Getting Ready for NAACCR XML in 2020

MARCH 27, 2019

Q&A

Please submit all questions concerning the content of the webinar through the Q&A panel

- Questions will be addressed at the end of the presentation.

If you experience technical difficulties please call us at 217 698 0800 x 111

Presenters

Isaac Hands, MPH
- Lead Software Architect, Kentucky Cancer Registry

Joe Rogers, MS
- Team Lead IDSAT, Division of Cancer Prevention and Control

Fabian Depry, MS
- Senior Systems Analyst, IMS

Jeff Reed
- American College of Surgeons, Senior Database Administrator for the NCDB database

NAACCR XML Readiness for 2020

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
  The fixed-width format will no longer be defined in Vol. II starting Jan. 2020
- Get Informed
  https://naaccr.org/xml
- Get Help
  https://www.naaccr.org/forums/forum/naaccr-xml-standard/

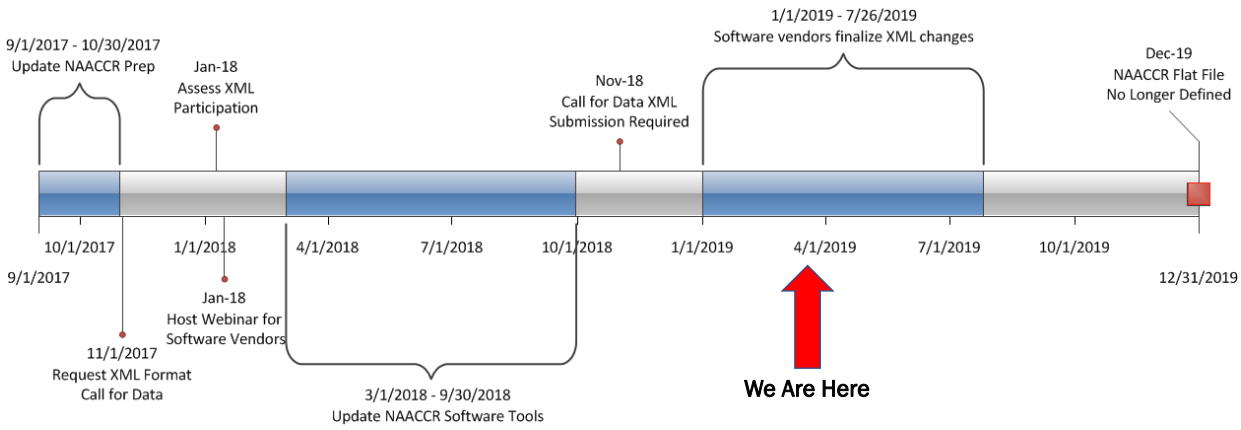
5

Overview

- Timeline for XML Implementation
- How did we get here and where are we?
- What is changing?
- What is the same?
- NPCR Presentation (Joe Rogers)
- SEER/IMS Presentation (Fabian Depry)
- ACoS/NCDB Presentation (Jeff Reed)
- Statements from other software vendors
- Q&A

6

Timeline for XML Implementation



7

How did we get Here?

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

2006 – CAT Workgroup proposes a pilot project to test utility of CDA

2007-2009 – CDA Pilot Project, Lantana Group Hired

2010 to 2013 – investigate other HL7-family options (greenCDA, FHIR)
No suitable option found: Reset expectations, re-organize efforts after 8 years…

August 2014 – New task force formed to create a custom XML Data Exchange Standard

June 2015 – NAACCR XML draft standard released

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

8

Work on NAACCR XML 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 (Including EDITS Support)

November 2017 - 26 / 69 (38%) central registries submitted XML for NAACCR Call for Data

October 2018 – Proposal for v1.4 to reduce length of names and define delimited output

November 2018 - **43 / 69 (86%)** central registries submitted XML for NAACCR Call for Data

December 2018 – Match*Pro release with native NAACCR XML Support

9

What is changing?

- Starting with NAACCR Volume II version 20, character starting positions will no longer be defined
- Starting with Vol. II v20, hospital registry software will submit XML to central registries
- Central registries will submit XML to national agencies (already happening)
- State-specific and other custom data items will be defined in User Dictionaries
- If you need a flat record from a Vol. II v20 file, convert into a delimited file using software tools provided by NPCR and SEER.
- EDITS for Vol. II v20 files will run natively on XML

10

What is not changing?

- NAACCR Volume II will still define data item lengths
- NAACCR Volume II will still define data item numbers
- You can still create flat data files for easy processing, but they will be delimited instead of fixed-width
- NAACCR Vol. II v18 (and earlier) files can still be fixed width

11

Why now?

- Standardized XML across the NAACCR Community is needed now
  - Treatment
  - Recurrence
  - Facility Admissions
  - Omics Data
  - Custom study-specific items
  - Narrative Text for Machine Learning
- Backlog of 2018 issues has created concern among registries, slowing the pace of data item changes in 2020
- We have waited long enough, it will not be easier when data item changes resume after 2020

12

NPCR Presentation

Joseph D. Rogers (Joe), MS
Team Lead: Informatics, Data Science, and Applications Team (IDSAT)
Cancer Surveillance Branch, Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention Chamblee Campus

13

Cancer Prevention Works

CDC / NPCR Support Plans for NAACCR XML

Joseph D. Rogers, M.S.
Team Lead: Informatics, Data Science, and Applications Team (IDSAT)
Cancer Surveillance Branch
Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Health Promotion
Centers for Disease Control and Prevention

NAACCR XML in 2020
March 27, 2019

Current CDC/NPCR NAACCR XML Support

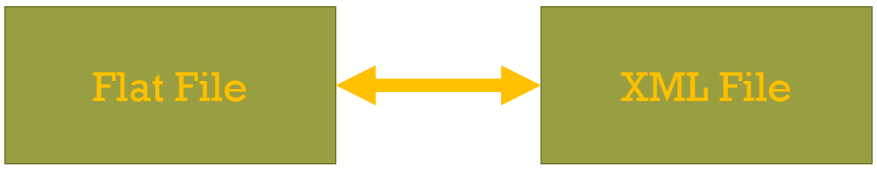
- CDC/NPCR currently provides an API and user interface utility, XMLExchange Plus:

  (https://www.cdc.gov/cancer/npcr/tools/registryplus/xml-exchange-plus.htm)

- The API can be used to validate a NAACCR XML file and to run edits.
- The API can be integrated in the registry's applications.

Current CDC/NPCR NAACCR XML Support, continued

- Based upon NAACCR standard dictionaries, conversion support is provided in both directions (conversion from a standard NAACCR XML file format to the NAACCR flat file format and back again).

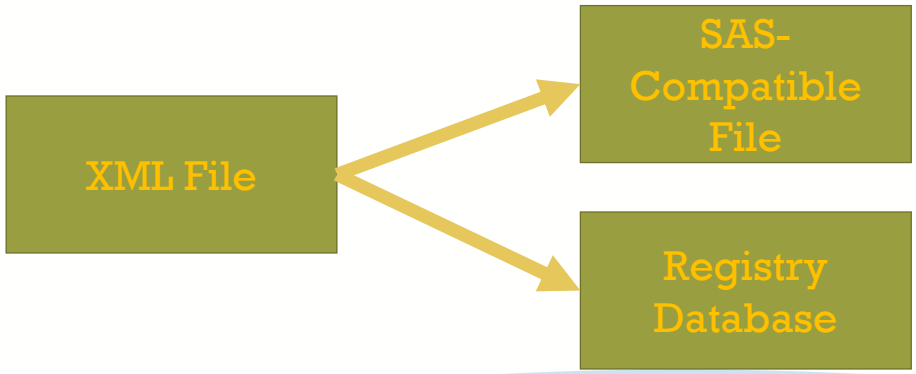


Planned Enhancements to CDC/NPCR NAACCR XML Support

- Validation, editing, and record review
  - NAACCR XML validation will include checks for non-standard and non-printable characters.
  - Editing features will include record-level character editing (analogous to using a text editor on a NAACCR flat ASCII file).
  - Record scanning functionality will be included.

Planned Enhancements to CDC/NPCR NAACCR XML Support, continued

- NAACCR XML support for analytic applications and database systems will be included.

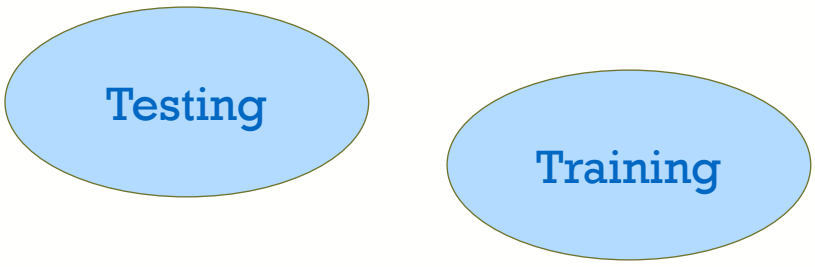


Planned Enhancements to CDC/NPCR NAACCR XML Support, continued

- Conversion support features
  - The user will be given the option to create a flat ASCII file.
  - The user can choose a delimiter character or no delimiter (column delimited).
  - Record columns (fields) will be sequenced using the NAACCR Data Item Numbers
  - Output files will be based on NAACCR XML fields (all NAACCR fields or a subset). NAACCR XML standards allow for all the NAACCR fields or a subset.

2020 Implementation Issues / Concerns

- Extensive API testing is needed.
- Any issues arising with interstate data exchange must be identified and resolved.
- Training requirements must be identified and a training plan developed.



2020 Implementation Issues / Concerns, continued

- Even if there are no changes to data submission standards or NAACCR 2018 standards for the coming year, XML implementation will require software updates.
- It will be challenging for registries to implement software upgrades while catching up on unprecedented backlogs caused by the delays in 2018 implementation.
- In their workflow, many cancer registrars review files before data transfer or upon receipt when edit errors are detected. There are also instances where modifications are needed directly to the data files. Registrars will need time to understand the new format to make the modifications.

2020 Implementation Issues / Concerns, continued

- NPCR Position for 2020
  - NPCR is committed to supporting the necessary transition from the NAACCR flat file format to the XML format.
  - NPCR will provide tools and utilities to support the new format.
  - However, NPCR believes registries need to focus during the coming year on eliminating backlogs.

Thank you!

Go to the official federal source of cancer prevention information:

www.cdc.gov/cancer

Cancer Prevention Works

Reliable. Trusted. Scientific.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

SEER/IMS Presentation

Fabian Depry

Information Management Services, Inc.

24

NIH NATIONAL CANCER INSTITUTE
Surveillance, Epidemiology, and End Results Program

Fabian Depry
Information Management Services, Inc.
March 27, 2019

SEER Software & NAACCR XML

Transitioning SEER software to the new NAACCR XML format.

Software Considered in this Presentation

SEER*DMS *-- SEER Data Management System* ❖ Marina Matatova (NCI) ❖ Linda Coyle, Chuck May (IMS)	**SEER Data Viewer** *-- Software to visualize, analyze and process data files* ❖ Carol Kosary (NCI) ❖ Fabian Depry (IMS)
SEER*Abs *-- SEER Abstracting Tool* ❖ Serban Negoita, Carol Kosary (NCI) ❖ Fabian Depry, Katherine Kirby (IMS)	**Match*Pro** *-- Linkage software* ❖ Lynne Penberthy (NCI) ❖ Will Howe, Don Green, Nicki Schussler (IMS)
SEER*Edits *-- SEER Registry Tool for SEER Submissions* ❖ Serban Negoita, Peggy Adamo (NCI) ❖ Jennifer Stevens (IMS)	**NAACCR XML Utility Tool** *-- Converts fixed-column files to XML and vice-versa* ❖ Carol Kosary (NCI) ❖ Fabian Depry (IMS)
SEER*Prep *-- Utility program to create SEER*Stat databases* ❖ Angela Mariotto, Rocky Feuer (NCI) ❖ Steve Scoppa (IMS)	*Development of the XML and Registry Management Software by IMS, Inc. for the NCI SEER Program is funded under contract HHSN261201500003B.*

SEER*DMS (SEER Data Management System)

- **Importing NAACCR XML Data**
  - Feature released in April 2016.
  - First registry to import a NAACCR XML data file was in September 2016.
  - Currently 6 SEER registries routinely import NAACCR XML files in production.
- **Exporting NAACCR XML Data**
  - Supported since February 2017.
  - System extracts in the NAACCR layout support Fixed-column and XML options.

SEER*Abs (SEER Abstracting Tool)

- Support for extracting NAACCR XML files added in May 2016.
- Shared Java library is used to write both XML and Fixed-column formats.
  - Configuration allows switching between both formats.
- No changes were needed in the data entry forms.
- Main reason for switching:
  - No more State Requestor Items.
  - Registry items defined in a user-dictionary that can be shared with other software.
- SEER*Abs is available online free of charge (https://seer.cancer.gov/seerabs/).

SEER*Edits (SEER Submission Tool)

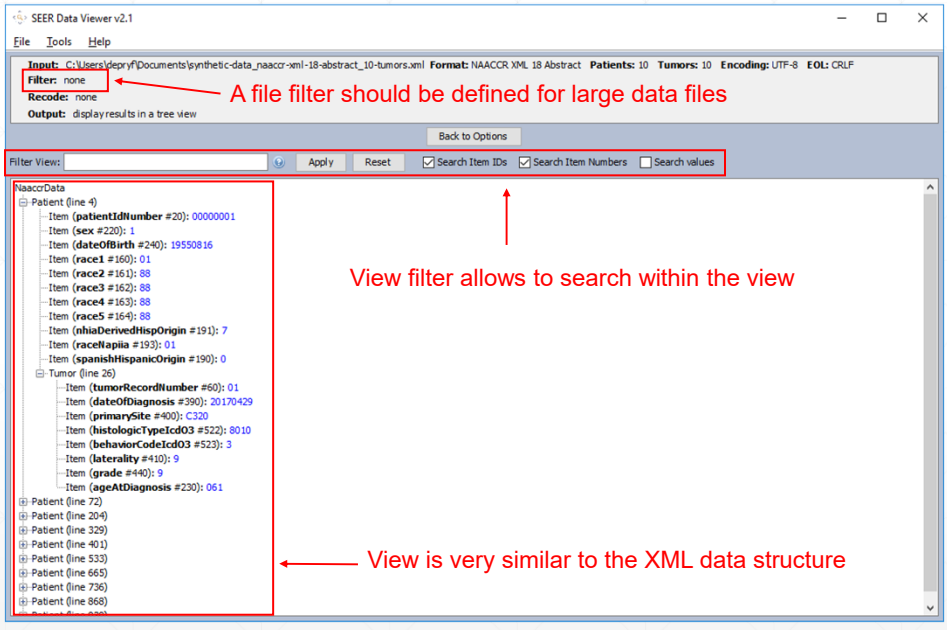
- Support for reading/writing NAACCR XML files added in September 2018.
- November 2018 SEER submission allowed both Fixed-Columns and XML formats;
  - Only one registry submitted in NAACCR XML.
  - Post-process tools maintained by IMS are still being converted to support XML
- The software is only used by SEER registries for the SEER submission.

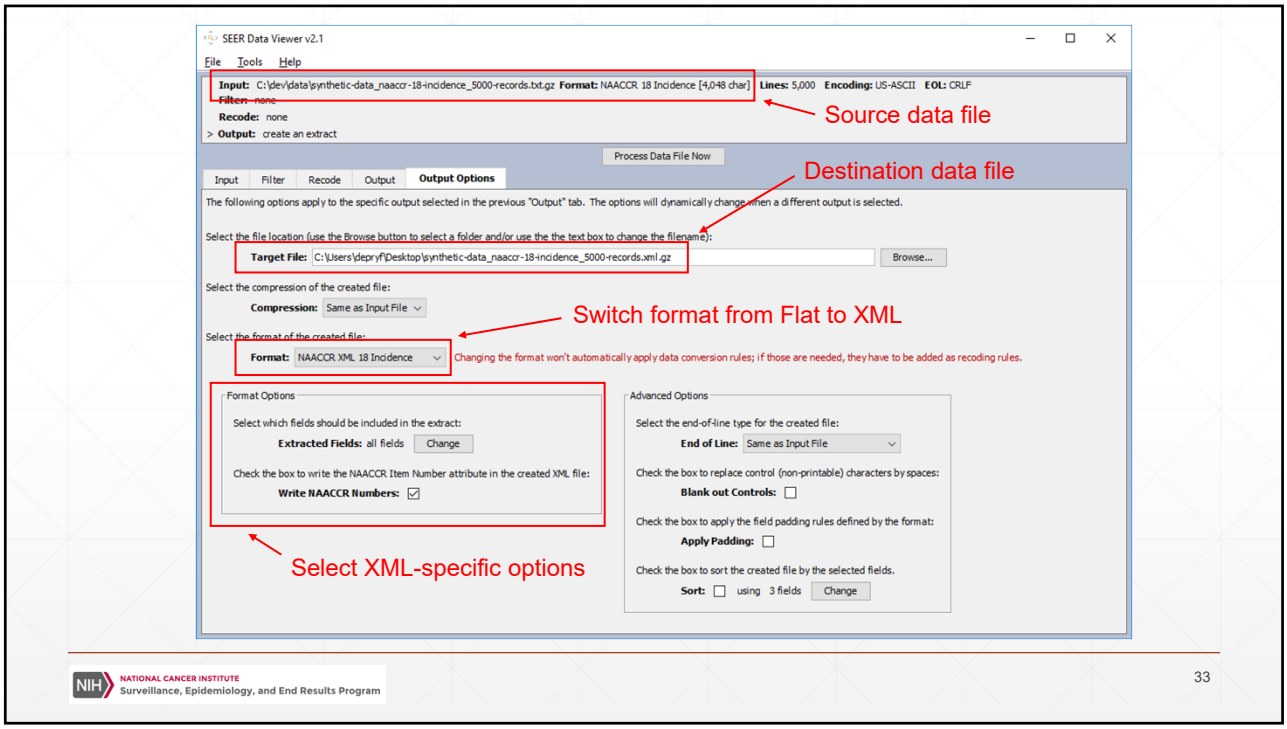
SEER*Prep / SEER*Stat

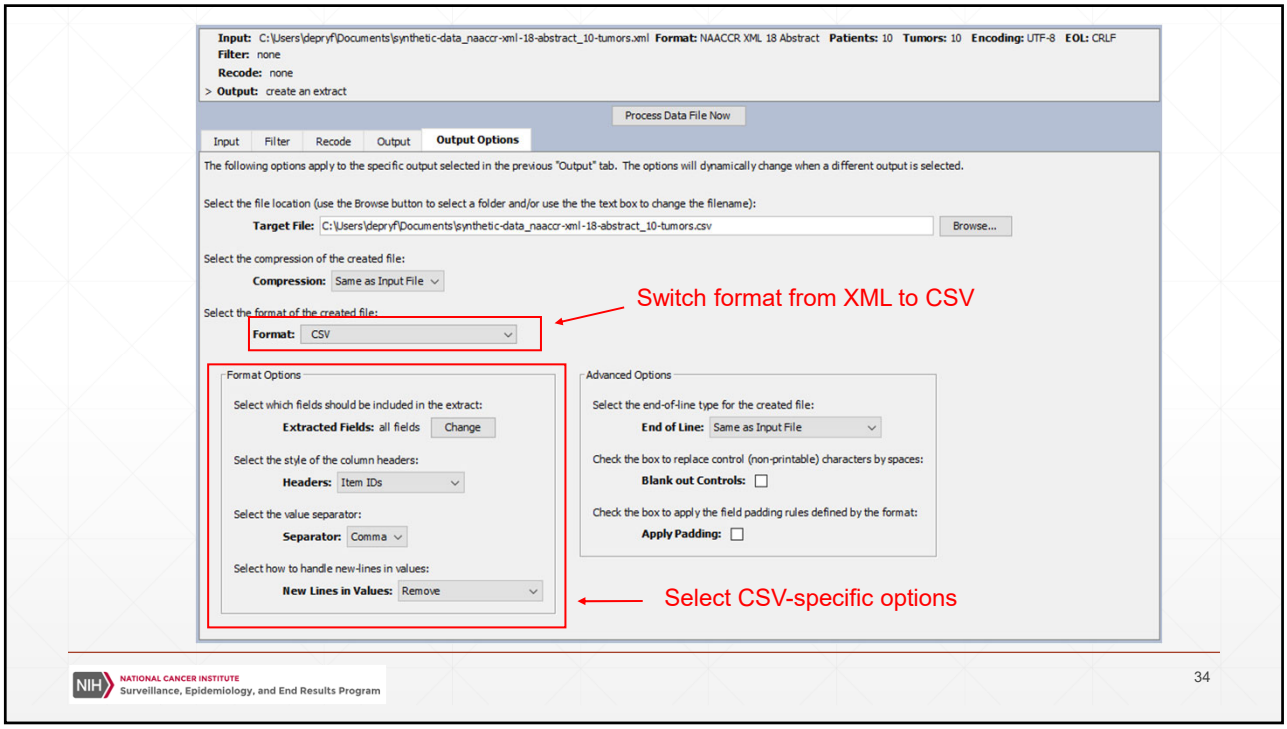
- SEER*Prep is used to transform input data files into a SEER*Stat database.
  - Only SEER*Prep is affected by NAACCR XML.
- Major upgrade to SEER*Stat, including a change in system architecture, is under way
- NAACCR XML will be supported when the new version of SEER*Stat is released.
- SEER*Prep might be updated earlier to support NAACCR XML (under evaluation).

SEER Data Viewer

- Utility tool used to analyze and recode data files.
  - File formats are a core concept of the software; adding support for NAACCR XML was very labor intensive.
- Support for NAACCR XML was added in January 2019.
- Software is publicly available (https://seer.cancer.gov/tools/dataviewer/).
- **This tool has features to help you deal with the NAACCR XML transition:**
  - Preview data in a table or a tree.
  - Re-create XML as Fixed-Column and vice-versa.
  - Create CSV from XML to use in SAS.

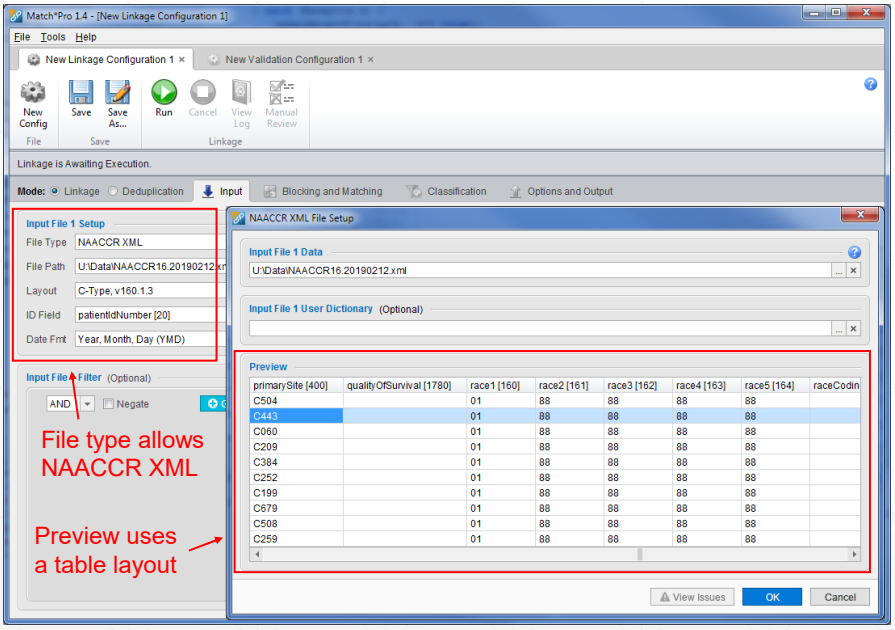



SEER Data Viewer v2.1
File Tools Help
Input: C:\dev\data\synthetic-data_naaccr-18-incidence_5000-records.txt.gz Format: NAACCR 18 Incidence [4,048 char] Lines: 5,000 Encoding: US-ASCII EOL: CRLF
Filter: none
Recode: none
> Output: create an extract
Source data file
Process Data File Now
Input Filter Recode Output Output Options
Destination data file
The following options apply to the specific output selected in the previous "Output" tab. The options will dynamically change when a different output is selected.
Select the file location (use the Browse button to select a folder and/or use the the text box to change the filename):
Target File: C:\Users\depryf\Desktop\synthetic-data_naaccr-18-incidence_5000-records.xml.gz
Browse...
Select the compression of the created file:
Compression: Same as Input File
Switch format from Flat to XML
Select the format of the created file:
Format: NAACCR XML 18 Incidence
Changing the format won't automatically apply data conversion rules; if those are needed, they have to be added as recoding rules.
Format Options
Select which fields should be included in the extract:
Extracted Fields: all fields
Change
Check the box to write the NAACCR Item Number attribute in the created XML file:
Write NAACCR Numbers:
Select XML-specific options
Advanced Options
Select the end-of-line type for the created file:
End of Line: Same as Input File
Check the box to replace control (non-printable) characters by spaces:
Blank out Controls:
Check the box to apply the field padding rules defined by the format:
Apply Padding:
Check the box to sort the created file by the selected fields.
Sort: using 3 fields
Change


Input: C:\Users\depryf\Documents\synthetic-data_naaccr-xml-18-abstract_10-tumors.xml Format: NAACCR XML 18 Abstract Patients: 10 Tumors: 10 Encoding: UTF-8 EOL: CRLF
Filter: none
Recode: none
> Output: create an extract
Process Data File Now
Input Filter Recode Output Output Options
The following options apply to the specific output selected in the previous "Output" tab. The options will dynamically change when a different output is selected.
Select the file location (use the Browse button to select a folder and/or use the the text box to change the filename):
Target File: C:\Users\depryf\Documents\synthetic-data_naaccr-xml-18-abstract_10-tumors.csv
Browse...
Select the compression of the created file:
Compression: Same as Input File
Switch format from XML to CSV
Select the format of the created file:
Format: CSV
Format Options
Select which fields should be included in the extract:
Extracted Fields: all fields
Change
Select the style of the column headers:
Headers: Item IDs
Select the value separator:
Separator: Comma
Select how to handle new-lines in values:
New Lines in Values: Remove
Select CSV-specific options
Advanced Options
Select the end-of-line type for the created file:
End of Line: Same as Input File
Check the box to replace control (non-printable) characters by spaces:
Blank out Controls:
Check the box to apply the field padding rules defined by the format:
Apply Padding:

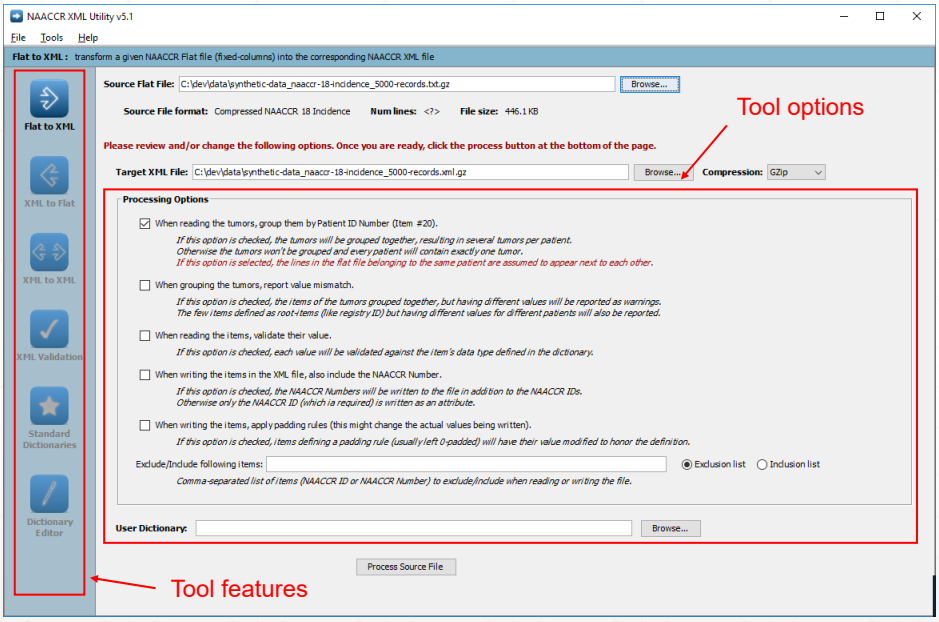
Match*Pro

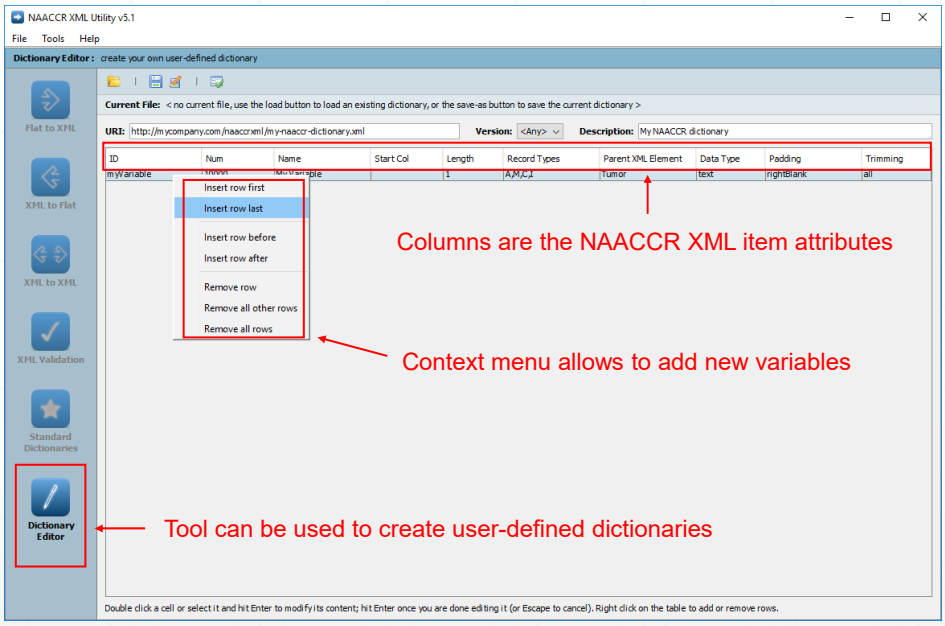
- Match*Pro is released under the Division of Cancer Control & Population Sciences program.
- Used to perform data files linkage.
- Support for reading NAACCR XML was added in December 2018.
- Next version will be released with support for validating data files (based on regular expressions or pick-lists), will also work on NAACCR XML files.
- Software is publicly available (https://surveillance.cancer.gov/matchpro).



NAACCR XML Utility Tool

- Designed by IMS under the SEER program.
- Referenced from the NAACCR website
  https://www.naaccr.org/xml-data-exchange-standard/#Software
- Publicly available online, along with other NAACR XML resources
  https://github.com/imsweb/naaccr-xml/wiki
- Features:
  - Fixed-column to XML; and XML to Fixed-column conversions.
  - XML data file validation.
  - User-defined dictionary creation.





Conclusion

- SEER*DMS, SEER*Abs, and SEER*Edits fully support NAACCR XML
- The **SEER Data Viewer** fully supports NAACCR XML **and is an excellent resource for anyone** who needs to view, recode, or convert XML data files
- Tools for processing SEER submission data are reviewed each year. IMS started converting these tools in 2016. Many have been converted to support NAACCR XML; others will be converted or retired (the processing will be handled by other mechanisms).
- The removal of Fixed-column in 2020 will not be an issue for the SEER ecosystem; work to support the XML standard started in 2015. IMS staff meet on a regular basis to review progress with an eye on the 2020 goal.

ACoS / NCDB

Jeff Reed
American College of Surgeons
Senior Database Administrator for the NCDB database

41

Cancer PROGRAMS
AMERICAN COLLEGE OF SURGEONS

Loading NAACCR XML Data Into A Relational Database

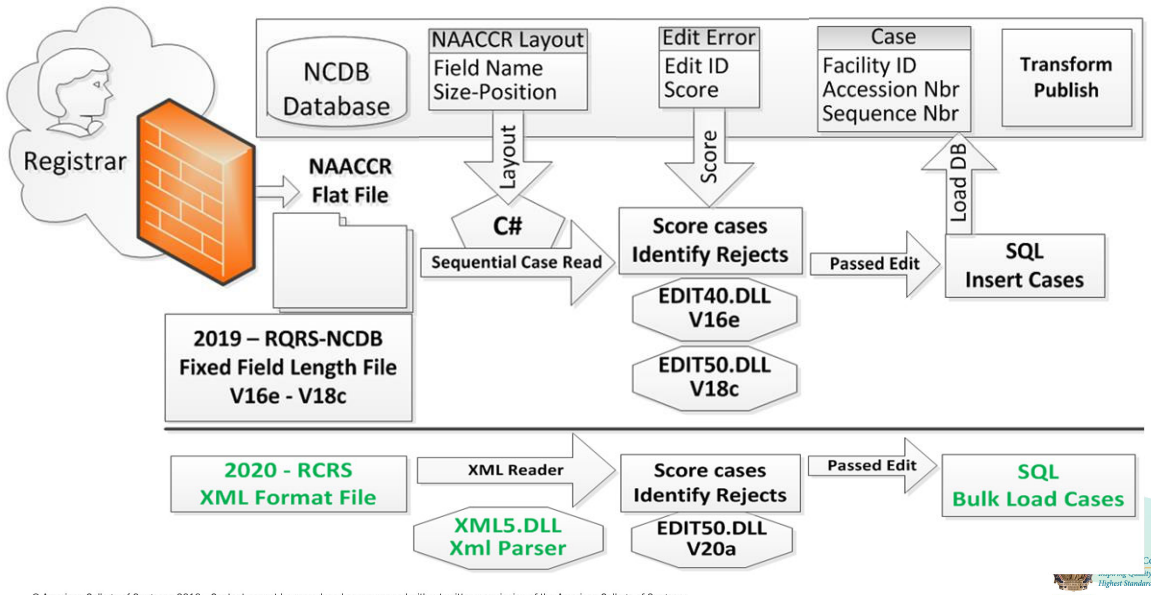
Jeff Reed
Database Architect - ETL Specialist
American College of Surgeons

American College of Surgeons
Inspiring Quality:
Highest Standards, Better Outcomes
100+years

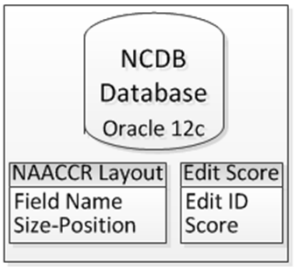
Topics

- SQL Loading of NAACCR Data Into NCDB
- Relational Data - Parents and Their Children
- New XML Format as Relational Data
- Using The CDC's XML5.DLL to Load Data
- Data Loader Application Prototype

SQL Loading of NAACCR V16–V18 Data



Oracle Processing Support Tables



NAACCR LAYOUT - http://datadictionary.naaccr.org/?c=7

VERSION	ORACLE_FIELD_NAME	START	LENGTH	NAACCR_COLUMN_NAME	NAACCR_NUM
180	RECORD_TYPE	1	1	"Record Type"	10
180	REG_TYPE	2	1	"Registry Type"	30
180	NAACCR_VERSION	17	3	"NAACCR Record Version"	50
180	NPI_REG_ID	20	10	"NPI--Registry ID"	45
180	FACILITY_ID	30	10	"Registry ID"	40
180	TUM_REC_NUM	40	2	"Tumor Record Number"	60
180	PATIENT_ID	42	8	"Patient ID Number"	20

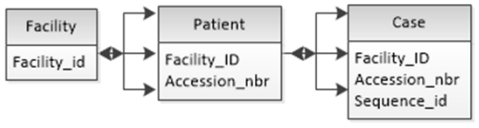
EDIT SCORING - https://www.cdc.gov/cancer/npcr/tools/edits/edits50.htm

EDITID	EDITNAME	V16NCDB	V16RQRS
N0123	Addr at DX--Postal Code (NAACCR)	200	0
N0124	RX Summ--Surg Prim Site (COC)	1	0
N0125	RX Summ--DX/Stg Proc (COC)	1	(null)
N0126	Sequence Number--Hospital (COC)	200	200
N0127	Rad--Location of RX (COC)	1	0
N0129	Primary Payer at DX (COC)	1	(null)
N0131	Reason for No Radiation (COC)	1	0
N0134	RX Summ--BRM (COC)	1	(null)

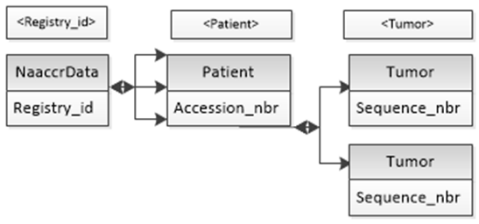
Relational Data Model

Flat File

REGISTRY_ID	ACCESSION_NBR	SEQUENCE_NBR
00690542	201900596	01
00690542	201900596	02



XML File



Parents and their Children

XML Header - Naaccr Data Dictionary

```
<NaaccrData baseDictionaryUri="http://naaccr.org/naaccrxml/naaccr-dictionary-180.xml"
 recordType="I" timeGenerated="2018-08-01T06:28:05.282-05:00"
 specificationVersion="1.3" xmlns="http://naaccr.org/naaccrxml">

   <Item naaccrId="registryId">20000289</Item>
```

Patient Data

```
<Patient>
    <Item naaccrId="birthplaceCountry">USA</Item>
    <Item naaccrId="vitalStatus">1</Item>
    <Item naaccrId="addrCurrentPostalCode">60585</Item>
    <Item naaccrId="addrCurrentState">IL</Item>
    <Item naaccrId="addrCurrentCity">PLAINFIELD</Item>
```

Tumor Data

```
<Tumor>
    <Item naaccrId="csSiteSpecificFactor5">987</Item>
    <Item naaccrId="derivedAjcc7M">010</Item>
    <Item naaccrId="derivedSs2000">2</Item>
    <Item naaccrId="seerRecordNumber">01</Item>
    <Item naaccrId="csSiteSpecificFactor4">987</Item>
    <Item naaccrId="derivedAjcc7N">020</Item>
```

Implementing the XML50.DLL

Cancer PROGRAMS
AMERICAN COLLEGE OF SURGEONS

C#

1. Load XML50.DLL
Create Delegates

2. SQL - Get Lists
get**Header**Fields(hList)
get**Patient**Fields(pList)
get**Tumor**Fields(tList)

3. Read Through XML Data File
Get Patient
Get tumor

Initialize DLL

```
Initialize(ref xmlId, xmlDictionary, xmlUserDictionary);
OpenXmlDataFile(xmlId,xmlFileName,owner, readitem_cb,progress_cb);
```

Loop Through Patients Tumors

```
ReadNextPatient(xmlId, ref t_is_eof)
GetPatientTumorsCount(xmlId, ref tumors_count)
ReadTumor(xmlId, t_tumor_ordinal)
```

Get Field Data from XML

```
GetItemDataByNaaccrId(xmlId, 'Record Type');
GetItemDataByNaaccrNum(xmlId, 40);

CloseXmlDataFile(xmlId);
```

XML Dictionary

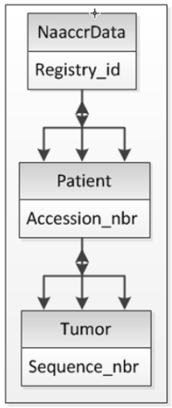
```
<NaaccrDictionary dictionaryUri="http://naaccr.org/naaccrxml/naaccr-dictionary-180.xml
        naaccrVersion="180"
        specificationVersion="1.3"
        description="NAACCR 18 base dictionary"
        xmlns="http://naaccr.org/naaccrxml">
    <ItemDefs>
        <ItemDef naaccrId="recordType"
            naaccrNum="10"
            naaccrName="Record Type"
            startColumn="1"
            length="1"
            recordTypes="A,M,C,I"
            parentXmlElement="NaaccrData"/>
        <ItemDef naaccrId="registryType"
            naaccrNum="30"
            naaccrName="Registry Type"
            startColumn="2"
            length="1"
            recordTypes="A,M,C,I"
            parentXmlElement="NaaccrData"
            dataType="digits"/>
```


American College of Surgeons
Inspiring Quality: Highest Standards, Better Outcomes
100+years

Load SQL From XML

Cancer PROGRAMS
AMERICAN COLLEGE OF SURGEONS

XML Data



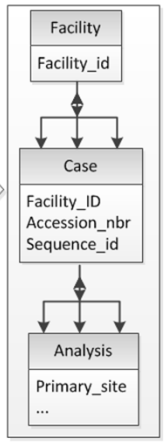
Create Delimited File

```
»20000289»20130608»1»
»20000289»20140318»1»
»20000289»20170317»1»
»20000289»20180516»1»
»20000289»20130516»1»
»20000289»20180110»1»
»20000289»20120430»1»
»20000289»20140131»1»
»20000289»20180630»1»
»20000289»20180326»1»
»20000289»20160111»1»
»20000289»20120826»1»
»20000289»20131223»1»
»20000289»20171125»1»
»20000289»20161202»1»
»20000289»20151214»1»
»20000289»20180110»1»
»20000289»20150517»1»
»20000289»20110724»1»
```

Create SQL Load File for bulk Load

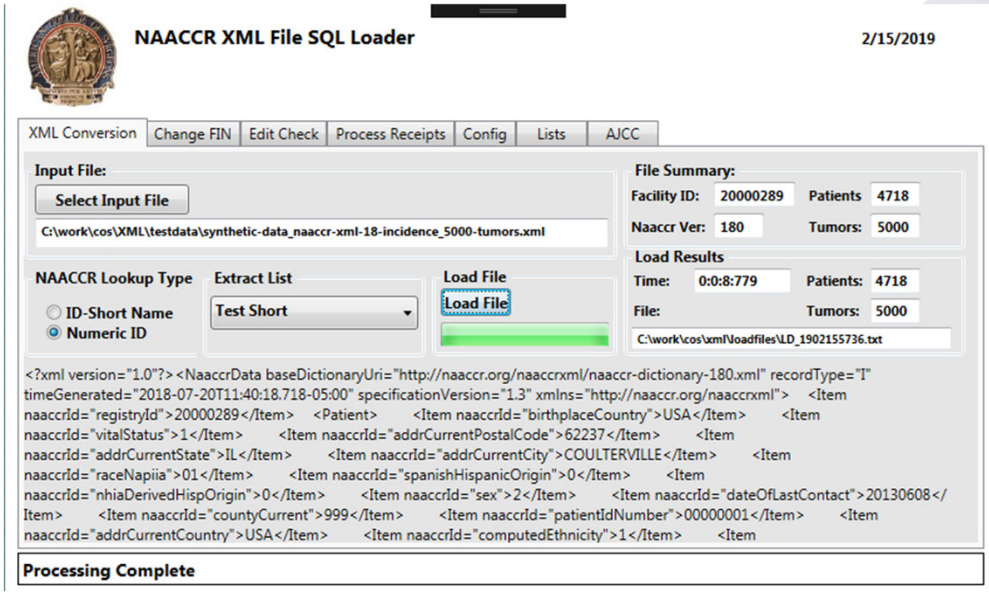
```
load data
INFILE 'C:\LD_1902134400.txt'
INTO TABLE NAACCR_LOAD
APPEND
FIELDS TERMINATED BY '»' (
FACILITY_ID char(10),
LAST_CONTACT_DT char(8),
VITAL_STATUS char(1),
CAUSE_OF_DEATH char(4),
BIRTHPLACE_STATE char(2),
BIRTHPLACE_COUNTRY char(3),
PATIENT_ID char(8),
COMP_ETHN char(1),
COMP_ETHN_SRC char(1),
CURR_COUNTRY char(3),
SEX char(1),
RACE char(2),
```

SQL Data



American College of Surgeons
Inspiring Quality: Highest Standards, Better Outcomes
100+years

XML NAACCR Data Loader Prototype



XML NAACCR Data Loader Prototype

Thank You

The NCDB Database

Cancer PROGRAMS
AMERICAN COLLEGE OF SURGEONS

The National Cancer Database (NCDB)

- Jointly sponsored by the American College of Surgeons and the American Cancer Society
- A clinical oncology database sourced from hospital registry data that are collected in more than 1,500 Commission on Cancer (CoC)-accredited facilities.
- Data represent more than 70 percent of newly diagnosed cancer cases nationwide and more than 34 million historical records.

American College of Surgeons
Inspiring Quality: Highest Standards, Better Outcomes
100+years

METRIQ®

"Elekta will be ready with our METRIQ registry product release in 2020 in support of the NAACCR requirements for transmission via XML format."

- Elekta, January 2019

52

"ERS is committed to incorporate and support NAACCR's XML State Export initiative and has been expecting the transition over the last year or more. We are excited to hear that NAACCR is moving forward in a measured manner with the States as well. ERS and CRStar expects to support this capability per the original timeline of December 2019, if not sooner. We look forward to partnering with NAACCR to test and implement this capability accurately and securely."

- ERS, Inc., January 2019

53

"C/NET Solutions has been a member of the NAACCR XML Workgroup for some time and was an early adopter of XML technology. We are fully prepared for the changeover (in 2020) to the XML format for transmitting new cases, corrections, and other record types. We have participated (and presented results of) a pilot program using XML format as a basis for internet-based transmit in concert with the California Cancer Registry (CCR). To ensure as smooth as transition as is possible, we will actively pursue other pilots and test programs throughout the registry community as the deadline approaches."

- Bert Heuer, Engineering Manager, C/NET Solutions, Public Health Institute, March 2019

54

ONCO
ONCOLOGY DATA MANAGEMENT SYSTEMS

"Onco is committed to supporting the NAACCR XML Data Standard and we look forward to the possibilities that it creates for the future of cancer reporting."
- Onco, Inc., March 2019

**This slide did not make it into the original presentation on March 27, 2019*

Q&A

Please type questions into the Zoom Q&A Panel

If your question is not answered during the webinar, 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

56

THANK YOU!