THE NAACCR CDA PILOT PROJECT
- The California Experience

Barry Gordon, PhD
Chris Rogers
C/NET Solutions, Berkeley CA
6/18/09
Pilot test – first build a model
Failure is always an option
Why CDA/XML?

- Loooloooong Text is Sooollllooo Inefficient
- Flat files are so hard to import
Flat file example (2010)

... 5-25-XX Left upper jugular chain adenopathy with mult. large nodes, one mobile axillary node; spleen& liver palpable.
... [Followed by 893 spaces]
Flat file problems

What if bookshelves had fixed length slots sized for the thickest book imaginable?

What if books had no titles to help find what you want?
<item no="220" loinc="21840-4" name="Sex [at Diagnosis]" >2</item>

<item no="2520" loinc="22029-3" name="Text: DX Proc: PE" section="Text-Diagnosis"> 5-25-XX Left upper jugular chain adenopathy with mult. large nodes, one mobile axillary node; spleen& liver palpable</item>
CDA Example

- `<observation classCode="OBS" moodCode="EVN">`
- `<templateId root="NAACCR-tId-2520">`
  `<code codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC" code="22029-3" displayName="Text: DX Proc: PE">`
  `<statusCode code="completed">`
  `<value xsi:type="ST">5–25–XX Left upper jugular chain adenopathy with mult. large nodes, one mobile axillary node; spleen& liver palpable.</value>`
How to implement CDA transmits at the hospital?

- Current CNExT transmit dataflow vs.
- New CDA-XML Transmit dataflow
Intermediate XML

<item no="220" name="Sex [at Diagnosis]" …>2</item>
<item no="230" name="Age at Diagnosis" …>023</item>
<item no="240" name="Birth Date" …>08091980</item>
<item no="250" name="Birthplace" …>031</item>
<item no="310" name="Text: Usual Occup" …>Arial Acrobat</item>
<item no="320" name="Text: Usual Industry" …">Air</item>
CDA Output Snippet

<!-- NAACCR-tId-characteristics-at-diagnosis -->
- <organizer moodCode="EVN" classCode="CLUSTER">
  <templateId root="NAACCR-tId-characteristics-at-diagnosis" />
  <statusCode code="completed" />
</organizer>

<!-- NAACCR-tId-220 -->
- <component>
  - <observation classCode="OBS" moodCode="EVN">
    <templateId root="NAACCR-tId-220" />
    <code codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC" code="21840-4" display_name="" />
    <statusCode code="completed" />
    <value xsi:type="CD" codeSystem="CS-OID-220" codeSystemName="NAACCR Sex" code="2" display_name="Female" />
  </observation>
</component>
CDA Testing Steps

1. Create test CNExT cases and export CDA
2. Use national test cases and export CDA
3. Run Alschuler’s online validator
4. Check CDA output for accuracy and completeness
5. Byte-compare normal flat file with CDA->flat output
6. (Import into Eureka was not tested)
### Compare CNExT transmits versus CDA→flat

**CNExT NAACCR flat transmit**

<table>
<thead>
<tr>
<th>NY00123 000</th>
<th>20188888888660 202308091980</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDA -&gt; flat . XSL</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>00123 000</td>
<td>288888888 0 2023</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00019DOB</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00020 AGE</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00022 SEX</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00138 SPANISH</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00550 RACECSO</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00549 RACECSC</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F03076 RACE5</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F03073, etc. RACE2 - 4</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00021 RACE1</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00023 MARSTAT .</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F05473 COUNTY</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00015 DXZIP</td>
</tr>
<tr>
<td></td>
<td>+-&gt; F00014 DXSTATE</td>
</tr>
</tbody>
</table>
CDA testing results

- State-specific fields can work in CDA
- Many problems along the way to get all NAACCR fields to transmit
- End-to-end comparisons of flat files before and after CDA show some errors
How long did it take to implement?
CDA– How hard to Implement the first time?

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Days</th>
<th>FTE's</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding the tools</td>
<td>10</td>
<td>0.04</td>
</tr>
<tr>
<td>Analysis/Design</td>
<td>18</td>
<td>0.07</td>
</tr>
<tr>
<td>Programming</td>
<td>61</td>
<td>0.24</td>
</tr>
<tr>
<td>Testing &amp; Feedback</td>
<td>42</td>
<td>0.17</td>
</tr>
<tr>
<td>Customizing</td>
<td>12</td>
<td>0.05</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>132</td>
<td>0.53</td>
</tr>
</tbody>
</table>
How hard is state customization?

- California is a worst case: N=80 extra data items
- Added 80 fields to mapping table
  - Used fake NAACCR and LOINC ID’s
- Modified XML lookup tables used by XSL
- Didn’t have to modify complex XSL (!)
## How Big Are These Records?

### Average # characters per transmitted case

<table>
<thead>
<tr>
<th></th>
<th>Flat</th>
<th>XML</th>
<th>CDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>6,696</td>
<td>23,065</td>
<td>68,763</td>
</tr>
<tr>
<td>2010</td>
<td>22,824</td>
<td>28,400</td>
<td>112,737</td>
</tr>
</tbody>
</table>
Is one file per case the best way to send CDA?

What is lost by using HL7–compliant XML rather than CDA?
Advantages of XML or CDA

- Empty fields are skipped
- Text formatting is allowed and helps readability
  - e.g. indentations
10/4/94  Excis lesions L ant & post thigh, R inframammmary chest: No significant finding. 10/18/94 Excis other lesions found on further exam 12/30/94 Re-excised lesions thigh & back: No finding 01/24/95 Re-excised L thigh & R chest: No fin

VERSUS

10/4/94  Excis lesions L ant & post thigh, R inframammmary chest: No significant finding. 10/18/94 Excis other lesions found on further exam 12/30/94 Re-excised lesions thigh & back: No finding 01/24/95 Re-excised L thigh & R chest: No fin
My Recommendations

- CDA was interesting, but...

- For interoperability, pursue XML with standard tags
Thanks to:

- California Department of Public Health and the National Program of Cancer Registries, for helping fund the work

- Chris Rogers, for doing the work

- Alschuler Associates, for developing the CDA tools.

- Our other partners in the study.