## Definition of NAACCR XML Components

Element: NAACCRFile

Children: Dictionary (0:1), Abstract (0:n), Patient (0:n)

This is the top level of the hierarchy.

Attribute: extractID (optional)

This is an identifier unique to the sender to identify the extract.

Attribute: NAACCRVersion (required)

This is the number of the NAACCR version which defines the standard items.

Attribute: xsdVersion (optional, fixed=1.0)

This is the XSD version. It allows for later changes in the XSD.

Attribute: datePrepared (dateTime, required)

This is date-time when the extract was prepared.

Attribute: sender (required)

This is the name of the facility which prepared the extract.

Attribute: senderCode (optional)

This is the facility ID of the sender. Note that values of this field may come from multiple code systems depending on the facility type.

Attribute: description (optional)

This is a description of the extract or the criteria used to select cases.

Attribute: geography (optional)

This is a code for the geographic area as to which the cases relate. Typically this would be a state, but it could be a sub-state, multi-state or provider-based registry.

Attribute: dateFrom (date, optional)

Attribute: dateTo (date, optional)

These are the first and last dates of the period used to select cases for the extract.

Element: Patient

Parent: NAACCRFile

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

This represents a single patient. The patient is required to have at least one Tumor.

Element: Tumor

Parent: Patient

Child: Item (0:n)

This represents a single tumor experienced by the patient.

Element: CancerRecord

Parent: NAACCRFile

Child: Item (0:n)

This represents a single version of a single abstract.

Attribute: documentID (string,optional)

Atttribute: timeStamp (dateTime, optional)

The documentID is a globally unique ID for the abstract or other data source. The timeStamp, which is the date-time of last modification, allows the recognition of different versions of the abstract based on the change processes that it has undergone (e.g. M-records, central registry edits, geocoding, death certificate linkage). While documentID and timestamp together can be used as a PHIR-compatible versioned document,use in this manner is not required; registries are free to edit item values without creating a new version or changing the timeStamp field. Whether a versioned document contains all items of the abstract or only the changed items is an implementation issue but should be agreed upon by the sender and recipient.

Attribute: isVoid (boolean, optional, default=false)

Attribute: voidBy (optional)

Attribute: voidDate (date, optional)

Attribute: voidReason (optional)

These fields allow the voiding of an abstract version for reasons such as duplicate entry, unreportability, or having been superseded. The date, responsible person name or ID, and reason may be supplied if isVoid=true.

Element: Item

Parents: Patient, Tumor, CancerRecord

This is the value for a single NAACCR item for its parent object. The value is carried in the data portion of the tag.

Attribute: itemName (optional)

This is the name of the item, which is based on the NAACCR field name but complies with XSD ID syntax. If any records in the XML file use itemName, the file must contain a dictionary ItemDef defining the metadata for that item.

Attribute: itemNo (integer, required)

This is the assigned item number of the item.

Attribute: isMissing (required, boolean, default=false)

This indicates that the value of this item is one of the item's defined missing values. This attribute is intended to simplify the work of downstream data analysts. Use of this field is optional with the facility and its use should is anticipated only with agreement in advance between the sender and the recipient.

Attribute: isDeprecated (required, boolean, default=false)

This indicates that the value of this item is not to be relied upon but is maintained for historical purposes. It is applicable only to items belonging to a cancer record. Item values may be deprecated because they have been superseded in a later version of the abstract, because they belong to a voided abstract, because they are inconsistent with other items deemed to be reliable, or because they have been definitively ruled out during consolidation. Values of a retired item need not but may be marked as deprecated. Use of this field is optional with the facility and its use should is anticipated only with agreement in advance between the sender and the recipient.

Element: Dictionary

Parent: NAACCRFile

Children: ItemDefs (0:n)

The dictionary is an optional element designed for both human and computer use. It is a structured data dictionary which allows single-field validation. It is anticipated that a base version will be distributed along with each NAACCR version for optional customization by state registries. Use of a dictionary is required when items are identified by itemName.

Attribute: dictionaryVersion (optional)

This allows versioning of the dictionary.

Attribute: releaseDate (date, optional)

This is the effective date of the dictionary.

Attribute: description (optional)

This is a description of the dictionary

Element: ItemDefs

Parent: Dictionary

Children: ItemDef (0:n)

This element is used to hold itemDefs for the file. It is anticipated that in the future the dictionary may contain other element types (such as document definitions) at the same level as ItemDefs.

Element: ItemDef

Parent: ItemDefs

Child: ValidValue (0:n)

This element contains information about the NAACCR item.

Attribute: itemName (required, XSD ID)

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

Attribute: itemNo (integer, required)

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

Attribute: dataType (required, enumeration)

This is the data type of the item. This is used for validation and is superseded by the regex attribute if present. 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: regex (optional)

This is a regular expression to be used to validate the data in the field; it is an alternative to datatype, as each datatype has a pre-defined regular expression for validation. For example, NAACCR item RecType (10) could have regex= "^[ICAM]$".

Attribute: parentElement (required, in {Patient, Tumor, CancerRecord})

This indicates whether the item applies to a patient, a tumor, or only to a cancer record; a cancer record will contain items whose parentElement is Patient or Tumor as well as CancerRecord.

Attribute: longName (optional)

This is the NAACCR descriptive name for the item.

Attribute: section (optional)

This is the name of the NAACCR section (e.g., "Demographic") in which the item appears

Attribute: fixedStart (integer, optional)

Attribute: itemLength (integer, optional)

This is the starting column and length of the item in the NAACCR flat file.

Attribute: retiredVersion (optional)

This identifies whether the item was included in a previous NAACCR version but is no longer used, or which has never been implemented. This permits data from such items included in an input file to be identified. The value is the NAACCR version as of which the item's use was discontinued ("v12", "v12.2"), or an indicator that the actual version is not known ("vUnk"). Such items will not be included in an output file without prior agreement between the sender and the recipient.

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.

Attribute: shadow (boolean, optional, default=false)

This identifies whether the item is a "shadow" item, that is, one which is a composite of several underlying items. Shadow items should not be used in data transmissions, but should be available on a read-only basis for edits and other purposes within a registry application.

Element: ValidValue

Parent: ItemDef

This is a method for identifying valid and missing values of an item. An ItemDef can use ValidValue elements to list explicitly its valid and missing values. If present, they may be used for validating the values of the item.

Attribute: code (required)

This is the value of the item used in Item elements.

Attribute: description (optional)

This is a description of the meaning of the value.

Attribute: isMissing (boolean, optional, default=false)

This indicates whether the value is considered missing.

## 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 itemNo=2500 /> or <Item itemNo=2500></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 using the XSD will be limited to assuring that only valid item names are used. Additional single-field validation should be done in the receiving program, such as (1) assuring that a an item appears under the appropriate parent element; (2) assuring that an item appears at most once within a patient or cancer record; (3) assuring that an item is used only in record types for which it is appropriate; (4) assuring that data matches the regex pattern for the item or datatype; and (5) assuring that item values are contained within the ValidValues for the item (if provided). Multi-field validations and required items will be handled by the existing EDITS facility.

Order of Items

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

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 dictionary when they are transmitted;
* changes not be made to the XSD (although the XSD may be included in a more comprehensive XSD)