Classification: Official

NHS England

Pathology Standards Implementation – SNOMED CT PaLM Mapping Project – Options Paper

Document filename: SNOMED CT PaLM Mapping Project Options Paper			
Programme	Pathology Standards Implementation	**Project**	SNOMED CT PaLM Mapping Project
Programme Head	Pete Turnbull		
Owner	Karim Nashar	**Version**	1.0
Authors	Gogs Byrn, Karim Nashar, Pete Garrett	**Version date**	15/04/2025

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
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
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
Read v2 Pathology Bounded Code List (Read 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

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 define and evaluate the options around supporting labs to map 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.

1.2 Audience

The intended audience for this document includes:

- NHS England Pathology Standards Programme Board
- NHS Pathology Standards Governance Board
- NHS England Diagnostics Digital Capability Programme

1.3 Related documents

This document should be read in conjunction with the following:

- DAPB4101 Information Standards Notice (Amd 63/2023)
- SNOMED CT PaLM Mapping Project – Best Practice*
- SNOMED CT PaLM Mapping Project - Tooling Requirements Specification*

*available via NHS England's Pathology Standards and Implementation website

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 reportables 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. This has involved formulating and testing a range of mapping strategies, testing and evaluating a range of mapping tools, establishing requirements for a tool that could provide all the desired functionality to support the use-case, and evaluating who is best placed to perform the mapping.

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

NB. It is important to acknowledge that much of the best practice established whilst undertaking this project can be applied universally to terminology mapping. Consequently, when considering the recommendations (6), a wider view beyond the scope of supporting DAPB4101 is encouraged, as terminology mapping supports interoperability across many aspects of digital healthcare, and the most challenging option offers more benefits if applied to supporting programmes of work beyond pathology and laboratory medicine, histopathology being an obvious example.

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 - 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 Costs and risks of not supporting mapping

On top of the principle that mapping to SNOMED CT PaLM is beneficial (2.1), and that not supporting mapping risks a limited uptake of the terminology, there are two main data quality issues that SNOMED CT PaLM aims to reverse that are prominent in legacy Read PBCL representations and mappings:

- Differentiation in how local terminology is mapped to national codes
- Differentiation in meaning and interpretation of national codes

If support and guidance around mapping to SNOMED CT PaLM is not provided to labs, these coding inconsistencies will perpetuate.

This is not because SNOMED CT PaLM is ambiguous like the PBCL, rather that without support and guidance, users may struggle to map optimally and consistently. For example, users might not apply appropriate rule constraints in mapping tools to help them identify correct map targets. Therefore, tools may identify targets based on the *closest* meaning instead of targets with the *exact* meaning. Such anomalies are commonly seen in current lab to Read PBCL maps, adding to the issues caused by the inherent ambiguity of the PBCL. Consequently, support and guidance will help users quickly identify correct map targets and where there is doubt, help establish appropriate map targets.

2.7 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 general practice.

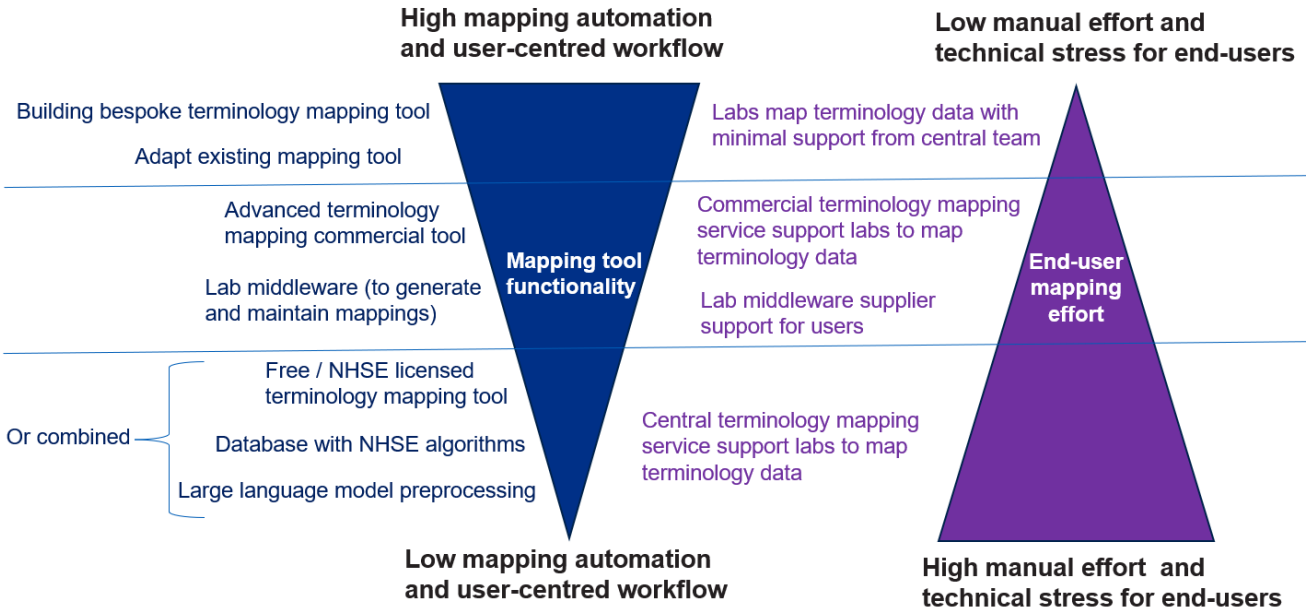
Increasing this to the top 300 would account for 99% of the data. Consequently, offering even the most basic mapping guidance and support would return huge value.

2.8 Dynamics of the mapping process

Section 3. summarises the requirements for mapping. The document then splits into the two main influential strands when considering the options to support the requirements:

1. Section 4. Mapping Tools - Options evaluates the different tools tested
2. Section 5. Who Does the Mapping? - Options evaluates the support available

At the most fundamental level, summarised in the figure below, the PaLM Mapping Project team understood the dynamic of the relationship between **1** and **2** to be one of inverse proportion. That is, as the capabilities and functionality of the mapping tool increases, the level of technical knowledge and support required for the lab user reduces. This was confirmed during testing.



3. Requirements

The scope of 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.

3.1 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

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

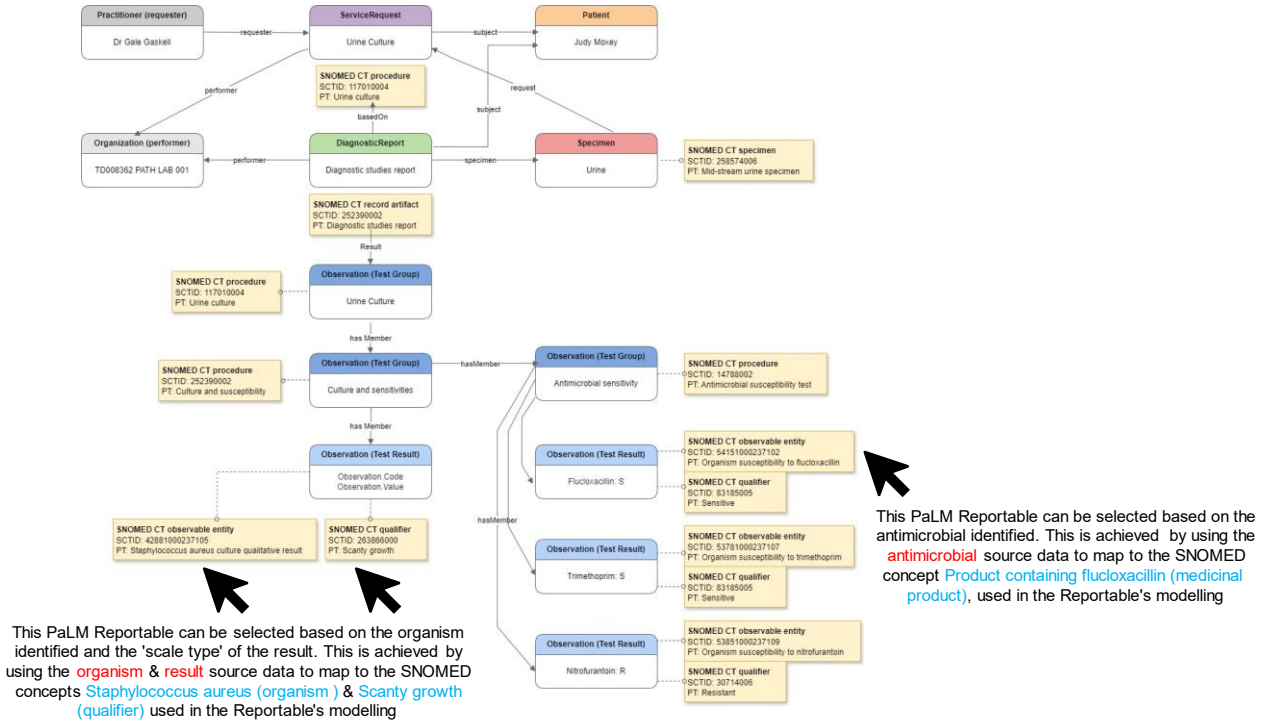
This requirement supports the goal of having fully atomic, structured, coded pathology reports, valuable in terms of enhancing data quality. The process involved is simpler than mapping single reportables, but the implementation challenge is harder because there are several target value sets that require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

3.3 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.



3.4 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.

4. Mapping Tools - Options

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 tools leverage the power of automation to reduce effort and provide functionality to support workflow and governance. This section evaluates the options.

4.1 Using free / NHS England licensed tools

CSIRO's Snapper tool and SNOMED CT International's Snap2SNOMED tool are free* to use (*the cost is covered by NHS England).

Both tools are built on the same back end, so offer the same technical automapping performance. Unfortunately, testing found this to be unsatisfactory, and neither tool could support two key requirements.

Regarding usability, Snapper was found to be poor, but Snap2SNOMED performed much better, having a more intuitive user interface, and far better workflow functionality.

Consequently, whilst these tools are free, there are limiting factors. However, Snap2SNOMED's usability offers the potential for it to be used to review maps created using other methods (4.6 / 4.7) and for governance.

Advantages:

- Free to use
- Available immediately
- Snap2SNOMED suitable for review and governance

Disadvantages:

- Unsatisfactory technical automapping performance
- Snapper has poor usability
- Does not identify multiple SNOMED CT PaLM reportables in complex microbiology reports
- Does not support automatic creation of new SNOMED CT PaLM content
- Relies on external bodies to keep data updated

4.2 Using a commercial tool

Several specialist tools are commercially available that offer good technical automapping performance and usability. In testing, one tool performed particularly well and met all but one of the key requirements. Moreover, some tools offer the provision of a support service. However, licensing would incur cost, increasingly so for multiple licenses.

Advantages:

- Good technical automapping performance
- Good usability
- Adaptable to requirements
- Available immediately
- Potential provision of support service

Disadvantages:

- License costs, multiplied if used across multiple individual labs
- Does not support automatic creation of new SNOMED CT PaLM content

4.3 Building a bespoke NHS England tool

Designed and built to the SNOMED CT PaLM Mapping Project's requirements specification, a bespoke tool would offer good technical automapping performance, good usability and deliver all the requirements. One key benefit would be the provision of an intuitive user interface designed specifically to support the pathology use-case.

However, creating such a tool would require significant money, time, and effort, both to build and maintain, costing far more than licensing or adapting a tool; therefore, returning little value for this singular use case. Furthermore, developing the capability to support the automatic creation of new SNOMED CT PaLM content would be difficult, thereby increasing development cost and time. However, the commercial benefit would be owned by the NHS.

Advantages:

- Would offer good technical automapping performance
- Would offer good usability
- Would deliver all the key requirements
- Commercial benefit would be owned by the NHS

Disadvantages:

- Significant cost to build and maintain
- Significant time to build
- Does not return good value for money
- Recreates functionality already available in other tools

4.4 Adapting a free / NHS England licensed tool

If tools such as those referenced in 4.1 were adapted, this could deliver the same outcome to building a bespoke tool (4.3) but may require less money, time, and effort.

However, the tools in this category that were tested all had technical limitations, so money would be spent developing basic requirements in addition to more challenging ones. Also, the commercial benefit would be owned by the supplier, not the NHS.

Advantages:

- Could deliver all the key requirements
- Would cost less to develop than a bespoke tool
- Would take less time to develop than a bespoke tool

Disadvantages:

- Initial tools tested had poor technical automapping performance and usability
- Cost to build
- Time to build
- Returns limited value
- Commercial benefit would be owned by the supplier, not the NHS

4.5 Using lab middleware (to generate and maintain maps)

Existing system suppliers are developing their lab middleware to support DAPB4101. This involves creating the FHIR message that represents the pathology report and mapping single reportables to SNOMED CT PBCL; a straightforward mapping task enabled via NHS England's existing, assured Read PBCL to SNOMED PBCL maps, currently used by GP systems to support the PMIP EDIFACT flow.

This middleware could also be developed to generate and maintain maps to SNOMED CT PaLM. This would support widespread adoption of SNOMED CT PaLM due to the suppliers' market share and remove the mapping burden for the vast percentage of labs.

However, the scope of each middleware product's development is yet to be established, as are development timescales, mapping methodologies, and importantly, what each supplier would charge for this service.

NB. Middleware is still required to support the flow of FHIR/SNOMED CT, regardless of whether it is used to generate maps to SNOMED CT PaLM or provide any other key mapping requirements. I.e., if map tables are created via another method, they can be used by middleware to generate DAPB4101 conformant pathology reports.

Advantages:

- Would streamline the process
- Would ensure widespread adoption of SNOMED CT PaLM

Disadvantages:

- Cost
- Increases labs' reliance on suppliers
- Suppliers lack SNOMED CT PaLM expertise to optimise the implementation
- Does not support mapping to other SNOMED CT concepts that populate supplemental data elements in the Pathology FHIR specification (tbc)
- Does not identify multiple SNOMED CT PaLM reportables in complex microbiology reports (tbc)
- Does not support automatic creation of new SNOMED CT PaLM content

4.6 Using a spreadsheet to manually map small volumes

A spreadsheet can be used to create map tables manually, using NHS England's nationally available, free to use SNOMED CT browser to search for target SNOMED CT PaLM reportables and to copy/paste their descriptions and concept IDs into a spreadsheet line-by-line. The spreadsheet can then output a map table in a suitable format. This is a simple, solution for mapping small volumes of 'high use' single reportables (e.g., the top 20) at minimal financial cost.

This approach does not support governance and workflow, but with small volumes, this is less of an issue. However, the approach does not support any other key requirements.

Advantages:

- Simple for mapping small volumes of single reportables
- No cost

Disadvantages:

- Only suitable for mapping small volumes of single reportables
- Risk of manual error
- Does not support mapping to other SNOMED CT concepts that populate supplemental data elements in the Pathology FHIR specification
- Does not identify multiple SNOMED CT PaLM reportables in complex microbiology reports
- Does not support provenance, governance, and assurance
- Does not support automatic maintenance
- Does not support automatic creation of new SNOMED CT PaLM content

4.7 Using a database plus NHS England developed algorithms

By leveraging the information contained in labs' existing Read PBCL/SNOMED PBCL maps together with labs' UoM/Specimen data, and subsequently applying artefacts, algorithms, and semantic parsing instructions developed by the PaLM Mapping Project team, a database can be used to map large volumes of single reportables, quickly, and at low financial cost.

However, this requires technical expertise and does not support governance and workflow, requiring the additional use of a tool like Snap2SNOMED. Also, this approach does not support any other key requirements.

Advantages:

- Good technical automapping performance
- Available immediately
- No mapping tool license costs

Disadvantages:

- Relies on a singular mapping strategy (using existing Read PBCL maps)
- Dependent on dedicated NHSE support (5.2)
- Requires additional tool for governance and workflow (e.g., Snap2SNOMED)
- Does not support mapping to other SNOMED CT concepts that populate supplemental data elements in the Pathology FHIR specification
- Does not identify multiple SNOMED CT PaLM reportables in complex microbiology reports
- Does not support automatic maintenance
- Does not support automatic creation of new SNOMED CT PaLM content

4.8 Using Large Language Models

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, and the specimen from the existing Read PBCL/SNOMED PBCL map). 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; something testing found to be valuable in the ensuing mapping process.

However, when asked to perform more deterministic tasks where the outcome is fixed (i.e., mapping local reportables to SNOMED CT PaLM directly) LLMs proved unreliable, often 'hallucinating' map targets, even when specifically prompted not to. Consequently, testing found LLMs less suitable for mapping than other methods. Moreover, LLMs do not support most of the requirements.

Advantages:

- Useful in cleansing source data prior to mapping
- Deals with acronyms well
- Can make predictions about content where it is not apparent (e.g., lab test sample)

Disadvantages:

- Probabilistic in nature, rather than deterministic, thereby less suitable for mapping
- Prone to hallucinations
- Requires explicit prompts
- Dependent on dedicated NHS England support (5.2)
- Requires additional tool for governance and workflow (e.g., Snap2SNOMED)
- Increases risks around data quality
- Does not identify multiple SNOMED CT PaLM reportables in complex microbiology reports
- Does not support automatic maintenance
- Does not support automatic creation of new SNOMED CT PaLM content

4.9 Table of functionality

This table illustrates the key (*potential) functionality of each option.

Detailed analysis of how individual mapping tools performed is recorded in the SNOMED CT PaLM Mapping Project Mapping tool evaluation scoring chart, available upon request to authorised NHS England staff via pathology.standardsandimplementation@nhs.net

	Maps single reportables to SNOMED PaLM	Maps supplemental data elements in the Pathology FHIR spec. to SNOMED concepts	Identifies multiple SNOMED PaLM reportables in complex microbiology reports	Records provenance re. concept maps	Supports automatic map maintenance	Supports automatic creation of new SNOMED PaLM content
Free/licensed tools	√	√	x	√	√	x
Commercial tools	√	√	√	√	√	x
Bespoke NHSE tool*	√	√	√	√	√	√
Adapted free/licensed tools*	√	√	√	√	√	√
New middleware*	√	tbc	tbc	√	√	x
Spreadsheet	√	x	x	x	x	x
Database	√	x	x	x	x	x
LLM	√	√	x	x	x	x

5. Who Does the Mapping? - Options

The mapping process workflow involves creation, review, and maintenance. As labs are ultimately responsible for their data, they must assure their own maps, 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 NHS England or by commercial suppliers. This section evaluates the options.

5.1 Individual labs

Workforce analysis shows that NHS lab staff are currently under-resourced and time poor. Consequently, expecting 122 labs to voluntarily take on the entire mapping burden unsupported is unrealistic. If labs attempted to meet all the requirements (3), this would require a specialist mapping tool and a level of supporting service.

Asking labs to map a small volume of ‘high use’ single reportables in a spreadsheet (4.6) however, is a realistic prospect, requiring minimal support, which could return significant value at minimal cost.

Advantages:

- With intuitive tooling and guidance, this approach empowers the labs to create, assure, and maintain their own maps (6.1)
- If mapping small volumes, achievable at minimal cost and support, (6.4)

Disadvantages:

- Difficult and costly if mapping anything other than small volumes
- Without intuitive tooling and guidance, increases burden and technical stress on lab staff
- Burden potentially discourages widespread adoption of SNOMED CT PaLM
- Without intuitive tooling and guidance, risks impacting data quality through inconsistent mapping approach

5.2 Central NHS England mapping service

Testing found the PaLM Mapping Project team successfully able to automate the creation of high-quality maps for single reportables by using LLMs to cleanse source data (4.8) and a database to apply algorithms, artefacts and semantic parsing instructions developed in-house (see 4.7). These maps were then clinically assured, proving the end-to-end process.

Consequently, if scaled up, such a team could be commissioned to offer a mapping service to all 27 Pathology Networks.

Advantages:

- Ensures widespread adoption of SNOMED CT PaLM
- Reduces burden on lab staff
- Promotes data quality through consistent mapping approach
- Requires a single license if a commercial tool is used (6.2)
- Need not rely on a commercial tool or tooling development (6.3)
- Reduces labs' reliance on suppliers

Disadvantages:

- Cost and risk borne by NHS England
- Service provision would require governance
- Potential ongoing maintenance dependency (tbc)

5.3 Middleware suppliers

If commissioned, middleware suppliers could support mapping (4.5).

Advantages:

- Ensures widespread adoption of SNOMED CT PaLM

Disadvantages:

- Cost
- Increases labs' reliance on suppliers
- Suppliers may not develop to the requirements

6. Recommendations

Following evaluation of the options detailed above, combined with insight around the challenge of implementing DAPB4101 in a timely, economical, realistic, and achievable manner, the PaLM Mapping Project team recommend the following for each level of scope, as detailed in section 3.

NB. It is important to note that the project scope did not include establishing the exact financial cost of each option. Therefore, when assessing trade-offs between speed, cost, control, and functionality, these recommendations were factored around the simple principle that is cheaper, quicker, and less risky to buy, rather than build software.

6.1 Full scope – Individual labs mapping using an adapted free / NHS England licensed tool

If all the requirements are to be met, including the 'stretch' goal of automatic creation of new SNOMED CT PaLM content, the PaLM Mapping Project recommend adapting a free / NHS England licensed tool, ideally where NHS England already has a good working relationship with a contracted supplier / external owner.

Such a tool would then be provided to all 27 Pathology Networks, the intention being that it would allow them to perform all mapping tasks themselves, with NHS England's Diagnostic Digital Capability Programme leveraging their relationship with the Networks to gain buy in. Once assured, map tables could be implemented by middleware to support the DAPB4101 reporting flow and maintained using the tool.

This would ensure widespread adoption of SNOMED CT PaLM. Moreover, mapping would be done in a consistent manner, thereby helping ensure data quality. Also, it would offer the Networks a level of autonomy.

However, the challenge, cost, and risk in developing such software should not be underestimated and would require significant input from NHS England.

This option would be encouraged if it was determined that the national rollout of SNOMED CT PaLM was likely to begin in the mid to long term, as this would allow more time to liaise with developers, implement the requirements, and test the tool.

6.2 Large scope – Central NHS England service helping labs map using the best commercial tool

A dedicated central service using the commercial tool that scored highest in our evaluation could support labs to map across all 27 Pathology Networks.

The plan would involve the central team gathering appropriate source data from each Network in turn, creating the initial maps and then submitting them for labs' clinical review. Once assured, map tables could be implemented by middleware to support the DAPB4101 reporting flow. The intention would be to collaborate with NHS England's Diagnostic Digital Capability Programme in order to leverage their relationship with the Networks to gain buy in, and to ensure that clinical assurance is done in a timely fashion.

Doing so would ensure quick and widespread adoption of SNOMED CT PaLM. Moreover, mapping would be done in a consistent manner, thereby helping ensure data quality. Importantly, it would remove much of the burden from time-poor, under resourced lab staff and would reduce labs' supplier costs. Whilst the focus would be on mapping single reportables to SNOMED CT PaLM (3.1), there is potential that this approach could extend to meeting other key requirements (3.2 / 3.3).

However, the license costs for this tool and supporting service have yet to be established.

Furthermore, there are questions over whether this is a time-bound, singular exercise designed solely to embed SNOMED CT PaLM, or whether NHS England will offer ongoing maintenance support and own the associated costs and risks.

This option would be encouraged if it was determined that the national rollout of SNOMED CT PaLM was likely to begin in the short-term, as it would swiftly and effectively assist labs to map to SNOMED CT PaLM.

6.3 Reduced scope - Central NHS England service helping labs map single reportables using LLMs, a database plus NHS England-developed algorithms, and Snap2SNOMED

This involves an internal stepwise process combining LLM preprocessing of source data (4.8), map creation using a database plus NHSE developed algorithms (4.7), and Snap2SNOMED for clinical review and governance.

A dedicated central service using this approach could support the mapping of single reportables to SNOMED CT PaLM (3.1) across all 27 Pathology Networks.

The plan would involve the central team gathering appropriate source data from each Network in turn, creating the initial maps and then submitting them for labs' clinical review. Once assured, these map tables could be implemented by middleware to support the DAPB4101 reporting flow. The map tables could subsequently be maintained by labs using either Snap2SNOMED or Snapper, as both leverage terminology services. The intention would be to collaborate with NHS England's Diagnostic Digital Capability Programme in

order to leverage their relationship with the Networks to gain buy in, and to ensure that clinical assurance is done in a timely fashion.

Doing so would ensure quick and widespread adoption of SNOMED CT PaLM. Moreover, mapping would be done in a consistent manner, thereby helping ensure data quality. Importantly, it would remove much of the burden from time-poor, under resourced lab staff. Additionally, there would be no licence costs and, it would reduce labs' supplier costs.

However, this would only deliver the key requirement to map single reportables (3.1), and would be a time-bound, singular exercise designed solely to embed SNOMED CT PaLM, with responsibility for ongoing maintenance sitting with labs.

6.4 Minimum scope - Individual labs manually mapping small volumes of 'high-use' single reportables using a spreadsheet

As previously stated, mapping just the 20 most used reportables in lab to GP reporting to SNOMED CT PaLM would account for half of all the pathology and laboratory medicine lab test result data recorded in Primary Care (2.7).

It would be quick and straightforward for labs to do this manually using a spreadsheet (4.6), and simple, step by step guidance could be provided via NHS England's Pathology Standards and Implementation website. Once assured, these map tables could be implemented by middleware to support the DAPB4101 reporting flow.

This would be a minimal cost exercise and ensure adoption of the most used SNOMED CT PaLM reportables, thereby cementing the terminology as a viable product across the estate. However, it would only address the key requirement to map single reportables (3.1), with responsibility for ongoing maintenance sitting with labs.