Guiding the Way to XML Data Exchange Implementation

Isaac Hands, MPH
Lead Software Architect, Kentucky Cancer Registry
Chair, NAACCR XML Data Exchange Workgroup
Representative-at-Large, NAACCR Board of Directors
TL;DR

• Get Ready:
  NAACCR fixed-width format is going away in January 2020

• Get Informed:
  https://naaccr.org/xml

• Get Help:
  https://www.naaccr.org/forums/forum/naaccr-xml-standard/
Overview

• Timeline for XML Implementation
• Data Exchange Standard Background
• Path to XML
• Overview of XML Data Exchange Standard
• Challenges
• Getting Help and Learning More
• Q&A
Timeline for XML Implementation

We Are Here

9/1/2017 - 10/30/2017
Update NAACCR Prep

10/1/2017

11/1/2017
Request XML Format Call for Data

Jan-18
Assess XML Participation

1/1/2018

Jan-18
Host Webinar for Software Vendors

4/1/2018

4/1/2018 - 9/30/2018
Update NAACCR Software Tools

7/1/2018

Nov-18
Call for Data XML Submission Required

1/1/2019

1/1/2019 - 7/26/2019
Software vendors finalize XML changes

7/1/2019

Dec-19
NAACCR Flat File No Longer Defined

12/31/2019

10/1/2019
XML Updates to Volume II

2020 changes will no longer define fixed-width format

• Version 18 (2018 changes): Fixed-width to XML conversion possible
• Version 19 (2019 changes): Fixed-width to XML conversion possible
• Version 20 (2020 changes): XML only, no fixed-width conversion

<table>
<thead>
<tr>
<th>Column #</th>
<th>Length</th>
<th>Item #</th>
<th>Item Name</th>
<th>XML NAACCR ID</th>
<th>PARENT XML ELEMENT</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 1</td>
<td>1</td>
<td>10</td>
<td>Record Type</td>
<td>recordType</td>
<td>NaaccrData</td>
<td>Record ID</td>
</tr>
<tr>
<td>2 - 2</td>
<td>1</td>
<td>30</td>
<td>Registry Type</td>
<td>registryType</td>
<td>NaaccrData</td>
<td>Record ID</td>
</tr>
<tr>
<td>3 - 16</td>
<td>14</td>
<td>37</td>
<td>Reserved 00</td>
<td>reserved00</td>
<td>Tumor</td>
<td>Record ID</td>
</tr>
<tr>
<td>17 - 19</td>
<td>3</td>
<td>50</td>
<td>NAACCR Record Version</td>
<td>naaccrRecordVersion</td>
<td>NaaccrData</td>
<td>Record ID</td>
</tr>
</tbody>
</table>
NAACCR Call for Data Submission

XML is required for November 2018 Submission to NAACCR Call for Data
• NAACCRPrep can be used to create the XML

• 26 / 69 (38%) central registries submitted XML in 2017 NAACCR CFD
Data Exchange Standard Background

“...designed to facilitate electronic transmission of cancer registry data among registries...”

<table>
<thead>
<tr>
<th>Data Transmission Syntax</th>
<th>Fixed-Width</th>
<th>XML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Model</td>
<td>Tumor</td>
<td>Patient / Tumor / Custom Extensions</td>
</tr>
<tr>
<td>Coding Standards</td>
<td></td>
<td>AJCC Cancer Staging Manual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Canadian Cancer Registry Data Dictionary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Commission on Cancer (STOR, FORDS, ROADS)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collaborative Stage Data Collection System</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NAACCR Standards for Cancer Registries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SEER Program Code Manual (NIH)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHO ICD-O Third Edition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>National Program of Cancer Registries (CDC)</td>
</tr>
</tbody>
</table>
Data Exchange Use Cases

• Provide standardized data from reporting sources to central registries

• Share data on residents of other states/provinces from one central registry to another

• Report data from diverse facilities or states/provinces contributing to a combined study
NAACCR Fixed-width Format

Developed since ~1988, 20 revisions

A18000000000000100000009SEVERNA PARKMD21146 999 9999999501 8801074...

<table>
<thead>
<tr>
<th>Incidence Record</th>
<th>Confidential Record</th>
<th>Full Case Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>4048 characters</td>
<td>6154 characters</td>
<td>24194 characters</td>
</tr>
<tr>
<td>Demographic, Tumor and Staging, Treatment, and Follow-up</td>
<td>+ Patient Identifiers and Physicians, Pathology Metadata</td>
<td>+ Text</td>
</tr>
</tbody>
</table>
Limitations of Fixed-Width Standard

• Lack of Extensibility
  • New items
  • Retired items
  • Custom items

• Poor Readability

• Empty Space is Wasted
Path to XML

2005 – Cancer Abstract Transmission Workgroup (CAT) evaluates different syntaxes and standards for data exchange

2006 – CAT Workgroup proposes a pilot project to test utility of CDA as the new data exchange method for NAACCR
  • **Guiding Principle:** Semantic Interoperability with Medical Record

2007-2009 – CDA Pilot Project, Lantana Group Hired
Lessons Learned from CDA Pilot Project

Difficult to Understand
• Should be understandable & easy to communicate

Difficult to Maintain
• Should be maintainable by NAACCR community

Poor Performance
• Size & speed should not suffer too much
NAACCR XML Guiding Principles

Understandable

Maintainable  Performant

NAACCR XML
Path to XML Continued...

2010 to 2013 – investigate other HL7-family options (greenCDA, FHIR)
Reset expectations, re-organize efforts after 8 years...

August 2014 – New task force formed to create a custom XML Data Exchange Standard, independent of HL7-family requirements

June 2015 – NAACCR XML draft standard released

September 2015 – NAACCR Board approves v1.0 of XML standard
XML Work Since Approval

February 2016 - V1.0 Java software tools and libraries released

November 2016 - NAACCR XML Pilot Project
   CDC, C/NET, Onco Inc, Rocky Mountain Data Systems, CA, TX, UT Central Registries

June 2017 - V1.0 XMLExchange Plus software and libraries released

August 2017 – Current version of XML Standard released (v1.3)

Ongoing – NAACCR XML Workgroup Meets twice a month
Overview of XML Data Exchange Standard

XML Elements:
- `<NaaccrData>`
- `<Patient>`
- `<Tumor>`
- `<Item>`

XML Attributes:
- `naaccrId`
- `naaccrNum`

```xml
<NaaccrData>
  <Item
    naaccrId="registryType"
    naaccrNum="30">2</Item>
  <Item ...
  <Patient>
  <Item
    naaccrId="nameLast"
    naaccrNum="2230">SMITH</Item>
  <Item ...
  <Tumor>
  <Item
    naaccrId="primarySite"
    naaccrNum="400">C180</Item>
  <Item ...
```
Data Model

<NaaccrData> - Items that apply to all patients and tumors
(ex. Record Type, Registry Type, Record Version)

<Patient> - Items that apply to all tumors for a patient
(ex. Race, Social Security Number, Date of Birth)

<Tumor> - Items that apply to a particular tumor
(ex. Primary Site, Diagnosis Date, Treatments)
Supporting Files

XSD
• enforces structural features of the XML document such as, element names, required attributes, parent-child relationships

Base Dictionary
• defines valid data items for each level of the data model (Patient, Tumor)
• maintained and controlled by NAACCR

User Dictionary
• defines custom Data Items, use-case dependent
• maintained by anyone, shared between senders and receivers
Key Benefits of NAACCR XML

• Extensibility

• Easy to maintain by NAACCR Community

• Simplified transition from fixed-width format

• Easy to create, consume, and process
Software Supporting NAACCR XML

Currently Supporting NAACCR XML
• C/NET Solutions
• Oncolog (Onco, Inc.)
• XMLExchange Plus (NPCR)
• Rocky Mountain Data Systems
• SEER*DMS, Submission Tool, Abstracting Tool, (IMS, Inc.)
• CPDMS (Kentucky Cancer Registry)

NAACCR XML Support Under Development
• METRIQ (Elekta Inc.)
• Web Plus (NPCR)
• SEER Data Viewer (IMS, Inc.)
Other Software for NAACCR XML

• User-friendly interfaces for working with NAACCR XML
• Software libraries for reading, writing, validating NAACCR XML
• Convert from XML to Fixed-width and back, graphical dictionary editor
• Run EDITS on XML files (NPCR XMLEExchange Plus)

IMS naaccr-xml GitHub project (Java)
  • https://github.com/imsweb/naaccr-xml

NPCR XMLEExchangePlus (Windows)
  • https://www.cdc.gov/cancer/npcr/tools/registryplus/xml-exchange-plus.htm
Challenges

• XML in SAS: Reading large, complete files and writing out XML

• Tools for viewing, searching, and browsing large XML files

• Modifying EDITS infrastructure to accommodate the Patient/Tumor data model

• Managing User Dictionaries across registries (Custom Data Items)
Getting Help and Learning More

Familiarize yourself with the standard:
https://www.naaccr.org/xml

Post on the forum:
https://www.naaccr.org/forums/forum/naaccr-xml-standard/

Try out Software Tools and Programming Libraries:
https://www.naaccr.org/xml#Software
Q&A

Please type questions into the WebEx Q&A Panel

If your question is not answered, please post on the forum: 
https://www.naaccr.org/forums/forum/naaccr-xml-standard

The forum is actively monitored by the NAACCR XML Workgroup and NAACCR Staff