

XML - truths and myths

Eric Browne

Clinical Information Project
eric.browne@dhs.sa.gov.au

Introduction

- “~~To XML or not to XML~~ — that is the question”

Aim

- ~~What to XML? & How to XML?~~

Aim

- What information to represent? ✓
- How to describe the information? ✓
- How to constrain the information? ✓

Aim

- Q. What Information?
- A. Record what clinician
 - saw
 - did
 - judged
 - ordered
 - advised

TACTICAL

better..

- Provide information to:
 - enable shared care

TACTICAL=> STRATEGIC

best!

- Provide information to:
 - optimize the health of each individual

STRATEGIC

consultation note

- S1 “The fat cat sat gingerly on the mat.
- S2 I think she was hurting.
- S3 Her teeth were green.
- S4 I think she had been fighting with Tom.”

XML 1

```
<article author="felix" date="04/04/04">
  S1 S2 S3 S4
</article>
```

XML 2

```
<article author="felix" date="04/04/04">
  <sentence> S1 </sentence>
  <sentence> S2 </sentence>
  <sentence> S3 </sentence>
  <sentence> S4 </sentence>
</article>
```

XML 3

```
<article author="felix" date="04/04/04">
  <observation> S1 </observation>
  <observation> S2 </observation>
  <observation> S3 </observation>
  <evaluation> S4 </evaluation>
</article>
```

XML 4

```
<article author="felix" date="04/04/04">
  <sentence type="observation">
    <subject> "the fat cat" </subject>
    <verb> "sat gingerly on" </verb>
    <object> "the mat" </object>
  </sentence>
  ...
</article>
```

XML 5

```

<article author="felix" date="04/04/04">
  <sentence type="observation">
    <subject>
      <noun>"cat" </noun>
      <article type="definite"> "the" </article>
      <qualifier>"fat" </qualifier>
    </subject>
    ...
  </sentence>
  ...

```

XML 5

```

<article author="felix" date="04/04/04">
  <sentence type="observation">
    <subject>
      <subject.noun> "cat" </subject.noun>
      <subject.article type="definite"> "the" </subject.article>
      <subject.qualifier>"fat" </subject.qualifier>
    </subject>
    ...
  </sentence>
  ...

```

Faithful record? ✓
Achieve tactic? ✓

Tactic reminder

- Record what clinician
 - saw
 - did
 - judged
 - ordered
 - advised

But..

What about the goal?

Domain vocabulary

```

<article author="felix" encounter-date="02/04/04">
  <subject>
    <sex>"female" </sex>
    <age>"unknown" </age>
    <BMI>"above-normal" </BMI>
  </subject>
  <problem severity="mild">"pain" </problem>
  <symptom type="subjective">"sits gingerly" </symptom>
  <diagnosis confidence="60">"injury" </diagnosis>
  <intervention>
    <service>
      <recall due="today + 7"/>
    </service>
  </intervention>
  ...

```

Tags express domain knowledge

Just tags?

```

<sex>"female" </sex>           <sex>"1" </sex>
<age>"unknown" </age>         <age>"99 9" </age>
<BMI>"above-normal" </BMI>    <BMI>"3 1kgpm2" </BMI>

```

explicit implicit

Value domain

Schemas?

- Specify labels for data in terms of some vocabulary.
- Can also specify some forms of relationships.
- Can also describe Universe of Discourse in terms of concepts, concept attributes, value domains, datatypes, other constraints - but, solely for **annotating data streams and message parsing/validating**.
- <Not useful> for <communicating with users> or for use by programs that need to manipulate data, or for software engineers designing and building systems, etc..

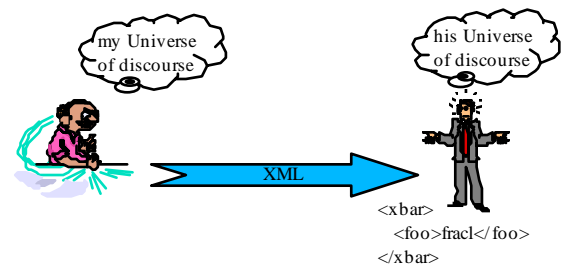
Well known schema languages

- DTDs - (Document Type Definitions)
- RELAX NG
- Schematron
- XML-Schema
- RELAX NG + Schematron
- RELAX NG + Schematron + XML-Schema datatypes
- others

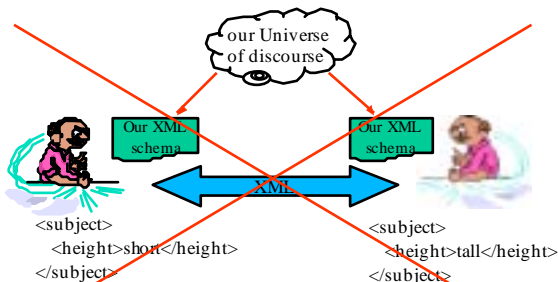
After choosing the schema language

- We still need to produce the schema(s)

Whose schema?



Universal health schema?



Yet another approach

- Let's forget about XML ! ← too tactical
- Let's forget about HL7 RIM ! ← too tactical
- Let's forget about openEHR Reference Model ! ← too tactical
- Let's focus on clinical concepts ← strategic
 - their properties
 - their constraints
 - their value domains
 - their vocabularies

Clinical concepts

- Identify ⇐ **Ontology of concepts**
- Analyse ⇐ **domain experts; think tanks**
- Define & Describe ⇐ **openEHR Archetypes + vocabularies + ISO11179?**
- Discuss ⇐ **broad stakeholder groups**
- Agree ⇐ **ditto**
- Promulgate ⇐ **AIHW Knowledgebank**

The byproduct

- Analysing, defining and describing real clinical concepts will determine the requirement for:-
 - datatypes
 - vocabularies
 - constraint language requirements
- In turn will help identify:-
 - suitability of reference models
 - suitability of implementation technologies

Health-specific requirements

- Concept specialisation ← **Archetypes ✓**
- Support for vast amount of knowledge
 - vocabularies
 - expressiveness e.g. precision and vagueness in text
- Rich temporal (and spacial) semantics
 - date, time, intervals, overlapping intervals, sequences, areas of interest
- Uncertainty, confidence, accuracy
- Ability to group and associate across information instances
- Support for analog, digital, text, multidimensional
- Spacial and temporal subsampling, supersampling, etc.
- A universal common goal.

And remember..

Even when you're knee deep in XML..

Never loose sight of the goal !