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**
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.
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.*
**Visit 1** Date:

Informed consent form signed?  □ no  □ yes
Participant meets main inclusion and exclusion criteria?  □ no  □ yes

<table>
<thead>
<tr>
<th>Medical History</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diagnosis</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
</tbody>
</table>

Gestational Age Weeks:  
Days:  
Birth weight (if known):  
OR  □ Check if birth weight unknown

<table>
<thead>
<tr>
<th>Medical Examination</th>
</tr>
</thead>
</table>
| Temperature pre-vaccination axillary:  °C  
Length:  cm  
Weight:  Kg  
Head circumference:  cm  

Fit for Immunisation:  □ no  □ yes

**Figure:** fragment of a case report form

**main modus of data collection** in clinical studies:

- different CRFs for e.g. demographic information, base line variables, diagnosis, consent, treatment information, first follow-up after treatment, regular long-term follow up
- not only variable definitions, also context is important
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

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

**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

<table>
<thead>
<tr>
<th>Field</th>
<th>Type</th>
<th>Range</th>
<th>Multiplicity</th>
<th>prepopulation</th>
<th>range and functional</th>
<th>existence</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenClinica</td>
<td>?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>CancerGrid</td>
<td>variable definition</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>caDSR Forms</td>
<td>CDE reference</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

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.

### Table: Process and Presentation Constraints

<table>
<thead>
<tr>
<th></th>
<th>Control Flow</th>
<th>Submission</th>
<th>Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>process</td>
<td>constraint</td>
<td>numbering</td>
</tr>
<tr>
<td></td>
<td>order on</td>
<td>language</td>
<td>(inferred</td>
</tr>
<tr>
<td></td>
<td>form / study</td>
<td>for guards</td>
<td>from control</td>
</tr>
<tr>
<td></td>
<td>level</td>
<td>and</td>
<td>flow or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>conditions</td>
<td>explicit)</td>
</tr>
<tr>
<td></td>
<td>roles</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>skip logic /</td>
<td>N</td>
<td>signature</td>
</tr>
<tr>
<td>CDISC-ODM</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenClinica</td>
<td>? / N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CancerGrid</td>
<td>skip logic /</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>skip logic</td>
<td></td>
</tr>
<tr>
<td>caDSR Forms</td>
<td>N / N</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>
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
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

Form Metamodel

is related to

Data Capture Model

instance of

Form Model

instance of

Data

instance of
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
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/](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/](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](http://ulice.web.cern.ch)