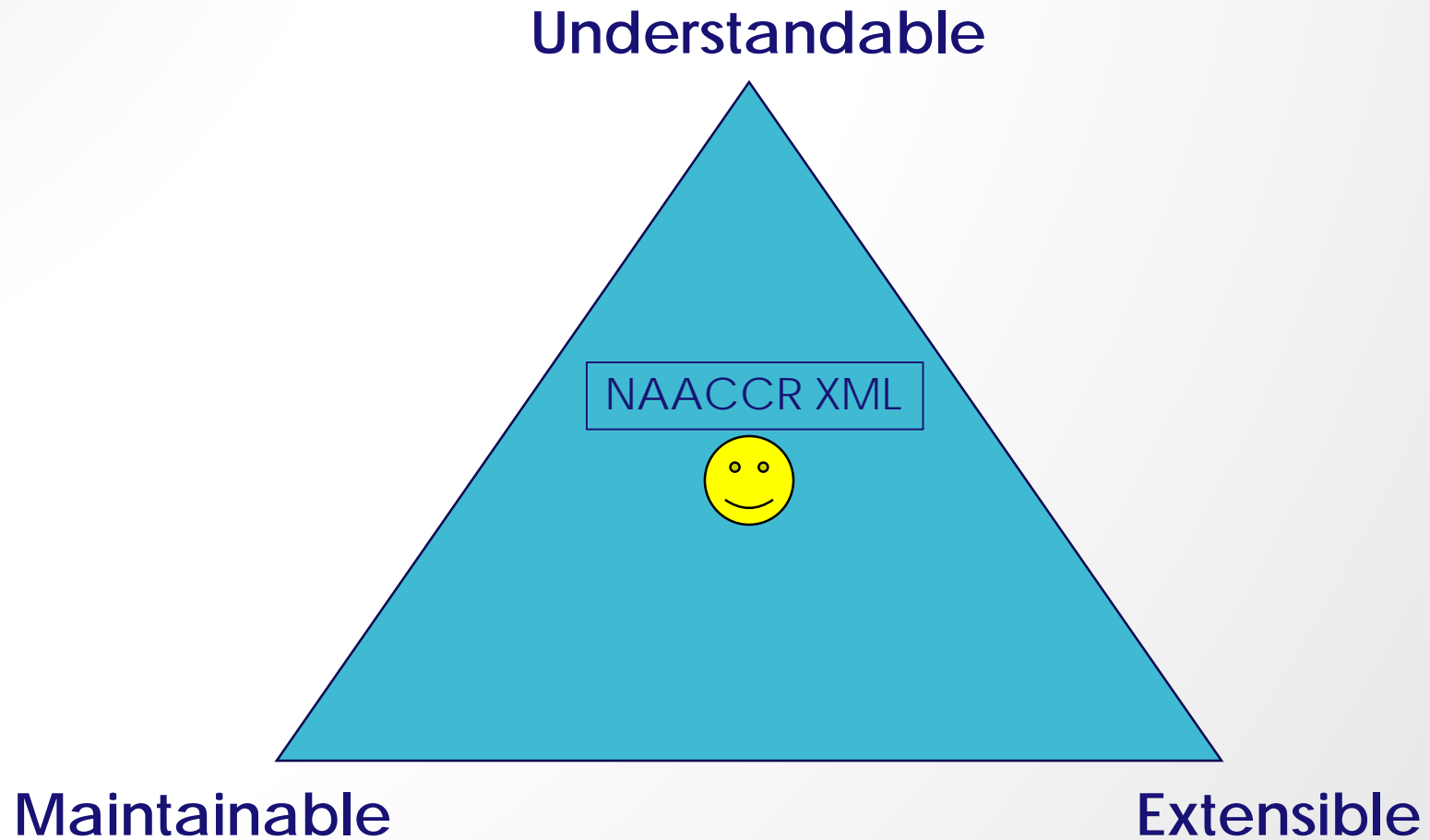


STRIKING A BALANCE



SCOPE OF DRAFT STANDARD

Specifically, a Cancer Registry Data Exchange Standard

Use Cases include:

- Registry to Registry
- Registry to Submission Agency
- Registry to Researcher

* **Not** for generic health record transactions *

DRAFT STANDARD BASICS

XML Elements:

- <NaaccrData>
- <Patient>
- <Tumor>
- <Item>

XML Attributes:

- naaccrId
- naaccrNum

```
<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 ...
    </Tumor>
    <Tumor>
      <Item naaccrId="primarySite" naaccrNum="400">C340</Item>
      <Item ...
    </Tumor>
  </Patient>
  <Patient> ... <Tumor>... </Patient><Patient>... <Tumor>...
</NaaccrData>
```

NAACCR XML ELEMENTS ARE NESTED

<NaaccrData>

- Items that apply to all patients and tumors
(Record Type, Registry Type, Record Version)

<Patient>

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

<Tumor>

- Items that apply to a particular tumor
(Primary Site, Diagnosis Date, Treatments)

<NaaccrData>

- Root level XML element that contains a separate <Patient> element for each patient record
- Has some “helper” attributes to make it easier to know how to parse a NAACCR XML file:
 - **recordType** – tells which NAACCR Items are valid in this file
 - **baseDictionaryUri** – tells which version of the NAACCR XML Dictionary to use (14, 15, etc.)

<Patient>

- Contains all NAACCR Items specific to a patient and all <Tumor> elements for a patient
- If consolidated patient data is known, a single <Patient> can contain multiple <Tumor> elements
- If consolidated data is unknown, then <Patient> elements will contain a single <Tumor> element

<Tumor>

- Always nested underneath a <Patient> element
- Contains all NAACCR Items specific to a tumor for a particular patient
- If consolidated data is known, multiple <Tumor> elements can be placed underneath a single <Patient>

<Item>

- All NAACCR Volume II Data Items are encoded as <Item> elements:

```
<Item naaccrId="nameLast" naaccrNum="2230">SMITH</Item>
```

- naaccrNum attribute is optional
- naaccrId is required
- All valid naaccrId values are defined in a NAACCR XML Base Dictionary file, derived from the Vol. II standard names

<Item> CONVENTIONS

- Missing Items can be truly missing
 - <Item></Item>
 - <Item> </Item>
 - <Item/>
 - Not in file at all
- The order of <Item> elements does not matter within their parent Element
- Line termination is optional in a NAACCR XML file
- Newline characters within Text are preserved

NAACCR XML BASE DICTIONARY

- How do we convert from NAACCR XML to a flat file and back ?
- How do we know if an <Item> belongs under the <NaaccrData>, <Patient>, or <Tumor> element ?
- Where are the naaccrId values defined ?
- How do we know which <Item> elements are valid for a particular record type ?

```
<NaaccrDictionary naaccrVersion="150" ... >
  <ItemDefs>
    <ItemDef naaccrId="registryType"
      naaccrNum="30"
      naaccrName="Registry Type"
      startColumn="2"
      length="1"
      recordTypes="A,M,C,I"
      parentXmlElement="NaaccrData"/>
    <ItemDef naaccrId="dateOfBirth"
      naaccrNum="240"
      naaccrName="Date of Birth"
      startColumn="196"
      length="8"
      recordTypes="A,M,C,I"
      parentXmlElement="Patient"/> ...
```

BASE DICTIONARY (CONT.)

- Formalized Volume II Definitions in an XML file
- Includes simple data type validation rules
- Base Dictionary name is a required attribute in each NAACCR XML file:
`<NaaccrData baseDictionaryUri="..."`
- Provides all information necessary for conversion of XML to flat file and back
startColumn, **length**, **trim** and **padding** attributes

ASSIGNING AN ITEM TO PATIENT OR TUMOR

- Base Dictionary has parentXmlElement attribute for every data item
Tumor, Patient, or NaaccrData
- <Item> elements can only be a child of one element, defined by parentXmlElement
- Assignment based on NAACCR Consolidated Item Workgroup, some discrepancies among registries

XML SCHEMA DEFINITION (XSD)

- W3C Standards-compliant XML specification
- Used by Schema Validation tools to enforce many structural features of an XML document such as:
 - element names, required attributes, etc.
- Used by XML mapping tools for relational databases
- NAACCR XML Draft Standard provides 2 schemas:
 1. XSD schema for a NAACCR XML data file
`naaccr_data.xsd`
 2. XSD schema for the NAACCR Dictionary
`naaccr_dictionary.xsd`

VALIDATION

3 levels of increasing validation specificity:

- 1. Element and Attribute Validation**

XSDs

- 2. Data Type and Nesting Structure**

Dictionary

- 3. Coding and Context**

EDITS (after conversion to flat file)

STATE REQUESTOR ITEMS

- Many registries put data in NAACCR Item 2220 (State/Requestor Item)
- Some registries assign NAACCR Item numbers to these custom Items (09500 – 09999) or (10000 – 10499)
- How do we preserve arbitrary custom NAACCR Items when converting from XML to flat file and back ?
- How do we allow registries to define their own <Item> definitions without changing the NAACCR-defined data dictionary ?

NAACCR XML USER DICTIONARY

- Users can define their own Dictionary to replace State/Requestor Items with Custom Items
- User Dictionary Items must fit within the length of the NAACCR Item they are replacing
- A Default User Dictionary is provided that preserves State/Requestor Items

```
<NaaccrDictionary naaccrVersion="150" ... >
  <ItemDefs>
    <ItemDef naaccrId="stateRequestorItems"
      naaccrNum="2220"
      naaccrName="State/Requestor Items"
      startColumn="2340"
      length="1000"
      recordTypes="A,M,C,I"
      parentXmlElement="Tumor"/>
    <ItemDef naaccrId="reserved00"
      naaccrNum="37"
      naaccrName="Reserved 00"
      startColumn="3"
      length="14"
      recordTypes="A,M,C,I"
      parentXmlElement="Tumor"/>
```

...

USER DICTIONARY (CONT.)

- **Base** Dictionary is static, defined by NAACCR for each Record Version
- **User** Dictionary is customizable, but only for State/Requestor and Reserved Items
- Custom User Dictionaries can be shared between registries to communicate data definitions

ARBITRARY EMBEDDED XML DATA

- What about data that does not fit into a State/Requestor Item ?
- How can we annotate a NAACCR XML file with other XML documents ?
 - ePath reports
 - EMR Records
 - Biomarker data

NAACCR XML EXTENSIONS

- Any valid XML can be embedded into a NAACCR XML Document as a NAACCR XML Extension
- Data Definitions for Extensions are handled outside of the Draft Standard
- Informal way to include arbitrary data without breaking the rest of the record
- Extensions will be lost in a conversion to flat file

LIMITATIONS AND FURTHER WORK

- NaaccrData/Patient/Tumor designations
- Rely on flat file for EDITS
- Many registry tools only work with flat file
- Need to synchronize efforts with NAACCR Vol. II annual updates
- Need to develop timeline for implementation
- Software development and support

THIS IS ONLY A DRAFT STANDARD

- Need feedback from community, testing help
- Conversion from flat file to XML and back is key
- Next steps:
 1. Gather feedback to create Draft v2
 2. Develop implementation timeline
 3. Present to NAACCR Standards and Registry Development Committee
 4. Present to NAACCR Board

MORE INFORMATION

<http://naaccrxml.org>

- XML Conversion Utility
- Implementation Guide
- FAQ
- XSDs
- Sample XML
- Test Results



THANK YOU

Questions ?

Isaac Hands, MPH

Kentucky Cancer Registry

NAACCR XML Task Force