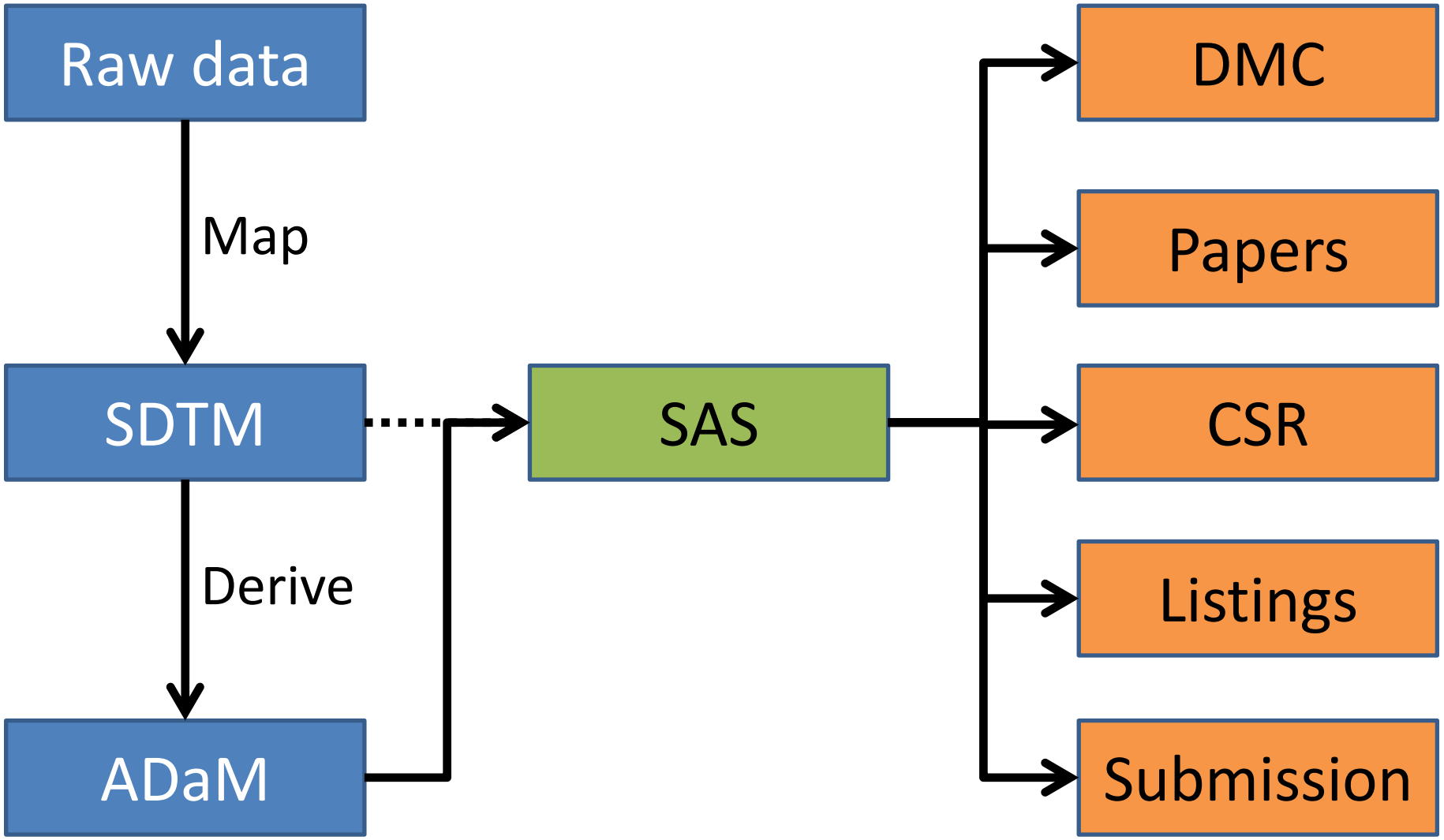
# Implementing CDISC Standards in large academic clinical trials

Will Stevens
Clinical Trial Service Unit
Nuffield Department of Population Health
Oxford
July 5th 2017

# Context

- Several large clinical trials (>10000 participants)
- Some smaller trials
- Starting from about 2010 CDISC standards began to be adopted, beginning with SDTM
- This talk is about my work in implementing SDTM, ADaM and define.xml in three trials since 2013. Most of the work was done for the REVEAL trial (30,000 participants, 4.1 year median FU).

# Workflow


Raw data
Map
SDTM
Derive
ADaM
SAS
DMC
Papers
CSR
Listings
Submission

# Tools and languages

- Study database and analysis databases hosted using Ingres SQL
- SQL is also used for data transformations
- But native language features aren't expressive enough
- SQL wrapped in a lightweight XML-based macro system
- PHP used on top of this to preprocess XML

# Example

```
<!-- Assign a visit number based on SVSTDY -->
<extremum-value min-or-max="min">
  <src-table>tv join sv</src-table>

  <src-key>sv.usubjid</src-key>
  <src-key>sv.svrefid</src-key>

  <test-exp>abs(tv.visitdy - svstdy)</test-exp>

  <column>tv.visitnum</column>
</extremum-value>
```

# Documentation

Added

Changed

Removed


***ADMEDDRA***

**Modified: 2017-06-09**

This table contains time to event for all MedDRA
categorised as follows: NONFATAL and SAE. F
NONFATAL and AE. For each combination of f
the time to event for all events (i.e. all events wit

| Column | Type | |
|---|---|---|
| usubjid | integer | Unique subject identifier for all stu |
| fasfl | char | Full analysis set population flag |
| siteid | varchar | Study site identifier |
| sitegr3 | varchar | Site group for grouping 3 (coarse r |
| sitegr3n | integer | Site group ID for grouping 3 (coars |
| param | varchar | MedDRA term corresponding to th |
| paramcd | integer | MedDRA SOC, HLGT or HLT code, |
| parcat1 | varchar | Level of the MedDRA code in PARA |
| parcat1n | integer | Numeric code corresponding to PA |

# Event classes

| Type of Major Vascular Event | Niacin–Laropiprant (N=12,838) | Placebo (N=12,835) |
|---|---|---|
| | *no. of participants with event (%)* | |
| Major coronary event | | |
| → Nonfatal myocardial infarction | 402 (3.1) | 431 (3.4) |
| Death from coronary cause | 302 (2.4) | 291 (2.3) |
| → Any major coronary event | 668 (5.2) | 694 (5.4) |
| Stroke | | |
| Nonhemorrhagic stroke | 389 (3.0) | 415 (3.2) |
| Hemorrhagic stroke | 114 (0.9) | 89 (0.7) |
| Any stroke | 498 (3.9) | 499 (3.9) |
| Revascularization procedure | | |
| Coronary revascularization | 591 (4.6) | 664 (5.2) |
| Noncoronary revascularization | 236 (1.8) | 258 (2.0) |
| Any revascularization procedure | 807 (6.3) | 897 (7.0) |
| → Any major vascular event | 1696 (13.2) | 1758 (13.7) |

Effects of Extended-Release Niacin with Laropiprant in High-Risk Patients. N Engl J Med 2014; 371:203-212.
DOI: 10.1056/NEJMoa1300955

# Web Ontology Language (OWL)

Major vascular event =

MI or Haemorrhagic stroke or Ischaemic stroke or Revascularization

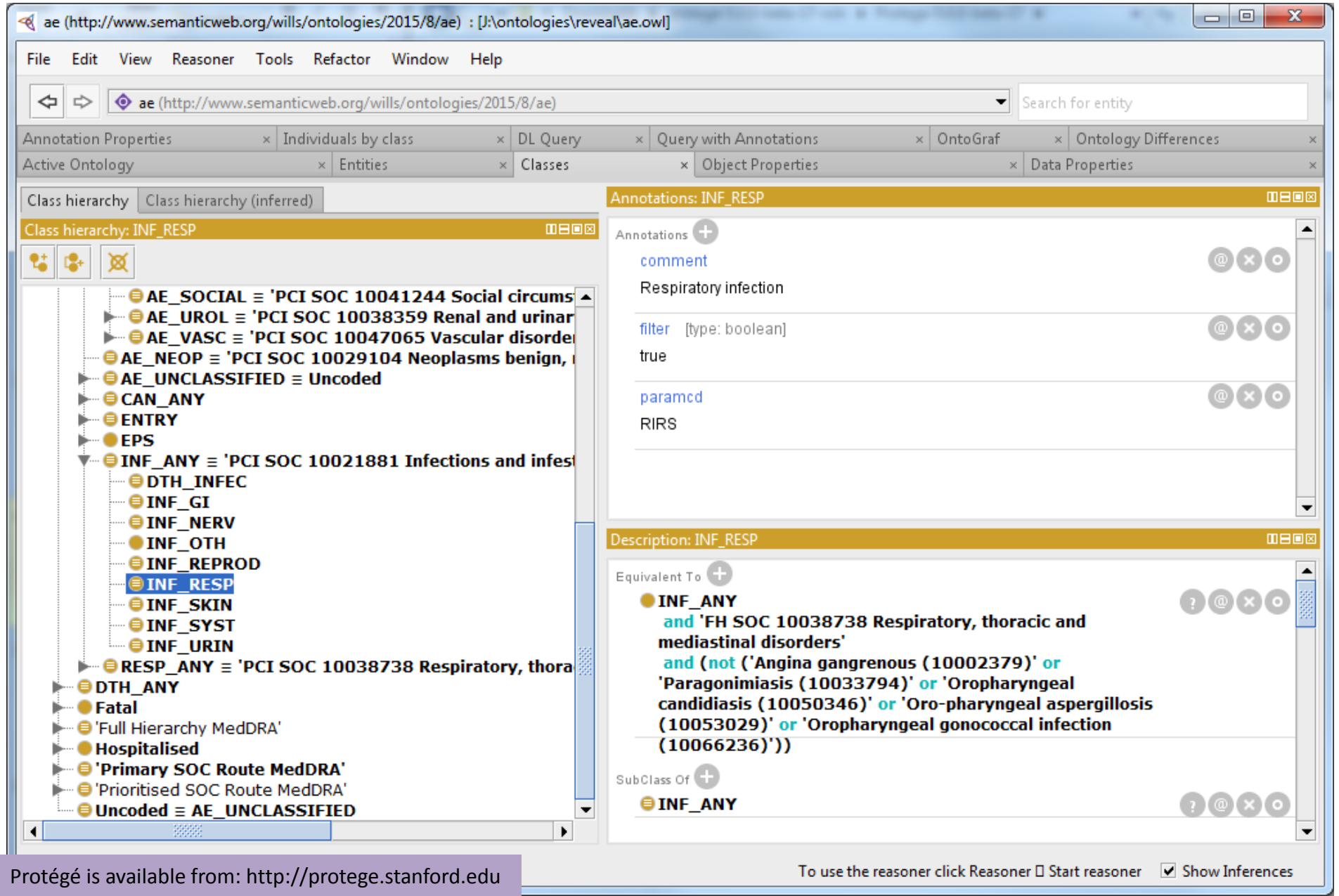
or ( Coronary heart disease and Fatal )

```
<EquivalentClasses>
   <Class IRI="#MAJ_VASC_EV"/>
   <ObjectUnionOf>
      <Class IRI="#MI"/>
      <Class IRI="#STR_HAEM"/>
      <Class IRI="#STR_ISCH"/>
      <Class IRI="#REVASC"/>
      <ObjectIntersectionOf>
         <Class IRI="#CHD"/>
         <Class IRI="#Fatal"/>
      </ObjectIntersectionOf>
   </ObjectUnionOf>
</EquivalentClasses>
```

```
<DisjointClasses>
   <Class IRI="#STR_HAEM"/>
   <Class IRI="#STR_ISCH"/>
</DisjointClasses>
```

Web Ontology Language (OWL). Available at: http://www.w3.org/2004/OWL/

# Protégé : A free, open source user interface for OWL documents




Protégé is available from: http://protege.stanford.edu

# Testing

- After producing SDTM domains, and after deriving ADaM datasets, tests are carried out on the data.
- Some tests apply across multiple domains/datasets. E.g. Does the domain contain at least some data? Are the --SEQ numbers correct?
- Some overlap with Pinnacle 21 tests.
- Some tests are written in response to errors found during development (to prevent reappearance of error, or recurrence of error elsewhere).

# Testing - examples

- Phase/period/visit columns – check that they are consistent with dates obtained from another source.
- When ADaM datasets should only contain a subset of data from SDTM, check that nothing unexpected is present.
- Check that sums and totals of the same data represented in different places match.

# Future developments

- Many of the submission components are also useful during the course of the trial: annotated CRF and define.xml should be made available early on in the trial.
- How does performance of SQL data transformations compare with SAS?
- Vector-based RDMS could potentially reduce storage requirements and speed up data transformation time.

# Strengths and weakness of SDTM and ADaM

- Overall, CDISC standards seem to be beneficial for us. It makes it easier to transition between one trial and another than if each trial (perhaps developed and conducted by a different team of people) has its own way of representing data.
- There is a lack of specificity in SDTM and ADaM, and often scope for different interpretations.
- No (or little) entity-level modelling of data in the way that software developers are generally used to doing.

# Strengths and weakness of SDTM and ADaM

- Standardised tabulation (SDTM), standardised analysis datasets (ADaM) and (with CDASH) standardised data at the point of collection. These are all at the edges of the trial: no standardised modelling of structured clinical trial data.
- Is this a strength or a weakness?

As standards develop over time, it is convenient if they are modular (well defined interfaces, but otherwise potentially independent of everything else), so that a change in one place doesn't require changes everywhere else.

Evolvability should be taken into account when putting together a system of standards. Design of the standards should consider: which parts of the standard can change without affecting other parts? What are the possible future changes?