Public Health England

Protecting and improving the nation's health

Cancer Outcomes and Services Data set
Version 7.0

Pathology - Technical Guidance

About Public Health England

Public Health England exists to protect and improve the nation's health and wellbeing, and reduce health inequalities. It does this through world-class science, knowledge and intelligence, advocacy, partnerships and the delivery of specialist public health services. PHE is an operationally autonomous executive agency of the Department of Health.

Public Health England
Wellington House
133-155 Waterloo Road
London SE1 8UG
Tel: 020 7654 8000
www.gov.uk/phe
Twitter: *@PHE_uk*
Facebook: *www.facebook.com/PublicHealthEngland*

This version produced January 2016

Publications gateway number: 2016212

Table of Contents

Version control

Version	Date	Details
V7.0	22/06/16	Final to align with COSD data set and pathology schema version 7.0 schema

1. Introduction

This document provides technical guidance to support providers of cancer services and IT software developers in the submission of the cancer outcomes and services pathology data set. It should be read in conjunction with:

- Information standards notice: reference SCCI 1521 Amd 01/2016
- COSD specification v7.0
- XML pathology schema documentation v7-0
- COSD pathology user guide v7.0

Users may also wish to read COSD v7.0 implementation guide which provides further support in implementation of the changes to the standard.

See NCIN website[1] for access to these documents and other information.

Note: NCIN has now become part of the National Cancer Registration and Analysis Service (NCRAS), which is part of Public Health England (PHE). The NCIN website will be re-branded shortly to reflect these changes but will continue to publish additional information and updates.

The cancer outcomes and services data set is a compiled data set which provides the standard for secondary uses information required to support implementation and monitoring of 'Improving Outcomes: a Strategy for Cancer' (IOSC)[2] and subsequently the 'Achieving World-Class Cancer Outcomes (A Strategy for England 2015-2020)[3].

Providers of cancer services are required to provide a monthly return on all cancer patients diagnosed from 1 January 2013 using this data set. The data are stored in the National Cancer Registration and Analysis Service (NCRAS) database (ENCORE). Submissions are made by each provider to the relevant NCRAS branch office for uploading to ENCORE.

Data may be extracted from a number of different electronic sources and submitted as separate files. The required format for submissions for COSD Pathology is XML. These and other details of the submission should be included in the COSD data transfer agreement agreed between the provider Trust and the NCRAS.

Data is collated and mechanisms for transmission of data from providers to NCRAS offices have been extended to carry the COSD data items. On receipt patient level data is validated and linked with existing records as appropriate.

Reporting mechanisms exist to provide consistent feedback on submissions and additional reports to local providers[4]. Both patient identifiable and anonymised data (where appropriate) will be made available for analysis and reporting purposes[5].

[1] http://www.ncin.org.uk/collecting_and_using_data/data_collection/cosd
[2] https://www.gov.uk/government/publications/the-national-cancer-strategy
[3] http://www.hscic.gov.uk/isce/publication/scci1521
[4] https://nww.cancerstats.nhs.uk/users/sign_in
[5] https://www.cancerdata.nhs.uk/dashboard/

2. Purpose and audience

The purpose of this document is to provide instruction to informatics personnel within provider organisations and IT software suppliers regarding file creation and submission of COSD Pathology data. It should be read in conjunction with the documents listed above.

This document describes the standards for file submission, including the XML construction and file naming to facilitate uploading onto ENCORE.

3. Help and support

For technical queries relating to the creation of these files please contact your local NCRAS office in the first instance. (See data transfer agreement for local details).

- For general queries regarding the data set contact:
  - COSDenquiries@phe.gov.uk
- For queries regarding the Data Dictionary contact:
  - datastandards@nhs.net
- For queries regarding the schema contact:
  - information.standards@hscic.gov.uk

Further help and tutorials on XML can be found on the following website: www.w3schools.com/xml/default.asp.

4. General submission principles:

- Pathology providers must submit all data relating to patients for whom they have either reported on and authorised the original pathology report or commented on (second opinion) as a specialist centre (creating a supplementary report)
- Submitted files must be sent by secure file transfer methods as agreed with your regional NCRAS office
- Files should reach the regional NCRAS office by the 25th working day following the end of month of the 'trigger event' for each diagnosed cancer (See user guide and Implementation guide for details of trigger events)
- Each file may include records for more than one tumour group
- Individual records must contain the section [Core - Patient Identity Details (Linkage)]
- Providers should aim to complete all the relevant data items as soon as possible, however as long as the mandatory fields [Core - Patient Identity Details (Linkage)] are completed the record can be submitted
- Records in each submission should include all applicable sections where possible
- Demographic section should be submitted by each provider on the first submission of a record
- For updated records only the updated/amended sections and core linkage items need to be submitted
- The data submission files must comply with the COSDPathology XML schema specifications pack*.

Note: All data MUST be submitted to the NCRAS in structured XML format.

5. General file formatting principles

The required format for submission of data for the COSD will be extensible markup language (XML) as specified in the COSD Pathology XML schema specification pack. This contains all of the schema documents listed below and all embedded schema referenced in the above documents. This schema pack also contains the data type schema, containing the formats for submitted data. The XML schema pack is available on request from the NHS Data Model and Dictionary Service or by emailing COSDenquiries@phe.gov.uk.

XML schemas (XSD) have been designed for the whole COSD Pathology data set including where required the site specific pathology elements required by COSD. In addition:

- These schemas define the expected structure of the XML submissions.
- Schema design is segmented into separate schema defining the different sections of COSD data specification.
- These schemas are embedded in a hierarchical manner into a single schema.
- These schemas contain information on the expected values for a data element.

Data submitted in XML format will be required to conform to the schemas for the appropriate cancer site, or core for all registerable conditions.

Within the schema, it is unfortunate that ‘core’ is a 4 letter word with 2 different meanings:

1. ‘Core’ is the subset of COSD data items (and their corresponding sections) that are common to all site groups.
    a. If you are submitting a lung record, then you need to provide the core and lung specific data items. This is perhaps the more obvious definition.
2. ‘Core’ is also the non-specific site group that is used when a record does not belong to any of the other well defined groups.

This explains the non-obvious naming convention of site groups and catch-all: BreastCore, CNSCore ... UpperGICore, UrologyCore; and CoreCore.

When you see CoreCore, think ‘OtherCore’ or even ‘OtherCommon’. We are hoping to address this in the next version of the data set.

The top level schema in the hierarchical structure is COSDPathologyCOSD_XMLSchema-v7-0.xsd.

The following list outlines the tumour specific site and associated schema, embedded within the above schema:

Tumour site	Schema
Core	COSDPathologyCore_XMLSchema-v7-0.xsd
Breast	COSDPathologyBreast_XMLSchema-v7-0.xsd
Central Nervous System	COSDPathologyCNS_XMLSchema-v7-0.xsd
Colorectal	COSDPathologyColorectal_XMLSchema-v7-0.xsd
Children, Teenagers and Young Adults	COSDPathologyCTYA_XMLSchema-v7-0.xsd
Gynaecology	COSDPathologyGynaecology_XMLSchema-v7-0.xsd
Haematology	COSDPathologyHaematology_XMLSchema-v7-0.xsd
Head and neck	COSDPathologyHeadNeck_XMLSchema-v7-0.xsd
Lung	COSDPathologyLung_XMLSchema-v7-0.xsd
Sarcoma	COSDPathologySarcoma_XMLSchema-v7-0.xsd
Skin	COSDPathologySkin_XMLSchema-v7-0.xsd
Upper GI	COSDPathologyUpperGI_XMLSchema-v7-0.xsd
Urology	COSDPathologyUrology_XMLSchema-v7-0.xsd

Character set: UTF-8
Information on permitted formats are contained in the data type schema

For clarity a sample of a single record for a core only item is included in Appendix 1.

6. Data set and XML record structure

6.1 The COSD element

The root element of a COSD XML file is **<COSD>** and only one is permitted and required per submission i.e. in each individual xml file. There are 6 child elements that need to be provided within the root element, these are:

- **<Id root="*uuid*" />**
  - The root attribute will be a universal unique identifier (UUID) for the submission. This takes the form of an 8-4-4-4-12 hexadecimal characters e.g. DEAEDCC2-76AA-411E-B994-8FDD98C3FFFA.
  - A UUID Library should be used to create it.
- **<OrgCodeSubmitter extension="*orgcode*"/>**
  - The extension attribute should contain the NACS code of the submitting Provider.
- **<RecordCount value="*count*" />**
  - The value attribute should identify the number of **<COSDRecord>** elements being supplied with this submission.
- **<ReportingPeriodStartDate>** and **<ReportingPeriodEndDate>**
  - These elements give the time period for the data submission. This is normally the "trigger event" date range and takes the (ISO) format YYYY-MM-DD.

- **<FileCreationDateTime>**
  - This is the timestamp of when the submission was generated and takes the format YYYY-MM-DDTHH:MM:SS e.g. 1900-01-01T10:11:12.

6.2 The COSDRecord element

The **<COSDRecord>** element is a child element of the **<COSD>** element. This element identifies a single record within the submission. As with the **<COSD>** element it has an **<Id>** element with a UUID value attribute. This is a unique identifier for the record and does not need to be preserved across submissions. i.e. COSDRecords pertaining to the same record in future submissions do not need to have the same UUID.

The COSDRecord also contains an element that identifies the site specific data set of that record and it is that element that contains data.

For example a Sarcoma COSDRecord is represented in COSD XML like:

```
<COSDRecord>
    <Sarcoma>
        <SarcomaCore>
            <!-- Sarcoma Core sections -->
        </SarcomaCore>
    </Sarcoma>
</COSDRecord>
```

It should therefore become clear to the reader that core (non-specific site) records have no content group:

```
<COSDRecord>
    <Core>
        <CoreCore>
            <!-- Core Core sections -->
        </CoreCore>
    </Core>
</COSDRecord>
```

NOTE: Site specific augment core group sections where necessary.

For example, a CoreCorePathology section may contain:

```
<CoreCorePathology>
    <InvestigationResultDate>...</InvestigationResultDate>
    ...
    <NeoadjuvantTherapyInd code="..."/>
</CoreCorePathology>
```

Whereas a BreastCorePathology section may also contain:

```
<BreastCorePathology>
    <InvestigationResultDate>...</InvestigationResultDate>
    ...
    <NeoadjuvantTherapyInd code="..."/>
    <BreastPathology>
        <CoreBiopsyNode code="..."/>
    </BreastPathology>
</BreastCorePathology>
```

7. Generating COSD XML

The hierarchical nature of the data set may lead providers to adopt an object orientated programming (OOP) approach to developing the COSD XML submission, but it is also possible to use a more procedural approach to generate the XML and to use condition logic to include or exclude specific data items.

This may be simpler to produce, but might be harder to maintain in the longer term and whilst OOP may be the more elegant solution, providers may need a pragmatic approach with constrained resources, existing skills/technologies, etc.

A working example of the OOP paradigm is given below. The code is far from complete, but demonstrates the key principle of inheritance. The use of CoffeeScript is merely for convenience; to all intents and purposes consider it pseudo code:

```
class CoreGenerator
  constructor: () ->

  buildCOSDRecord: (pathology) ->
    @buildCore(pathology)

  # Core section
  buildCore: (pathology) ->
    @buildCoreLinkagePatientId(pathology.patient)
    @buildCoreDemographics(pathology.patient)
    @buildCorePathology(pathology) for pathology in tumour.pathologies

  buildCoreLinkagePatientId: (patient) ->
    alert "build CoreLinkagePatientId”

  buildCoreDemographics: (patient) ->
    alert "build CoreDemographics "

  buildCorePathology: (pathology) ->
    alert "build CorePathology"

class BreastGenerator extends CoreGenerator
  # Core section
```

```
  buildCorePathology: (pathology) ->
    super pathology
    alert "build BreastCorePathology/BreastPathology"
    # TODO: add site specific fields here

generator = new BreastGenerator
generator.buildCOSDRecord(...)
```

8. Reserved characters

XML has reserved characters which should not be used in data submissions, these are:

Reserved Character	Meaning	Entity reference
>	Greater than	>
<	Less than	<
&	Ampersand	&
%	Percent	%

Note: if it is unavoidable that these reserved characters are used in data submission, they should either be replaced by the corresponding entity reference or encapsulated with the <![CDATA[{data text}]]> tag. Particular care should be taken with the pathology and imaging report text fields within the COSD.

8.1 File naming convention

The submission file **must** be named using the following convention:

XML file

- COSD_<FILE SOURCE>_<Submitting Org>_<Reporting Period Start Date>_<Reporting Period End Date>_<Date of file creation>.xml

Where**:**

- <COSD> is a fixed value
- <FILE SOURCE> is MDT or PAS or PATH or RIS (or other source description as agreed with NCRAS)
- <Submitting Org> is the Organisation Code (e.g. AB3) for the submitting organisation
- <Reporting Period Start Date> **must** always be in the format CCYY-MM-DD.
- <Reporting Period End Date> **must** always be in the format CCYY-MM-DD.
- <Date and time of file creation> Timestamp when the file was created in the format CCYY-MM-DDThh:mm:ss.

Example**:**

- The file name for organisation (X09) submitting its own MDT data for activity month June 2011 on the 5th September 2011 at 10:30:22 AM will be:
  - COSD_MDT_X09_2011-06-01_2011-06-30 2011-09-05T10_30 22.xml

Files MAY be zipped prior to transmission, in this case the file extension .zip will be acceptable.

Each file submitted MUST have a unique filename as generated by the above method.

9. Data submissions

XML data should be validated against the schema prior to submission.

XML data submissions should be given a new UUID in the <COSD> element, where submission are altered and re-submitted a new UUID should be applied. Each record in the submission should have a unique UUID as the root attribute for the <COSDRecord> element. New UUID must be created for resubmissions of data.

All data submissions must be transmitted between nhs.net email accounts (or alternative email accounts that are accepted as part of the nhs.net framework). The regional NCRAS offices will provide providers with the relevant recipient account in the data transfer agreement.

Monthly data submissions are required from each provider within 25 working days of the relevant month end. A schedule of submission deadlines is available on the NCIN website.

Trust should complete the COSD submission template which should be attached to each email submission, along with the data submission files themselves. This provides a summary overview of their submission and alerts the NCRAS to any special or extenuating circumstances which may have affected the submission that month. See Appendix 2.

10. Who will submit the data?

Files should be submitted by NHS providers of any adult or children cancer services.

The submission files should relate to a single provider and only for data that they own.

Providers must clarify arrangements or changes for submitting the data with their local NCRAS office

COSD data started in January 2013, and submissions to the NCRAS commenced from March 2013. Trusts were mandated to collect and submit Pathology data in XML from January 2016, with an agreed delayed start until July 2016, September 2016 submissions.

11. What data items should be submitted?

All applicable data items specified as either mandatory or required in the data set and XML schema documentation should be submitted as soon as available.

The mandatory, required or optional (M/R/O) column indicates the recommendation for the inclusion of data. This applies specifically to the XML files but should also be used to decide on data to be included through other message formats.

M = Mandatory: this data item is mandatory; the record cannot be submitted if the mandatory data items are not completed. The file will be rejected if mandatory items are absent and other data items are completed.

R = Required: this data item is required as part of NHS business rules and must be included where available or applicable, however, the section can be submitted without completing all the required items.

O = Optional: this data item can be included at the discretion of the submitting organisation and their commissioners as required for local purposes.

X: This data item will not normally be included in the direct submission from cancer service provider organisations and is excluded from the schema, (core - death details may however be included if required as this is covered in the schema).

12. Validation

The data will be validated in ENCORE according to a set of rules. If the data validation rules are not met, the whole or relevant parts (data set sections or records) of the extract may be rejected and returned to the provider.

An indication of the areas which require attention will be provided in the form of an error report or other explanation.

The provider will normally be expected to resolve any errors/issues or add missing data and re-extract the file for sending to the NCRAS within 5 working days, for re-validation. The turnaround time for validation, any re-submission and subsequent re-validation is necessarily short as delays will adversely affect the timeliness and quality of the data and the validity of conformance reporting.

A record of rejected files will be kept by the NCRAS as an audit trail and to support conformance monitoring; original files may be retained in order to make a comparison with subsequent files received.

13. Reporting

The NCRAS has developed standardised reports which are available to all providers submitting data through the NCRAS conformance portal (CancerStats)[6] only available through NHS N3 Network. Analytical reports are being developed separately[7].

Providers should continue to contact their regional NCRAS office to request any data they require which is not made available via standardised reporting.

[6] https://nww.cancerstats.nhs.uk/users/sign_in
[7] https://www.cancerdata.nhs.uk/dashboard/

Appendix 1 – sample lung XML record

```
<?xml version="1.0" encoding="UTF-8"?>
<COSD:COSD
xmlns:COSD="http://www.datadictionary.nhs.uk/messages/COSD-v7-0"
xsi:schemaLocation="http://www.datadictionary.nhs.uk/messages/COSD-v7-0
COSDCOSDXMLSchema-v7-0.xsd"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
<Id root="60E8A409-24B6-4ADC-9C96-0273F668BDC8"/>
<OrgCodeSubmitter extension="AB1"/>
<RecordCount value="1"/>
<ReportingPeriodStartDate>1900-01-01</ReportingPeriodStartDate>
<ReportingPeriodEndDate>1900-01-01</ReportingPeriodEndDate>
<FileCreationDateTime>1900-02-01T10:11:12</FileCreationDateTime>
<COSDRecord>
 <Id root="0C0219F4-B3A5-420F-9C85-AC01910191B7"/>
 <Lung>
   <LungCore>
     <LungCoreLinkagePatientId>
       <NHSNumber extension="1234567890"/>
       <LocalPatientId>AB123456</LocalPatientId>
       <NHSNumberStatusIndicator code="03"/>
       <Birthdate>1900-01-01</Birthdate>
       <OrgCodeOfProvider extension="AB1"/>
     </LungCoreLinkagePatientId>
     <LungCoreDemographics>
       <PersonFamilyName>
         <family>SMITH</family>
       </PersonFamilyName>
       <PersonGivenName>
         <given>JOHN</given>
       </PersonGivenName>
       <Address>
         <UnstructuredAddress>
           <streetAddressLine>123 Main Road</streetAddressLine>
         </UnstructuredAddress>
       </Address>
       <Postcode>
         <postalCode>AB1 2CD</postalCode>
       </Postcode>
       <Gender code="1"/>
     </LungCoreDemographics>
     <LungCorePathology>
        <InvestigationResultDate>1900-01-01</InvestigationResultDate>
        <TopographySNOMED code="T04030"/>
        <TopographySNOMED code="900000000000013009"/>
        <MorphologySNOMEDPathology code="M80003"/>
        <MorphologySNOMEDPathology code="900000000000548007"/>
        <NeoadjuvantTherapyInd code="Y"/>
     </LungCorePathology>
   </LungCore>
 </Lung>
</COSDRecord>
</COSD:COSD>
```

Appendix 2 – File submission template

Generic file source	File name	Number of records	Any reasons for variation from number expected YES/NO	Extenuating circumstances if previous column contains 'YES'	Other comments
MDT					
PATH					
PAS					
RIS					

* Generic file names as listed should align with the sources identified in the data transfer agreement. Any other file sources should be referenced using consistent terminology as agreed with the NCRAS.