# SNOMED CT support for Precision Medicine

## Introduction

> "The good physician treats the disease; the great physician treats the patient who has the disease" – Sir William Osler (1849-1919)

The global ambition in medicine, is to treat patients within the context of precision medicine, with a longer-term goal of delivering highly personalized care and treatment. It is important to understand exactly how these two initiatives are conceptually linked to personal data usually held in electronic health records and to plan how SNOMED CT, as a global (terminology) language of health can best support the transition to personalised medicine.

## Definition of precision medicine

Precision medicine has its roots in human genome analysis, which has led to a greater understanding of disease, its causes and treatment.

> Precision medicine is defined as "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person".
>
> https://medlineplus.gov/genetics/understanding/precisionmedicine/definition/

Precision medicine provides a greater understanding of disease and its treatment by using genomic information to inform best practice guidance, building on the evidence-based medicine principles which are well established in clinical practice. Personalised medicine provides the logical next step in this approach by using an individual's genetic code to predict disease and identify the optimal treatment for a specific individual. Providing personalised medicine has always been the aim of clinicians to deliver treatments, but it is only now that the technology is available to deliver truly personalised treatments based on a detailed understanding of the individual's clinical phenotype and genotype. The availability of whole genome sequencing, clinical data obtained by history taking and examination as well laboratory data and observations from wearable technologies has made it possible to identify specific causes of disease and optimise treatments for the individual, rather than providing the same treatment for all subjects. However, with this leap forwards in technology, there are some challenges, not least of which is successfully utilising the vast amounts of data that are now available.

## Personalized healthcare delivery

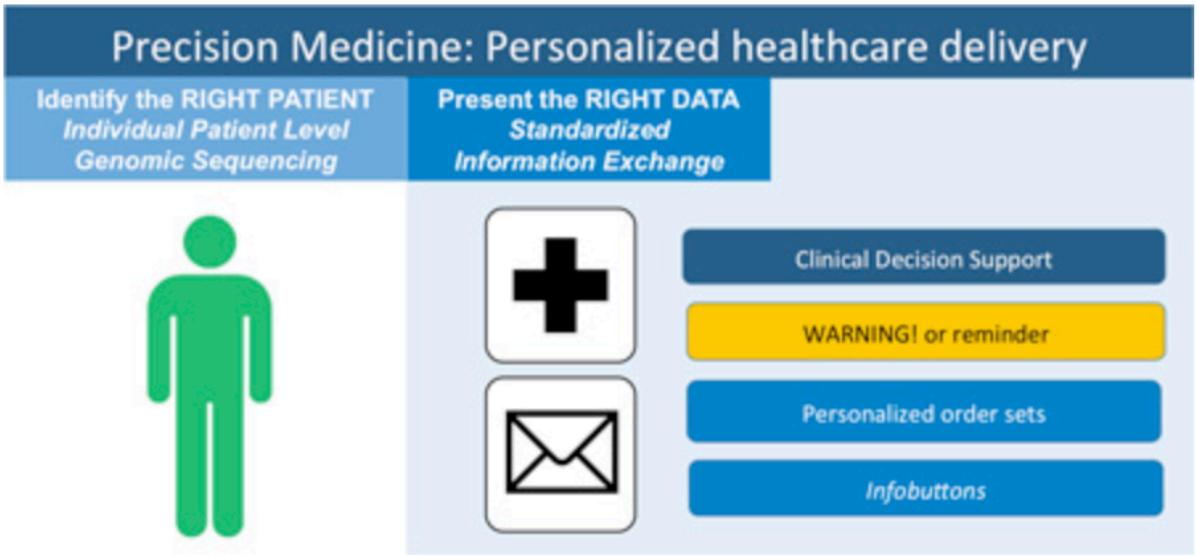


Fig.1.
Integrated Precision Medicine: The Role of Electronic Health Records in Delivering Personalized Treatment, Amy Sitapati et al, Wiley Interdiscip Rev Syst Biol Med. 2017 May; 9(3): 10.1002/wsbm.1378.

## Target audience

- All clinicians
- Genomics researchers
- EHR vendors
- CDSS vendors

## Use cases

### Examining the intersection of individual patient data with population health data

Population health data and large-scale data analytics are based on point of care recording by clinicians. To meet the requirements of clinicians at the point of care requires a (very) granular clinical terminology, supporting them to record detailed information on the patient, their clinical findings and treatment. This information can then be amalgamated into large data sets, that support research and discovery processes. When undertaking large analysis, an ontological basis to the terminology is essential to allow accurate and meaningful queries to be delivered.

### Use machine learning and development of features that predict people at risk of specific disease and open-up use of genetic screening

An important aspect of genomic medicine is to link the clinical phenotype to the genotype by showing how genetic variants and mutations predict the risk of developing disease and define treatments that prevent or delay its onset. The understanding of the processes involved requires large scale data analytics and the use of machine learning and artificial intelligence. These processes can be assisted through the use of ontologically based terminologies, to provide easier understanding of the data, and allows data to be organised in a meaningful way.

### Associating social determinants of health with genetic factors

Increasingly in healthcare, there is an acknowledgement of the importance of social determinants of health and their effect on disease identification and treatment. In view of this understanding, there is an increased appetite by researchers and data analysts to access social information in addition to the clinical information routinely collected as part of healthcare delivery. These social factors are particularly important in identifying and treating at-risk groups, for example as was seen in London, UK, the high incidence of COVID-19 in patients with a South East Asian population, the cause of which went beyond routine clinical comorbidities.

### Associating clinical phenotype with acquired or inherited genetic variation

The terminology used at the point of care must be able to differentiate between acquired and inherited diseases. This is important to assist clinicians in accurately recording diagnosis and is essential to accurately identify treatment options based on a clear diagnosis. This is also important to inform the data analysis process and allow phenotypic information to be associated with the correct disorder type. In view of large-scale data analysis, ideally, the terminology used should be able to support the analysis of the data in a computable manner.

### Providing researchers with granular clinical phenotype information from EHR

Genomic medicine is expanding our understanding of diseases, their causes and treatments. Researchers and data analysts seek to identify new and novel linkages that were not previously understood. This requires the point of care clinical data to provide as much granular information as possible, across the widest range of subject matter. This will include areas such as social determinants of health, access to healthcare, geographic and environmental factors and others. Utilizing a broad range of subject matter from the patient record allows analysts to better understand the cause and indications of disease and the effectiveness of specific treatments on patients (either singular or as cohorts).

### Detailing clinical inclusion/exclusion for research cohorts

The identification of patient cohorts is essential for the management of disease within clinical practice, and also in research. For research purposes, cohorts need to be identified using detailed criteria, and should be performed using processes that support rapid identification

of individual patients. This requires detailed information rather than aggregated data extracted from classifications.

### Detailing clinical inclusion/exclusion criteria within treatment algorithms

To deliver precision medicine requires a detailed description of the treatment pathway, along with additional clinical criteria. These treatment pathways are increasingly very detailed and require specification in a manner that can be easily accessed by clinicians. This normally takes the form of a computable specification that can be provided to the clinician to specify treatment criteria, which can then be applied to patient populations or individual patients quickly through the use of existing clinical decision support systems (CDSS).

## SNOMED functionality (current)

### Coverage of rare diseases – content and links to Orphanet

Through the collaboration with INSERM, SNOMED CT has increased the content of rare diseases, and has also fully modelled this content to support data analytics. The use of this content is further enhanced through the development and release of a jointly agreed mapping between SNOMED CT and Orphanet.

### Risk stratification

SNOMED CT includes definitional modelling of content to identify the pathway of disease, including inherited, acquired and severity, allowing the path of a disease to be tracked over time.

### Use of scales, assessments and performance status

The inclusion of scales and assessments in SNOMED CT allows clinicians to accurately measure the path of a disease based on symptoms and phenotypic findings, and also the effectiveness of treatments over time. This is essential by delivering accurate shareable data based on internationally agreed assessment scales and protocols.

### Use of substance and product dictionaries

The inclusion of substances and products in SNOMED CT are essential for the specification of treatment algorithms. They also support the identification of prescribed treatments and allow data analysis to be performed based on the clinical outcomes of specific treatments. In addition, the alignment of these hierarchies with recognised classifications enhances their utility by researchers, particularly in the genomic space.

### Substrate for messaging between systems

The SNOMED CT reference set specification allows value sets (subsets) to be specified easily, which can then be implemented quickly. This supports the rapid identification and implementation of patient cohorts, supporting research activity. There use is also essential in

CDSS systems to specify clinical criteria based on the hierarchical placement of concepts. The use of the reference set specification supports the transfer of specified data between systems, and also the sharing of query specifications that can be used across geographic locations.

### Substrate for patient and citizen facing applications and remote data collection

The SNOMED CT description policy allows the use of synonyms, which can include clinical synonyms and "patient friendly terms". The use of synonyms is important to support patient-facing applications, and their inclusion within mobile applications used by patients to provide clinical measurement or access to patient reported outcomes.

## Post coordination/expression constraint language

SNOMED CT has the ability to support the post coordination of clinical phrases, through an established and agreed specification. SNOMED CT expression language allows clinicians to express in greater detail than traditional precoordinated terminology would support. This is particularly important in areas of clinical practice where the understanding of disease is still being developed, and precoordinated terms cannot as yet be globally agreed. In the context of laboratory and imaging specialties, this process is useful within synoptic reports to provide a coded and computer processable view of the rich information included in the reports. This approach allows a greater level of detail to be made available in coded form, which can then be used for a variety of different purposes. The use of SNOMED CT expressions also supports the analysis of large data sets for research purposes, which can leverage many different axes based on the attributes available in the concept model, for example subsumption testing.

## SNOMED functionality (future)

### Linkages (maps) to genomic terminologies (HPO, OMIM, MONDO)

Genomic research organisations have historically used terminologies from within the genomic space. These include genomic terminologies such as HPO (Human Phenotype Ontology), OMIM (Online Mendelian Inheritance of Man) and MONDO. Their long-term usage has led to the development of robust research processes, which require a mapping/linkage to be provided between them and SNOMED CT to allow researchers to fully access SNOMED CT and its structure. The mapping/linkage will be required to be bi-directional to allow access to data from EHR systems and specification of treatment algorithms for implementation in CDSS systems in the EHR.

## Precision Medicine process - real world usage

The delivery of precision and personalised medicine has broadly three fundamental requirements. These are the ability to access genomic data, specifically sequencing data, the ability to access an individual's clinical record, which must include a broad range of clinical and supporting information and the ability to specify a treatment algorithm to be

implemented with an electronic health record (EHR). SNOMED CT is ideally placed to support the access of the data required and the specification of treatment algorithms. The access of data within the clinical record is a fundamental requirement, but this alone will not deliver the required outcome. To achieve the promised benefits of precision and personalised medicine EHRs are critical, but this must include a focus on interoperability. Integration of the knowledge gained through genomics research can only be actionable by front line clinicians if it is provided in a manner that supports routine patient care, and the actionable data can only be specified if the full extent of the patient record can be accessed by researchers to allow optimal treatments to be identified. EHRs continue to evolve, and now routinely included sophisticated decision support. The evolution of clinical terminologies based on ontological principles serve to support the communication of electronic data between systems.

The data within an EHR includes a rich diversity of information. The core clinical information will record the major clinical occurrences through an individual's life. These will include diagnosis, treatments, investigations and medication. But in addition to these are important additional information, which may include social determinants of health, economic factors and links to other family members records. As understanding of the impacts of these on clinical care has increased through research, so have the requirements to access greater levels of granular information. In the current landscape, this can be challenging with the implementation of multiple terminologies within systems. The terminologies need to have a broad coverage not simply focused on traditional clinical areas of interest. Interoperability is crucial to ensure all information can be accessed irrespective of how it has been stored, both in terms of EHR systems, and in terms of genomic researchers utilising the information who may have requirements to run research processes assuming genomics terminologies specifically.

Of specific importance is use of information within the laboratory and imaging domains. These supply a rich source of data, which inform the diagnosis and treatment of patients. In both disciplines the rich information is presented in the form of synoptic reports, which include structured coded data and free text. Both can be accessed for patient management and research purposes, with a requirement for free text to be summarised into coded form where the information is used for data analytic purposes, both at point of care and for secondary usage.

To deliver precision and personalised medicine requires researchers to specify and provide the details of optimal treatment regimes, and in addition, that those same regimes can be specified in a manner that will support clinicians to deliver the optimal treatments at the point-of-care within EHR systems. This means that algorithms need to be specified in a way that exisiting clinical decision support systems (CDSS) can implement. This requires using ontologies to allow sections within the terminology to be identified, rather than specifying all the individual codes required which would prove very challenging and inaccurate.

To support both precision medicine and genomic research requires the ability to share data between organisations. This may refer to the exchange of patient data for research purposes, either as single records or as patient data sets. This requires the data to be in a format which supports interoperability of either approach. To implement clinical treatment algorithms in a consistent manner across many healthcare organisations, requires a clinical terminology which has global acceptability, and can work with other standards, particularly, but not limited to, communication standards.

To support the transfer information, the terminology used in communication standards such as HL7 FHIR should be globally accepted and be available to all organisations involved. This will include clinical, research and academic organisations. The terminology should also allow for the definition of subsets to support the identification of patient cohorts based on specific criteria which may include aspects beyond the normal range of clinical terms.

## Clinical settings

### *Clinical trials*

Clinical trials require the identification of patient cohorts based on detailed criteria. These criteria can include both clinical and other types of specification, such as social and economic factors. Therefore, it is important that a single terminology includes all axes to allow a query to be run across a database of patient clinical data in a manner that easily implementable to provide rapid feedback. The use of SNOMED CT for this purpose is ideal due to its concept coverage, and also the use of expression constraint language allowing detailed shareable queries to be produced and utilised.

### *Treatment delivery/review*

Within the “real world” EHR setting, SNOMED CT concept coverage provides a rich source of concepts to be selected by clinicians to allow comprehensive and detailed recording of patient information. This is essential to provide a rich source of data for researchers to access. The ontological basis of SNOMED CT also supports the utilisation of concepts to define easily implementable treatment protocols that can then be implemented in CDSS systems to provide functional clinical direction to be provided to clinicians via the EHR at the point of care.

The use of SNOMED CT for clinical risk stratification through the use of staging scales, performance status, co-morbidities, social determinants will also directly support the implementation of treatment algorithms by accurately identifying decision points in the care delivery process. This is essential to delivery safe and optimal clinical care. The scope of concept coverage is essential to meet this requirement both in terms of identification for research purposes and also in the delivery of precision medicine.

Whilst the granularity of SNOMED CT is obviously important, the structure of SNOMED CT also plays an important part. This allows the concepts to be used in the correct context within structured data recording, ensuring unambiguous usage of concepts within an information model. Through the use of agreed global standards, this allows data to be interoperable between systems, both within countries and across borders.

The inclusion of observables within SNOMED CT allows clinicians to accurately and unambiguously record the outcomes of interventions. The modelling of observables supports the detailed analysis of data at a population-level, an also allows the content to be used within treatment algorithms to define key decision points based on the outcome of individual treatment directives.

As the understanding of disease increases there is an increasing recognition of the broader factors that increase both the prevalence of disease and the outcomes of treatments. A key aspect is the effect of social factors, and a requirement to record social determinants of health as a key information item within a patient record. SNOMED CT already has content in this area and is actively work with global stakeholders to expand the coverage of social determinants of health. The availability of this content both provides a broader insight for researchers and may also be used within treatment algorithms as decision points.

The utility of SNOMED CT for data analytics can be used at various levels. The obvious large-scale usage relates to population-based health and genomic research specifically. But the same utility can provide actionable insights to front-line clinicians by providing rapid feedback through the use of automated or semi-automated report generation for both clinician and patient use. Access to feedback in real-time allows clinicians to respond quickly to changes in patient status, which specifically supports detailed automated treatment algorithms to be available the CDSS systems with the EHR, driven by artificial intelligence functionality.

For the treatment of long-term conditions including cancer and rare diseases, SNOMED CT can provide a longitudinal view of phenotypic changes in the patient's condition, allowing treatment to be tailored to the patient's requirements in real time. This is delivered through the granularity of SNOMED CT content, and through the utilisation of its ontological basis to leverage hierarchical structure in clinical decision algorithms. These algorithms can leverage the structured data within a EHR to support automated decision making in complex treatment approaches. These same approaches can be utilised to measure quality improvement in individual patients, and within healthcare organisations to measure quality of delivery and outcome. The ability to record and code aspects of patient clinical history is essential to the delivery of healthcare, by making the information available to clinicians at the point of care, and also for longitudinal analysis of patient conditions. This is particularly important in long term conditions, where clinical changes can occur at different rates throughout the period of disease. SNOMED CT provides both a rich granularity of content, whilst also providing a robust history mechanism to allow access to content which has been deprecated over time.

## Summary

SNOMED CT is well placed to provide key support to deliver precision and personalised medicine by supporting to both EHR data collection, analysis of data and provision of implementable treatment algorithms. The granularity of SNOMED CT in EHR systems provides a rich source of data both clinical and other supporting factors. Access to this rich information source is essential to provide the insight for Genomics researchers to ensure a greater level of understanding of the contributing factors to disease diagnosis and treatment. SNOMED CT is also well placed to work alongside existing genomics terminologies to support the discovery process in research by leveraging the ontological basis of the terminology. As a robust, well documented global clinical terminology standard, SNOMED CT is well placed to provide a source of clinical terminology that is interoperable across systems globally. This is important in all clinical fields, but of particular importance in research where the patient cohorts maybe spread geographically. The collaboration with global standards bodies such as HL7, ensures that SNOMED CT can be used effectively and safely to transfer patient data between healthcare and research organisations. For these purposes, the ability of reference set mechanism to specify subsets is particularly important to support both the implementation of communication standards and to identify clinically valid content to be applied in specific fields within the message, such as HL7 FHIR. The history mechanism within SNOMED CT is also fundamental to support the analysis of patient data over time. Once a treatment algorithm is developed and agreed, SNOMED CT can provide the common computable language for CDSS to implement within clinical platforms and EHR systems. The use of SNOMED CT will allow for accurate definition, whilst also supporting longer term changes in clinical terming as a result of clinical advances, without the need for extensive algorithm rewrites every time a new term is added into practice, or a term is withdrawn.

Authors - Ian Green & Charles Gutteridge
26/02/2020