XML + HL7: Does this spell help or trouble for registries?

Barry Gordon, PhD
C/NET Solutions, Calif. Cancer Registry
June 9, 2005

Outline

- XML is what?
- Why do we care?
- XML features
- CDC/HL7 Use
- Possible NAACCR use
- Tools for working with XML
- What's next?

C/NET SOLUTIONS

Acknowledgement

- Many ideas presented here were
  - Stolen from or
  - Thought up while working on funds from:
    - CDC-NPCR

C/NET SOLUTIONS

XML is what?

- A markup language using tags
- Originally created by WWW to disseminate structured data/docs over the web
- Now supported in MS Office 2003
- Makes information portable (cross-platform)
- Separates data and content
- Enforces rules (via schemas)
- Allows many output formats and styles for same data.

C/NET SOLUTIONS

XML - What it's not

- Not just for the Web
- Not a database
- Not a programming language

C/NET SOLUTIONS

XML Features

- Self-documenting tags
- Browser-viewable
- Many views of the same data

C/NET SOLUTIONS

XML tags example

```
<?xml version="1.0" encoding="UTF-8" ?>
- <books>
  - <book>
      <title>The XML Plot</title>
      <author>Brown, Browny</author>
      <publisher>Singleday</publisher>
      <price>$24.95</price>
      <contentType>Fiction</contentType>
      <format>Hardback</format>
      <isbn>0385504209</isbn>
    </book>
  </books>
```

Viewed in web browser

```
  <?xml version="1.0" encoding="UTF-8" ?>
- <books>
  - <book>
      <title>The XML Plot</title>
      <author>Brown, Browny</author>
      <publisher>Singleday</publisher>
      <price>$24.95</price>
      <contentType>Fiction</contentType>
      <format>Hardback</format>
      <isbn>0385504209</isbn>
    </book>
  </books>
```

Viewed with styles

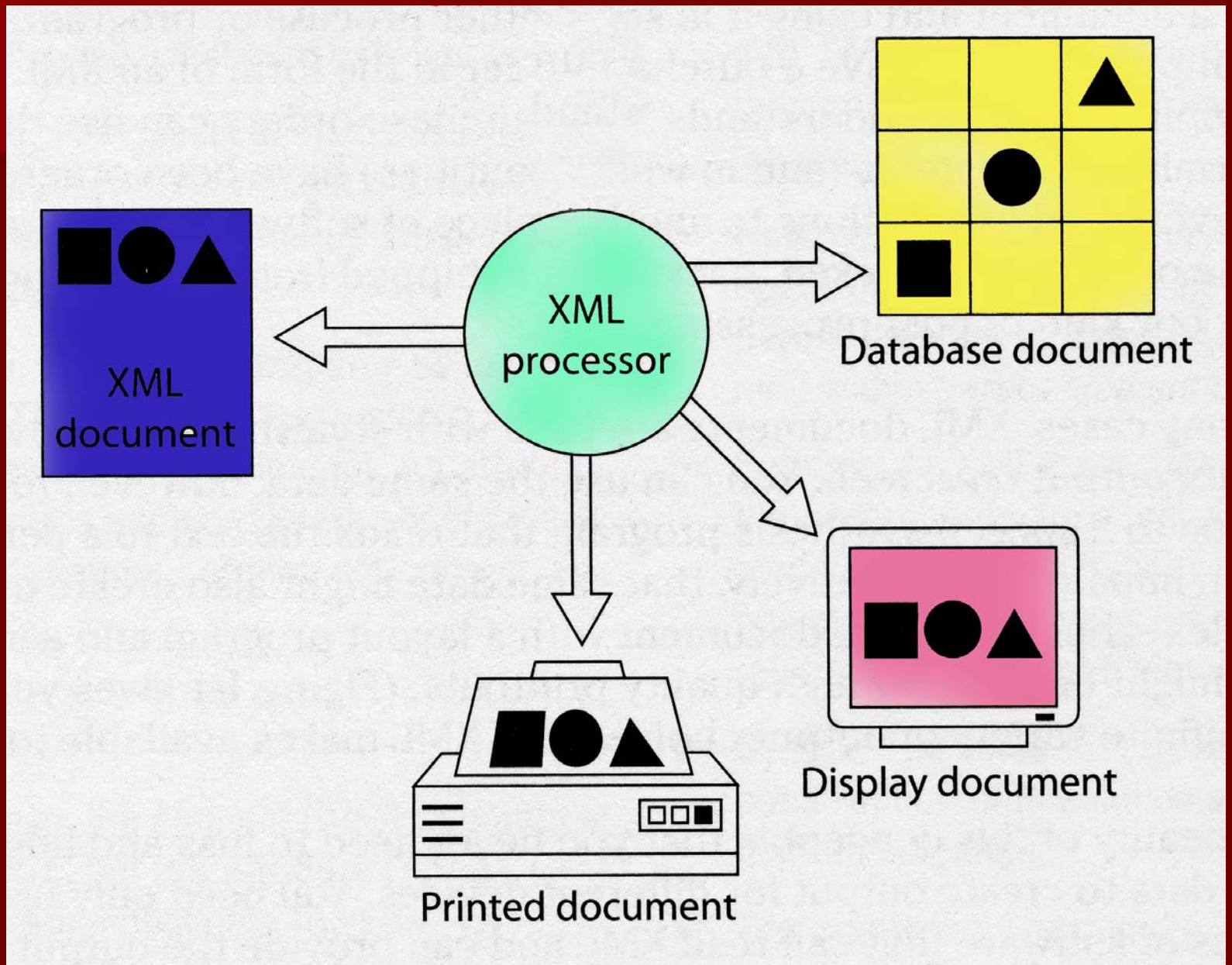
The XML Plot

Brown, Browny
Singleday
$24.95
Fiction
Hardback
0385504209

State of Fear of XML

Crichton, Mikey
Self
free
Fiction
Hardback
99999999

Outputs from XML (transforms)



Many views of the same data

- Via XSLT transforms
- Similar to using css style sheets files to change web styles
  - Example: csszengarden.com

C/NET SOLUTIONS

XML: Why do we care?

- 'CHEWUPS'
  - CDC's choice for state transmits (NEDSS)
  - HL7 standard format
  - Easier to parse and program than flat files
  - Web-browser viewable
  - Updates easier than fixed-length formats
  - Popular with business/IT
  - Self-documenting and constraining

C/NET SOLUTIONS

Advantage #1: CDC's Use of XML

- NPCR uses XML to generate CS tables and documentation
- NEDSS/PHIN use
  - For lab reporting
  - Using HL7 version 3
  - Created STF format

C/NET SOLUTIONS

NEDSS Use of XML

- Part of PHIN network
- Used to exchange state <> CDC data
- ebXML packages send XML + other data
  - National Cancer Research Network (NCRN)
  - Over 90 reportable conditions.

C/NET SOLUTIONS

Simplified Transitional Format

The *Simplified Transitional Format* (STF) has been developed by the CDC in an effort to *simplify* the *transition* by states to the HL7 Version 3 Message for NCRN.

C/NET SOLUTIONS

Advantage #2: HL7 provides standards for XML Formats

- HL7 Version 2.4, 2.5
- HL7 Version 3

C/NET SOLUTIONS

V3 CDC infectious disease report trigger section [in XMLSpy]

```
<?xml version="1.0" encoding="UTF-8"?>
<trigger>
    <publicHealthCaseEvent classCode="CASE" moodCode="EVN
">
        <code code="10210" codeSystem="
2.16.840.1.114222.4.11.183" codeSystemName="PH_PHC_TYPE"
displayName="Tetanus"/>
        <effectiveTime>
            <low value="20050213000000"/>
        </effectiveTime>
```

Advantage #3: XML is Easy to Program

- Cross-platform
- Integrates with MS Office 2003
- Many other tools are available.

C/NET SOLUTIONS

Tools

- Commercial vendor tools
  - XML Spy
  - Orion Symphonia/Rhapsody
  - Lots of others…
- Upcoming C/NET XML/NAACCR translator

C/NET SOLUTIONS

Advantage #4: XML is Web browser-viewable

- Some illustrations from possible NAACCR messages
- Non-standard compared with NAACCR-standardized.

C/NET SOLUTIONS

Non-NAACCR XML implementation

NAACCR0.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<cases>
    <case>
        <Item10144>Ahab</Item10144>
        <Item202005>Captain</Item202005>
        <Item30304444>01012005</Item30304444>
        <Item4045432>C430</Item4045432>
    </case>
</cases>
```

Text | Grid | Schema/WSDL | Authentic | Browser

NAACCR example using HL7/ NAACCR standards

```
<case>
    <Name–Last>value="Ahab" uid="NAACCR2230"
datatype="ST"</Name–Last>
    <Name–First>Captain</Name–First>
    <AddrCurrent–PostalCode value="37201" uid="NAACCR1830"
datatype="ST"/>
    <Sex value="1" codeSystem="2.16.840.1.114222.4.11.206"
uid="NAACCR0220" datatype="CV"/>
    <Date-of-Diagnosis>01012005</Date-of-Diagnosis>
    <Primary-Site>C430</Primary-Site>
</case>
<case>
```

Example NAACCR fields

```
value="Ahab" uid="NAACCR2230" datatype="ST"
value="Captain" uid="NAACCR2240" datatype="ST"
value="C403" uid="NAACCR0400"
codeSystem="2.16.840.1.114222.4.11.206" datatype="CV"
```

C/NET SOLUTIONS

Advantage # 5: Format updates are easier

- Easy compared to fixed-length formats.
  - New items, length changes easy
- Coding tables changes are version-tracked
- Allows rich text

C/NET SOLUTIONS

Advantage #6: XML is popular with programmers

- ebXML and XML used in business-to-business web services
- Programmers see it as a simple way to store and send data
- But maverick implementations hurt connectivity
- Example of rich-text formatting of path results

C/NET SOLUTIONS

Non-XML Pipe-delimited version of path text

OBX|1|ST|SPSUR||||||||D
OBX|2|TX|SPSUR||(NOTE)~~SURGICAL PATHOLOGY REPORT~~~SPECIMEN #: Sxx-xxxx~SPECIMEN(S)~1: Small bowel biopsy.~2: Gastric biopsy.~3: Distal esophagus.~4: Mid esophagus.~~PREOPERATIVE~Gastroesophageal reflux disease. Gastritis.~~POSTOPERATIVE~Distal antral gastritis. Distal erosive esophagitis.~~~FINAL PATHOLOGIC DIAGNOSIS~~1.   SMALL INTESTINE, BIOPSY (X3):~MUCOSA OF SMALL INTESTINE WITH NO PATHOLOGICAL CHANGE.~2.   STOMACH, BIOPSY:~MILD CHRONIC GASTRITIS OF BODY MUCOSA OF STOMACH.~ORGANISMS RESEMBLING H. PYLORI ARE NOT PRESENT ON THE ROUTINE OR~SPECIFIC~IMMUNOPEROXIDASE STAINED SLIDES.~3.   DISTAL ESOPHAGUS, BIOPSY (X3):~MILD CHRONIC REFLUX ESOPHAGITIS.~4.   MID ESOPHAGUS, BIOPSY (X2):~STRATIFIED SQUAMOUS EPITHELIAL LINING OF MUCOSA OF ESOPHAGUS WITH~NO~PATHOLOGICAL CHANGE

C/NET SOLUTIONS

XML report with formatted text


```
<OBX.1>1</OBX.1>
<OBX.2>ST</OBX.2>
<OBX.3>
    <CE.1>35660-0</CE.1>
    <CE.2>Final Diagnosis</CE.2>
    <CE.3>LN</CE.3>
</OBX.3>
<OBX.5>
```


1. Small intestine, biopsy (x3): mucosa of small intestine with no pathological change.
2. Stomach, biopsy: Mild chronic gastritis of body mucosa of stomach. organisms resembling h. Pylori are not present on the routine or specific immunoperoxidase stained slides.
3. Distal esophagus, biopsy (x3): Mild chronic reflux esophagitis.
4. Mid esophagus, biopsy (x2): Stratified squamous epithelial lining of mucosa of esophagus with no pathological change


```
</OBX.5>
```

Advantage #7: Self-documenting data structure verified by schemas

- Schemas define and validate message structures
- Schemas are written in XML too

C/NET SOLUTIONS

XML Schema snippet for path message

```
<?xml version="1.0" ?>
<!-- XML Composer generated schema, Orion Systems NZ Ltd -->
- <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:dt="urn:schemas-microsoft-com:datatypes">
  - <ElementType name="ORUR01" model="closed" order="seq" content="eltOnly">
      <element type="MSH" minOccurs="1" maxOccurs="1" />
      <element type="PID" minOccurs="1" maxOccurs="1" />
      <element type="PD1" minOccurs="0" maxOccurs="1" />
      <element type="NK1" minOccurs="0" maxOccurs="*" />
      <element type="NTE" minOccurs="0" maxOccurs="*" />
      <element type="PV1" minOccurs="0" maxOccurs="1" />
      <element type="PV2" minOccurs="0" maxOccurs="1" />
      <element type="ORC" minOccurs="0" maxOccurs="1" />
      <element type="OBR" minOccurs="1" maxOccurs="1" />
      <element type="OBX" minOccurs="0" maxOccurs="*" />
```

SOLUTIONS

XML and cancer registries?

- Would fit with ebXML and PHIN standards
- Some states already starting to use XML
- Allows rich text formatting
- Easily updated when field definitions change
- Verifies format and content structure automatically using schemas

C/NET SOLUTIONS

What's next for California?

- Develop and test HL7/XML 'rich text' transmit
- Build XML translator toolbox
  - XML schemas for E-path, case transmit
  - Stand-alone conversion between XML, HL7 pipes, and NAACCR classic.
  - Table-driven for easy revision.

C/NET SOLUTIONS

What's Next Nationally?

- IT Committee
  - Canvas states who are starting to implement in XML
  - Discuss HL7 V2/V3 choices
  - Slowly develop framework that can help guide registry implementations
  - Make sure there is no XML 'mandate' – provide translation tools
- Keep on talkin'…..
  - Email: barryg@askcnet.org

C/NET SOLUTIONS