Classification: Official

NHS England

Pathology Standards Implementation – SNOMED CT PaLM Mapping Best Practice

Document filename: SNOMED CT PaLM Mapping Best Practice			
Programme	Pathology Standards Implementation	**Project**	SNOMED CT PaLM Mapping Project
Owner	Karim Nashar	**Version**	1.0
Authors	Gogs Byrn, Karim Nashar, Peter Garrett, Andrew Perry, Medi Harsini	**Version date**	25/04/2025

Contents

Acknowledgements

The PaLM Mapping Project team wish to acknowledge the contributions of Dr Mark Sleeman and Tamara Forster from Berkshire and Surrey Pathology Service, Andy Harris, clinical informatician, and Dr Jonny Pearson and Dr Alice Waterhouse from NHS England's Data Science team. Their input has been vital to the success of the project, and we are grateful for their subject matter expertise, time, and effort.

Glossary

Term	Definition
Attribute Value Pair	A characteristic of the meaning of a SNOMED CT concept. An attribute is assigned a value *(attribute value pair)* when used in the definition of a concept
Component	The substance or chemical constituent being measured in a pathology and laboratory medicine test
Concept Map	A (one way) mapping from a set of codes to one or more other codes, in this context, SNOMED CT concepts
DAPB4101	An Information Standard that supports the interoperable sharing of pathology and laboratory medicine test results to General Practices
Data Element	In this context, data items in a pathology report that help define patient test results
Expression Constraint Language (ECL)	The SNOMED CT computable language which in this context, is used to constrain SNOMED CT map targets
Laboratory Information Management System (LIMS)	Software with features that support a modern laboratory's operations
Pathology FHIR Specification	A standards framework used to define the structure of pathology and laboratory medicine test reports, mandated for use in DAPB4101
PMIP EDIFACT	A messaging integration by which GP practices receive structured pathology test results from labs, which DAPB4101 has been designed to replace
Property	The quantifiable or measurable characteristic of a pathology and laboratory medicine test

Read v2 Pathology Bounded Code List (PBCL)	A subset of Read v2 codes for use in laboratory to General Practice. Semantically equivalent SNOMED CT PBCL concepts have since been created *(see below)*
Reportable	A laboratory and pathology medicine test result code - a single data item contained in the report identifying the specific laboratory test to which a value(s) and interpretation can be assigned
Requestable	A laboratory and pathology medicine test request code - a single data item contained in the report identifying the specific laboratory test(s) originally requested
SNOMED CT PaLM (Pathology and Laboratory Medicine)	SNOMED CT concepts split into two variants that represent pathology and laboratory medicine test request and result codes in more granularity than SNOMED CT PBCL
SNOMED CT PBCL	SNOMED CT concepts that are semantically equivalent to Read PBCL codes *(see above)*.
UoM	Units of measure
Value Set	A selection of a set of codes for use in a particular context
Workflow	A structured sequence of tasks and processes designed to achieve a specific goal, outlining roles, order of operations, and necessary resources

1. Introduction

1.1 Purpose

The purpose of this document is to articulate best practice for labs to follow when mapping source terminology data representing patient test results to SNOMED CT PaLM and to other coded data elements in the Pathology FHIR Specification, in support of the DAPB4101 Pathology and Laboratory Medicine Reporting Information Standard.

Please note, the principles outlined in this document can be applied more broadly, to support mapping to other terminologies and code sets used in other areas of healthcare. Therefore, the pathology and laboratory medicine terminology mapping requirement can be considered a use-case to draw out the key principles of best practice for automated, user-guided terminology mapping.

1.2 Audience

The intended audience for this document includes:

- Pathology lab IT Managers
- Pathologists / Biomedical Scientists involved in the clinical review of maps
- Suppliers of terminology mapping software
- Stakeholders seeking transferable best practice guidance on terminology mapping

1.3 Related documents

This document should be read in conjunction with the following:

- DAPB4101 Information Standards Notice (Amd 63/2023)
- NHS England Data Science Team - Advanced Analysis Components to Support SNOMED PaLM Mapping Project

1.4 Executive summary

Created by NHS England, SNOMED CT PaLM is a terminology that more accurately represents pathology and laboratory medicine patient test result codes (‘reportables’) than the legacy, outdated Read PBCL codes currently in use. Together with the Pathology FHIR Specification, SNOMED CT makes up the new DAPB4101 Information Standard that supports lab to GP reporting, and is endorsed by the Royal Colleges, the BMA, and the Institute of Biomedical Science. The implementation of SNOMED CT PaLM will bring many benefits, not least pandemic preparedness, with its ability to add new national reportables for new lab tests.

Due to the limitations of many Laboratory Information Management Systems (LIMS) SNOMED CT PaLM cannot yet be used natively by most labs, so mapping of 'local' reportables to SNOMED CT PaLM is required.

ISO, the International Organization for Standardisation defines mapping as *'the process of associating concepts from one terminological resource to concepts in another terminological resource and defining their equivalence in accordance with a documented rationale and a given purpose.'* However, whilst prevalent in digital healthcare, mapping can be laborious, can introduce error, can result in variation, and requires ongoing maintenance.

Consequently, NHS England's PaLM Mapping Project team have sought to establish the best and most efficient way that labs can map to SNOMED CT PaLM and to other coded data elements in the Pathology FHIR Specification; these being key data items that allow the pathology report to be represented as structured data, and how best to create and maintain assured map tables and value sets. Chief activity centred on formulating, testing, and evaluating a range of mapping strategies and mapping tools. The complete objectives, scope, benefits, risks, and constraints of the project are captured in the SNOMED CT PaLM Mapping Project Brief.

Chart Position	Specialism	Hospital Local Code	Hospital Local Description	Hospital UoM	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
1	Biochemistry	CREA	Creatinine	umol/L	1107001000000108	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum
2	Biochemistry	NA	Sodium	mmol/L	1107871000000107	Substance concentration of sodium in serum (observable entity)	Sodium substance concentration in serum
3	Biochemistry	K	Potassium	mmol/L	1107761000000109	Substance concentration of potassium in serum (observable entity)	Potassium substance concentration in serum
4	Haematology	HB	Haemoglobin	g/L	1107511000000100	Mass concentration of haemoglobin in blood (observable entity)	Haemoglobin mass concentration in blood
5	Biochemistry	ALB	Albumin	g/L	1105861000000106	Mass concentration of albumin in serum (observable entity)	Albumin mass concentration in serum
6	Haematology	PLT	Platelets	10*9/L	1108041000000107	Platelet count in blood (observable entity)	Platelet count in blood
7	Haematology	WBC	White Cell Count	10*9/L	1110441000000100	White blood cell count in blood (observable entity)	White blood cell count in blood
8	Haematology	MCV	MCV	fL	1491000237105	Mean corpuscular volume of erythrocytes in blood (observable entity)	Erythrocytes MCV (mean corpuscular volume) in blood
9	Haematology	NEUT	Neutrophils	10*9/L	1108071000000101	Neutrophil count in blood (observable entity)	Neutrophil count in blood
10	Haematology	LYMP	Lymphocytes	10*9/L	67541000237108	Count of lymphocytes in blood (observable entity)	Lymphocyte count in blood
11	Haematology	EOSI	Eosinophils	10*9/L	1107391000000104	Eosinophil count in blood (observable entity)	Eosinophil count in blood
12	Haematology	MCH	MCH	pg	1022471000000107	Mean corpuscular haemoglobin (observable entity)	MCH - Mean corpuscular haemoglobin
13	Haematology	RBC	RBC	10*12/L	1022451000000103	Red blood cell count (observable entity)	Red blood cell count
14	Haematology	MONO	Monocytes	10*9/L	1107991000000100	Monocyte count in blood (observable entity)	Monocyte count in blood
15	Haematology	BASO	Basophils	10*9/L	1106091000000103	Basophil count in blood (observable entity)	Basophil count in blood
16	Haematology	HCT	Haematocrit	Ratio	1111571000000101	Haematocrit volume fraction of blood (observable entity)	Haematocrit volume fraction of blood
17	Biochemistry	ALP	ALP	U/L	1106051000000106	Enzyme activity of alkaline phosphatase in serum (observable entity)	Alkaline phosphatase enzyme activity in serum
18	Biochemistry	ALT	ALT	U/L	1106081000000100	Enzyme activity of alanine aminotransferase in serum (observable entity)	Alanine aminotransferase enzyme activity in serum

It is important to acknowledge that much of the best practice established whilst undertaking this project can be applied universally to terminology mapping. Consequently, a wider view beyond the scope of supporting DAPB4101 is encouraged, as terminology mapping supports interoperability across many aspects of digital healthcare.

2. Background Information

2.1 Why use SNOMED CT PaLM?

Pathology and laboratory medicine tests are crucial to the diagnosis and management of disease. Over a billion lab tests are performed by the NHS each year, and to illustrate the extent to which the data generated underpins healthcare, lab test reportables account for over a third of all the SNOMED CT recorded in GP records each year. Consequently, the terminology that supports reporting is of fundamental importance to patient care.

The Pathology Standards Programme was set up at NHS Digital under the CCIO7 initiative, with a remit to deliver a terminology to replace Read PBCL, as this outdated terminology generates less reproducible and comparable lab test result data than is needed. Moreover, whilst Read PBCL is a national terminology, it does not provide a common language representing the same meaning across all labs. Through extensive consultation and collaboration with the professional community, the new terminology was iteratively developed, with the first release of SNOMED CT PaLM in 2023.

In terms of data quality, SNOMED CT PaLM reportables are a considerable upgrade on Read PBCL as they provide a common language that can be interpreted the same way across labs, conform to well-defined editorial principles, and carry modelled components that logically define each concept. This empowers machine processing, in turn strengthening clinical decision support and data analysis. Significantly, SNOMED CT PaLM also allows for the creation of new reportables that can be released in a short space of time, whereas Read PBCL does not; a situation that caused major problems during the COVID-19 pandemic. Moreover, the granularity of SNOMED CT PaLM offers the means to represent multiple reportables carried in complex microbiology reports, which are currently only represented using text strings.

Regarding the wider picture, the government has the stated aim of using *"AI tools to streamline public services, eliminate delays through improved data sharing, and reduce costs."* AI systems are only as good as the data they are trained on, and as pathology reporting data constitutes such a high percentage of NHS data, mapping to SNOMED CT PaLM will help provide better data quality in a significant area of healthcare diagnostics, creating a virtuous circle with quality data allowing AI systems to make better decisions.

2.2 Strategic drivers

SNOMED CT PaLM is part of the DAPB4101 Pathology and Laboratory Medicine Reporting Information Standard that aligns to:

- NHS England - Transformation Directorate Data saves lives
- NHS England - Diagnostics: Recovery and Renewal
- DHSC - The future of healthcare: our vision for digital, data and technology in health and care policy

2.3 Laboratory landscape

There are currently 122 NHS labs in England, administered by 27 NHS Pathology Networks, who support lab to GP reporting. There are also a small number of privately run labs who provide this service. Due to several factors, labs do not currently use a single, standardised terminology to represent reportables 'natively' in their LIMS (2.5). Moreover, there are multiple LIMS providers, and multiple legacy LIMS instances. Consequently, there is significant variance in reportable source data across the estate.

2.4 Historical mapping

Over twenty years ago, to support interoperability between labs and GP practices via the PMIP EDIFACT messaging integration, labs were required to manually map their local reportables to Read PBCL codes. This labour-intensive process was repeated by every lab in England, resulting in variation in the type of source data used to map, and the methods employed, thereby impacting data quality. The variation was exacerbated by the inherent ambiguity found in many Read PBCL codes, leading to varied interpretation per lab. Additionally, there was no best practice around governance, assurance, and maintenance, resulting in a lack of provenance information and questions over the reliability of maps.

2.5 Why is mapping required?

The use of SNOMED CT PaLM as a standardised terminology used natively in LIMS, thereby negating the need for mapping, is recognised by NHS England as a long-term goal. However, analysis shows this is unlikely to happen in the short to medium-term as it would involve systematic change requiring extensive clinical and technical assurance, plus implementation of system upgrades across the entire estate, which would be labour intensive, time consuming, and costly to the NHS. Moreover, in recent years, many labs have commissioned new LIMS, thereby tying them into long-term contractual agreements around system capabilities. Any change to said agreements would therefore require re-negotiation, incurring cost.

Furthermore, NHS England consider SNOMED CT PaLM to be a 'reference' terminology rather than an 'interface' terminology, and as such, less suitable for the type of 'user-friendly', character-limited human-readable descriptions the professional community prefers; a situation that lends itself to mapping.

In light of these factors, DAPB4101 mandates SNOMED CT as an interoperability solution, i.e., it supports the sharing of pathology test results between healthcare settings, rather than mandating that it be used at source. Consequently, SNOMED CT PaLM reportables do not have to be generated and flow natively from LIMS, if maps from local reportables to SNOMED CT PaLM are subsequently used to transform the data when sharing the report to the GP system. This can be achieved using 'middleware', as per the process that supports PMIP EDIFACT.

Now that SNOMED CT PaLM is available, a similar mapping effort to that involved in mapping to Read PBCL is required. NHS England are seeking to help make this process as straightforward, robust, efficient, and economical as possible, using automation where appropriate; thereby easing the burden on under-resourced, time-poor lab staff.

2.6 Mapping effort/benefit ratio

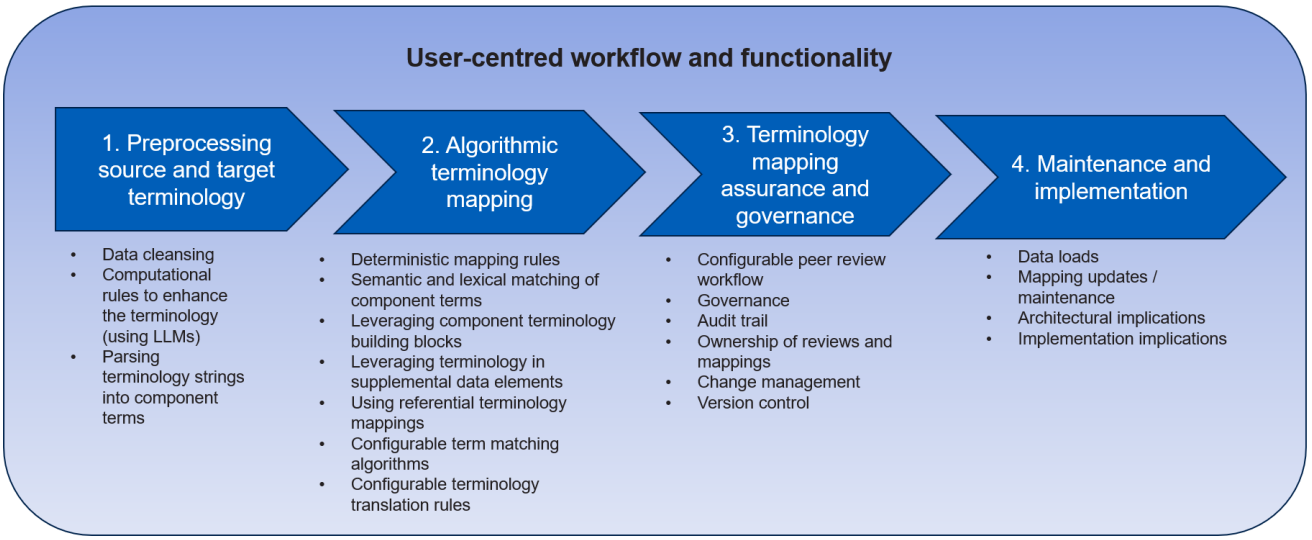
NHS England's SNOMED CT Code Usage in Primary Care data provides a clear illustration of the value of mapping even a small number of SNOMED CT PaLM Reportables. Mapping just the 20 most used reportables in lab to GP reporting would account for 50% of the pathology and laboratory medicine lab test result data recorded in Primary Care. Increasing this to the top 300 reportables would account for 99% of the data. Consequently, offering even the most basic mapping support would return huge value.

3. Mapping overview

This section provides an overview of the principles, workflow, and strategies that support the terminology mapping process.

3.1 Mapping principles and workflow

The overarching principles, tasks, and processes derived from the pathology and laboratory medicine terminology mapping requirement that are understood to be applicable in the wider terminology mapping context are represented in the diagram below.



The four main steps are underpinned by interface and tooling functionality that together enable a user-centred workflow.

The document will first cover the pathology and laboratory medicine terminology mapping requirement (4), followed by an explanation as to what makes a SNOMED CT PaLM reportable (5). Workflow will then be covered in the following sections:

1. Preprocessing source and target terminology (**Error! Reference source not found.**)
2. Algorithmic terminology mapping (7)
3. Terminology mapping assurance and governance (8)
4. Terminology mapping maintenance and implementation (9)

3.2 Mapping strategies

Terminology mapping should be as deterministic as possible. Probabilistic methods such as those employed by large language models (LLMs) can still be useful in handling variation or

missing data but for the pathology and laboratory medicine terminology mapping requirement, they are considered second tier (10.2).

Consequently, the three main strategies recommended are:

- Semantic and lexical mapping
- Leveraging terminology component building blocks
- Using referential terminology mapping artefacts

A basic overview of these strategies is provided later in Section 7, together with practical examples of associated techniques that support the pathology and laboratory medicine terminology mapping requirement.

During testing, the PaLM Mapping Project team demonstrated that a combination of strategies optimised the automation of terminology mapping outputs for expert review and assurance.

The PaLM Mapping Project team collaborated with NHS England's Data Science team to ensure that preprocessing terminology steps and deterministic mapping techniques were most suited to the pathology and laboratory medicine terminology mapping requirement. The Advanced Analysis Components to Support SNOMED PaLM Mapping Project paper produced by the Data Science team outlines the role of advanced analytics components to support mapping and provides key considerations relevant to the project. The paper is referenced heavily throughout this document and provides further detail around each mapping strategy.

3.2.1 Leveraging terminology component building blocks

At this point, it is useful to provide a basic explanation of how leveraging terminology component building blocks can support mapping, as this strategy is integral to both the preprocessing of source and target terminology, and to algorithmic terminology mapping.

This strategy involves leveraging the structure of the data and applying translation rules to facilitate mapping.

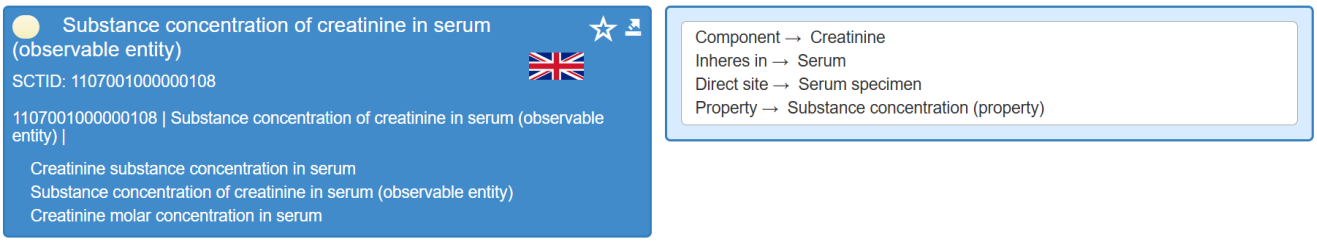
As detailed in (5.1), SNOMED CT PaLM reportables carry coded attributes relating to a lab test's property, component, specimen, and technique. These can be viewed as terminology component 'building blocks'. Equivalent local source data exists that represents each building block.

This example shows the source data relating to a lab's local 'serum creatinine' reportable.

Hospital UoM	Hospital Local Reportable code	Hospital Specimen code	Hospital Technique code
umol/l	creatinine	serum	n/a
property	component	specimen	technique

NB. **umol/L** defines the property as 'a **'substance concentration'** (7.2.2)

The equivalent SNOMED CT PaLM reportable is shown below:



By applying techniques at both the preprocessing and algorithmic mapping stage that leverage these data elements (**Error! Reference source not found.** and 7), the mapping output is greatly enhanced.

4. Mapping requirements

The scope of pathology and laboratory medicine terminology mapping covers several requirements, detailed below. It is important to distinguish between them, because the process involved requires different levels of effort, the challenge of implementing each one into live use differs, and each returns a different level of value.

4.1 Mapping single reportables to SNOMED CT PaLM

This is the core requirement, returning the most value. Fortunately, it is also the most straightforward, and implementing maps into live use is realistic and achievable.

Hospital A Local Code	Hospital A Local Description	Hospital A UoM	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
CREA	Creatinine	umol/L	1107001000000108	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum
NA	Sodium	mmol/L	1107871000000107	Substance concentration of sodium in serum (observable entity)	Sodium substance concentration in serum
K	Potassium	mmol/L	1107761000000109	Substance concentration of potassium in serum (observable entity)	Potassium substance concentration in serum
HB	Haemoglobin	g/L	1107511000000100	Mass concentration of haemoglobin in blood (observable entity)	Haemoglobin mass concentration in blood
ALB	Albumin	g/L	1105861000000106	Mass concentration of albumin in serum (observable entity)	Albumin mass concentration in serum
PLT	Platelets	10*9/L	1108041000000107	Platelet count in blood (observable entity)	Platelet count in blood
WBC	White Cell Count	10*9/L	1110441000000100	White blood cell count in blood (observable entity)	White blood cell count in blood

4.2 Mapping to other SNOMED CT concepts that populate supplemental data elements in the Pathology FHIR specification

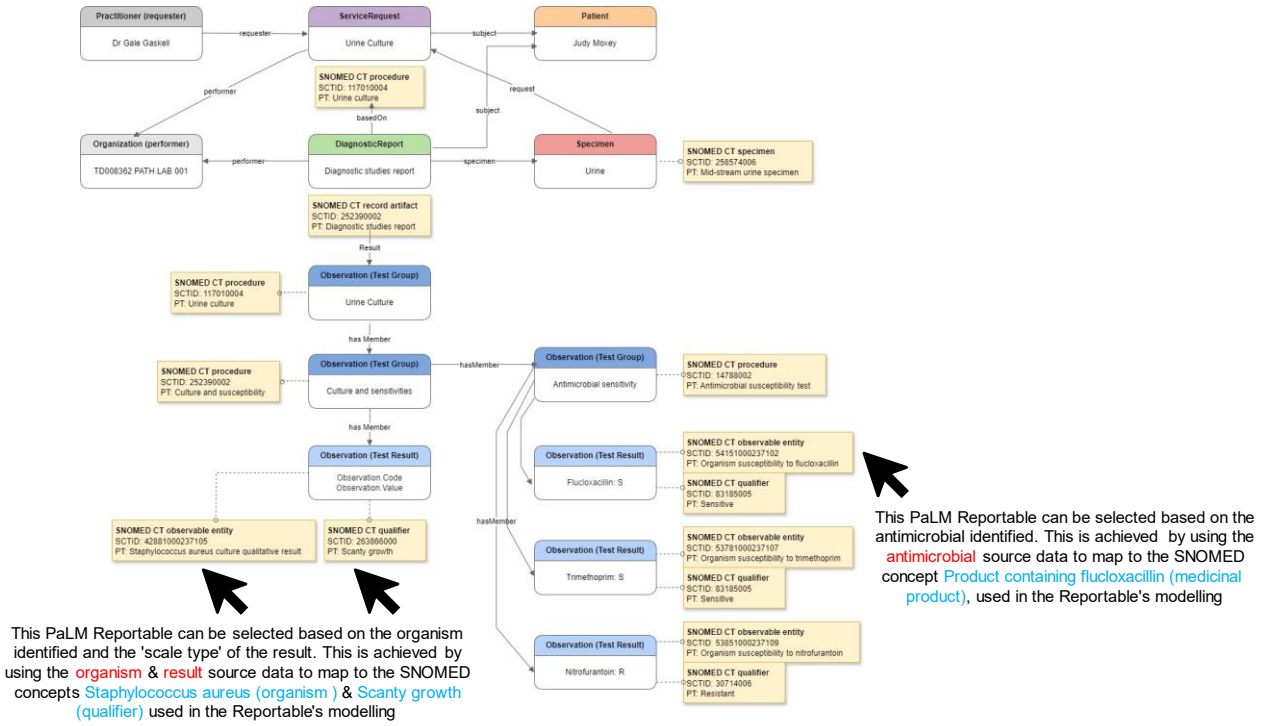
This requirement supports the goal of having fully atomic, structured, coded pathology reports, valuable in terms of enhancing data quality. The process involved is simpler than mapping single reportables, but the implementation challenge is harder because there are several target value sets that require agreement and sign-off by the clinical profession.

Hospital A Specimen local code	Hospital A Specimen local description	SNOMED Specimen Concept ID	SNOMED Specimen FSN	SNOMED Specimen Synonym
CG13	Endocervical swab (STD PCR)	444787003	Swab of endocervix (specimen)	Swab of endocervix
CG14	Eye swab (STD PCR)	445160003	Swab of eye (specimen)	Swab of eye
CG15	First void urine	698276005	First stream urine specimen (specimen)	First void urine sample
CG16	Rectal swab (STD PCR)	258528007	Rectal swab (specimen)	RS - Rectal swab
CG17	Throat swab (STD PCR)	258529004	Throat swab (specimen)	TS - Throat swab
CG18	Urethral swab (STD PCR)	258530009	Urethral swab (specimen)	US - Urethral swab
CG19	Vaginal swab (STD PCR)	258520000	Vaginal swab (specimen)	VS - Vaginal swab
CG20	Vulval swab (STD PCR)	258523003	Vulval swab (specimen)	Vulval swab

4.3 Identifying multiple SNOMED CT PaLM reportables for complex microbiology reports

Whilst a valid requirement, due to its singular nature, this delivers less overall value. The process involved is more difficult than for the above requirements and automated techniques

are yet to be fully assured. Automation would require a specialist mapping tool, and there is an argument that the same output could be more easily achieved by building map tables manually. Implementing the change into live use will likewise be difficult, due to testing and assurance requirements.



4.4 Additional requirements

4. recording provenance, governance, and assurance information of concept maps
5. automatic map maintenance
6. automatic creation of new SNOMED CT PaLM content

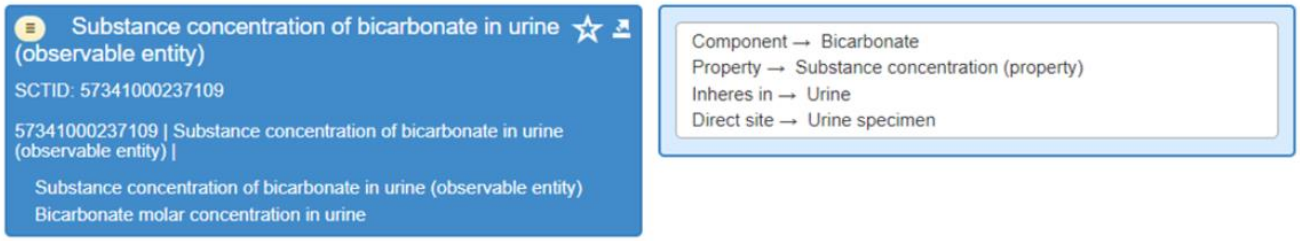
These requirements deliver value towards ongoing maintenance and development of SNOMED CT PaLM. **4** and **5** are straightforward to achieve and implement, using existing functionality and best practice. In contrast, **6** is considered a ‘stretch’ goal, as it is difficult to achieve and implement, requires technological solutions to be created, tested, and assured, and would impact existing best practice.

5. What makes a SNOMED CT PaLM reportable?

This section provides key information about SNOMED CT PaLM concepts.

5.1 Building blocks

A SNOMED CT PaLM reportable is a single data item in a pathology report identifying the specific lab test to which a value / interpretation can be assigned. SNOMED CT 'Observable entity' concepts are used, and it's useful to think of these concepts as asking a question; in this example, *'what is the result of the substance concentration of bicarbonate in urine test?'*



Key information about a lab test is carried inherently in a SNOMED CT PaLM reportable, with concepts carrying coded attributes relating to a test's property, component, specimen, and where stated, technique. These elements are also seen in the human-readable description of a SNOMED CT PaLM reportable's 'fully specified name'. By establishing the equivalent lab source data, one can subsequently identify the correct target SNOMED CT PaLM reportable to map to (**Error! Reference source not found.**).

5.2 Coverage

There are currently around 3,900 SNOMED CT PaLM reportables covering the following specialisms:

- Biochemistry
- Haematology
- Immunology
- Microbiology – this includes bacteriology, virology and serology
- Transfusion medicine

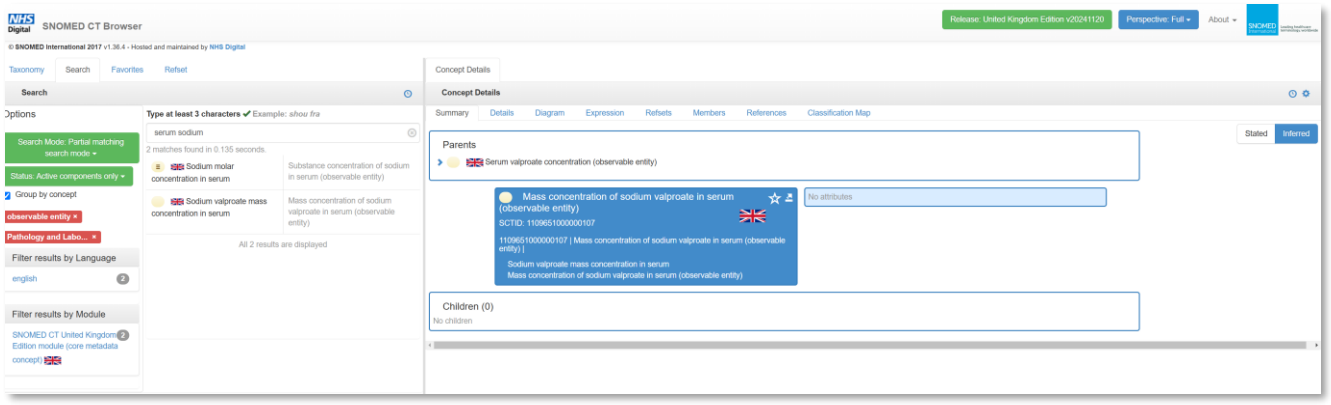
NB. Being pathology and laboratory medicine-specific, SNOMED CT PaLM does not cover genomics and diagnostic imaging.

5.3 How to access SNOMED CT PaLM

To aid implementation, all SNOMED CT PaLM reportables are published in a SNOMED CT reference set (Refset):

1853551000000106 Pathology and Laboratory Medicine observable entity simple reference set (foundation metadata concept)

Users can also browse SNOMED CT PaLM content using the search function in NHS England's SNOMED CT Browser.



For labs who wish to integrate SNOMED CT PaLM directly into their information systems, the Refset is included in the SNOMED CT UK Clinical Edition; available via NHS England's TRUD website, or via NHS England's Terminology Server (detailed information on how to access SNOMED CT content is available via NHS ENGLAND's Delen website).

However, most mapping tools enable this Refset to be accessed without the requirement to integrate SNOMED CT directly. Dependent on a tool's functionality, constraining map targets to this Refset may require the user to enter the following Expression Constraint Language (ECL) expression:
^ 1853551000000106 |Pathology and Laboratory Medicine observable entity simple reference set (foundation metadata concept)|

5.4 Single reportables vs 'groupers'

SNOMED CT PaLM reportables are all single reportables, i.e., they each represent a single lab test result code. Therefore, they do not represent panels/batteries, profiles, complex reports, or other such 'groupers'. Instead, multiple SNOMED CT PaLM reportables would be contained in such grouped pathology reports.

For example, a complex microbiology report such as a urinary microscopy, culture and sensitivities test report might contain several SNOMED CT PaLM reportables such as:

- Qualitative result of Staphylococcus aureus culture (observable entity)
- Susceptibility of organism to flucloxacillin (observable entity)

- Qualitative result of hyaline casts in mid-stream urine microscopy (observable entity)

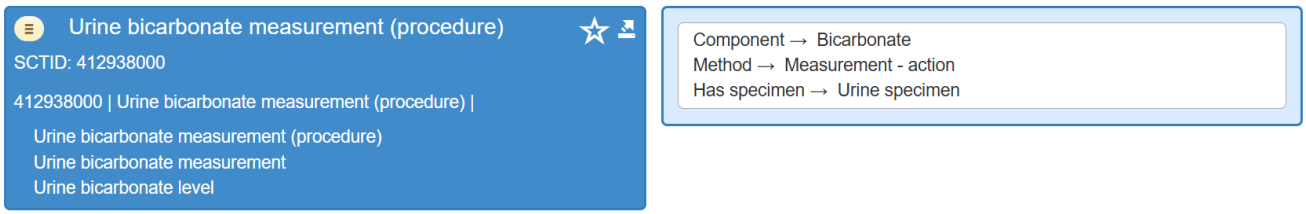
How such SNOMED CT PaLM reportables can be identified as appropriate for use in grouped reports depends on the type of report. For panels and profiles, this is straightforward, as local single reportables are already identified in the report and can be mapped accordingly. For complex microbiology reports, a method by which appropriate SNOMED CT PaLM reportables can be identified is detailed in Section 7.5.

5.5 Supplemental data elements in the Pathology FHIR Specification

To support the sharing of atomic, coded data in pathology reports, other data elements in the Pathology FHIR Specification that supplement SNOMED CT PaLM reportables can be populated with different types of SNOMED CT codes (e.g., lab test result values such as Positive (qualifier value). The mapping process involved is detailed in Section 7.4.

5.6 SNOMED CT PaLM test request codes (requestables)

In addition to reportables, SNOMED CT PaLM also provides concepts that represent lab test request codes ('requestables'). These are single data items in the report identifying the specific test(s) requested. SNOMED CT 'Procedure' concepts are used for this purpose.



To aid implementation, all SNOMED CT PaLM requestables are published in a SNOMED CT reference set (Refset):
1853561000000109 Pathology and Laboratory Medicine procedure simple reference set (foundation metadata concept)

NB. To date, the main focus of standardisation to support DAPB4101 has been on reportables, and the focus of this document is on mapping source terminology data that represents patient test results. However, similar mapping principles apply to mapping source requestable data.

6. Preprocessing source and target terminology

To optimise the mapping process, initial preprocessing of source and target terminology data is required. This involves data cleansing and applying computable rules to enhance the data, which can be facilitated by parsing terminology into component building blocks. This section explores these topics.

6.1 The key challenge

Underlying differences between human-readable descriptions applied to source and target reportables can make it difficult for mapping tools to establish correct target maps. When representing a reportable, for a lab's local use-case, a single word describing the lab test's component is often sufficient because users understand the additional context (i.e., a test's property, specimen, and technique). However, for a national use-case like SNOMED CT PaLM, a clear and unambiguous description is required, stating exactly what is being measured, how it will be measured, and where the sample was taken so that codes can be processed by third parties who do not understand the local context.

The table below illustrates the significant differences between typical local reportables' human-readable descriptions and those used SNOMED CT PaLM.

Hospital A Local Description	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
Creatinine	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum
Sodium	Substance concentration of sodium in serum (observable entity)	Sodium substance concentration in serum
Potassium	Substance concentration of potassium in serum (observable entity)	Potassium substance concentration in serum
Haemoglobin	Mass concentration of haemoglobin in blood (observable entity)	Haemoglobin mass concentration in blood
Albumin	Mass concentration of albumin in serum (observable entity)	Albumin mass concentration in serum
Platelets	Platelet count in blood (observable entity)	Platelet count in blood

Consequently, preprocessing of source data that serves to generate a SNOMED CT PaLM-like string for ingestion into a mapping tool is of great benefit because the closer the source data resembles a SNOMED CT PaLM description, the easier it is to map.

6.2 Data cleansing

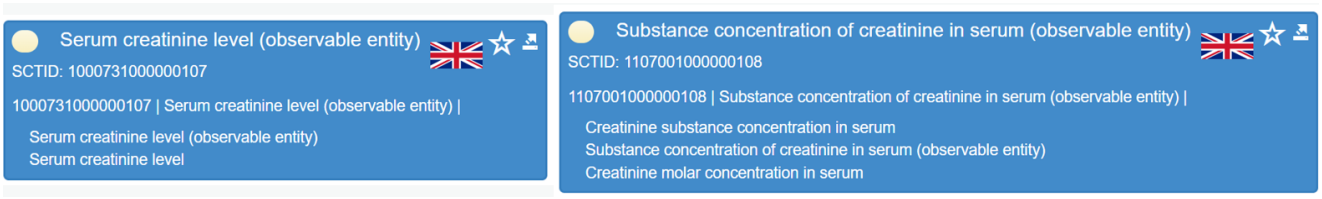
Data cleansing involves specifying the structure of the data, specifying translation rules, and enhancing the richness of the source terminology, and where required, the target terminology. Techniques for doing so are provided in section 6.5.

NB. SNOMED CT PaLM is an explicit, unambiguous target terminology, so no target terminology enhancements are required.

6.3 Parsing terminology into component building blocks

Parsing involves splitting source and target codes' human-readable descriptions into their component parts, which are equivalent to the coded attributes or building blocks described in section 3.2.1. This can be used to facilitate mapping by matching each of the parsed components in the source and target codes.

For example, the SNOMED CT PBCL concept "Serum creatinine level" can be parsed into the components: Property=Level, Component=Creatinine, Specimen=Serum. In the same way the SNOMED CT PaLM concept "Substance concentration of creatinine in serum" can be parsed into the components: Property=Substance concentration, Component=Creatinine, Specimen=Serum. The parsed components can be used to create a candidate map from the SNOMED CT PBCL concept "Serum creatinine level" to SNOMED CT PaLM concept "Substance concentration of creatinine in serum".



6.4 Value of source data cleansing

Many mapping tools use string based matching algorithms such as Levenshtein distance to establish map targets. Levenshtein distance measures the minimum edits (insertions, deletions, substitutions) needed to transform one string into another (see the Advanced Analysis Components to Support SNOMED PaLM Mapping Project paper for further details). Using this technique, the example below illustrates the value of source data cleansing to enhance a mapping tool's ability to establish correct map targets.

In this example, the local reportable **'Creatinine'** is to be mapped to SNOMED CT PaLM. The lab test's property and specimen have been established via other sources (6.5.1) and the correct map target should be Substance concentration of creatinine in serum.

Hospital A Local Description	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
Creatinine	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum

In its raw form, the number of changes between the source description and the target description is 37. In a list of SNOMED CT PaLM reportables that contain the string ‘creatinine’, the desired target appears fifth, clearly demonstrating how the source description alone is not enough to maximise the chances of getting a correct match.

rank	SNOMED PaLM description	Levenshtein distance
1	Urine albumin:creatinine ratio	21
2	Urine C-peptide/creatinine ratio	22
3	Creatinine renal clearance in 24 hours	28
4	Clearance ratio of calcium to creatinine	30
5	Substance concentration of creatinine in serum	37
6	Substance concentration of creatinine in fluid	37
7	Substance concentration of creatinine in urine	37

The table below demonstrates shows that when property and specimen information about the lab test is processed into the string, the Levenshtein distance decreases to zero, thereby establishing the appropriate target:

Information	**SNOMED CT PaLM description**	**Levenshtein distance**
Component	Creatinine	37
Component and Specimen	Creatinine in serum	28
Component and Property	Substance concentration of creatinine	9
Component, Property and Specimen	Substance concentration of creatinine in serum	0

The table below demonstrates how processing this additional information into source descriptions when testing the mapping of the ‘top 300’ reportables improved overall returns:

Source Information	**Correct map target**
Component	41%
Component and Property	58%

Component and Specimen	68%
Component, Property, and Specimen	70%

6.5 Techniques for source data cleansing

6.5.1 Establishing source data via existing Read PBCL / SNOMED PBCL maps

Leveraging the information in existing Read PBCL / SNOMED CT PBCL maps (7.3.2.2) provides an alternative means to establish appropriate source data.

Hospital Local Code	Hospital Local Description	Hospital UoM	Read PBCL Code	Read PBCL Description	SNOMED PBCL Concept ID	SNOMED PBCL Description	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
CREA	Creatinine	umol/L	44J3.	Serum creatinine	1000731000000107	Serum creatinine level	1107001000000108	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum

In this example, the component and UoM source data are already available as source data, whilst the specimen type **serum** can be determined from the mapped Read PBCL/SNOMED CT PBCL descriptions. The same process can help establish a lab test's property where a unit of measure in unavailable.

6.5.2 Using LLMs to specify the structure of the data, specify translation rules, and enhance the richness of the input terminology

A Large Language Model (LLM) is an advanced artificial intelligence system trained on vast amounts of text data to understand and generate human-like language.

In the context of mapping to SNOMED CT PaLM, testing found LLMs useful as a means to cleanse labs' source data and convert it into a format that facilitated mapping. This was achieved via LLMs ability to expand acronyms and extract key elements required for mapping (e.g., identifying a lab test's component from the source code description, the test's property from associated UoM data (7.2.2), and the specimen from the existing Read PBCL/SNOMED PBCL map (7.3.2.2). In combination, these elements were used to instruct the LLM to create a SNOMED CT PaLM-like string for ingestion into a mapping tool as source data.

A prompt is an instruction given to a LLM to guide it to a desired response. In this context, a prompt was used in testing to cleanse source data to create a SNOMED CT PaLM-like string and is available in the Appendix (12.3). To avoid LLM 'hallucination' there is a need to engineer the prompt as per the instructions below:

- Add context outlining what you are trying to achieve and the context for use
- Explain the inputs - including file names and column headings
- Explain the outputs - including column headings
- Explain the processing steps required to get to the output, which typically include:

1. Expanding any acronyms associated with a lab test
2. Identifying the lab test component from the source description
3. Identifying the lab test property from the associated UoM
4. Identifying the lab test specimen by:

I. Extracting from the existing Read PBCL/SNOMED CT PBCL map

II. Predicting the specimen when not otherwise available

III. Using a specimen mapping table

5. Identifying the technique
6. Combining the information into a form that can be processed by a terminology mapping tool
7. Keeping expanded forms of any acronym
8. Keeping names of any strings that do not meet the standard SNOMED CT PaLM pattern

7. Algorithmic mapping

The main recommended algorithmic techniques applicable to wider terminology mapping are listed in the bullets below and then exemplified and specified via the pathology and laboratory medicine terminology mapping use-case in subsequent subsections. These are the fundamental techniques the PaLM Mapping Project team investigated to optimise automated terminology mapping outputs during testing:

- Semantic and lexical mapping (7.1)
- Using terminology component building blocks (7.2)
- Configurable mapping to additional data elements that supplement source data (7.2.1)
- Using configurable terminology translation algorithms (7.2.2)
- Using referential terminology mapping artifacts (7.3)

This section details how these techniques can be used to facilitate mapping single reportables to SNOMED CT PaLM (i.e., mapping terminology descriptions pertaining to a single code between two terminology systems). However, the same underlying techniques can facilitate both mapping to supplemental data elements in the Pathology FHIR Specification (7.4) and identifying multiple SNOMED CT PaLM reportables for complex microbiology reports (7.5).

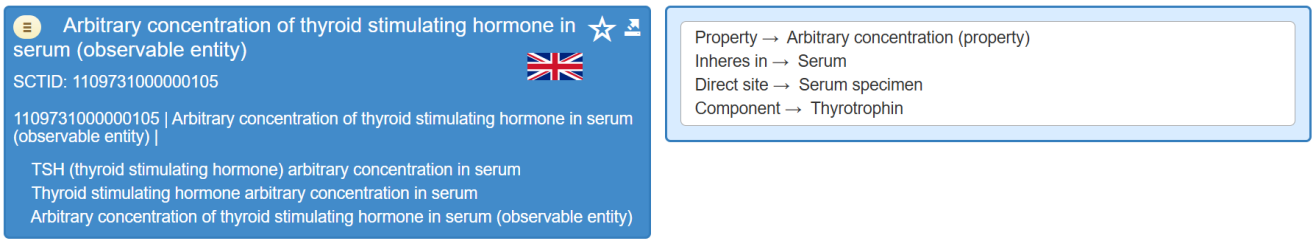
7.1 Semantic and lexical mapping of code descriptions

Most mapping tools use string based matching algorithms to establish map targets (see the Advanced Analysis Components to Support SNOMED PaLM Mapping Project paper for details).

In this example, due to the close similarity between source and target human-readable descriptions, the local reportable can easily be mapped to the SNOMED CT PaLM target.

Hospital Local Code	Hospital Local Description	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
ESR	Erythrocyte Sedimentation Rate	Erythrocyte sedimentation rate (observable entity)	ESR - erythrocyte sedimentation rate

As seen in the example below, SNOMED CT PaLM reportables carry more than one human-readable description: the ‘fully specified name’, the ‘preferred term’ (which often includes acronyms), and in some instances, additional synonyms. All these should be used to facilitate mapping.



Acronyms are prevalent in pathology and laboratory medicine, so the ability of mapping tools and LLM's to process acronyms is particularly valuable.

To assist review, a mapping tool can establish the level of equivalence between terms being matched (e.g., equivalent, more specific, less specific etc.), thereby offering a level of confidence in the validity of the proposed map.

7.2 Using terminology component building blocks

These techniques leverage the strategy defined in Section 3.2.1.

7.2.1 Configurable mapping to additional data elements that supplement source data

This example shows the source data relating to a lab's local 'serum creatinine' reportable.

Hospital UoM	Hospital Local Reportable code	Hospital Specimen code	Hospital Technique code
umol/l	creatinine	serum	n/a
property	component	specimen	technique

NB. **umol/L** defines the property as 'a **'substance concentration'** (7.2.2)

By configuring an algorithm and then semantically/lexically mapping these source data elements to SNOMED CT PaLM reportables' descriptions, or to the descriptions of the SNOMED CT concepts used in their modelling, as seen in the example below, this will establish the matching SNOMED CT PaLM reportable.

7.2.2 Using configurable terminology translation algorithms

During testing, the PaLM Mapping Project team created a map table of common UoM defined as lab test properties used in the modelling of SNOMED CT PaLM, as per the example below:

Common UoM	SNOMED PaLM Property attribute
%	percentage
10*12/L	count
g/L	mass concentration

Whilst not a clinically assured artefact, this table was used in testing as a referential artefact to facilitate mapping. Testing found that applying an algorithmic translation of UoM to lab test property enabled the mapping of large volumes of reportables, and when combined with a specimen type extracted via labs' source data or from the mapped Read PBCL description (6.5.1), this became a powerful method of identifying SNOMED CT PaLM map targets.

7.3 Using referential terminology mapping artifacts

This section explores how loading referential terminology concept maps into a tool can greatly enhance the automated mapping output.

7.3.1 Using other labs' curated maps to SNOMED CT PaLM

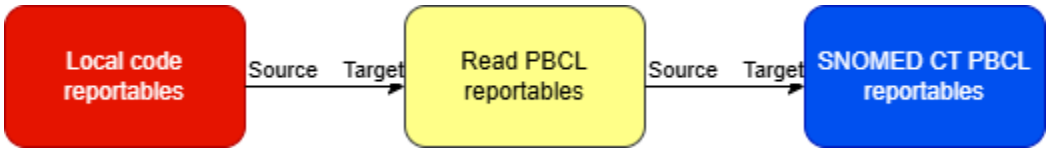
Where one labs' local reportables have been mapped to SNOMED CT PaLM, that information can help establish a second labs' mapping, especially when existing maps to Read PBCL/SNOMED CT PBCL are factored in.

In this example, Hospital A and Hospital B have mapped their local reportables to the same Read PBCL/SNOMED CT PBCL code. Hospital A are unsure which SNOMED CT PaLM reportable to map to, whereas Hospital B have established that their SNOMED CT PaLM map is Qualitative result of Salmonella species culture (observable entity). Consequently, Hospital B's map could be used to evaluate whether the same SNOMED CT PaLM reportable is appropriate for Hospital A.

Hospital A Local Code	Hospital A Local Description	Read PBCL Code	Read PBCL Description	SNOMED PBCL Description	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
CCUL	Culture Report	4J17.	Sample culture	Sample culture	?	?	?
Hospital B Local Code	**Hospital B Local Description**	**Read PBCL Code**	**Read PBCL Description**	**SNOMED PBCL Description**	**SNOMED PaLM Concept ID**	**SNOMED PaLM FSN**	**SNOMED PaLM Preferred Term**
3CSAL	Salmonella Culture	4J17.	Sample culture	Sample culture	44211000237105	Qualitative result of Salmonella species culture (observable entity)	Salmonella species culture qualitative result

7.3.2 Using local reportable to Read PBCL maps

Labs reporting results to general practice all use their own local reportable to Read PBCL maps to support the PMIP EDIFACT flow (2.4). The NHS England assured mapping table from Read PBCL to SNOMED CT PBCL is then used by GP systems to ingest these reports and record lab test result codes as SNOMED CT in the patient record.



Whilst there is a degree of variance in the quality of labs' existing Read PBCL maps (2.4), they remain a key resource to facilitate mapping to SNOMED CT PaLM. Leveraging the information contained in these maps can help in three ways:

7.3.2.1 Using Read PBCL / SNOMED CT PBCL to SNOMED CT PaLM maps

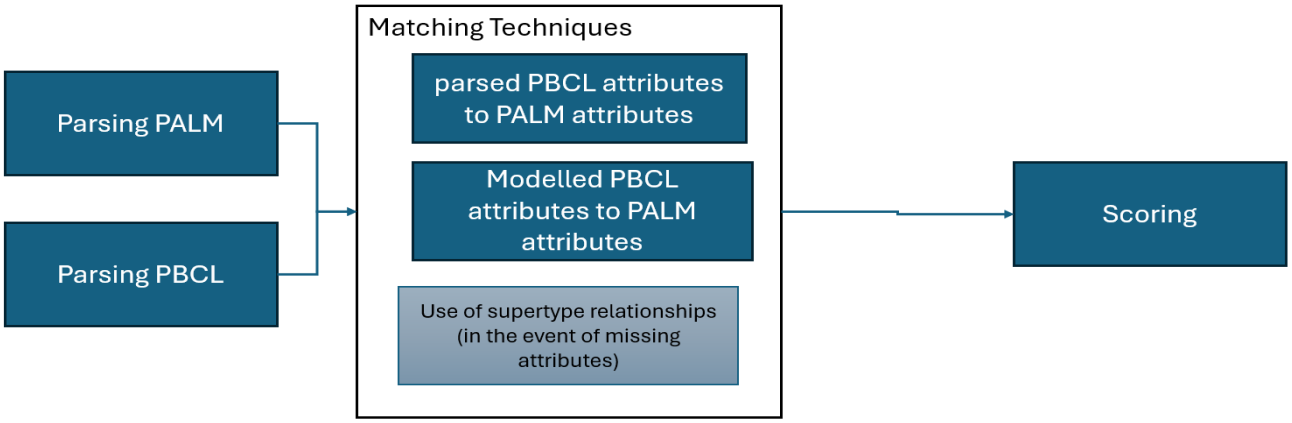
During testing, the PaLM Mapping Project team created a map table between Read PBCL / SNOMED CT PBCL and SNOMED CT PaLM.

Chart Position	Specialism	Read PBCL Code	Read PBCL Description	SNOMED PBCL Concept ID	SNOMED PBCL Description	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term	Alternatives (e.g., UoM agnostic, or where specimen is unclear)
1	Biochemistry	44J3.	Serum creatinine	1000731000000107	Serum creatinine level	1107001000000108	Substance concentration of creatinine in serum (observable entity)	Creatinine substance concentration in serum	
2	Biochemistry	44I5.	Serum sodium	1000661000000107	Serum sodium level	1107871000000107	Substance concentration of sodium in serum (observable entity)	Sodium substance concentration in serum	
3	Biochemistry	44I4.	Serum potassium	1000651000000109	Serum potassium level	1107761000000109	Substance concentration of potassium in serum (observable entity)	Potassium substance concentration in serum	
4	Haematology	423..	Haemoglobin estimation	1022431000000105	Haemoglobin estimation	1107511000000100	Mass concentration of haemoglobin in blood (observable entity)	Haemoglobin mass concentration in blood	Haemoglobin substance concentration in blood
5	Biochemistry	44M4.	Serum albumin	1000821000000103	Serum albumin level	1105861000000106	Mass concentration of albumin in serum (observable entity)	Albumin mass concentration in serum	
6	Haematology	42P..	Platelet count	1022651000000100	Platelet count	1108041000000107	Platelet count in blood (observable entity)	Platelet count in blood	
7	Haematology	42H..	Total white cell count	1022541000000102	Total white cell count	1110441000000100	White blood cell count in blood (observable entity)	White blood cell count in blood	White blood cell count in urine
		42H7.	Total white blood count						
8	Haematology	42A..	Mean corpuscular volume (MCV)	1022491000000106	MCV - Mean corpuscular volume	1491000237105	Mean corpuscular volume of erythrocytes in blood (observable entity)	Erythrocytes MCV (mean corpuscular volume) in blood	
9	Haematology	42J..	Neutrophil count	1022551000000104	Neutrophil count	1108071000000101	Neutrophil count in blood (observable entity)	Neutrophil count in blood	
10	Haematology	42M..	Lymphocyte count	1022581000000105	Lymphocyte count	67541000237108	Count of lymphocytes in blood (observable entity)	Lymphocyte count in blood	
11	Haematology	42K..	Eosinophil count	1022561000000101	Eosinophil count	1107391000000104	Eosinophil count in blood (observable entity)	Eosinophil count in blood	
12	Haematology	428..	Mean corpusc. haemoglobin(MCH)	1022471000000107	MCH - Mean corpuscular haemoglobin	1022471000000107	Mean corpuscular haemoglobin (observable entity)	MCH - Mean corpuscular haemoglobin	

This artefact was generated by repurposing a tool created to support the modelling of SNOMED CT PBCL and SNOMED CT PaLM that uses a bespoke set of semantic parsing instructions and a MySQL database.

The diagram below shows the process involved, with each SNOMED CT PBCL and SNOMED CT PaLM reportable's human-readable descriptions split into component parts (as per the example given in section 6.3) before algorithms were applied against SNOMED CT

attribute values used in the modelling of SNOMED CT reportables to identify appropriate map targets. These were then scored in terms of appropriate matching.



Whilst not a clinically assured artefact, these concept maps were used in testing as a referential terminology mapping artefact to facilitate mapping, achieved by using labs' local reportable to Read PBCL concept maps as a primary source input. The Read PBCL codes were subsequently mapped to SNOMED CT PBCL reportables, which in turn were mapped to SNOMED CT PaLM targets, as seen in the diagram below, a process further explored in Section 8.3.1.

7.3.2.2 Establishing source data via Read PBCL / SNOMED CT PBCL maps

As previously explored (6.5.1) Read PBCL / SNOMED CT PBCL maps provide a means to establish source data.

In this example, the local reportable 'PLT' / 'Platelets' has been mapped to Read PBCL / SNOMED CT PBCL. The additional word **'count'** in the Read PBCL and SNOMED CT PBCL descriptions establishes the lab test's **property**. This additional information can be used to semantically match to the SNOMED CT PaLM reportable.

Hospital Local Code	Hospital Local Description	Read PBCL Code	Read PBCL Description	SNOMED PBCL Concept ID	SNOMED PBCL Description	SNOMED PaLM Concept ID	SNOMED PaLM FSN	SNOMED PaLM Preferred Term
PLT	Platelets	42P..	Platelet count	1022651000000100	Platelet count	1108041000000107	Platelet count in blood (observable entity)	Platelet count in blood

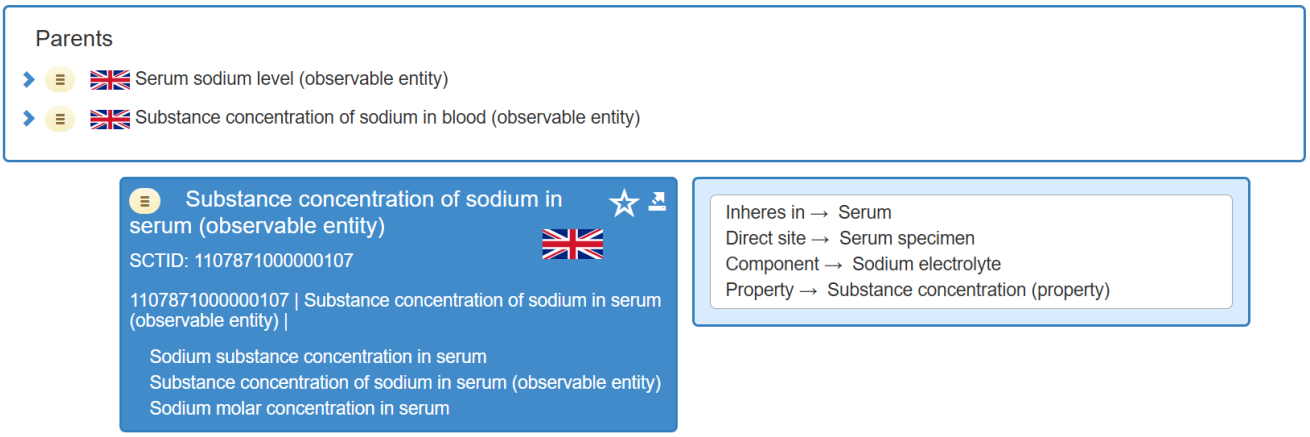
NB. Testing found this particularly powerful in helping to help determine specimen types.

7.3.2.3 Using SNOMED CT taxonomic relationships

In this example, the local reportable ‘NA’ / ‘Sodium’ has been mapped to Read PBCL and the equivalent SNOMED CT PBCL reportable Serum sodium level (observable entity).

Hospital Local Code	Hospital Local Description	Read PBCL Code	Read PBCL Description	SNOMED PBCL Concept ID	SNOMED PBCL Description
NA	Sodium	44I5.	Serum sodium	1000661000000107	Serum sodium level

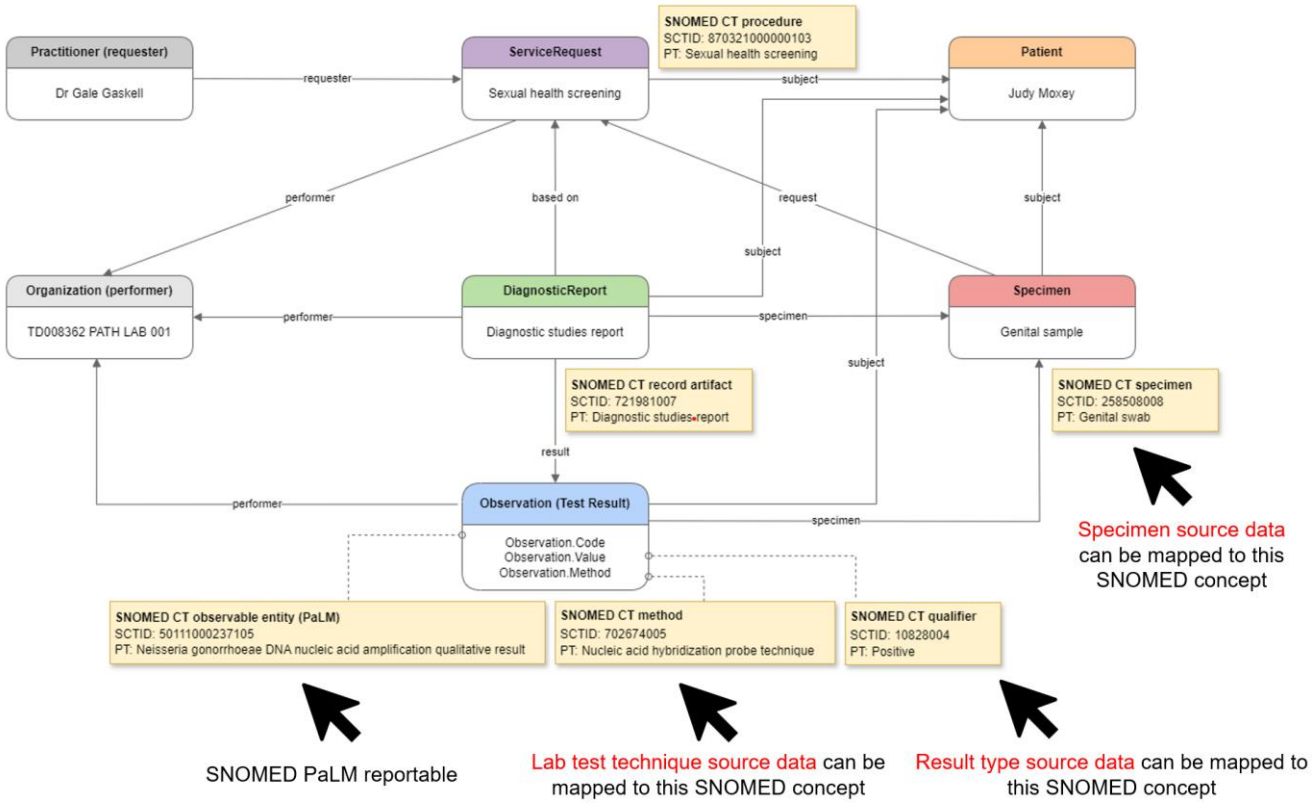
Via the hierarchical SNOMED CT ‘Is a’ relationship, the SNOMED CT PaLM reportable Substance concentration of sodium in serum (observable entity) is a ‘child’ of Serum sodium level (observable entity) as seen in the diagram below:



Consequently, via algorithms that leverage SNOMED CT’s logical concept model, it can be established that the SNOMED CT PaLM concept is a potentially suitable map target. This is particularly useful as a secondary check to compliment other mapping techniques.

7.4 Techniques for mapping to supplemental data elements in the Pathology FHIR Specification

This diagram illustrates the various data elements contained in a DAPB4101 conformant Neisseria gonorrhoeae DNA detection test report, as defined by the Pathology FHIR specification. The SNOMED CT PaLM reportable Qualitative result of Neisseria gonorrhoeae deoxyribonucleic acid nucleic acid amplification (observable entity) represents the lab test result code.



Reportables such as these do not inherently carry information about the **specimen** and **lab test technique**. Consequently, to ensure this coded information is carried in the report, lab source data representing the specimen and technique can be mapped to SNOMED CT **Specimen** and **Procedure** concepts respectively. Likewise, lab source data representing the **result type** can be mapped to an equivalent SNOMED CT **Qualifier value** concept.

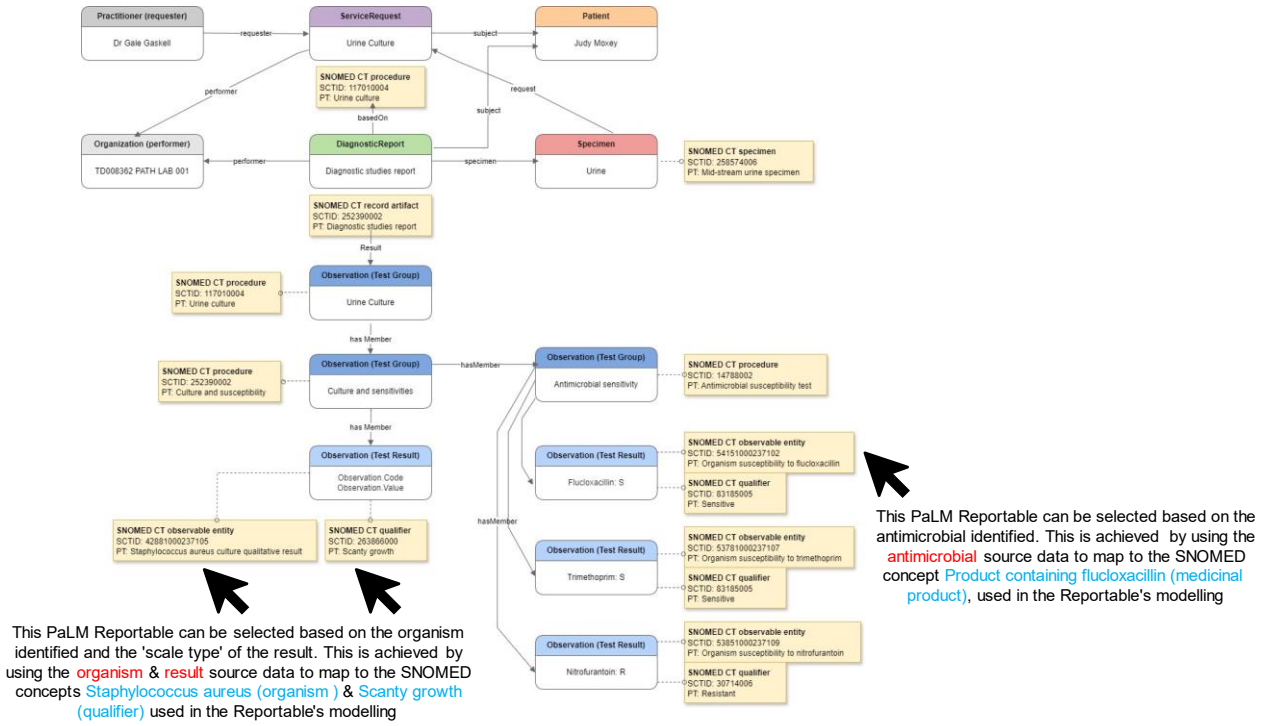
Although the mapping techniques remain the same, the process involves creating target value sets of applicable SNOMED CT concepts, necessitating access to wider SNOMED CT content than the PaLM (Pathology and Laboratory Medicine) observable entity simple reference set. How to access this content is similar to the process described in Section 5.3, and likewise, mapping tools can facilitate this. Dependent on the tool's functionality, constraining map targets may require the use of simple ECL expressions, e.g., < 123038009 |Specimen (specimen)|

7.5 Techniques for identifying multiple SNOMED CT PaLM reportables for complex microbiology reports

As described in section 5.4, SNOMED CT PaLM reportables are single reportables, and multiple SNOMED CT PaLM reportables are contained in complex microbiology reports.

The diagram below illustrates the various data elements contained in a DAPB4101 conformant urine culture and sensitivities complex report, as defined by the Pathology FHIR specification, and includes several SNOMED CT PaLM reportables. The challenge for labs is to identify which SNOMED CT PaLM reportables are appropriate for inclusion.

This can be done by mapping ancillary lab source data used in the report to SNOMED CT concepts used in the modelling of SNOMED CT PaLM reportables (these concepts being the reportables' 'attribute values'), then applying algorithms that leverage SNOMED CT's concept model to identify which SNOMED CT PaLM reportables use this particular modelling, thereby identifying them as appropriate for inclusion.



In this example, lab source data representing the **organism** and **result** can be mapped to SNOMED CT **Organism** and **Qualifier value** concepts to help identify the SNOMED CT PaLM reportable Qualitative result of Staphylococcus aureus culture (observable entity). Likewise, lab source data representing the **antimicrobial** can be mapped to SNOMED CT **Medicinal product** concepts to help identify the SNOMED CT PaLM reportable Susceptibility of organism to flucloxacillin (observable entity).

The process involves creating target value sets of applicable SNOMED CT concepts (7.4) and subsequently applying algorithms. E.g., by first mapping **flucloxacillin** source data to the SNOMED CT concept Product containing flucloxacillin (medicinal product) an algorithm can then identify all SNOMED CT PaLM reportables where Product containing flucloxacillin (medicinal product) is used as the Towards (attribute) value; thereby identifying the SNOMED CT PaLM reportable Susceptibility of organism to flucloxacillin (observable entity).

8. Mapping creation, assurance, and governance

The mapping workflow involves creation, review, and maintenance. For a given clinical domain, it is recommended that the review and assurance of automated mappings is undertaken by subject matter experts (SMEs) and that all information regarding review and ownership of mapping is captured.

For the pathology and laboratory medicine terminology mapping requirement, as labs are ultimately responsible for their data, they must clinically assure their own maps to SNOMED CT PaLM, so would always be involved in the review, but other parts of the process could be performed or supported, in part or in full, by third parties.

8.1 Roles and responsibilities

The personnel required for a mapping project could include:

- Mapping service
- Clinical / lab IT SMEs
- Mapping support service

It is important to note that the involvement and capacity of the mapping service and mapping support service can be minimised in proportion to the capabilities and functionality built into mapping software to empower the lab user to drive the process.

8.1.1 Responsibilities of mapping service

The aim of this service is to create the maps and manage the process from start to finish. Ideally, it should minimise the workload placed upon clinical SMEs so they can focus on assurance and clinical sign off.

Pathology lab IT Managers are well placed to provide this service as they understand local processes and procedures and how these are reflected in their labs' reportable data. Alternatively, a central NHS England service or commercial vendor could provide it.

8.1.2 Responsibilities of clinical / lab IT SMEs

Pathologists / Biomedical Scientists currently working in labs are required to review and assure maps created and provide clinical sign off.

Moreover, if existing maps to Read PBCL are used as a strategy to map to SNOMED CT PaLM (7.3.2), they are able to provide assurance that these maps are current and correct.

8.1.3 Responsibilities of mapping support service

To deal with the technical demands of more complex mapping, commercial mapping software may be required. The vendor may also offer a mapping support service, providing custom algorithm generation and custom workflows to facilitate mapping and subject matter expertise in terminology mapping.

8.2 Change management and version control

As detailed in section 9, maps will change over time. Two elements of functionality that support this are change management and version control. Change management supports clinical audit, as it enables users to see changes to maps, when these were made, and by whom. Version control underpins change management and supports the production process between authoring, review, publishing and maintenance, as users can see the status of a map at a particular point in time.

8.3 Managing the mapping process

This section explores some of the issues involved in the mapping process via examples of using two different techniques to map a lab's local reportables to SNOMED CT PaLM.

8.3.1 Managing the process of using referential terminology maps

This process involves a chain of referential terminology maps (7.3). Firstly, local reportable to Read PBCL maps, then NHS England's assured Read PBCL to SNOMED CT PBCL maps, and finally SNOMED CT PBCL to SNOMED CT PaLM maps to reach appropriate SNOMED CT PaLM map targets. As such, it requires an appropriate level of assurance to ensure reliability and minimise errors.

As a chain of maps are key to the output, as seen in the diagram below, this increases the requirement for manual and technical assurance by SMEs at appropriate steps to address sources of error.

- Local code to Read PBCL maps

The most likely sources of error are the maps between local codes and Read PBCL. During testing, the PaLM Mapping Project team found both duplicate local codes and variation in local code to PBCL mappings between Pathology Networks. As noted in section 2.4, errors could have been introduced at inception, particularly if the curation and assurance process was not captured. Further variation may have been introduced if either informatic or clinical

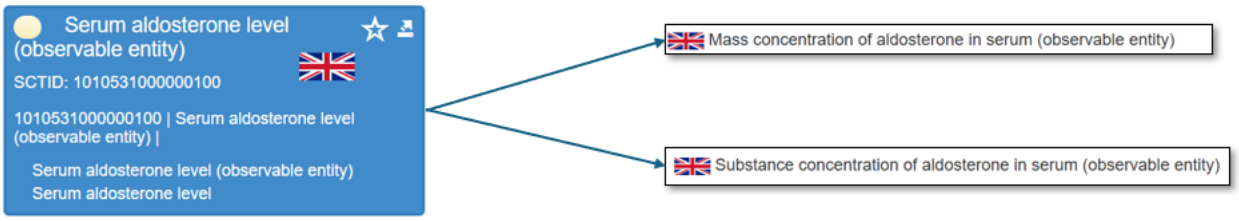
processes have changed since the maps were created. Consequently, the maps might not reflect current practice and so require clinical review.

- Read PBCL to SNOMED CT PBCL maps

As NHS England's assured, many to one maps are used, this step should not require clinical review but may require technical review to ensure mapping has occurred correctly.

- SNOMED CT PBCL to SNOMED CT PaLM maps

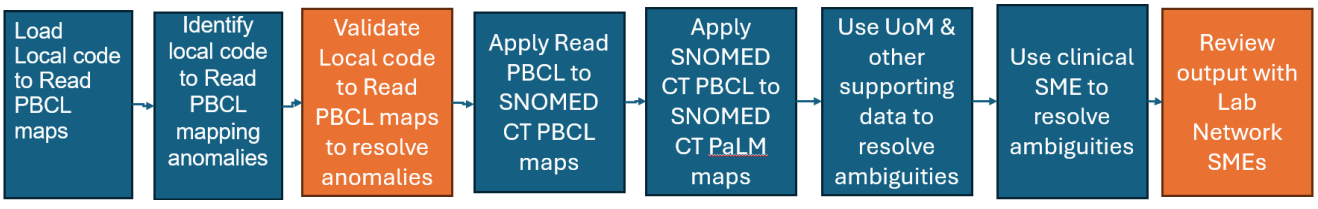
One to one mapping between SNOMED CT PBCL reportables and SNOMED CT PALM reportables is not always possible. As an example, SNOMED CT PBCL content often uses the unspecific term "level" to represent a lab test property, something interpreted and represented in more granularity in SNOMED CT PaLM. A typical consequence of this is that such SNOMED CT PBCL reportables could map to either "substance concentration" or "mass concentration" SNOMED PaLM reportables, as per the example below:



Consequently, some of these ambiguities require resolving via UoM data (or other supporting data), whilst others will require clinical subject matter expertise to resolve.

The mapping process would therefore involve:

- Load local code to Read PBCL maps
- Identify local code to Read PBCL mapping anomalies*
- Validate local code to Read PBCL maps to resolve anomalies with SMEs
- Apply Read PBCL to SNOMED CT PBCL maps
- Apply SNOMED CT PBCL to SNOMED CT PaLM maps
- Use UoM and other supporting data to resolve ambiguities*
- Use clinical SME to resolve ambiguities
- Review output with laboratory network SMEs



*These validation steps could be performed algorithmically.

8.3.2 Managing the process of using terminology component building blocks

This process involves using source data to create terminology component building blocks (7.2), to generate SNOMED CT PaLM-like strings for ingestion into a mapping tool, that will in turn map to SNOMED CT PaLM targets.

In this example, the source data is gleaned from:

- Property – Lab source data
- Component – Local code description
- Specimen – Mapped Read PBCL / SNOMED CT PBCL description

Whilst this method is generally more reliable than relying on using referential terminology maps alone, there are some instances where it can prove problematic, requiring manual and technical assurance from SMEs.

- Missing specimen type

Where a specimen type cannot be derived from the Read PBCL / SNOMED CT PBCL description, expert domain knowledge is required to predict the mostly likely option. For example, the Read PBCL code “Epithelial cell count” (4KH0.) is generally measured in urine, whilst the Read PBCL code “Lymphocyte count” (42M..) is generally measured in blood.

- Where there is an incorrect map between the Local code and the Read PBCL code

This requires clinical validation and fixing.

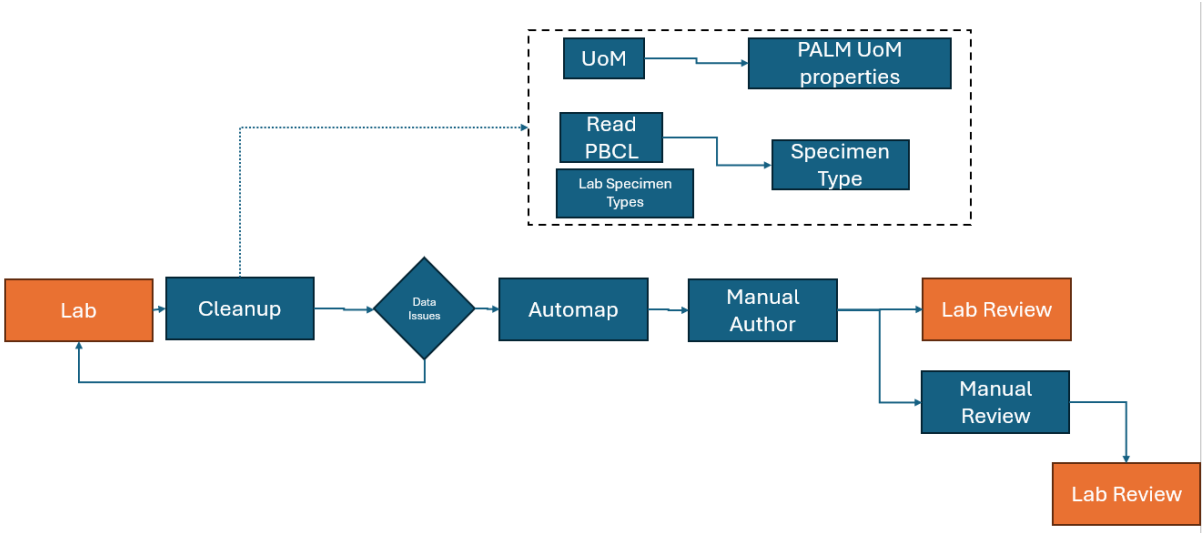
- Where the mapped Read PBCL code is broader in meaning than the local code (e.g., certain microbiology reportables)

This requires clinical validation and fixing.

The mapping process would therefore involve:

- Collate all required source data
- Cleanse the data
- Rectify any data issues and combine into ‘master codes’ with SNOMED CT PaLM-like strings “{property} of {component} in {specimen}”
- Load ‘master codes’ as source data
- Map to SNOMED CT PaLM

- Use clinical SME to resolve ambiguities
- Review output with SMEs



9. Mapping maintenance and implementation

This section looks at considerations for map maintenance and implementation. In addition to the pathology and laboratory medicine terminology mapping requirement, these are widely applicable to mapping other national interoperable terminologies. Key considerations are:

- End-user mapping maintenance (9.1)
- Architectural requirements (**Error! Reference source not found.**)
- Implications for standardising mapping development and implementation – specific guidance for the lab domain (9.3)

9.1 End-user mapping maintenance

As a national list of reportables, SNOMED CT PaLM will grow as new lab tests are added, and changes will be made through ongoing quality assurance. Therefore, it is not a static artefact. Consequently, maps to SNOMED CT PaLM should not be static, and will need updating as SNOMED CT PaLM changes. Moreover, if a lab's practices change (e.g., a test that was performed in plasma changes to being performed in serum), the map to SNOMED CT PaLM will need to change.

Therefore, to ensure data integrity it is necessary to

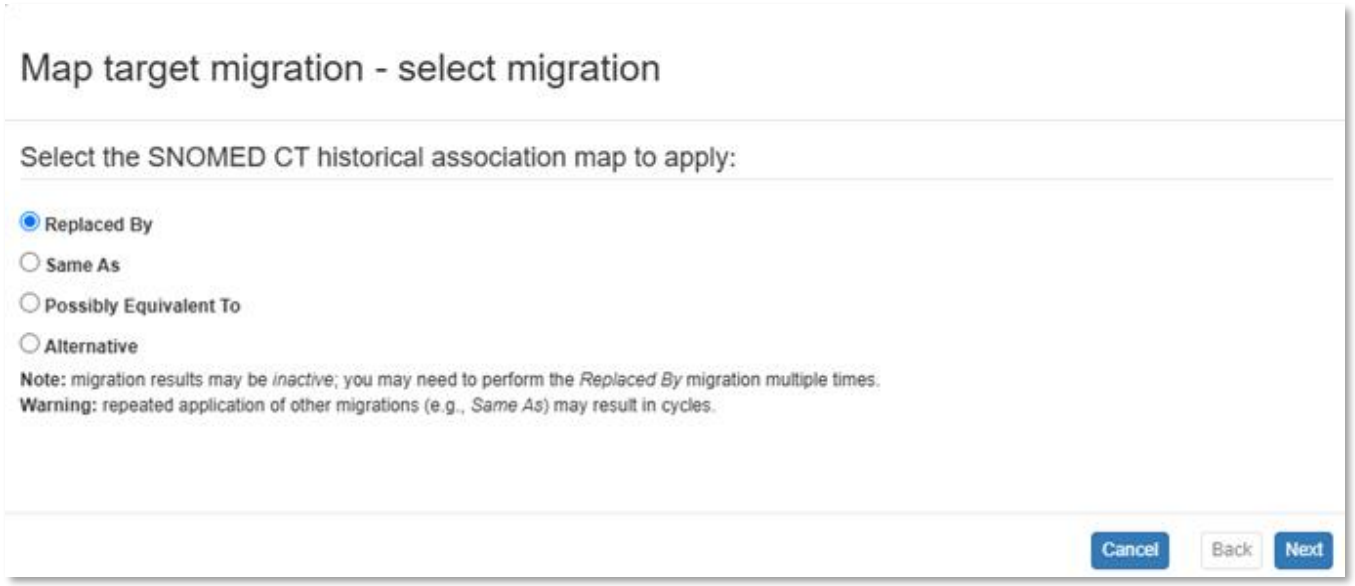
- Replace retired SNOMED CT PaLM concepts with their replacements
- Map to more suitable content created in later versions of SNOMED CT PaLM

9.1.1 SNOMED CT Historical association reference set mechanism

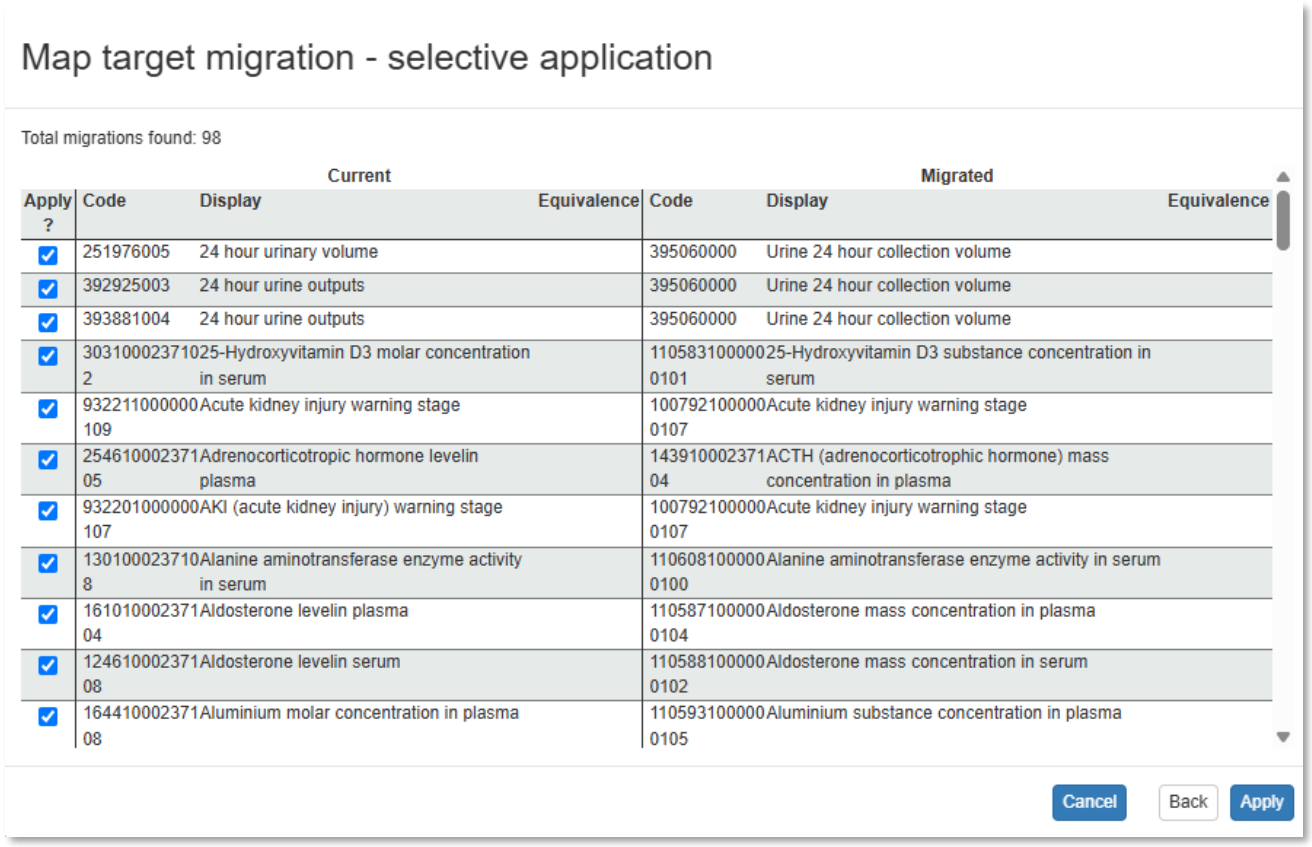
Finding replacement SNOMED CT content can be done through SNOMED CT's Historical association reference set. This technical artefact enables users to find replacements for retired components (in this case, SNOMED CT PaLM reportables) by detailing the association between them. Details on these associations are available from SNOMED International.

The key considerations are that the majority of the changes will be categorised as 'SAME AS'. In such cases, this generally doesn't require review (8.1.2) as the replacement concept is identical to the one retired. Other relationship types such as 'POSSIBLY EQUIVALENT TO' do require more consideration and would need to be reviewed.

Mapping tools can leverage the information with the Historical association reference set to automate maintenance, as shown in the example below:



A user might choose to automate the changes (where the 'SAME AS' association is in effect) or manually review, allowing the user to choose a different SNOMED CT PaLM reportable if appropriate, as shown in the example below:



For governance purposes, the tool should then record the changes with an audit history so users can understand what changes were made, when, and by whom. Any changes would

need to be captured and recorded for governance and provenance before being rolled out in live systems.

9.2 Architectural requirements

The architectural requirements became more visible as the PaLM Mapping Project team worked through the functional and technical requirements of terminology mapping tooling. These requirements are mainly specific to:

- Data loading
- Maintaining latest releases of interoperable terminology code systems
- Maintaining and sharing concept maps and value sets for referential terminology optimisation, training language models to support mapping, and for lab reuse and implementation
- The capability to point to other SNOMED CT extensions to glean available content to pull into the national SNOMED CT release instead of duplicating creation of SNOMED CT content
- Publishing and releasing concept maps and value sets in the international standard FHIR data model format

Details of the architectural design can be found in Appendix 12.4

9.3 Implications for standardising the development and implementation of mapping – specific guidance for the lab domain

There are two main options for developing, maintaining, and extending local terminology to standard interoperable terminology mappings for users to consider. These options are based on the belief that one-to-many mappings are an implementation burden and introduce risk, as this lends itself to inconsistencies between labs.

1. Where there is ambiguity in historical terminology mappings, and many-to-one mappings (e.g., local reportable to Read PBCL maps), labs should create new local reportables to map to more granular SNOMED CT PaLM reportables.
   As part of this process, it is recommended to cluster duplicate / replicated codes (both source and target codes) for the expert reviewer to quickly identify any potential historical errors / duplicates / inconsistencies in an individual lab's code list, or a Pathology Network's combined code list, and to remap the old many-to-one mappings to specific SNOMED CT PaLM reportables.

If there is no available SNOMED CT PaLM reportable, labs should not apply a generic code mapping. Instead, labs should request new SNOMED CT PaLM content to accurately match the source content.

2. Labs should standardise data entry lists as structured terminology component building blocks (3.2.1) that render as granular SNOMED CT PaLM content in the reported interoperable message. I.e., create a professionally led system input standard for the ‘minimum information about a lab test result’ per lab discipline. This standardises the interoperable ‘synthetic primary key’ created in the middleware that auto-maps to SNOMED CT PaLM (**Appendix 12.4)**
NB. This is arguably the more challenging and costly approach compared to option 1, as it requires system interface changes along with a professional standardisation effort.

These options are vital to successful terminology mapping implementation because for the pathology and laboratory medicine terminology mapping requirement, and other use-cases, they ensure remapping to more specific reusable, comparable SNOMED CT terminology.

As a case in point, SNOMED CT PaLM removes ambiguities around specimen types. A prime example being that a coded blood specimen no longer represents blood materials like acellular blood, plasma, or serum; as seen in legacy representations of lab data such as Read PBCL. This makes old representations of ‘whole blood’ redundant, because in SNOMED CT PaLM, ‘blood’ means ‘whole blood’. Blood specimens are differentiated and specified in the result types of SNOMED CT PaLM and so must be mapped to correspond with the exactness of the source lab analyser’s output. This precision is key to enhancing the accuracy and comparability of patient test result information, both by human interpretation and machine-readable interpretation.

10. Mapping software

Terminology mapping is a labour-intensive process, requiring both technical knowledge and subject matter expertise to ensure the quality and clinical safety of the output. Terminology mapping software leverages the power of automation to reduce effort and provide functionality to support workflow and governance.

10.1 Mapping tools

Several specialist tools are available that offer good technical mapping capability and good usability. These tools use a variety of techniques to support mapping (7) and can establish the level of equivalence between terms being matched, thereby offering a level of confidence in the validity of the proposed map (7.1). Mapping tools' also offer flexible functionality to support workflow and governance (8) and can facilitate access to target terminologies and code sets (5.3).

10.2 Large Language Models

Section 6.5.2 details how LLM's are valuable in cleansing source data. However, as LLM are probabilistic in nature (3.2), during testing, when asked to perform more deterministic tasks where the outcome is fixed (i.e., mapping local reportables to SNOMED CT PaLM) LLM's were found to be unreliable, often 'hallucinating' map targets, even when specifically prompted not to. Consequently, LLM's were deemed less suitable for mapping than other methods.

10.3 Mapping tool requirements specification

Detailed requirements for building a terminology mapping tool that supports the pathology and laboratory medicine terminology mapping use-case are documented in the SNOMED CT PaLM Mapping Project Requirements Specification. The requirements were developed in parallel to testing a variety of mapping strategies and mapping tools.

The specification has been published together with an options paper on SNOMED CT PaLM mapping for NHS England Programmes and Pathology Networks to consider, available upon request to authorised staff via pathology.standardsandimplementation@nhs.net.

11. Next steps

The work undertaken by the PaLM Mapping Project team has focused on mapping structured and semi-structured source coded data to standardised interoperable coded data to meet the pathology and laboratory medicine terminology mapping requirement. Essentially, this involves the replacement of legacy Read PBCL terminology with SNOMED CT PaLM.

Next steps would be to expand the investigation into technologies that could standardise unstructured data and clinical narrative text as interoperable structured coded data. NHS England's Data Science Team have already been investigating applying techniques such as named entity recognition to histopathology narratives to extract and tokenise key information in the patient test report. Their Advanced Analysis Components to Support SNOMED PaLM Mapping Project paper explains 'extension beyond current requirements' for the mapping next steps. Such techniques could potentially be integrated into the terminology mapping process to enhance the automated conversion of such data to corresponding SNOMED CT structured coding.

Investigating and applying new technologies emerging in data science and AI to national implementation of terminology mapping such as the pathology and laboratory medicine terminology mapping requirement, that optimise automation coupled with governance and clinical assurance functionality is predicted to further alleviate technical stress and release clinical capacity.

Please direct any enquiries about the pathology and laboratory medicine terminology mapping requirement to pathology.standardsandimplementation@nhs.net

12. Appendix

12.1 Definitions of Data Elements used in DAPB4101

- **Laboratory test report** - this contains all the data items relating to a laboratory test(s), represented by a FHIR UK Core R4 bundle.
- **Laboratory test result code ('reportable')** - a single data item contained in the report identifying the specific laboratory test to which a value(s) and interpretation can be assigned, represented by SNOMED CT Observable entity concepts.
- **Laboratory test request code ('requestable')** - a single data item contained in the report identifying the specific laboratory test(s) initially requested that led to the generation of results. These will also represent test panels/batteries within the reports. represented by SNOMED CT Procedure concepts.
- **Report identifier** - a single data item contained in the report identifying the FHIR message as a diagnostic studies report. To be represented by the SNOMED CT concept 721981007 Diagnostic studies report (record artifact).
- **Laboratory test result value** – a single data item contained in the report, often accompanied by a unit of measure, representing the information determined by making the observation. Currently represented by numerics/text strings, to be represented by numerics or codes contained in FHIR UK Core value sets where possible or, where appropriate, text strings.
- **Unit of measure (UoM)** - a single data item contained in the report accompanying a laboratory test result value. Currently represented by text strings, to be represented by Unified Code for Units of Measure (UCUM) codes where possible or, where appropriate, text strings.
- **Laboratory test result interpretation** - a single data item contained in the report, representing a categorical assessment of an observation value (high/low/normal for example). Currently represented by text strings, to be represented by codes contained in FHIR UK Core value sets where possible or, where appropriate, text strings.
- **Laboratory test specimen** – a single data item contained in the report identifying the specimen upon which the test was performed. Currently represented by text strings, to be represented by SNOMED CT Specimen/Substance/Morphologic abnormality/Body structure/Physical object concepts, or where appropriate, text strings.

12.2 Forms of Test Results

Test results are reported using a variety of forms:

- **Quantitative result**

The result is expressed as a number, usually with an associated unit of measure. Comparators may be used to indicate that the actual value is greater than or less than the stated value. A range of values may be reported instead of a single value.

Examples of quantitative test results include:

- Mass concentration of albumin in serum: 47 g/L
- eGFR (estimated glomerular filtration rate) using creatinine Chronic Kidney Disease Epidemiology Collaboration equation per 1.73 square metres: >90 mL/min/1.73m2

- **Semi-quantitative result**

The result is expressed using a descriptor to indicate the relative degree of positivity or negativity based on a scale. The scale is not always formally defined. Semi-quantitative results are widely used in microbiology, particularly for reporting microscopy and culture test results to indicate the amount or level of growth of organisms.

Examples of semi-quantitative test results include:

- Organism susceptibility to nitrofurantoin: RESISTANT

- **Qualitative result**

The result is expressed using a descriptor to indicate positivity or negativity. The descriptors are usually defined as pairs, for example: positive/negative, detected/not detected, isolated/not isolated. In this respect, qualitive results can be seen as a subset of semi-quantitative results in that they contain only two scale points to indicate positivity or negativity.

Examples of qualitative test results include:

- Qualitative result of Hepatitis B surface antigen in serum: NEGATIVE
- Qualitative result of methicillin resistant Staphylococcus aureus ribonucleic acid nucleic acid amplification: NOT DETECTED

- **Quantitative result combined with an interpretation**

In some cases, a result may contain a combination of quantitative and non-quantitative elements. The non-quantitative element is sometimes expressed separately as an interpretation to provide a categorical assessment of the quantitative value.

Examples include:

- Count of lymphocytes in blood: 0.70 10*9/L: LOW
- Arbitrary concentration of Rubella immunoglobulin G antibody in serum: >10 IU/ml: DETECTED

- **Narrative result**

The result is presented as text.

For example:

- Aerobic culture: No growth detected after 5 days incubation

12.3 Prompt used in testing for LLM source data cleansing

Objective

You are purposed with breaking down a set of local lab pathology data into its component elements in order to provide an input to map to a set of SNOMED CT concepts which will be provided. The model should be able to enable us to break down the concepts in order to help another tool generate highly accurate mappings while maintaining clarity and confidence in its suggestions.

Input Requirements

The GPT will receive the local laboratory codes in the spreadsheet “GPTImportSheet”. The format is:

- Local Code: Unique identifier for the local lab test (e.g., CREA)
- Local Description: Description of the test (e.g., “Creatinine”)
- UoM: Unit of measure (e.g., “umol/L”)
- Read PBCL Code (optional): Legacy Read PBCL code (e.g., 44J3.)
- Read PBCL Description (optional): Description corresponding to the Read PBCL code

Output Requirements

Deepseek will return the output as a table with the following column headers:

- Local Code: From Input Unique identifier for the local lab test (e.g., CREA)
- Local Description: From Input Description of the test (e.g., “Creatinine”)
- UoM: From Input Unit of measure (e.g., “umol/L”)
- Read PBCL Code (optional): From Input Legacy Read PBCL code (e.g., 44J3.)
- Read PBCL Description (optional): From Input Description corresponding to the Read PBCL code

Deepseek will interpret the above and populate the following fields:

- Component
- Property
- Sample
- Combined

Component field

This is the specific component which we are measuring. This can be a substance such as calcium, protein such as albumin, a cell such as lymphocyte. This information is generally in the local description or Read PBCL description. Some processing of these fields can be required such as expanding any acronyms, such as MCH for mean corpuscular haemoglobin.

Property field

This is the measurement technique which will be used to measure the component examples, which can be substance concentration, molar concentration, percentage etc. The technique which can be used can be obtained through the use of the Units of Measure Field.

There is a sample mapping table between the UoM and their properties in "Units_of_measure"

Sample field

This is the sample in which the component is measured and can be blood, serum, urine etc. This information can be picked up from the local description or Read PBCL description. Where no example exists then place a guess based on the most likely substance, for example "blood" for "basophils" or urine for epithelial cells.

Use only the fields found in "Direct Site Attributes in PALM-20250225-183230"

Combined field

Combine the other three calculated fields into the following format "{property} of {component} in {sample}" for example "Substance concentration of magnesium in serum" or "Count of lymphocytes in blood". Keep the expanded form of the acronym in this field.

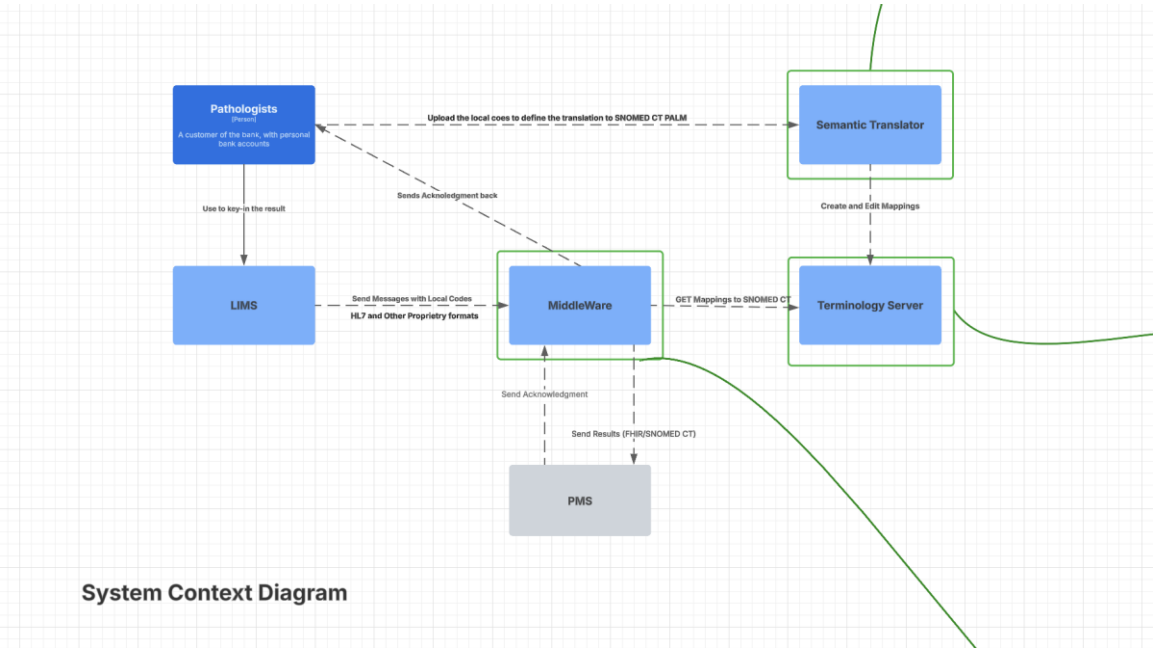
12.4 Ontology and semantic translation architecture

12.4.1 Context Diagram:

At the highest level, the system consists of:

- Labs: Send HL7 messages with local codes.
- Middleware: Translates local codes to SNOMED CT using a Terminology Server.

- Terminology Server: Provides SNOMED CT mappings.
- GP Systems: Receive the translated messages.
- UI Component: Allows pathologists to manually map local codes to SNOMED CT.

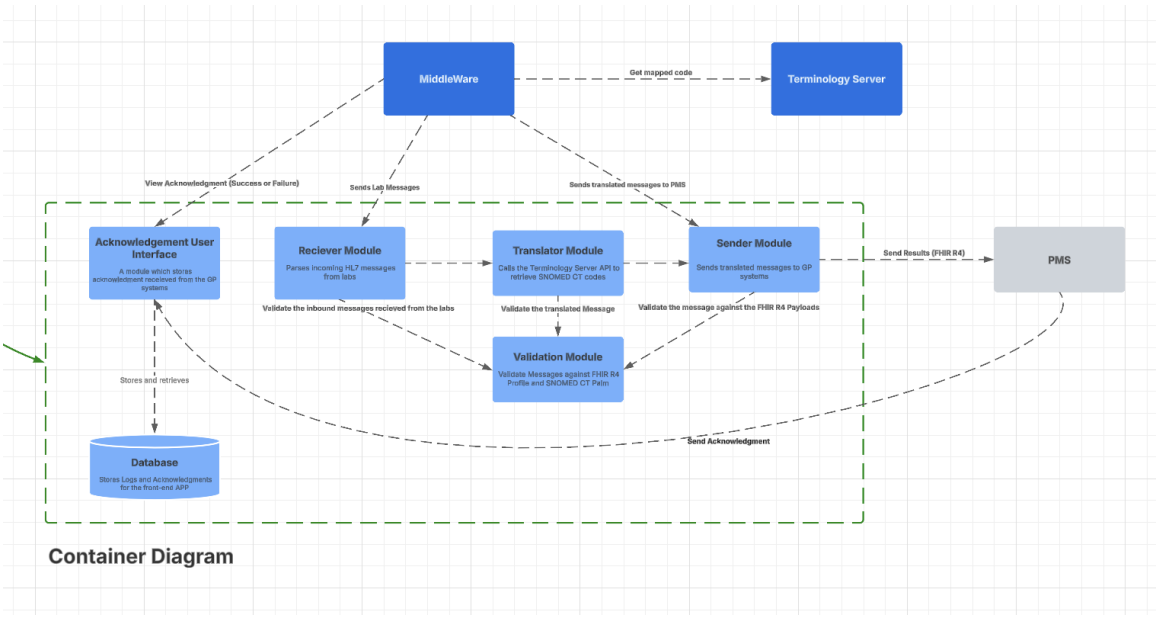


12.4.1 Container Diagram

Breakdown of core system components:

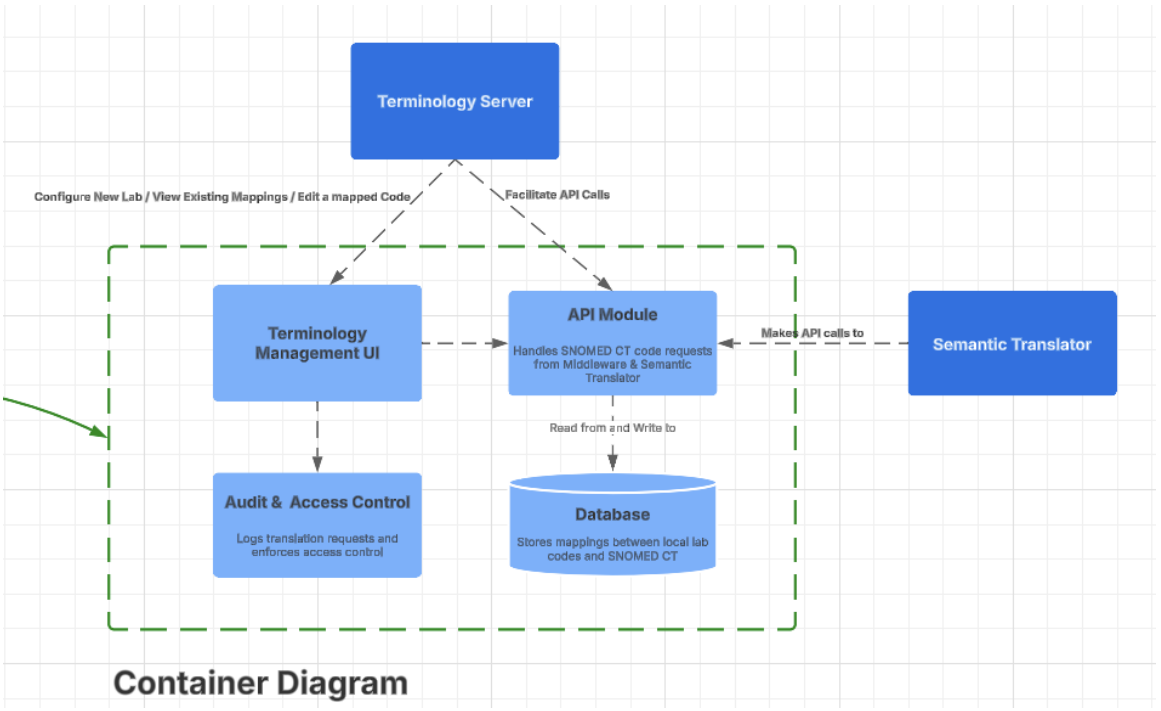
12.4.1.1 Middleware

- Receiver Module: Parses incoming HL7 messages from labs.
- Translator Module: Calls the Terminology Server API to retrieve SNOMED CT codes.
- Sender Module: Sends translated messages to GP systems.
- Error Handling & Logging: Tracks failures and logs translation attempts.
- Cache Layer (Optional): Stores recent translations to reduce API calls.



12.4.1.2 Terminology Server

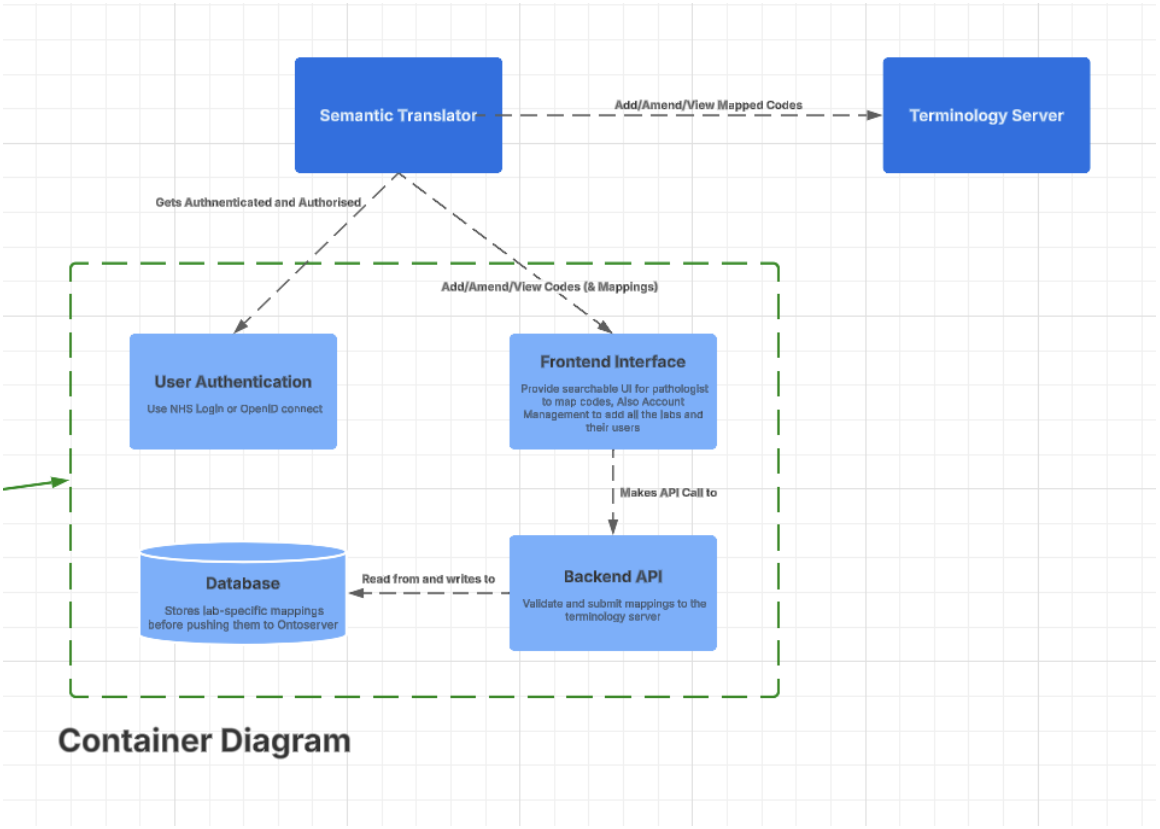
- API Module: Handles SNOMED CT code requests from Middleware.
- Database: Stores mappings between local lab codes and SNOMED CT.
- Terminology Management UI: Allows pathologists to manually update mappings.
- Audit & Access Control: Logs translation requests and enforces access control.



12.4.1.3 Semantic Translator

- Frontend Interface: Provides a searchable UI for pathologists to map codes.
- Backend API: Validates and submits mappings to the Terminology Server.

- Database: Stores lab-specific mappings before pushing them to Ontoserver.
- User Authentication: Uses NHS Login or OpenID Connect.



12.4.2 Security & Performance Considerations

12.4.2.1 Security

Security requirements at high level requires to implement authentication and authorisation for API access. The authorization plays more pivotal roles to ensure a pathologist from a source lab can only add/amend/view his own lab codes/mappings.

12.4.2.2 Performance

- Optimise API response times.
- Implement caching for frequent terminology lookups.
- Ensure scalability across 5,000 NHS labs.

12.4.3 Message flow design

Step 1: Lab Sends a Message (Proprietary or HL7 Format)

Labs may send messages in different formats (HL7 v2, JSON, XML, CSV, or proprietary formats).

Each message must contain:

- **Sending Facility** (e.g., Facility A vs. Facility B may impact translation).
- **Local Codes** (Lab's internal test codes).
- **Unit of Measure** (e.g., mg/dL, mmol/L—important for conversion logic).
- **Relevant Metadata** (Loinc, System IDs, or any extra identifiers).

Example Incoming Proprietary Message (JSON format)

```
{
  "sendingFacility": "Facility A",
  "patientId": "987654",
  "testCode": "L12345",
  "testName": "Blood Glucose",
  "unitOfMeasure": "mg/dL",
  "value": "5.4",
  "otherMetadata": {
    "loincCode": "2339-0",
    "labSystem": "LabSoftX"
  }
}
```

Example Incoming HL7 Message

```
MSH|^~\&|LabSystem|Facility
A|Middleware|NHS|20250224090000||ORU^R01|12345|P|2.5.1

PID|||987654^^^NHS||Doe^John||19800101|M|||123 Test
St^^London^^SW1A2AA|||

OBR|1|1234|5678|BLOOD GLUCOSE^LOCAL|||20250224090000

OBX|1|NM|L12345^LOCAL|1|5.4|mg/dL|||N|||20250224090000
```

Step 2: Middleware Receives & Normalises Message

Middleware identifies the format (HL7, JSON, XML, etc.) and converts it to a common structure.

- Extracts Sending Facility, Local Code, Unit of Measure, and Metadata.
- Normalises units (if needed).
- Sends **structured request** to the Terminology Server.

Standardised Request to Terminology Server (FHIR ConceptMap API)

```
GET /ConceptMap/$translate?code=L12345

    &system=http://local.lab.codes

    &context=sendingFacility=Facility A

    &unitOfMeasure=mg/dL
```

Step 3: Terminology Server Determines SNOMED CT Mapping (Conditional Translation)

- The **Terminology Server** does not perform simple 1-1 mappings. Instead, it **evaluates context**:
- **Local Code + Sending Facility + Unit of Measure** may influence mapping.
- **Other metadata (LOINC codes, system ID, lab-specific codes)** may also contribute.
- May require **rule-based logic** to determine the correct **SNOMED CT PALM code**.

Example: How Context Changes the Mapping

Local Code	Sending Facility	Unit of Measure	SNOMED CT PALM Mapping
L12345	Facility A	mg/dL	43396009 (Blood Glucose Measurement)
L12345	Facility B	mmol/L	271062007 (Glucose level in plasma)
L67890	Facility C	mg/L	1049381000000107 (Protein Level Test)

FHIR ConceptMap Response (Conditional Mapping Applied)


```
{
  "resourceType": "Parameters",
  "parameter": [
    {
      "name": "match",
      "resource": {
        "code": "43396009",
        "system": "http://snomed.info/sct",
        "display": "Blood Glucose Measurement"
      }
    }
  ]
}
```


Step 4: Middleware Updates Message & Sends to GP System

- Middleware **updates the original message** with the translated **SNOMED CT codes**.
- Constructs **HL7 v2 message** (if GP system requires it) or **FHIR resource**.

Example: FHIR Observation Sent to GP System


```
{
  "resourceType": "Observation",
  "id": "blood-glucose",
  "status": "final",
  "code": {
    "coding": [
      {
        "system": "http://snomed.info/sct",
        "code": "43396009",
        "display": "Blood Glucose Measurement"
      }
    ]
  },
  "valueQuantity": {
    "value": 5.4,
    "unit": "mg/dL",
    "system": "http://unitsofmeasure.org",
    "code": "mg/dL"
  },
  "subject": {
    "reference": "Patient/987654"
  }
}
```

12.4.4 FHIR Data structures

FHIR ConceptMap (Core Mapping Storage)

- Stores local lab-to-SNOMED CT mappings.

- Enables automatic translation via `$translate` API.

FHIR ValueSet (Defining Valid Local Codes)

- Ensures that only valid local lab codes are used in mappings.

- Helps with UI dropdowns and validation.

FHIR ConceptMap API Calls

- Translate Local → SNOMED CT via GET `/ConceptMap/$translate`.

- Add New Mapping via POST `/ConceptMap`.