## Definition of NAACCR XML Data Elements

Element: NaaccrData

Children: 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: metadataVersion (required)

This is the version of the NaaccrMetadata element (q.v. in Metadata Elements section) used to generate the data. It includes 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. 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: 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. Note that the format is YY-MM-DD with all components required.

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: Document, 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. At least one of the attributes naaccrId or naaccrNum must be supplied.

Attribute: naaccrId (optional, NCName)

This is the name of the item, which is based on the NAACCR field name but complies with XSD NCName syntax.

Attribute: naaccrNum (integer, required)

This is the assigned item number of the item.

## Definition of NAACCR XML Metadata Elements

Element: NaaccrMetadata

Children: Dictionary (0:1)

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 previous section). It is anticipated that there will be more child elements than solely the dictionary, such as field lists for profile support. It is anticipated that a base version will be distributed along with each NAACCR version for optional customization by state registries.

Attribute: xsdVersion (optional, fixed=1.0)

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

Attribute: metadataVersion (required)

This allows versioning of the metadata. It should identify the NAACCR version to which it corresponds and any sets of local/requestor/system items it contains.

Attribute: releaseDate (date, optional)

This is the effective date of the metadata. Note that the format is YY-MM-DD with all components required.

Attribute: description (optional)

This is a description of the metadata specification.

Element: Dictionary

Parent: NaaccrMetadata

Children: ItemDef (0:n)

This element contains the complete list of items which may be used in a NaaccrData message.

Element: ItemDef

Parent: ItemDefs

Child: ValidValue (0:n), Description (0:n)

This element contains information about the NAACCR item.

Attribute: naaccrId (required, ID)

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 (required, enumeration)

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 RecType (10) could have regex= "^[ICAM]$".

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

This indicates whether the item applies to a patient or a tumor.

Attribute: naaccrName (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: startColumn (integer, optional)

Attribute: length (integer, optional)

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: 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: IsGroupItem (boolean, optional, default=false)

This identifies whether the item is a group item, that is, one which is a composite of several underlying items as defined in Appendix E to NAACCR volume II. This attribute must be true if it is referenced by the parentId of any ItemDef. Group 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.

Attribute: groupItemId (IDREF, optional)

This identifies the group item of which this item is a member. All items which constitute a group item must have this field populated.

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. The contents of the tag may be formatted text relating to the use of a particular value. It can include CR, LF and tab; it is recommended that XHTML be used for formatting. Use of this element can permit a program to offer the user detailed help messages.

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.

Element: Description

Parent: ItemDef

This is a method for supplying descriptive comments relating to the itemDef. The contents of the description may be formatted text, including CR, LF and tab; it is recommended that XHTML be used for formatting. Use of this element can permit a program to offer the user detailed help messages.

Attribute: topic (optional)

This attribute allows the use of different types of description, such as “definition,” “usage”, “rationale”.

Attribute: sortOrder (integer, optional)

This attribute allows the ordering of description elements within the ItemDef.

## 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 Metadata level; and at the EDITS level. Messages which fail XSD validation will be rejected in their entirety; messages which fail Metadata validation for a particular Patient will reject all data for that 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;
4. metadata attribute values for dataType and parentXMLElement have permissible values; and
5. metadata attribute values for groupItemId reference items defined in the dictionary

Metadata 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;
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; and
5. each Item element has a data value which matches a ValidValue of the ItemDef if ValidValues are provided in the metadata.

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.

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)

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.