure
+ displayInformation: string [0..*]
+ annotation: LocalAnnotation [0..*]
this model requires at least one section container. Modelling some forms will require this to be anonymous (shrugs)
<<enum>>
AnswerRequired
+ yes
+ no
+ warn
SectionRepeat
+ minOccurs: string
+ maxOccurs: string {\d*|*}
ISO19763: ModelElementSetMapping
Annotation
+ predicate: AnnotationPredicate
+ object:identifier
formModel
+ submissionConstraint: Constraint [0..*]
+ consistencyConstraint: Constraint [0..*]
+ isTemplate: Boolean
hasSection {ordered}
1..*
Section
+ consistencyConstraint: Constraint [0..*]
+ relevanceConstraint: Constraint [0..*]
+ representation: Representation [0..1]
+ sectionRepeat: SectionRepeat
hasSection {ordered}
0..*
hasQuestion {ordered}
1..*
QuestionInPlace
+ question:Question
+ representation: Representation [0..1]
+ relevanceConstraint: Constraint [0..*]
+ answerRequired: AnswerRequired
+ defaultValue: string [0..1]
ValueDomain
+ entryConstraint: RegExp [0..*]
ISO11179: Value_Domain
Question
+ questionText: string
ISO11179: Data_Element
GeneralAnnotation
+ subject:identifier
LocalAnnotation
a general annotation provides a simple way to associate an identified structural item within this model with another identified item in the model or elsewhere - indicating derivation in design, a term in an ontology, or a component of a domain or data model
<<enum>>
AnnotationPredicate
+ sameAs
+ derivedFrom
+ broaderThan
+ narrowerThan
+ relatedTo
+ dependsUpon