## Definition of NAACCR XML Data Elements

Element: NaaccrData

Children: Item (0:n), Patient (0:n)

This is the top level of the hierarchy.

Attribute: baseDictionaryUri (anyURI, required)  
Attribute: userDictionaryUri (anyURI, optional, default=http://naaccr.org/naaccrxml/defaultUserDictionary.xml)

These attributes reference two distinct NaaccrDictionary elements (q.v. in Dictionary Elements section). dictionaryUri is required and corresponds to a particular version of the NAACCR standard. userDictionaryUri is optional and describes state, requestor and system items. The NAACCR default user dictionary defines reserved and local use portions of the NAACCR flat file as unformatted text. State, local and requestor items are defined in a custom user dictionary identified by the userDictionaryUri attribute which allocates space in the reserved and local use portions of the NAACCR flat file (q.v. Other Conventions section).

Attribute: timeGenerated (dateTime, required)

This is date-time when the extract was produced by the system. Note that its format is YYYY-MM-DDThh:mm:ss, with the T being the start of the time segment and all six components required.

Attribute: recordType (required, in {A,M,C,I})

This is the NAACCR record type to which the data corresponds.

Element: Patient

Parent: NaaccrData

Chldren: Tumor (1:n), Item (0:n)

This represents a single patient. The patient is required to have at least one tumor. While more than one tumor may be contained within a single patient, it is not required that all tumors relating to a patient be reported together.

Element: Tumor

Parent: Patient

Child: Item (0:n)

This represents a single tumor experienced by the patient.

Element: Item

Parents: NaaccrData, Patient, Tumor

This is the value for a single NAACCR item for its parent object. The value is carried in the data portion of the tag. White space is handled with the XML “replace” option.

Attribute: naaccrId (NCName, required)

This is the identification of the item, which is based on the NAACCR field name but is suitable for use as a variable name.

Attribute: naaccrNum (integer, optional)

This is the assigned item number of the item.

## Definition of NAACCR XML Dictionary Elements

Element: NaaccrDictionary

Children: ItemDefs (1:1)

This is the top level of the hierarchy. It represents the set of metadata necessary to validate and use the data in a NaaccrData element (q.v. in NAACCR Data section). It is anticipated that in the future there will be more child elements than solely the dictionary, such as field lists for profile support.

It is anticipated that a base dictionary will be distributed along with each NAACCR version. State, system and requestor items will go into a second dictionary to allow customization by state registries.

Attribute: dictionaryURI (anyURI, required)

This is the name by which the dictionary is identified. Although it is in the format of a web address or other URI, per the XML rules this does not imply that such a web address actually exists; the processing program is assumed to know the actual locations of the files it needs.

Attribute: naaccrVersion (optional)

This identifies the NAACCR version to which the dictionary corresponds. It is primarily intended for use with NAACCR-released base dictionaries.

Attribute: description (optional)

This is a description of the dictionary. It is intended to be used to describe the sets of state, requestor and system data items included in a custom user dictionary

Element: ItemDefs

Parent: NaaccrDictionary

Children: ItemDef (0:n)

This element contains the complete list of items which may be used in a NaaccrData message. Note that the NAACCR dictionary for a particular version will contain only those items in use in the specified NAACCR version. It is expected that states will comply strictly with NAACCR policy that, where it is desired to maintain data after an item has been retired, a new state or requestor item will be defined with a new name, id, number and flat file location.

Element: ItemDef

Parent: ItemDefs

This element contains information about the NAACCR item.

Attribute: naaccrId (ID, required)

This is the standard name of the item. It is unique across all ItemDefs.

Attribute: naaccrNum (integer, required)

This is the item number of the item. It is unique across all ItemDefs.

Attribute: dataType (optional, enumeration, default=string)

This is the data type of the item. This is used for validation and in most cases can be used to imply the value of the regex attribute. It carries information on justification and padding for the flat file. The acceptable values are as follows:

code: a field whose full length contains numeric characters  
 codeWithBlank: a code for which all blanks is a meaningful value  
 alpha: a field whose full length contains alphabetic characters  
 alphaWithBlank: an alpha for which all blanks is a meaningful value  
 string: variable-length text which starts with an alphanumeric character; in the NAACCR record, it is right-padded with blanks to fill its length  
 stringInitialBlankOK: variable-length text which starts with any character; in the NAACCR record, it is right-padded with blanks to fill its length  
 integer: a number, potentially with an assumed decimal point; in the NAACCR record, it is right-padded with blanks to fill its length  
 integerLeftZeroPad: a number, potentially with an assumed decimal point; in the NAACCR record, it is left-padded with zeroes to fill its length  
 date: a date in one of the following formats – YYYYMMDD, YYYYMM, YYYY

Attribute: regexValidation (optional)

This is a regular expression to be used to validate the data in the field; it supplements datatype when a non-standard regular expression is required for validation. For example, NAACCR item Record Type [10] could have regex= "^[ICAM]$".

Attribute: parentXmlElement (required, in {NaacrData, Patient, Tumor})

This indicates whether the item applies to a tumor, a patient, or all the patient-tumor data in the message. It is used to validate whether an item is in the correct portion of the message. The following items are NaaccrData items: NAACCR Record Version [50], Record Type [10], Registry ID [40], Registry ID NPI [45], Registry Type [30], Vendor Name [2170]. Patient items are all and only those items which per the NAACCR consolidation rules should have the same value across all tumors relating to the patient (e.g., Date of Birth [240]). All other items are Tumor items.

.

Attribute: naaccrName (optional)

This is the NAACCR descriptive name for the item.

Attribute: startColumn (integer, required)

Attribute: length (integer, required)

This is the starting column and length of the item in the NAACCR flat file. They are required if the metadata is to be used for importing or exporting NAACCR flat files.

Attribute: sourceOfStandard (optional)

This is the agency primarily responsible for promulgating the data element. Typical values might include "NAACCR12", "NPCR CER", "SYS".

Attribute: recordTypes (optional, in {I,C,A,M,L})

This is a comma-separated list which identifies the NAACCR record types that use this item.

## Other conventions regarding the NAACCR XML format

Empty fields

Empty fields (those which would be all blanks in the flat file) need not be transmitted, even if all blanks is a meaningful value. Empty fields can also be transmitted with empty Items (<Item naaccrNum=2500 /> or <Item naaccrNum=2500></Item>). Note that this is not the same as <Item naaccrNum=2500>{spaces}</Item>. At such time as we provide a capability to identify items as required for a particular record format, use of an empty Item will indicate a required data item with an all-blank value.

Line terminations

Lines in the XML file should terminate with carriage return-line feed or line feed.

Validation

Validation of NAACCR is done at three levels: at the XSD level; at the Dictionary level; and at the EDITS level. Messages which fail XSD validation will be rejected in their entirety; messages which fail Dictionary validation for a particular Patient will flag that Patient for correction or reject all data for the Patient; and messages which fail EDITS validation for a particular Patient will flag that Patient for follow-up and correction.

XSD Level Validation

The XSDs specify the structure of the Data and Metadata messages. XSD validation assures that:

1. elements are properly nested;
2. the number of elements of a particular type complies with the minimum and maximum specified in the XSD;
3. all required attributes are present; and
4. dictionary attribute values for recordType, dataType and parentXMLElement have permissible values.

Dictionary Level Validation

Software which receives NAACCR XML Data should use an appropriate metadata specification to perform validations to assure that:

1. each Item element has an NaaccrId and/or NaaccrNum which corresponds to a single ItemDef element in the metadata;
2. each Item element is contained within the element identified by the parentXMLElement attribute of the ItemDef;
3. at most one Item element corresponding to a single ItemDef appears within a Patient or Tumor element; and
4. each Item element has a data value which matches the regular expression implied by the datatype attribute or specified in the regex attribute of the ItemDef

EDITS Level Validation

EDITS level validation is handled by the existing EDITS facility. It assures that:

1. each Item element has a data value which is a valid value (duplicating the ValidValues capability);
2. each Item element required for a particular message is provided; and
3. multi-field validation conditions are met for a particular Patient or Tumor.

Order of Items

It is recommended that items under a parent element appear in the file in naaccrNum order.

Grouped Data Items

Grouped data items as defined in Appendix E of Volume II of the NAACCR standard will not be supported in XML. Grouped data items should be transmitted by transmitting their component items and the grouped data item value, if used, should be computed from them.

Local/Requestor/System Items

Non-NAACCR-specified items may appear in the XML file under the following conditions:

* local/requestor/system items be assigned names and numbers:
  + item names should be unique to the locality, requestor or vendor, typically through the use of prefixes (e.g. RI\_SmokingHistory, CER\_Height, RP\_AbsRefId)
  + item numbers should be assigned in the appropriate range, 9500-9999 for local/requestor use, 10000-10499 for system use (see Standards vol. I sec. 3.3) [
* local/requestor/system items appear in the user dictionary;
* changes not be made to the NAACCR standard dictionary or the NAACCR XSDs.

When creating a user dictionary, it is important to understand that any data in reserved or local use items will be lost if the columns defining them are not included in the user dictionary. For example, suppose that State/Requstor Items [2220] (cols. 2340-2439) has been used in the past to enter some free text, and it is desired to record a new requestor item, Use of Tanning Beds [9600] in column 2439. The new user dictionary must specify all of the following:

* the new Use of Tanning Beds item;
* a new state/requestor item representing the free text in cos. 2340-2438; and
* all the reserved areas from the default user dictionary, e.g. Reserved02 [530] (cols. 464-527).

Failure to define the new state/requestor item would result in the loss of all data in cols. 2340-2438.

Extensions

Non-NAACCR-specified attributes may be used with any element defined herein per the XSD. Non-NAACCR-specified elements may be added within any element defined herein except Item per the XSD.