NIH NATIONAL CANCER INSTITUTE

OFFICE OF COMMUNICATIONS AND PUBLIC LIAISON

PDQ XML Specification Document

FEBRUARY 26, 2025

1. INTRODUCTION

The National Cancer Institute is distributing the PDQ® (Physician Data Query) information in XML format since October 2002. *Table 1* summarizes the files that are currently provided to distribution partners.

Sample records for all the XML file types are listed in the appendix.

Table 1: PDQ XML Directory Names	
PDQ XML Directory	**Data Description**
DrugInfoSummary	Contains cancer drug information summaries from NCI. The summaries provide consumer-friendly information about cancer drugs and drug combinations.
GlossaryTerm	Contains terms that are part of NCI's Dictionary of Cancer Terms or NCI's Dictionary of Genetics Terms and includes definitions, and may include images and pronunciation audio
Multimedia	Contains files for graphics, illustrations, sound, etc.
Summary	Contains cancer information summaries in the following categories: Treatment, Screening/Detection, Prevention, Supportive and Palliative Care, Genetics, and Complementary and Alternative Medicine (CAM). Most of these summaries (with the exception of Genetics summaries) are provided in two formats. The Health Professional format contains detailed, technical information, and includes links to relevant citations; the Patient format contains information obtained from the Health Professional version, but written at a less technical, lay level, and contains links to glossary terms. Both types of summaries include images and figures.
Terminology	Contains the names, synonyms, and interrelationships of the terms used to index information in the PDQ system.

Please note the following display guidelines:

1) Records in the Terminology files are included for linking to other documents, and do not need to be displayed.

2. DISTRIBUTION

The PDQ XML files are being distributed via SFTP over the Internet.
New dissimination partners will receive an account and authenticate using an ssh-key setup in coordination with the National Cancer Institute (NCI). One user account will be provided for each organization. There will be a primary contact responsible for setting up and maintaining this account.

The data will be available on the NCI SFTP server with the address *CANCERINFO.NCI.NIH.GOV*. The documents are located under the directory */pub/pdq/full*. This directory will contain one directory for each document type listed in *Table 1.* Each of these directories will contain a number of files, each representing the document type indicated by the directory name. The files will be named according to their internal document ID number and each will be a valid XML document according to the DTD provided. The document ID numbers across all the directories will be unique.

For example:

The *GlossaryTerm* directory may contain the following files:
CDR111.xml, CDR115.xml, CDR119.xml, CDR293.xml, …

while the Terminology directory contains the files:
CDR112.xml, CDR114.xml, CDR120.xml, CDR189.xml, …

and the Summary directory contains the files:
CDR113.xml, CDR116.xml, CDR117.xml, CDR118.xml, CDR121.xml, …

Content partners will be able to download these files individually (compressed or uncompressed), download some of these files (uncompressed), or copy the entire contents of a directory as a compressed TAR file.

Please note that the multimedia files provided will be stored as binary files with a *.mp3, *.gif or *.jpg extension instead of being provided as a valid XML document.

In addition to the data directories and the TAR files we're providing a few auxiliary files which can help identifying changes to the set of documents. The *changes*-files (i.e. Summary.YYYYWW with YYYYWW indicating the year and week that the data had been created) identifies all modified documents for a given document type since the last update. The content of the files lists the document name and the identifier of the change (added, modified, or removed). As an example, you might see the content of a *changes-file* like:

```
...
111.xml:added
115.xml:added
119.xml:modified
```

293.xml:modified
323.xml:removed
...

We're also providing the two files *Summary.en* and *Summary.es.* These two files indicate which of the summary files are written in Englisch (EN) or Spanish (ES). The file named *YYYYWW.changes* contains a count of all changes/additions/deletions for the current update cycle.

3. GENERAL DATA FORMAT

PDQ data is distributed in XML (Extensible Markup Language) format. Wikipedia defines **XML** (**Extensible Markup Language**) as "a set of rules for encoding documents in a format that is both human-readable and machine-readable. The W3C's XML 1.0 Specification and several other related specifications—all of them free open standards—define XML." XML is a document description language that describes the structure of documents. PDQ data is accompanied by a Document Type Definition (DTD), a formal definition, of all elements in a document type (file). DTDs are used to validate each document, assuring data integrity and accurate record structure.

XML describes document structures with markup language. An XML document can either be well-formed or well-formed and valid. Valid XML documents must conform to XML specifications and an associated DTD. NCI will be distributing valid XML documents and a DTD for each document type (file).

XML documents include several components:

An XML processing instruction that identifies the version of XML being used, the way it is encoded, and whether it references other files:

```
<?xml  version='1.0' encoding='UTF-8'?>
```

A document type declaration typically contains the formal markup declarations in its internal subset, or references an external file containing the relevant markup declarations (DTD). The specification of a (internal or external) DTD is optional. However, the name specified in the DOCTYPE declaration must match the root element of the document:

```
<!DOCTYPE Country>
```

As an example, a fully-tagged (country) document, including a root element and all other markup nested within the root element could look like this:

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Country >
  <Country id="CDR0000043753">
  <CountryFullName>U.S.A.</CountryFullName>
  <Continent>North America</Continent>
  <PostalCodePosition>after PoliticalSubUnit_State</PostalCodePosition>
</Country>
```

For more information on XML, please review the World Wide Web Consortium (W3C) specification for XML:

Extensible Markup Language (XML) 1.0 (Fifth Edition) W3C Recommendation 26 November 2008.

http://www.w3.org/TR/2008/REC-xml-20081126/

4. COMMON ELEMENTS

PDQ XML documents may contain common elements, defined at the beginning of the DTD. These elements describe common data structures such as a paragraph, a person, or organization location. *Table 2* summarizes the common top-level elements. *Table 3* summarizes inline paragraph markup.

Table 2: Common Data Elements	
PDQ XML DTD	**Comments**
Section	
Title	Defined by *%TitleData* entity
AltTitle	Defined by *%TitleData* entity
KeyPoint	Defined by *%ParaElements* entity
Para	Defined by *%ParaElements* entity
LiteralLayout	Defined by *%ParaElements* entity
Table	
Title	Defined by *%TitleData* entity
TitleAbbrev	Defined by *%TitleData* entity
Tgroup	See DTD
ItemizedList	
ListTitle	Defined by *%TitleData* entity
ListItem	Defined by *%TextElements* entity
OrderedList	
ListTitle	Defined by *%TitleData* entity
ListItem	Defined by *%TextElements* entity
QandASet	
MarkedUpTitle	Defined by *%TitleData* entity
QandADiv	See DTD
QandAEntry	See DTD
Contact	
ContactName	PCDATA
ContactDetail	Defined by *%Location* entity
Section	See above
Diagnosis	
SpecificDiagnosis	PCDATA
DiagnosisParent	PCDATA
PostalAddress	
Street	PCDATA
City	PCDATA
CitySuffix	PCDATA
PoliticalSubUnitName	PCDATA
CountryName	PCDATA
PostalCode_ZIP	PCDATA
PostalCodePosition	PCDATA
Phone	PCDATA
TollFreePhone	PCDATA
Email	PCDATA
WebSite	PCDATA

Table 2: Common Data Elements	
PDQ XML DTD	**Comments**
MainTopics	
TermRef	PCDATA
SecondaryTopics	
TermRef	PCDATA
ReferenceSection	
Citation	Defined by *%ParaElements* entity
DateLastModified	PCDATA
DateFirstPublished	PCDATA
ProfessionalDisclaimer	Defined by *%TextSectionElement* entity
MediaLink	Defined by *%TextSectionElement* entity
Caption	Defined by *%TextSectionElement entity*
MediaRef	PCDATA

Table 3: Recommended Display of Inline Markup in CDR Documents	
Tag	**Rendering**
Emphasis	*Italics*
Strong	**Bold**
Superscript	T^1
Subscript	T_2
TT	`Fixed space font with preserved linebreaks`
GeneName	*Italics*
ScientificName	*Italics*
DrugName	No special formatting
ForeignWord	*Italics*
Reference	Underlined and linked to the References section (e.g. [1])
SummaryRef	Underline
ProtocolRef	Underline
GlossaryTermRef	Underline
LOERef	Underline
ExternalRef	Underline
DrugRef	Underline
Contact (in text)	Indented as follows: National Cancer Institute Office of Communications 31 Center Drive, MSC 2580 Bethesda, MD 20892-2580
ItemizedList	Display depends on the attribute values:
Compact = No	If *Compact=No* exists, include an additional line break between list items
Style = simple	No visual character/graphic before each ListItem

Table 3: Recommended Display of Inline Markup in CDR Documents	
Tag	**Rendering**
Style = bullet	Bullets before each ListItem: • XYZ
Style = dash	Dashes before each ListItem: – ABC – EFG
ListTitle	*Italics*
OrderedList	Display depends on the attribute values:
Compact = No	If *Compact=No* exists, include an additional line break between list items
Style = Arabic	a) xxx b) yyy
Style = Ualpha	A. This is a test B. This is not a test
Style = Lalpha	a. John Doe b. Joe Smith
Style = Uroman	I. Albert Einstein II. Galileo Galilei
Style = Lroman	i. James Bond ii. Austin Power
ListTitle	*Italics*

5. GUIDE FOR PDQ XML ELEMENTS BY DOCUMENT TYPE

Summary

The PDQ cancer information summaries are peer-reviewed, evidence-based summaries on topics including adult and pediatric cancer treatment, supportive and palliative care, screening, prevention, genetics, and complementary and alternative medicine. Most of the summaries are available in two different formats:

- Health Professional versions provide detailed information written in technical language and are fully referenced with links to PubMed abstracts.
- Patient versions are written in lay language and include links to the NCI Dictionary of Cancer Terms. Many patient summaries also include illustrations.

Many of the summaries are also available in Spanish. For more information, please see http://www.cancer.gov/types.
Dissemination partners can categorize the summaries by type and by audience using the *SummaryType* and *SummaryAudience* data elements. For example, all treatment summaries will have a *SummaryType* value of *Treatment* and a *SummaryAudience* value of *Health professionals* or *Patient***.** Each summary will also include links to terminology.

In addition, the *SummarySection* element identifies subsections of a summary. The *SectionMetaData* element provides information about the section, such as the types of information included in the section, e.g, prognosis; as well as the type of cancer.

Table 4 describes the data elements for the PDQ XML Summary in the DTD.

Table 4: Summary Data Elements

PDQ XML DTD	Comments
SummaryMetaData	Metadata about the summary to support search andretrieval, as well as categorization of summaries.
SummaryType	Treatment, Supportive Care, Screening, Prevention, and Genetics,Complementary and alternative medicine.
SummaryAudience	Patient summaries have a SummaryAudience value of *Patients*; Health Professional summaries have a SummaryAudience value of *Health professionals*.[1]
SummaryLanguage	Values for the SummaryLanguage element are *English* and *Spanish*.
SummaryDescription	This element is mainly NCI use on its Web site
SummaryURL	This elementcontains the URL for the summary on NCI's Cancer.gov Web site. Partners may want to use this to link to the NCI Web site to ensure that their users always have access to the most current version of the summary..
MainTopics (L)[2]	This element contains optional multiply occurring

[1] At present, patient summaries are not available for Genetics
[2] Elements followed by an (L) are linked elements. These elements will have an *id* Attribute that allows partners to obtain detailed information from linked source documents.

Table 4: Summary Data Elements	
PDQ XML DTD	**Comments**
	TermRef elements that describe the main topics of the summary...
TermRef	This element includes the ID and the term name from the PDQ Terminology records. For treatment summaries, the type of cancer term is included in this element.
SecondaryTopics (L)	This element identifies any additional topics that the summary may be related to..
TermRef	This element includes the ID and the term name from the PDQ Terminology records.Like the Main Topic element, the text content is derived from the preferred name of the Term record
SummaryAbstract	Short abstract with information about the content of the summary.
SummaryKeyWords	Parent element for individual keyword elements.
SummaryKeyWord	Keywords relevant for the current summary.
SummaryTitle	Defined by *%TitleData* entity
AltTitle	Alternate title with attributes *TitleType* and valid values *Short* and *Display*.
SummarySection	Each summary can have many sections and subsections.
SectionMetadata	Information about the section, such as the types of information included in the section, e.g, prognosis, as well as the type of cancer is included in this set of elements.
SpecificDiagnosis (L)	The text content of this element names the specific cancer or cancer-related condition that is discussed in the section. For example. the Treatment for Stage I Breast cancer section of the Breast cancer treatment summary will contain the term "stage I breast cancer'. Metadata can be useful to enhance search and organization of content. The term used here is also from the PDQ terminology file and the Ref attribute can be used to obtain more information from the Terminology file.
SectionType	The data in this element identifies the kind of information provided in the section. Treatment summaries consistently use SectionType metadata. Values for this element include "Treatment options by stage", "Treatment options for recurrent cancer", "Current clinical trials"
Title	Section title
AltTitle	Alternate title. New attribute *TitleType* with valid values of *Short* and *Display*.
KeyPoint	Data element summarizing key points made in the

Table 4: Summary Data Elements	
PDQ XML DTD	**Comments**
	section. Preferably display in a key points box.
SummarySection	Subsections may be recursively nested.
ReferenceSection	This element contains one or more citations that are referenced in the text. Citations may be to records in the National Library of Medicine's PUBMED database or to records in the Protocol or CTGOV file, or to Web sites and Web pages. The citations are formatted using a standard style derived from the American Medical Association Style Guide. Citations are typically used in health professional summaries only. There is one ReferenceSection for each top-level SummarySection. Attributes include PMID (PubMed ID), MEDLINEID, and ProtocolID to facilitate linking to PubMed for abstracts, and to PDQ protocols for protocol abstracts.
Citation	
TranslationOf	The element will link to the original version of the translated summary; used only for Spanish summaries
ReplacementFor	The element, if present, links to the document ID of the document that had been replaced by earlier updates. Multiple occurrences of replacements will be concatenated and space separated. The element has been deprecated.
PatientVersionOf	This element provides a link to the health professional version of the summary
DateLastModified	Date the summary had editorial changes made
DateFirstPublished	
ProfessionalDisclaimer	For health professional summaries only

In addition to structured data elements, cancer information summaries will be progressively enhanced with semantic and formatting markup in text sections. This markup can be used to format display formatting and to enhance search and retrieval. See *Table 3* for inline markup recommendations. (These inline markup elements may be used in other CDR document types also)

SummaryRef and SummaryLink: Summaries may contain references to subsections within the summary or to other summaries, using the *SummaryRef* and *SummaryLink* elements. These need to be displayed as clickable links to the summaries in a Web display. The *href* and *ref* attributes can be used to create links between summaries. The *SummaryRef* element also contains an attribute named *url* that contains the path of a referenced summary on the National Cancer Institute's website which may also be used to create a link pointing to the document on the NCI website.

For example, given the following SummaryRef element

```
<SummaryRef href="CDR0000062687#_331"
url="/types/colorectal/hp/colon-treatment-pdq">Table
5</SummaryRef>
```

the following link pointing to the patient version of the NCI's adult brain tumors summary can be created.

https://www.cancer.gov/types/colorectal/hp/colon-treatment-pdq#link/_331

GlossaryRef: Summaries may also contain links to the GlossaryTerm file using *GlossaryRef* elements. It is recommended that these be shown as clickable links that provide the user with a definition of the term.

MediaLink and ***MediaRef*** elements should be replaced with the referred media document.

LOERef: The PDQ Editorial Boards that maintain the health professional summaries use a formal ranking system of levels of evidence to help the reader judge the strength of evidence linked to the results being discussed in the summary. These are represented in the text within the *LOERef* element. The Attribute on this element allows links to terms in the GlossaryTerm file. These terms have a definition with a "health professional" attribute and are not appropriate for inclusion in a dictionary for lay audiences.

Terminology

The XML Terminology file contains one logical record for each term. The file includes terms for drugs, drug combinations, diagnoses, cancer stages and cellular types, therapy modalities, and genetic conditions. This information is listed in the SemanticType element. Semantic types are defined for PDQ and are provided to dissimination partners. The component drugs are subsumed by the TermRelationship element, where the combination terms have the specific drugs in the *RelatedTermName* element with the *RelatedTermType* value of *Has component*.

For example, the record for *cyclophosphamide/doxorubicin/fluorouracil* will have *doxorubicin*, *cyclophosphamide*, and *fluorouracil* listed as related terms, with the RelatedTermType of "*Has component"*.

A terminology record may also contain a *MenuInformation* element. This information could be used for dynamically generating menus and is intended to improve indexing and searching and retrieval. These menus will be used for NCI's Cancer.gov Website and dissimination partners may choose to use them for a similar purpose. Initially all the menu information will be related to search and retrieval of clinical trials.

Table 5 describes the data elements for the PDQ XML Term element in the DTD.

Table 5: Terminology Data Elements	
PDQ XML DTD	**Comments**
PreferredName	
OtherName	Combines SNAM and SYN; provides additional semantic detail
OtherTermName	
OtherNameType	Defines the type of OtherName. Valid values include *Acronym, Synonym, Lexical variant, Abbreviation, Common usage, Related string, Foreign brand name, US brand name, Obsolete name, Spanish, Subtype, Broader*
Definition	
DefinitionText	
DefinitionType	
TermTypeName	Indicates term usage in PDQ. Values include *Index term* (terms used to index data), *Header term, Protocol selection criteria, Semantic type, Obsolete term*.
SemanticType	Some old values in Type have been remapped. Table 6 lists all the Semantic Types used in PDQ.
TermRelationship	Wrapper for Parent and Related Terms (no child relationships)
ParentTerm	
ParentTermName	
ParentType	ISA
RelatedTerm	

Table 5: Terminology Data Elements	
PDQ XML DTD	**Comments**
RelatedTermName	
RelationshipType	Data element with values including *Has component, Associated gene, Associated genetic condition, Related protocol selection criteria*
RelatedWebsites	
MenuInformation	Optional element
MenuItem	Wrapper element for MenyType, MenuParent, DisplayName.
MenuType	Values are one of *Clinical Trials—CancerType, ClinicalTrials—Drug,* or *Cancer Information*.
MenuParent	Link information to Term Menu parent. This may not always be the same as the TermParent. Element will not exist for top level menus, e.g. cancer.
DisplayName	Optional element. Use *PreferredName* for display if this element does not exist.
DateFirstPublished	
DateLastModified	

Table 6: Semantic Types
PDQ XML Semantic Types
Disease/diagnosis
Cancer diagnosis
Cancer stage
Cancer grade
Genetic condition
Secondary related condition
Intervention/procedure
Therapeutic intervention/procedure
Cancer therapy modality
Supportive care modality
Diagnostic test/procedure
Preventative test/procedure
Drug/Agent category
Drug/Agent
Drug/Agent combination
Gene

Glossary

The XML Glossary file makes the Cancer.gov glossary/dictionary available in its entirety to our partners. The glossary/dictionary file can be used as a standalone file and a support file for glossary links in Summary documents.

PDQ partners who wish to display a dictionary/glossary of cancer terms for lay audiences should only use those terms that do not have the attribute *Dictionary*="Exclude" and should not pick up TermDefinitions with an attribute of *Audience*="HealthProfessional".

Table 7 describes the data elements in the PDQ XML Glossary/Dictionary DTD.

Table 7: Glossary Data Elements

PDQ XML DTD	Comments
TermName	
TermPronunciation	
TermDefinition	
DefinitionText	Most glossary terms have definitions geared towards consumers (lay audiences). In addition, there may be some terms that have a definition for a health professional audience. Definitions with the attribute of "Audience = HealthProfessional" should not be displayed in a consumer-oriented dictionary.
Dictionary	
Audience	
TermRef	
MediaLink	Defined by *%TextSectionElement* entity
EmbeddedVideo	
SpanishTermName	Initially, this element will not be available. Data may be available in the future.
SpanishTermDefinition	Initially, this element will not be available. Data may be available in the future.
DefinitionText	Most glossary terms have definitions geared towards consumers (lay audiences). In addition, there may be some terms that have a definition for a health professional audience. Definitions with the attribute of "Audience = HealthProfessional" should not be displayed in a consumer-oriented dictionary.
Dictionary	
Audience	
DateFirstPublished	
DateLastModified	
RelatedInformation	Wrapper element for links to related information
RelatedExternalRef	Link to an external Internet resource. The *UseWith* attribute specifies the language of the glossary term

Table 7: Glossary Data Elements	
PDQ XML DTD	**Comments**
	entry with which the resource is to be used. For example a link to a Web page in English may have *UseWith='en'* as an attribute value and the link would be presented with the English term definition.
RelatedSummaryRef	Link to an internal PDQ summary. The *UseWith* attribute specifies the language of the glossary term entry with which the summary is to be used. For example a link to a summary in English may have *UseWith='en'* as an attribute value and the link would be presented with the English term definition.
RelatedDrugSummaryRef	
RelatedGlossaryTermRef	

Multimedia

PDQ® is enhancing its content with multimedia (graphics, illustrations, audio etc). PDQ® documents, specifically summaries and glossary terms, include images such as anatomic illustrations, pictures of diagnostic and therapeutic procedures, diagrams, photographs, Xray photos, etc. and audio for the pronunciation of glossary terms. Eventually, PDQ may also include other types of multimedia such as video.

Three types of information will be published:

- Media image files
  Images will be published as GIF or JPEG/JFIF encoded files.
- Media audio (pronunciation) files
  Audio files will be published as MP3 encoded files.
- Links from XML to media files.
  Some documents will contain XML elements that refer to the Media file images.

These are described below.

Media Files

Images and audio are published as free standing binary files, in GIF, JPEG/JFIF, or MP3 encoded format.

Each media file is named with a unique file name consisting of a CDR document ID and a file type extension. The format of the name is:

CDRnnnnnnnnnn.jpg
or
CDRnnnnnnnnnn.gif
or
CDRnnnnnnnnnn.mp3

where “nnnnnnnnnn” is a 10-digit number, zero filled on the left, for example:

CDR0000012345.jpg
CDR0000386420.gif
CDR0000645789.mp3

New images produced for use in PDQ publications will often be produced in files with relatively large, high quality formats – the best quality JPEG or GIF format. The pixel dimensions of these images may be too large for inclusion in a Web page, and the file sizes may be too large for fast retrieval by users with slow connections to the Web.

These images are produced and distributed to licencees in these large, high-quality formats so that they may be used for other purposes besides Web display, and may be

easily re-used to make derivations. For example, the NCI Web site, where PDQ® information is presented to users will down-size (reduce the width and height), down-sample (reduce the number of pixels per inch), and compress (reduce the number of bytes required to encode the image) the images for display on our Web site.

Our partners can use any number of commercial or open source software applications to perform resizing/sampling/compression – either with individual images or in batches. It is expected that partners will not make other modifications to these images if they are being used in PDQ® content.

XML Links to Images

Links from an XML document to an image are made using a *MediaLink* element. MediaLink is a "*TextElement*", defined in the common data elements included in all document types. It can appear anywhere that a TextElement can appear, in any document type. Currently, MediaLinks are included in Summary and GlossaryTerm document types. There are no plans at this time to include them in other document types, but our document type definition permits them to appear anywhere and it is possible that some other existing document types, or future document types, may include them.

The media element structures are described in the following table.

Table 8: Media Data Elements

PDQ XML DTD	Comments
MediaLink	Link to a media file.
ref attribute	Required reference to the image object. The reference is a CDR ID, without a trailing ".jpg", ".gif", or ".mp3" extension. Example: ref="CDR0000387201"
type attribute	Optional mime type of the document content. Values are "audio/mpeg", "image/jpeg", or "image/gif".
alt attribute	Required alternative text, used in constructing HTML alt attributes.
inline attribute	Optional indication that an image is to be displayed as inline text. Might be used in some very exceptional case where a symbol must be inserted into text as if it were a character, but no Unicode character is available for it. Values are "Yes" or "No", with a default of "No" if no "inline" attribute is present.
language attribute	Optional attribute indicating the spoken language of and audio file or the language of any text that appears superimposed on (embedded in) the image. Valid values are ISO 2 or 3 letter language codes,

	such as “en”, “es”, “fr”, etc. If there is more than one language in an image, multiple codes may be specified with space separators. The default language for an image if no language attribute is present is the language of the XML document in which the MediaLink appears. In an English document, the default language for images is “en”. In a Spanish document, the default is “es”. A non-default may be used if the CDR author believes it is important to indicate that the language in an image is different from the language of the document. This should rarely occur, if ever.
id attribute	Optional target identifier for internal hyperlinks from elsewhere in the document via a *MediaRef* element. An id attribute must always be unique within a document. There is no default id.
Caption subelement	Optional caption describing an image. The caption is intended for display with, e.g., below, the image. It is a TextElement that may contain other TextElements. For example, a Caption may contain a GeneName, a GlossaryTermLink, a DrugName, a superscript, etc. Captions will be present on most, if not all images, but are not required and may not be needed in some circumstances.
language attribute	Optional language of the Caption. It is like the language attribute for the MediaLink. The default value is like that of MediaLink, i.e., the language of the enclosing document.
MediaRef element	A reference from inside a document to a MediaLink element in the same document. This would typically be used if text in the document refers to an image, for example: “See <MediaRef href=’F3’>Figure 3</MediaRef>” MediaRef is a TextElement that may appear anywhere in a document.
href attribute	Required reference to the “id” attribute of a MediaLink element in the same document.

DrugInformationSummary

The PDQ Drug Information Summaries provide consumer-friendly information about cancer drugs and drug combinations.

Summaries for individual cancer drugs cover the uses of these drugs, research results, possible side effects, approval information, and ongoing clinical trials. The list includes brand and generic names for the drugs.

Summaries for cancer drug combinations are listed by abbreviation or common name and are shown in capital letters. Each summary gives a list of the drugs that make up the combination and explains what the combination is used for. It also has links to summaries for individual drugs in the combination.

Table 9 describes the data elements in the PDQ XML Terminology DTD.

Table 9: DrugInformationSummary Data Elements

PDQ XML DTD	Comments
DrugInfoMetaData	Contains elements and attributes that describe data about the drug info summaries
DrugInfoType	Valid values: Brief, Detailed Note: Currently, there do not exist any 'Detailed' drug information summaries. Combination attribute 'Yes' to indicate a drug combination summary
DrugInfoAudience	Valid values: Patients, Health professionals The element describes the intended audience for drug information summaries. At this time, we only have summaries with a value of 'Patients'.
DrugInfoDescription	Text element for describing the purpose of the drug information summary. Sample text: This page contains brief information about bevacizumab and a collection of links to more information about the use of this drug, research results, and ongoing clinical trials.
DrugInfoURL	Element that contains the link title for the url for use on the Cancer.gov Web site, with the url as an attribute.
Manufacturers	
Manufacturer	Optional element to list manufacturer of drug. This element links to an organization. The element is currently not used.
FDAApproved	Element for indicating FDA approval status of a drug or a drug combination. Valid values: Yes, No, Combination approved, Individual drugs approved

Table 9: DrugInformationSummary Data Elements	
PDQ XML DTD	**Comments**
TerminologyLink	Link to the drug record in the terminology file.
GlossaryLink	Link to the drug record in the glossary file.
USBrandNames	
USBrandName	Element listing US Brand Names for drug. This element is currently not used.
Synonyms	
Synonym	This element is currently not used.
PronunciationInfo	
TermPronunciation	Pronunciation text
MediaLink	Pronunciation audio
FDAExternalRef	
DrugInfoTitle	Title of the DIS.
Section	
DrugInfoDisclaimer	Section elements Note telling user that the information is educational only and should be used in consultation with a health care professional.
DateFirstPublished	Date the DIS is first published to Cancer.gov Web site. This date is system generated.
DateLastModified	Date the DIS has changes made to it. This is a manually entered date.

6. SUPPORT INFORMATION

For questions on the administrative aspects of the PDQ XML data change, please contact the NCI Content Dissemination Services by e-mail at:

ncicontentdissemination@nih.gov

If you have technical questions about the conversion process to XML, or questions related to accessing the data, please contact our technical staff at:

NCIPDQoperator@mail.nih.gov

Responses to all inquiries will generally be provided within 48 hours.

This document, along with PDQ XML sample data and the PDQ DTD file are available on the SFTP server:

cancerinfo.nci.nih.gov

under the directory
/pub/pdq/docs

APPENDICES

Appendix A – Sample Document Transformation

The documentation directory */pub/pdq/docs* on the sFTP server contains a sample of a PDQ summary XML file, a sample Summary XSLT stylesheet, and the HTML output that the transformed XML file will produce.

There are several options you could us to run the example yourself on your computer an create the resulting HTML output.
Please note that the sample document contains a limited number of CSS styles. If you were to compare the resulting HTML document with one shown on displayed on the Cancer.gov website, the look and feel will be very different. The sample is intended to demonstrate the transformation from XML to HTML. For additional formatting changes you would want to create your own CSS stylesheets.

a) **Using a XSL processor**
   You could create the HTML output if you have access to an XSL processor - a software component that implements the XSL standard. The stylesheet (*PDQ-summary.xsl)* is applied to the sample summary (*PDQ-summary.xml*) using your processor and converted into the HTML document (*PDQ-summary.html*).
b) **Using a web browser**
   Alternatively to using an XSL processor, a web browser like Firefox will be able to apply the XSL transformation of the XML document. The following line has been added to the XML document which instructs the browser to use the named XSL stylesheet to convert the document:
   ```
   <?xml-stylesheet href="PDQ-summary.xsl" type="text/xsl" ?>
   ```
   Open the XML file in your browser – after copying both, the XML and the XSLT file to the same directory on your computer – with the following URL:
   `file:///<path-to-file>/PDQ-summary.xml`

   You would need to remove the above line from the XML document if you want to see the unprocessed, raw XML input file.

   Note that not all browsers allow to display a local XML file with a XSL stylesheet due to security concerns and Firefox only allows this after adjusting the "privacy.file_unique_origin" configuration setting. It should be possible, though to copy the files to a location that can be served by a web server. Accessing the files over HTTP will likely work as well.
c) **Using Python 3**
   Python 3 is able to run a local webserver that allows you to perform the steps that may be blocked by the web browser when following option (b).
   A. Copy the following files to a directory on a computer with Python 3 installed: The summary content (PDQ-summary.xml), the summary stylesheet (PDQ-summary.xsl), and favicon.ico.

B. In the directory with these files are located, run the Python command

```
python3 -m http.server
```

You will see the following output on your screen

```
Serving HTTP on :: port 8000 (http://[::]:8000/) ...
::1 - - [26/Feb/2025 15:05:05] "GET /PDQ-summary.xml HTTP/1.1"
200 -
::1 - - [26/Feb/2025 15:05:05] "GET /PDQ-summary.xsl HTTP/1.1"
200 -
```

C. Open the link below in you favorite browser to view the resulting HTML document:

```
http://localhost:8000/PDQ-summary.xml
```

Appendix B – Table Output Rendering Recommendations

The following recommendations are for *all* tables in *all* doctypes in *all* XSLT outputs.

This document has been modified to correct and amplify the description of how display of gridlines and frames for tables in PDQ documents should be handled. The specific changes include:

- Dropping of references to the presence of the ColSep attribute in the Row element (the ColSep attribute is not allowed for this element)
- Inclusion of instructions for handling ColSep and RowSep attributes in the ColSpec element
- Expansion of the documentation to describe in more explicit detail how the Frame, ColSep, and RowSep attributes interact with each other and how conflicting values of these attributes are resolved on elements at different levels of the table hierarchy.

Style Defaults

The following defaults should be used for outputs when no relevant attribute values are present.

Table Font

The font-size in a <Table> (all elements except the <Table><Title>) should be in a relatively smaller size (minus 1 or 2 points) than the font-size used for the text of the <Para> immediately preceding the <Table> (or used in the parent <SummarySection>element in a Summary document). Default to the same font-family as the preceding <Para> or the parent <SummarySection>.

Tfoot

Render in italics.
Vertical alignment should default to “Top”; overridden by the presence of the attribute, <Tfoot><Row><entry Valign= >.
Horizontal alignment should default to “Left”; overridden by the presence of the attribute, <Tfoot><Row><entry Align= >.

Thead

Render in boldface.
Vertical alignment should default to “Middle”; overridden by the presence of the attribute, <Thead><Row><entry Valign= >.
Horizontal alignment should default to “Center”; overridden by the presence of the attribute, <Thead><Row><entry Align= >.

Citation Links

Continue to number references in the sequence established within the parent <SummarySection> in a Summary document.

Glossary Links

Text should be underlined.

Title

Render in boldface. Font-size should be the same as the preceding <Para> or the parent <SummarySection> in a Summary document.

Table

Include one blank line after a table and before the next paragraph level element.

Tbody

Vertical alignment should default to “Top”; overridden by the presence of the attribute, <Tbody><Row><entry Valign= >.
Horizontal alignment should default to “Left”; overridden by the presence of the attribute, <Tbody><Row><entry Align= >.
Left-justified text should be indented 5 pts. From the left gridline.

Other Formatting Considerations

Frame

All tables should be framed in a 1-point line; overridden by the <Table Frame= > attribute.

Gridlines

All entries should be bordered with 1-point gridlines; superceded by the Frame attribute; overridden by the ColSep and RowSep attributes as described below.

For all outer edges of a table, the decision whether to display the bordering gridline is determined by the value of the Frame attribute of the <Table> element. The possible values for Frame are:

- All [the default]
- Bottom
- Sides
- Top

- TopBot
- None

For each entry which is not at the rightmost edge of the table, the decision whether to display the gridline at the right edge of the entry's cell is determined as follows.

- If the <entry> element itself has an explicit (non-empty) ColSep attribute, then the gridline is suppressed if the value of the attribute is zero ('0'), and displayed if the value is one ('1').
- Otherwise, if the <ColSpec> element corresponding to the rightmost column of the entry has an explicit ColSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- Otherwise, if the enclosing <Tgroup> element has an explicit ColSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- Otherwise, if the enclosing <Table> element has an explicit ColSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- If none of these elements has an explicit ColSep attribute, then the gridline is displayed.

For each entry which is not at the bottom edge of the table, the decision whether to display the gridline at the bottom edge of the entry's cell is determined as follows.

- If the <entry> element itself has an explicit (non-empty) RowSep attribute, then the gridline is suppressed if the value of the attribute is zero ('0'), and displayed if the value is one ('1').
- Otherwise, if the <Row> element corresponding to the lowest row of the entry has an explicit RowSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- Otherwise, if the <ColSpec> element corresponding to the rightmost column of the entry has an explicit RowSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- Otherwise, if the enclosing <Tgroup> element has an explicit RowSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- Otherwise, if the enclosing <Table> element has an explicit RowSep attribute, then the gridline is suppressed if the value of the attribute is zero, and displayed if the value is one.
- If none of these elements has an explicit RowSep attribute, then the gridline is displayed.

Column Spanning

An <entry> can span multiple columns. Use the NameSt= attribute to determine the leftmost column in the spanned columns; use the NameEnd= attribute to determine the rightmost column in the span. Spanned columns do not have internal gridlines.

Row Merging

An <entry> can be merged with other rows. Use MoreRows= attribute to determine how many more rows, in addition to the current row, this entry should occupy. Merged rows do not have internal gridlines.

Column Width

Use the <ColSpec ColWidth=> attribute to size columns appropriately. Typically, we will be using proportional sizing.

For example:

<ColSpec ColNum="1" ColName="col1" ColWidth="1*"/>
<ColSpec ColNum="2" ColName="col2" ColWidth="2*"/>

Column #2 is twice the width of Column #1.

For example:

<ColSpec ColNum="1" ColName="col1" ColWidth="1*"/>
<ColSpec ColNum="2" ColName="col2" ColWidth="3*"/>
<ColSpec ColNum="3" ColName="col3" ColWidth="2*"/>

Column #3 is twice as wide as Column #1; Column #2 is three times as wide as Column #1.

For example:

<ColSpec ColNum="1" ColName="col1" ColWidth="*"/>
<ColSpec ColNum="2" ColName="col2" ColWidth="2*"/>
<ColSpec ColNum="3" ColName="col3" ColWidth="*"/>

Column #2 is twice as wide as either Column #1 or Column #3.

Occasionally, you may find an absolute column width in points, inches, centimeters, millimeters, or picas. For example:

<ColSpec ColNum="1" ColName="col1" ColWidth="2.5in"/>
<ColSpec ColNum="2" ColName="col2" ColWidth="*"/>
<ColSpec ColNum="3" ColName="col3" ColWidth="*"/>

Column #1 is 2.5 inches wide. Columns #2 and #3 occupy the remaining space, split equally between the two columns.

Summary

Element	Attribute	Default Value
entry	NameSt	
entry	NameEnd	
entry	MoreRows	0
ColSpec	ColWidth	*
Tbody/Row/entry	Align	Left
Tbody/Row/entry	Valign	Top
Tfoot/Row/entry	Align	Left
Tfoot/Row/entry	Valign	Top
Thead/Row/entry	Align	Center
Thead/Row/entry	Valign	Middle
Table	Frame	All
Table Tgroup ColSpec Entry	ColSep	1
Table Tgroup ColSpec Row entry	RowSep	1

Component	Style	Default
CitationLink	Numbering	Number within <SummarySection> sequence
GlossaryTermLink	Typeface	Underlined
Table	font-family	Preceding Para font
Table	font-size	Smaller than preceding Para font size (-1 or 2 pts)
Table	Spacing	1 blank line after a table
Table/Title	font-family	Same as preceding Para font
Table/Title	font-size	Same as preceding Para font size
Table/Title	typeface	Bold
Tfoot	typeface	Italics
Thead	typeface	Bold
Tbody/Row/entry	indent if Align=Left	5 pts.

Appendix C – XML Sample Records

Summary

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Summary (View Source for full doctype...)>
- <Summary id="CDR0000062955">
  - <SummaryMetaData>
      <SummaryType>Treatment</SummaryType>
      <SummaryAudience>Patients</SummaryAudience>
      <SummaryLanguage>English</SummaryLanguage>
      <SummaryDescription>Expert-reviewed information summary about the treatment of
        breast cancer.</SummaryDescription>
      <SummaryURL
        xref="http://cancer.gov/cancertopics/pdq/treatment/breastCancer/HealthProf
        essional">Breast Cancer</SummaryURL>
    - <MainTopics>
        <TermRef ref="CDR0000038832">breast cancer</TermRef>
      </MainTopics>
    </SummaryMetaData>
    <SummaryTitle>Breast Cancer</SummaryTitle>
  - <SummarySection id="_125">
      <Title>General Information About Breast Cancer</Title>
    - <SummarySection id="_126">
      - <SectMetaData>
          <SectionType>Cancer description</SectionType>
          <SectionType>Organ description</SectionType>
        </SectMetaData>
        <KeyPoint id="_127">Breast cancer is a disease in which malignant (cancer) cells
          form in the tissues of the breast.</KeyPoint>
      - <MediaLink ref="CDR0000999999" alt="breast anatomy drawing" inline="No"
          thumb="No">
          <Caption>Breast Anatomy Drawing</Caption>
        </MediaLink>
      - <Para id="_128">
          The breast is made up of
          <GlossaryTermRef href="CDR0000046188">lobes</GlossaryTermRef>
          and
          <GlossaryTermRef href="CDR0000046441">ducts</GlossaryTermRef>
          . Each breast has 15 to 20 sections called lobes, which have many smaller
          sections called
          <GlossaryTermRef href="CDR0000046308">lobules</GlossaryTermRef>
          . Lobules end in dozens of tiny bulbs that can produce milk. The lobes, lobules,
          and bulbs are linked by thin tubes called ducts.
        </Para>
      - <Para id="_129">
          Each breast also has
          <GlossaryTermRef href="CDR0000045020">blood vessels</GlossaryTermRef>
          and
          <GlossaryTermRef href="CDR0000269462">lymph vessels</GlossaryTermRef>
          . The lymph vessels carry an almost colorless fluid called
```

<GlossaryTermRef href="CDR0000046305">lymph</GlossaryTermRef>
. Lymph vessels lead to
<GlossaryTermRef href="CDR0000257523">organs</GlossaryTermRef>
called
<GlossaryTermRef href="CDR0000045762">lymph nodes</GlossaryTermRef>
. Lymph nodes are small bean-shaped structures that are found throughout the body. They filter substances in lymph and help fight
<GlossaryTermRef href="CDR0000045364">infection</GlossaryTermRef>
and disease. Clusters of lymph nodes are found near the breast in the
<GlossaryTermRef href="CDR0000046510">axilla</GlossaryTermRef>
(under the arm), above the collarbone, and in the chest.
</Para>
- <Para id="_130">
The most common type of breast
<GlossaryTermRef href="CDR0000045333">cancer</GlossaryTermRef>
is
<GlossaryTermRef href="CDR0000045085">ductal carcinoma</GlossaryTermRef>
, which begins in the
<GlossaryTermRef href="CDR0000046476">cells</GlossaryTermRef>
of the ducts. Cancer that begins in the lobes or lobules is called lobular carcinoma and is more often found in both breasts than are other types of breast cancer.
<GlossaryTermRef href="CDR0000045313">Inflammatory breast cancer</GlossaryTermRef>
is an uncommon type of breast cancer in which the breast is warm, red, and swollen.
</Para>
</SummarySection>
- <SummarySection id="_140">
- <SectMetaData>
<SectionType>Diagnostic tests</SectionType>
</SectMetaData>
<KeyPoint id="_141">Tests that examine the breasts are used to detect (find) and diagnose breast cancer.</KeyPoint>
<Para id="_142">A doctor should be seen if changes in the breast are noticed. The following tests and procedures may be used:</Para>
- <ItemizedList id="_143" Style="bullet">
- <ListItem>
<GlossaryTermRef href="CDR0000045775">Mammogram</GlossaryTermRef>
: An
<GlossaryTermRef href="CDR0000045944">x-ray</GlossaryTermRef>
of the breast.
- <MediaLink ref="CDR0000999998" alt="photo of mammography" inline="No" thumb="No">
<Caption>Photo of Mammography</Caption>
</MediaLink>
</ListItem>
- <ListItem>
<GlossaryTermRef href="CDR0000045164">Biopsy</GlossaryTermRef>
: The removal of cells or tissues so they can be viewed under a microscope to check for signs of cancer. If a lump in the breast is found, the doctor may need to cut out a small piece of the lump. A
<GlossaryTermRef href="CDR0000046244">pathologist</GlossaryTermRef>

views the tissue under a microscope to look for cancer cells. Four types of biopsies are as follows:

- <ItemizedList id="_144" Style="dash">
  - <ListItem>
    <GlossaryTermRef href="CDR0000046411">Excisional biopsy</GlossaryTermRef>
    : The removal of an entire lump or suspicious tissue.
  </ListItem>
  - <ListItem>
    <GlossaryTermRef href="CDR0000046698">Incisional biopsy</GlossaryTermRef>
    : The removal of part of a lump or suspicious tissue.
  </ListItem>
  - <ListItem>
    <GlossaryTermRef href="CDR0000045657">Core biopsy</GlossaryTermRef>
    : The removal of part of a lump or suspicious tissue using a wide needle.
  </ListItem>
  - <ListItem>
    <GlossaryTermRef href="CDR0000045798">Needle biopsy</GlossaryTermRef>
    or
    <GlossaryTermRef href="CDR0000045691">fine-needle aspiration biopsy</GlossaryTermRef>
    : The removal of part of a lump, suspicious tissue, or fluid, using a thin needle.
  </ListItem>
</ItemizedList>
</ListItem>
- <ListItem>
<GlossaryTermRef href="CDR0000044668">Estrogen</GlossaryTermRef>
and
<GlossaryTermRef href="CDR0000044713">progesterone receptor test</GlossaryTermRef>
: A test to measure the amount of
<GlossaryTermRef href="CDR0000046076">estrogen</GlossaryTermRef>
and
<GlossaryTermRef href="CDR0000045158">progesterone</GlossaryTermRef>
(
<GlossaryTermRef href="CDR0000045713">hormones</GlossaryTermRef>
) receptors in cancer tissue. If cancer is found in the breast, tissue from the
<GlossaryTermRef href="CDR0000046634">tumor</GlossaryTermRef>
is examined in the laboratory to find out whether estrogen and progesterone could affect the way cancer grows. The test results show whether
<GlossaryTermRef href="CDR0000045110">hormone therapy</GlossaryTermRef>
may stop the cancer from growing.
</ListItem>
</ItemizedList>
</SummarySection>
- <SummarySection id="_271">
- <SectMetaData>
<SectionType>Prognostic factors</SectionType>
</SectMetaData>

<KeyPoint id="_272">**Certain factors affect prognosis (chance of recovery) and treatment options.**</KeyPoint>
- <Para id="_273">
The
<GlossaryTermRef href="CDR0000045849">**prognosis**</GlossaryTermRef>
(chance of recovery) and treatment options depend on the following:
</Para>
- <ItemizedList id="_274" Style="bullet">
- <ListItem>
The
<GlossaryTermRef href="CDR0000045885">**stage**</GlossaryTermRef>
of the cancer (whether it is in the breast only or has spread to lymph nodes or other places in the body).
</ListItem>
<ListItem>**The type of breast cancer.**</ListItem>
<ListItem>**Estrogen-receptor and progesterone-receptor levels in the tumor tissue.**</ListItem>
- <ListItem>
A woman's age, general health, and
<GlossaryTermRef href="CDR0000046296">**menopausal**</GlossaryTermRef>
status (whether a woman is still having
<GlossaryTermRef href="CDR0000045784">**menstrual periods**</GlossaryTermRef>
).
</ListItem>
- <ListItem>
Whether the cancer has just been
<GlossaryTermRef href="CDR0000046450">**diagnosed**</GlossaryTermRef>
or has
<GlossaryTermRef href="CDR0000045862">**recurred**</GlossaryTermRef>
(come back).
</ListItem>
</ItemizedList>
</SummarySection>
</SummarySection>
- <SummarySection id="_148">
<Title>**Stages of Breast Cancer**</Title>
- <SummarySection id="_149">
- <SectMetaData>
<SectionType>**Diagnostic tests**</SectionType>
</SectMetaData>
<KeyPoint id="_150">**After breast cancer has been diagnosed, tests are done to find out if cancer cells have spread within the breast or to other parts of the body.**</KeyPoint>
- <Para id="_151">
The process used to find out whether the
<GlossaryTermRef href="CDR0000045333">**cancer**</GlossaryTermRef>
has spread within the breast or to other parts of the body is called
<GlossaryTermRef href="CDR0000046597">**staging**</GlossaryTermRef>
. The information gathered from the staging process determines the
<GlossaryTermRef href="CDR0000045885">**stage**</GlossaryTermRef>
of the disease. It is important to know the stage in order to plan treatment.
</Para>

```
  </SummarySection>
- <SummarySection id="_253">
  - <SectMetaData>
      <SectionType>Classification by stage</SectionType>
    </SectMetaData>
    <KeyPoint id="_153">The following stages are used for breast cancer:</KeyPoint>
  - <MediaLink ref="CDR0000999997" alt="Tumor Size FINAL" inline="No"
      thumb="No">
      <Caption>Tumor Size FINAL</Caption>
    </MediaLink>
  - <SummarySection id="_154">
      <KeyPoint id="_155">Stage 0 (carcinoma in situ)</KeyPoint>
    - <Para id="_156">
        There are 2 types of breast
        <GlossaryTermRef href="CDR0000046488">carcinoma in
          situ</GlossaryTermRef>
        :
      </Para>
    - <ItemizedList id="_157" Style="bullet">
      - <ListItem>
          <GlossaryTermRef href="CDR0000045674">Ductal carcinoma in
            situ</GlossaryTermRef>
          (DCIS) is a noninvasive,
          <GlossaryTermRef href="CDR0000046220">precancerous</GlossaryTermRef>
          condition in which
          <GlossaryTermRef href="CDR0000044636">abnormal</GlossaryTermRef>
          <GlossaryTermRef href="CDR0000046476">cells</GlossaryTermRef>
          are found in the lining of a breast
          <GlossaryTermRef href="CDR0000046441">duct</GlossaryTermRef>
          . The abnormal cells have not spread outside the duct to other
          <GlossaryTermRef href="CDR0000046683">tissues</GlossaryTermRef>
          in the breast. In some cases, DCIS may become
          <GlossaryTermRef href="CDR0000045741">invasive
            cancer</GlossaryTermRef>
          and spread to other tissues, although it is not known at this time how to
           predict which
          <GlossaryTermRef href="CDR0000046324">lesions</GlossaryTermRef>
          will become invasive.
        </ListItem>
      - <ListItem>
          <GlossaryTermRef href="CDR0000046315">Lobular carcinoma in
            situ</GlossaryTermRef>
          (LCIS) is a condition in which abnormal cells are found in the
          <GlossaryTermRef href="CDR0000046308">lobules</GlossaryTermRef>
          of the breast. This condition seldom becomes invasive cancer; however,
           having lobular carcinoma in situ in one breast increases the risk of
           developing breast cancer in either breast.
        </ListItem>
      </ItemizedList>
    </SummarySection>
  - <SummarySection id="_158">
      <KeyPoint id="_159">Stage I</KeyPoint>
    - <Para id="_160">
```

```
            In
            <GlossaryTermRef href="CDR0000045142">stage I</GlossaryTermRef>
            , the
            <GlossaryTermRef href="CDR0000046634">tumor</GlossaryTermRef>
            is 2
            <GlossaryTermRef href="CDR0000354457">centimeters</GlossaryTermRef>
            or smaller and has not spread outside the breast.
          </Para>
        </SummarySection>
      - <SummarySection id="_161">
          <KeyPoint id="_162">Stage IIA</KeyPoint>
        - <Para id="_163">
            In
            <GlossaryTermRef href="CDR0000045146">stage IIA</GlossaryTermRef>
            :
          </Para>
        - <ItemizedList id="_164" Style="bullet">
          - <ListItem>
              no tumor is found in the breast, but cancer is found in the
              <GlossaryTermRef href="CDR0000045607">axillary lymph
                nodes</GlossaryTermRef>
              (the
              <GlossaryTermRef href="CDR0000045762">lymph nodes</GlossaryTermRef>
              under the arm); or
            </ListItem>
          - <ListItem>
              the tumor is 2
              <GlossaryTermRef href="CDR0000354457">centimeters</GlossaryTermRef>
              or smaller and has spread to the axillary lymph nodes; or
            </ListItem>
            <ListItem>the tumor is between 2 and 5 centimeters but has not spread to
              the axillary lymph nodes.</ListItem>
          </ItemizedList>
        </SummarySection>
      </SummarySection>
    </SummarySection>
  - <SummarySection id="_249">
    - <SectMetaData>
        <SectionType>Changes to summary</SectionType>
      </SectMetaData>
      <Title>Changes to This Summary (05/20/2004)</Title>
      <Para id="_250">The PDQ cancer information summaries are reviewed regularly
        and updated as new information becomes available. This section describes the
        latest changes made to this summary as of the date above.</Para>
      <Para id="_251">Editorial changes were made to this summary.</Para>
    </SummarySection>
    <PatientVersionOf ref="CDR0000062787" />
    <DateLastModified>2004-05-20</DateLastModified>
</Summary>
```

Terminology

Terminology without MenuInformation

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Term>
- <Term id="CDR0000039741" NCIThesaurusConceptID="C1234">
    <PreferredName>amifostine</PreferredName>
    <OtherName>
      <OtherTermName>WR-2721</OtherTermName>
      <OtherNameType>Abbreviation</OtherNameType>
    </OtherName>
    <OtherName>
      <OtherTermName>ethiofos</OtherTermName>
      <OtherNameType>Synonym</OtherNameType>
    </OtherName>
    <OtherName>
      <OtherTermName>ethyol</OtherTermName>
      <OtherNameType>Synonym</OtherNameType>
    </OtherName>
    <OtherName>
      <OtherTermName>gammaphos</OtherTermName>
      <OtherNameType>Synonym</OtherNameType>
    </OtherName>
    <SemanticType
    ref="CDR0000256166">Drug/agent</SemanticType>
    <RelatedWebsites xref="http://testwebsite.gov/
    amifostine">Amifostine</RelatedWebsites>
  </Term>
```

2) Terminology with MenuInformation

i) Parent Menu Record

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Term >
<Term id="CDR0000041060">
   <PreferredName>cancer</PreferredName>
   <SemanticType ref="CDR0000256086">Cancer
      diagnosis</SemanticType>
 - <MenuInformation>
    - <MenuItem>
        <MenuType>Clinical Trials—CancerType</MenuType>
      </MenuItem>
      <MenuItem>
        <MenuType>Cancer Information</MenuType>
      </MenuItem>
   </MenuInformation>
</Term>
```

ii) Child Menu Record

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Term >
<Term id="CDR0000291250">
  <PreferredName>myeloproliferative disorders</PreferredName>
  <SemanticType
    ref="CDR0000256085">Disease/diagnosis</SemanticType>
- <MenuInformation>
  - <MenuItem>
      <MenuType>Clinical Trials—CancerType</MenuType>
      <MenuParent
        ref="CDR0000041060">cancer</MenuParent>
      <DisplayName>Myeloproliferative disorders (including
        CML)</DisplayName>
    </MenuItem>
    <MenuItem>
      <MenuType>Cancer Information</MenuType>
      <MenuParent ref="CDR0000291249">hematologic
        cancers</MenuParent>
      <DisplayName>Myeloproliferative
        disorders</DisplayName>
    </MenuItem>
  </MenuInformation>
  <DateLastModified>2003-03-03</DateLastModified>
</Term>
```

iii) Child Menu Record

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Term >
<Term id="CDR0000039347">
  <PreferredName>chronic myeloproliferative
    disorders</PreferredName>
  <SemanticType ref="CDR0000256086">Cancer
    diagnosis</SemanticType>
- <TermRelationship>
  - <ParentTerm>
      <ParentTermName
        ref="CDR0000039800">hematopoietic/lymphoid
        cancer</ParentTermName>
      <ParentType>ISA</ParentType>
    </ParentTerm>
  </TermRelationship>
- <MenuInformation>
  - <MenuItem>
      <MenuType>Clinical Trials—CancerType</MenuType>
      <MenuParent
        ref="CDR0000041060">cancer</MenuParent>
      <DisplayName>Chronic myeloproliferative disorders (incl.
        CML)</DisplayName>
```

```
      </MenuItem>
       <MenuItem>
         <MenuType>Clinical Trials—CancerType</MenuType>
         <MenuParent
            ref="CDR0000041060">cancer</MenuParent>
         <DisplayName>Myeloproliferative disorders, chronic
            (incl. CML)</DisplayName>
      </MenuItem>
   </MenuInformation>
   <DateLastModified>2003-02-05</DateLastModified>
</Term>
```

GlossaryTerm


```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE GlossaryTerm (View Source for full doctype...)>
- <GlossaryTerm id="CDR0000046063">
    <TermName>mutation</TermName>
    <TermPronunciation>(myoo-TAY-shun)</TermPronunciation>
  - <TermDefinition>
      <DefinitionText>Any change in the DNA of a cell. Mutations may be
        caused by mistakes during cell division, or they may be caused by
        exposure to DNA-damaging agents in the environment. Mutations can
        be harmful, beneficial, or have no effect. If they occur in cells that
        make eggs or sperm, they can be inherited; if mutations occur in
        other types of cells, they are not inherited. Certain mutations may
        lead to cancer or other diseases.</DefinitionText>
      <Dictionary>Cancer.gov</Dictionary>
      <Audience>Patient</Audience>
    </TermDefinition>
  - <TermDefinition>
      <DefinitionText>A change in the usual DNA sequence at a particular gene
        locus. Mutations (including polymorphisms) can be harmful,
        beneficial, or neutral in their effect on cell function.</DefinitionText>
      <Dictionary>Genetics</Dictionary>
      <Audience>Health professional</Audience>
    </TermDefinition>
    <MediaLink ref="CDR0000708090" type="audio/mpeg" alt="Pronunciation of
     dictionary term "mutation"" language="en" id="_3" />
    <MediaLink ref="CDR0000708089" type="audio/mpeg" alt="Pronunciation of
     dictionary term "mutación"" language="es" id="_4" />
    <SpanishTermName>mutación</SpanishTermName>
  - <SpanishTermDefinition>
      <DefinitionText>Cualquier cambio en el ADN de una célula. Las
       mutaciones pueden ser causadas por errores durante la multiplicación
       de las células o por la exposición a sustancias del ambiente que dañan
       el ADN. Las mutaciones pueden ser nocivas, beneficiosas o no tener
       ningún efecto. Si se presentan en las células que producen los óvulos o
       los espermatozoides, pueden heredarse; si las mutaciones se presentan
       en otros tipos de células, no se heredan. Ciertas mutaciones pueden
       llevar a padecer de cáncer u otras enfermedades.</DefinitionText>
      <Dictionary>Cancer.gov</Dictionary>
      <Audience>Patient</Audience>
    </SpanishTermDefinition>
    <DateLastModified>2007-05-17</DateLastModified>
</GlossaryTerm>
```

DrugInformationSummary

i. Drug Information Summary


```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE DrugInformationSummary (View Source for full doctype...)>
- <DrugInformationSummary id="CDR0000669854">
- <DrugInfoMetaData>
  <DrugInfoType>Brief</DrugInfoType>
  <DrugInfoAudience>Patients</DrugInfoAudience>
  <DrugInfoDescription>This page contains brief information about denileukin
    diftitox and a collection of links to more information about the use of this
    drug, related news and research results, and ongoing clinical
    trials.</DrugInfoDescription>
  <DrugInfoURL
    xref="http://www.cancer.gov/cancertopics/druginfo/denileukindiftitox">D
    enileukin Diftitox</DrugInfoURL>
  <FDAApproved>Yes</FDAApproved>
  <TerminologyLink ref="CDR0000042325">denileukin diftitox</TerminologyLink>
  <GlossaryLink ref="CDR0000045454">denileukin diftitox</GlossaryLink>
- <USBrandNames>
  <USBrandName>ONTAK</USBrandName>
    </USBrandNames>
    </DrugInfoMetaData>
  <DrugInfoTitle>Denileukin Diftitox</DrugInfoTitle>
- <Section id="_1">
- <Para id="_2">
    Denileukin diftitox is approved by the
  <GlossaryTermRef href="CDR0000454785">Food and Drug Administration
    (FDA)</GlossaryTermRef>
    to treat a certain type of
  <GlossaryTermRef href="CDR0000046771">cutaneous T-cell
    lymphoma</GlossaryTermRef>
    . It is used in patients whose disease has not gotten better with other
    treatment or has
  <GlossaryTermRef href="CDR0000045862">recurred</GlossaryTermRef>
    (come back).
    </Para>
- <Para id="_3">
    Denileukin diftitox is also being studied in the treatment of other types of
  <GlossaryTermRef href="CDR0000045333">cancer</GlossaryTermRef>
    .
    </Para>
    </Section>
- <Section id="_FDA">
  <Title>Information from the FDA</Title>
- <Para id="_FDA_16.12">
  <ExternalRef xref="http://www.cancer.gov/cancertopics/druginfo/fda-
    denileukindiftitox">FDA Approval for Denileukin Diftitox</ExternalRef>
```

```
- Information from the FDA about the approval of this drug and the clinical
trials that led to the approval.
</Para>
</Section>
- <Section id="_NCI">
<Title>Information from the NCI</Title>
- <Para id="_NCI_1">
<ExternalRef
xref="http://www.cancer.gov/Templates/drugdictionary.aspx?CdrID=42325">Definition from the NCI Drug Dictionary</ExternalRef>
- Detailed scientific definition and other names for this drug.
</Para>
- <Para id="_NCI_10.15">
<ExternalRef xref="http://www.cancer.gov/clinicaltrials/results/denileukin-diftitox-CTCL0310">Clinical Trial Results: Drug Slows Progression of Cutaneous T-Cell Lymphoma</ExternalRef>
- Important clinical trial results for this drug, background information, and
how the trial was done.
</Para>
- <Para id="_NCI_12.27">
<ExternalRef
xref="http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted">Targeted Cancer Therapies</ExternalRef>
- Information about the use of this drug to treat cancer.
</Para>
</Section>
- <Section id="_ClinicalTrial_1">
<Title>Clinical Trials</Title>
- <Para id="_ClinicalTrial_2">
<ExternalRef
xref="http://www.cancer.gov/Search/ClinicalTrialsLink.aspx?id=42325&idtype=1">Clinical Trials for Denileukin Diftitox</ExternalRef>
- Check for trials from NCI's PDQ Cancer Clinical Trials Registry now
accepting patients.
</Para>
</Section>
- <DrugInfoDisclaimer>
- <Para id="_Disclaimer_4">
<Strong>Important:</Strong>
The drug information on this Web page is meant to be educational. It is not
a substitute for medical advice. The information may not cover all possible
uses, actions, interactions, or side effects of this drug, or precautions to be
taken while using it. Please see your health care professional for more
information about your specific medical condition and the use of this drug.
</Para>
</DrugInfoDisclaimer>
<DateFirstPublished>2010-05-05</DateFirstPublished>
</DrugInformationSummary>
```

ii. Drug Combination Summary

```
<?xml version="1.0" encoding="utf-8" ?>
<!DOCTYPE DrugInformationSummary (View Source for full doctype...)>
- <DrugInformationSummary id="CDR0000636197">
- <DrugInfoMetaData>
  <DrugInfoType Combination="Yes">Brief</DrugInfoType>
  <DrugInfoAudience>Patients</DrugInfoAudience>
  <DrugInfoDescription>This page contains brief information from the National
    Cancer Institute (NCI) about the drug combination called MOPP and lists the
    drugs included in the combination. Links to NCI's Drug Information
    Summaries about the individual drugs in the combination are included,
    when available.</DrugInfoDescription>
  <DrugInfoURL
    xref="http://www.cancer.gov/cancertopics/druginfo/MOPP">MOPP</DrugI
    nfoURL>
  <TerminologyLink ref="CDR0000041779">MOPP regimen</TerminologyLink>
  <GlossaryLink ref="CDR0000635892">MOPP regimen</GlossaryLink>
    </DrugInfoMetaData>
  <DrugInfoTitle>MOPP</DrugInfoTitle>
- <Section id="_1">
- <Table Frame="None" id="_3">
  <Title>Drugs included in the MOPP combination:</Title>
- <TGroup ColSep="1" Cols="2">
  <ColSpec Align="Right" ColName="col1" ColNum="1" ColWidth="1.00*" />
  <ColSpec Align="Left" ColName="col3" ColNum="2" ColWidth="8.47*" />
- <TBody>
- <Row>
- <entry Align="Right">
  <Strong>M</Strong>
    </entry>
  <entry Align="Left">= Mechlorethamine</entry>
    </Row>
- <Row>
- <entry Align="Right">
  <Strong>O</Strong>
    </entry>
- <entry Align="Left">
    =
  <ExternalRef
    xref="http://www.cancer.gov/cancertopics/druginfo/vincristinesulfate">Vi
    ncristine Sulfate</ExternalRef> (Oncovin)
    </entry>
    </Row>
- <Row>
- <entry Align="Right">
  <Strong>P</Strong>
    </entry>
  <entry Align="Left">= Procarbazine</entry>
    </Row>
```

```
- <Row>
- <entry Align="Right">
  <Strong>P</Strong>
    </entry>
  <entry Align="Left">= Prednisone</entry>
    </Row>
    </TBody>
    </TGroup>
    </Table>
- <Para id="_4">
    Each of the
  <GlossaryTermRef href="CDR0000348921">drugs</GlossaryTermRef>
    in this
  <GlossaryTermRef href="CDR0000045650">combination</GlossaryTermRef>
    is approved by the
  <GlossaryTermRef href="CDR0000454785">Food and Drug Administration
    (FDA)</GlossaryTermRef>
    to treat
  <GlossaryTermRef href="CDR0000045333">cancer</GlossaryTermRef>
    or conditions related to cancer.
    </Para>
- <Para id="_5">
    MOPP is used to treat
  <GlossaryTermRef href="CDR0000270800">Hodgkin
    lymphoma</GlossaryTermRef>
    . This combination may also be used with other drugs or treatments or to
    treat other types of cancer.
    </Para>
    </Section>
- <Section id="_NCI">
  <Title>Information from the NCI</Title>
- <Para id="_NCI_1">
  <ExternalRef
    xref="http://www.cancer.gov/Templates/drugdictionary.aspx?CdrID=4177
    9">Definition from the NCI Drug Dictionary</ExternalRef>
    - Detailed scientific definition and other names for this drug.
    </Para>
    </Section>
- <DrugInfoDisclaimer>
- <Para id="_Disclaimer_4">
  <Strong>Important:</Strong>
    The drug information on this Web page is meant to be educational. It is not
    a substitute for medical advice. The information may not cover all possible
    uses, actions, interactions, or side effects of this drug, or precautions to be
    taken while using it. Please see your health care professional for more
    information about your specific medical condition and the use of this drug.
    </Para>
    </DrugInfoDisclaimer>
  <DateFirstPublished>2009-09-18</DateFirstPublished>
    </DrugInformationSummary>
```

Appendix D – PDQ Data Type Definition (DTD)

The latest PDQ DTD file (*pdq.dtd)* is located on the SFTP server in the directory */pub/pdq/docs*.

Appendix E – Changes to Specification Document

Date	Sec	Change
Apr-03	1	Revised paragraph one and two.
	2	Revised paragraph one of *FTP Distribution* section.
	3	Updated sample Country record.
	4	Added *Compact=No* attribute for OrderedList in Table 3.
	5	Summary: Added paragraph for optional *ReplacementFor* attribute.
	5	Organization: Added *FormerName* element in Table 6.
	5	Terminology: Added paragraph for *MenuInformation*.
	5	Terminology: Added *MenuInformation* element and child elements in Table 8.
	5	Glossary: Added optional *Dictionary=Exclude* attribute to *GlossaryTerm* element
	5	Glossary: Added optional *Audience=HealthProfessional* attribute to TermDefinition and SpanishTermDefinition.
	App B	Replaced summary sample record.
	App C	Replaced protocol sample record.
	App D	Replaced organization sample record.
	App E	Replaced person sample record.
	App F	Replaced terminology sample record.
	App F	Added terminology sample records including *MenuInformation* fragment.
	App G	Replaced glossary sample record.
	App H	Replaced genetics professional sample record.
	App I	Replaced country sample record.
	App J	Replaced political subunit sample record.
	App K	DTD change: Added optional Compact attribute for OrderedList element
	App K	DTD change: Added optional FormerName element to OrganizationNameInformation element for Organization documents
	App K	DTD change: Added optional MenuInformation elements for Term documents
	App K	DTD change: Added optional ReplacementFor element for Summary documents
Sep-03	5	Glossary: Expanded TermDefinition in GlossaryTerm to include new elements for Audience and Dictionary
	App G	Replaced glossary sample record.
	App K	DTD change: Changes to GlossaryTerm TermDefinition and SpanishTermDefinition

Feb-04	5	CTGovProtocol: Added section for new CTGovProtocol document type
	App A	Updated Table Output Rendering Recommendations. Expended on rendering of grid lines.
	App L	Renamed Appendix L (Changes to Specification Document) to Appendix M
	App K	Renamed Appendix K (PDQ Data Type Definition) to Appendix L
	App K	Added Appendix K, CTGovProtocol Sample Record
	App L	DTD change: Added 'None' as a valid value for the Frame attribute of the Table element.
Oct-04	2	Added change in directory location for vendor files. Added information for multimedia files.
	5	Multimedia: Added section for new Multimedia document type.
	6	Added change in directory location for vendor files.
	App B	Replaced Summary Sample Record to include MediaLinks
	App L	DTD change: Included media elements.
Mar-05	5	Summary: Added new elements SummaryDescription and SummaryURL
	5	Terminology: Added new element RelatedWebsites and attribute NCIThesaurusConceptID.
	App B	Modified Summary Sample Record to include new elements
	App F	Modified Terminology Sample Record to include new elements.
Sep-05	App L	DTD change: Added optional QandAEntry attribute.
Nov-05	5	Removing references to former ELHILL data format.
	5	Protocol: Added description for new elements.
	App L	DTD change: Added several new elements with version 1.55; DateSubmittedtoPDQ, ClinicalTrialsGovID, DateLastVerified, RegistryInfo, StudyDesign, StartDate, Outcome, ExpectedEnrollment, RelatedPublications. Moved and renamed element ProtocolTitle to PDQProtocolTitle, created new element ProtocolTitle, added new *ref* attribute for OverallContactBackup, and removed attribute LegacyPDQID and element ReimbursementApproval.
Mar-07	5	Summary: Added AltTitle element to supplement SummaryTitle. New AltTitle attribute *TitleType* with values *Short* and *Display* added.
	5	Summary, Glossary: Added new SummaryRef attribute *url.*
	5	Protocol: Added new *Gender* element and changed element *LeadOrgProtocolID* to be optional.
	5	CTGovProtocol: The CTGovDisclaimer element has been changed from a text to a para element.
	5	Glossary: Added new wrapper element *RelatedInformation* with two child elements *RelatedExternalRef* and *RelatedSummaryRef*, added new element *TermRef*.
	5	Added element *DrugRef* to common text elements.
	App L	Replaced DTD version 1.65 with version 1.67
Jan 2008	5	Protocol: Added new *CompletionDate* element.
	5	Protocol: Added new Safety attribute to Outcome element.
	5	Summary: Made top-level AltTitle element mandatory.

	App L	Replaced DTD version 1.67 with version 1.71
Nov 2008	1,2	Minor text correction regarding GeneticsProfessional files.
Sep 2010	5	Removing reference to Person documents which are not distributed anymore
		Adding new document type DrugInformationSummary.
	App L	DTD change: Added DrugInformationSummary document type
	App L	DTD change: Removed Person document type.
Jan 2014	2, 6	Updating contact information
Mar 2016	1, 5, App C	Removing documentation for active and closed protocols.
	App A	Added section on newly included sample transformation documents.
Apr 2018	1	Replacing NCI logo Deleting documentation for removed document types
	2, 3, 4	Misc. updates and text changes
	5	Deleting documentation for removed document types Misc editorial changes
	App C	Deleting sample records for removed document types
	App D	Removing content of DTD which can be found in the *docs* directory.
Oct 2022	5	Deleting documentation for removed GeneticsProfessionals.
	App C	Deleting sample record for removed GeneticsProfessionals.