# Using NAACCR XML with Relational Databases

Isaac Hands, MPH
Lead Software Architect, Kentucky Cancer Registry
Chair, NAACCR XML Data Exchange Workgroup
NAACCR Annual Conference 2016

# Background: NAACCR XML Data Exchange Standard

- NAACCR XML Version 1.0 Approved by NAACCR Board Nov. 2015

- Patient-Centered Data Model

```
<Patient>
    <Item naaccrNum="dateOfBirth">19600117</Item>
    <Item naaccrId="sex">1</Item>
    …
    <Tumor>
        <Item naaccrNum="dateOfDiagnosis">20100314</Item>
        <Item naaccrNum="primarySite">C340</Item>
        …
```

# Background: NAACCR XML Data Exchange Standard

- Allows for User-Defined fields that fit within fixed-width record
  - State Requestor Fields, NPCR Fields, Reserved Fields
  - Create User-Defined Data Dictionary with Item Definitions
- Allows XML “Extensions” that go beyond the fixed width file
- Software tools have been created to read and write NAACCR XML
  - Created by IMS, Inc.
  - Available on GitHub as open source project
- Great advantages over fixed-width format, the future of NAACCR Data Exchange

# Challenges Transitioning to NAACCR XML

- All cancer registries use relational database systems, not obvious how to map XML data into a relational model
- A lot of custom software built around the fixed-width format
- Data analytics and statistical reporting is based on “flat” data, dealing with XML may be unfamiliar (SAS)

# Typical Data Exchange Scenarios

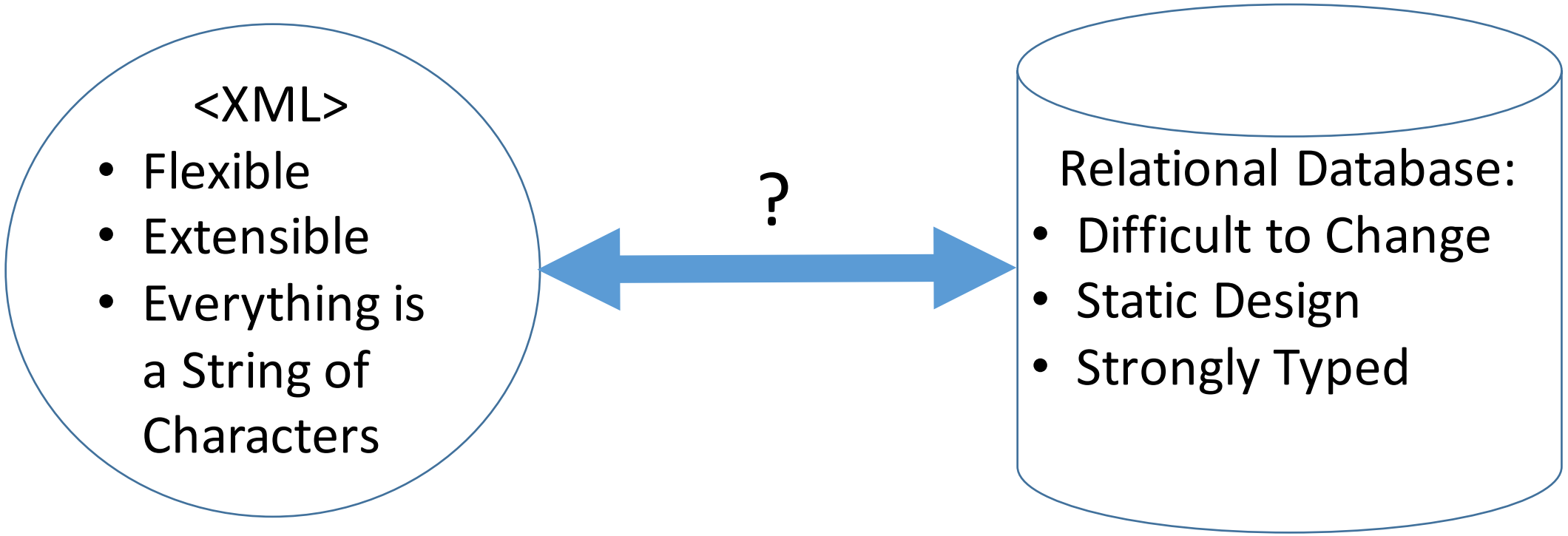
Scenario 1: Sender and receiver have very different data management systems and storage methods

- Registry to Out-of-State Registry
- Registry to Data Submission Agency

Scenario 2: Sender and receiver want to setup an ongoing data exchange with the same method of storing data

- Long-term research study

# Mapping XML to Databases



# Common Methods of Mapping XML to a Relational Database

| Method | Pro | Con |
|---|---|---|
| **1. Direct Storage of XML**<br>• Graph databases<br>• XML columns | • Automated<br>• Fast<br>• Flexible | • Vendor Lock-In<br>• Not Good for a Diverse Community<br>• Can be difficult to analyze data |
| **2. Schema-driven Mapping**<br>• XSD<br>• DTD<br>• RelaxNG | • Mostly Automated<br>• Self-Documenting<br>• Supported | • Loss of flexibility<br>• Limited by schema standard |
| **3. Custom Software** | Do whatever you want | Expensive<br>• Support<br>• Maintenance |

# Objectives

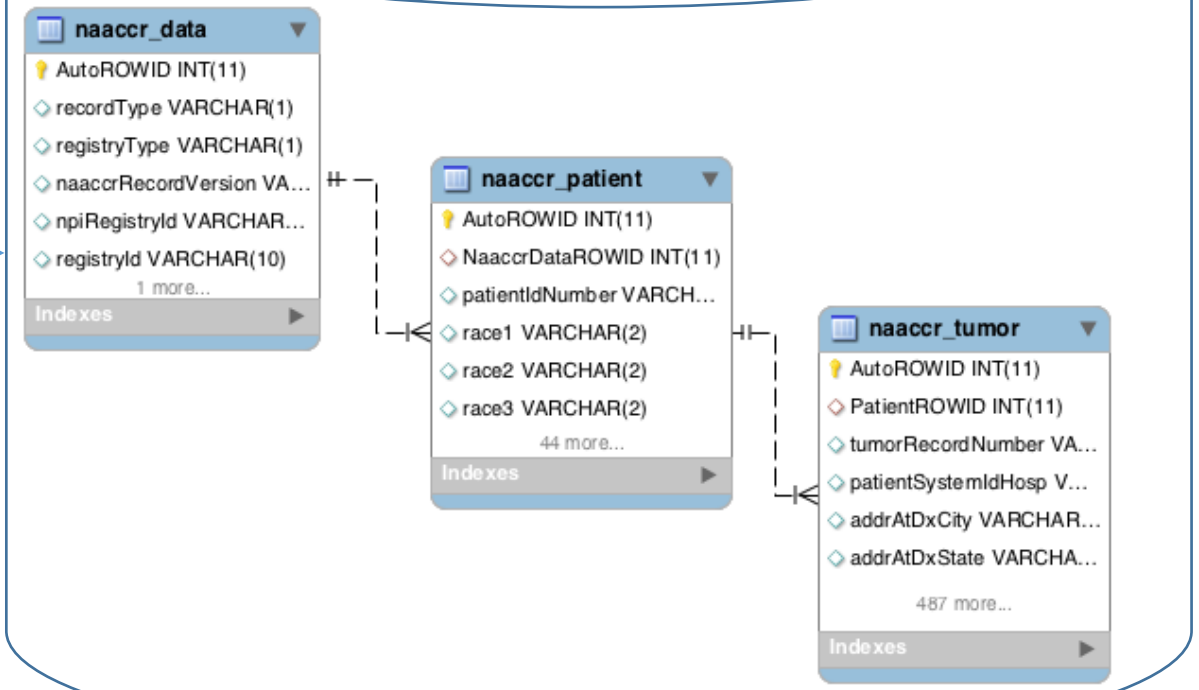
1. Read a NAACCR XML file and write to a relational database
   - Auto-Create database if it does not exist (along with tables and columns)
2. Create a NAACCR XML file from a relational database
   - Use SQL to specify patients or tumor to extract from database
   - Performance: Batch reads from database for large queries
3. Support custom data elements and extensions
4. Use existing NAACCR XML software tools as much as possible
5. Support Multiple Relational Databases

# Methods: Define a database table model

3 Item Containers in NAACCR XML

<NaaccrData>
<Item>
<Item>…
<Patient>…
<Item>
<Item>…
<Tumor>…
<Item>
<Item>…

At least 3 Tables in Database,
Items Become Columns



# Methods: Define a way to configure the database table model

<**PatientTable** *connection*="cpdms-default-mysql” *tableName*="naaccr_patient”>

...

<**TumorTable** *connection*="cpdms-default-mysql” *tableName*="naaccr_tumor”>

<**NaaccrItemColumn** *naaccrId*="dateOfDiagnosis" *columnName*="DiagDate"/>

<**ExtraColumn** *columnName*=”PatID” *foreignTableColumnName*="AutoID"/>

<**TableListener** *javaClass*="edu.uky.kcr.naxml.listener.CpdmsTableListener"/>

<**ExtraTable** *connection*="cpdms-postgres” *tableName*="case_other”>

# DEMO

# Results

- Several relational database systems supported:
  - PostgreSQL, MySQL, Sqlite
- Custom NAACCR Data Elements are supported
- Existing NAACCR XML Software needed modifications

# Conclusions

- Custom software for mapping from NAACCR XML to Relational Databases is a viable option
- NAACCR could oversee creation and maintenance of standard software tools
- Other relational databases could be supported
- Existing NAACCR XML Software could be expanded to support this solution
- Graphical User Interface would be nice

# Contact Information

Isaac Hands

Kentucky Cancer Registry

Chair, NAACCR XML Data Exchange Workgroup

isaac@kcr.uky.edu