

Models for Forms

Daniel Abler, Charles Crichton, James Welch, Jim Davies, Steve
Harris

University of Oxford

October 24, 2011

Contents:

- 1 Context: Clinical studies
 - Case report forms and study design

- 2 Overview of current practice
 - Current Practice
 - Comparison

- 3 Foundations and requirements
 - Data quality
 - Requirements
 - Compositionality
 - Additional Questions

- 4 Conclusions

Context: Clinical Studies

- Patient-oriented clinical research includes studies of human diseases, therapies and interventions.
- Clinical studies are conducted to allow for evaluation of health interventions regarding their *safety* and *efficacy*.
- Objective, design, methodology and statistical considerations are described in a *trial protocol*: determines data collection
- Data analysis requires *homogeneous data capturing practices* over duration of the study and among study partners.

Context: Clinical Studies

However...

- Data typically captured by different groups of researchers. Evolving knowledge requires new questions to be asked and CRFs to be adapted.
- Integration of data from independent studies is difficult or impossible due to *incompatible data collection and/or insufficient documentation*.

Current Practice

CDISC-ODM Clinical Data Interchange Standards Consortium (CDISC)'s Operational Data Model (ODM).

Documentation standard for clinical trials.

DDI Data Documentation Initiative (DDI).

Archival standard for social science data.

OpenClinica, RedCAP Excel based form models for defining forms. Used in software for clinical trial support (OpenClinica or RedCAP respectively).

Cancergrid, caDSR Form Builder Informatics support for biomedical studies focusing on re-use of common data elements to promote data interoperability across studies.

Also paper-based systems, spreadsheets, lightweight databases, etc.

Identification and Logical Structure

- identification of data components in order to refer to data
- identification or groups of data components to express logical structures

	identifiers and scope	versioning	grouping	relations and hierarchy	structure multiplicity constraint	structure annotation
CDISC-ODM	study level	Y	Y	Form, Item, ItemGroup,	Y	N
OpenClinica	form level	N	Y	CRF, section, group, item	?	N
CancerGrid	form level	N	Y	Form, FormModel, Control, IncludedVariable / Section / Table	Y	N
caDSR Forms	module level	Y	Y	Module, Question	Y	N

Table: Identification and logical structure

Data Constraints

- constraints on values entered against single data component
- relation between values entered against different data components
- constraints (used as submission guards) become universal properties of data set

	Field				across Fields	Structures	
	Type	Range	Multiplicity	prepopulation	range and functional	existence	definition
CDISC-ODM	?	Y	?	?	?	Y	Y
OpenClinica	?	Y	Y	Y	N	Y	Y
CancerGrid	variable definition	N	N	N	N	N	Y
caDSR Forms	CDE reference	N	N	N	N	Y	Y

Table: Data Constraints

Process or presentation constraints

- Process constraints (“form logic”) determines visible content and data components of the form
- Presentation aspects may influence interpretation of collected data.
- Both, process and presentation of form may influence usability of the form and thus quality of resulting data.

	Control Flow			Submission		Presentation		
	process order on form / study level	roles	constraint language for guards and conditions	submission conditions	special submission guards	numbering (inferred from control flow or explicit)	rendering options / paper forms / interviews	layout instructions / rendering instructions (checkbox, dropdown,...)
CDISC-ODM	skip logic / Y	?	skip logic	N	signature	explicit	N	N
OpenClinica	? / N	N	?	?	N	explicit	N	Y
CancerGrid	skip logic / N	N	skip logic	N	N	N	N	N
caDSR Forms	N / N	N	N	N	N	N	N	N

Table: Process and Presentation Constraints

Form-based data collection and data quality

Three aspects of Data Quality [Strong et al., 1997]

- *correctness*:
the extent to which values entered correspond to the intended interpretation
- *completeness*:
the extent to which the data collected is complete
- *comprehensibility*:
the extent to which the data comes with adequate documentation

Three form-design impacts on data quality

- Guiding user with data input
- Validation prior to submission
- Association of resulting data with appropriate metadata

A Domain Specific Model for Forms

Domain specific modelling [DSM 2011 Preface]

*A domain-specific modeling language follows abstractions and conventions of the domain, while preserving the meaning (semantics) of those models that is consistent with the domain. This approach allows the system models to simultaneously represent the **design**, **implementation**, and **documentation** of the system.*

A language of forms

- planning and coordination of data collection activity
- generation of data collection artifacts
- separating form design from implementation (loose-coupling)
- documentation of data collected

Required features for a language of forms

- Support the construction of forms for large clinical studies
- Separation of concerns: Structure, Presentation and Validation
- Versioning of all form elements
- Questions to relate to external resources
- Richer datatypes for individual question responses
- Alternative rendering of questions
- Data capture workflow (Submission / notification / scheduling)
- Compositionality

What does composition mean?

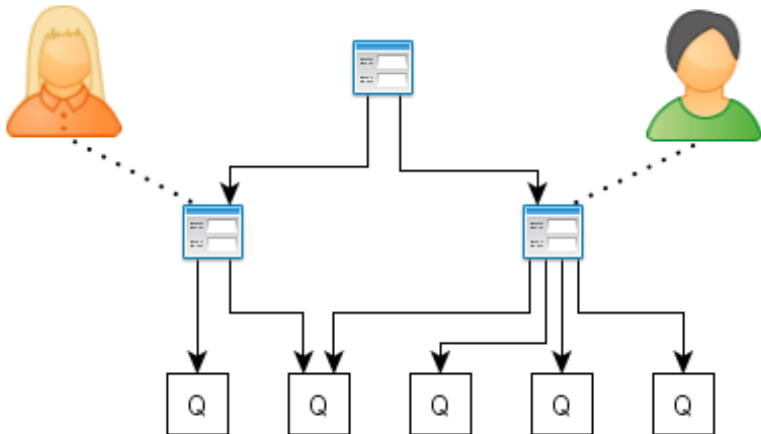
- We can create larger form components by composing a number of smaller form components
- Questions, Sections, Forms, Sub-studies, Studies

Aspects of composition

- Identification and logical structure
- Data constraints
- Process / presentation constraints

Why is compositionality important?

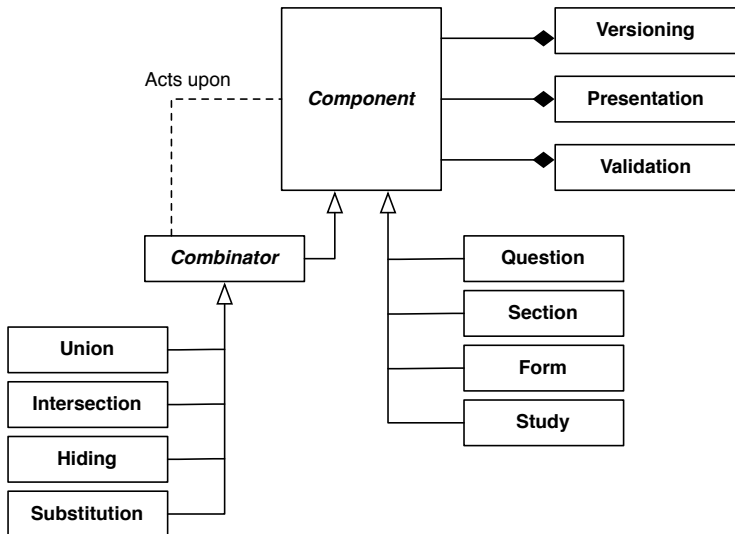
Meta-analysis is the composition of multiple studies and their results



Composed forms may not be well-formed

- The constraints on sub-studies, for example required question ordering, might conflict.
- Validation constraints might be incompatible
- Thus non-constructive composition operators are required, for example to hide questions.
- Our forms language needs to include a wider range of composition operators, not just Union, but also Intersection, Hiding, Substitution

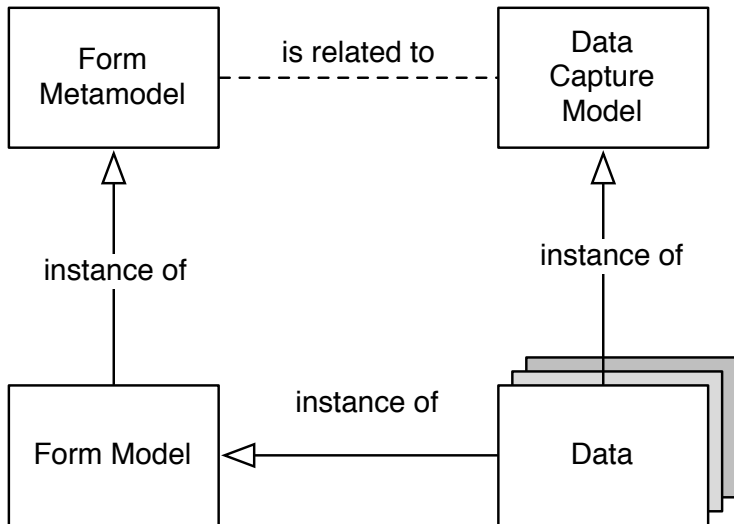
Metamodel



Comparability of components: forms or studies

- Can we say that different studies, or at least parts of them are comparable?
- Is there a notion of 'sufficiently similar' we can use?

Data capture formats



Additional Questions

- Referencing between models, and between components
- Expression and constraint languages for structure, validation and presentation
- Dynamic features: study workflows, presentation constraints, submission
- Balance between separation of concerns and clarity of model

Conclusion

Data standards for clinical research data collection forms: current status and challenges [Richesson, R. L.& Nadkarni, P.]

Currently, no universal CRF-design standards exist, though conventions and some 'best' practices do. [...] Data-capture standards can facilitate efficacious development and implementation of new studies, element reuse, data quality and consistent data collection, and interoperability.[...] Of more immediate and widespread (pan-disease) **relevance** are **standardization efforts toward** the development of **sound processes and workflow for CRF and CRF section** development, as well as **data collection and validation.**

Conclusion

- Identified a need for a domain specific language
- Determined requirements from work in clinical studies
- Compared existing work and current practice to identify key features
- Key features: Compositionality and Data Capture

Acknowledgements

PTCRi



Particle Therapy Cancer Research Institute

Webpage: <http://www.ptcri.ox.ac.uk/>

PARTNER



Particle Training Network for European Radiotherapy

Particle Therapy Marie Curie Early Initial Training Network Fellowship of the European Community's Seventh Framework Programme under contract number (PITN-GA-2008-215840-PARTNER).

Webpage: <http://www.ptcri.ox.ac.uk/>

ULICE



Union of Light Ion Centres in Europe (ULICE)

Co-funded by the E. C. within the Framework Programme 7 Capacities Specific Programme. (Grant Agreement 228436).

Webpage: <http://ulice.web.cern.ch>