The Value of Semantic Interoperability for Healthcare

Orion Health White Paper
Michael Hosking, CHIA, B, OccTh
Clinical Product Specialist
022018

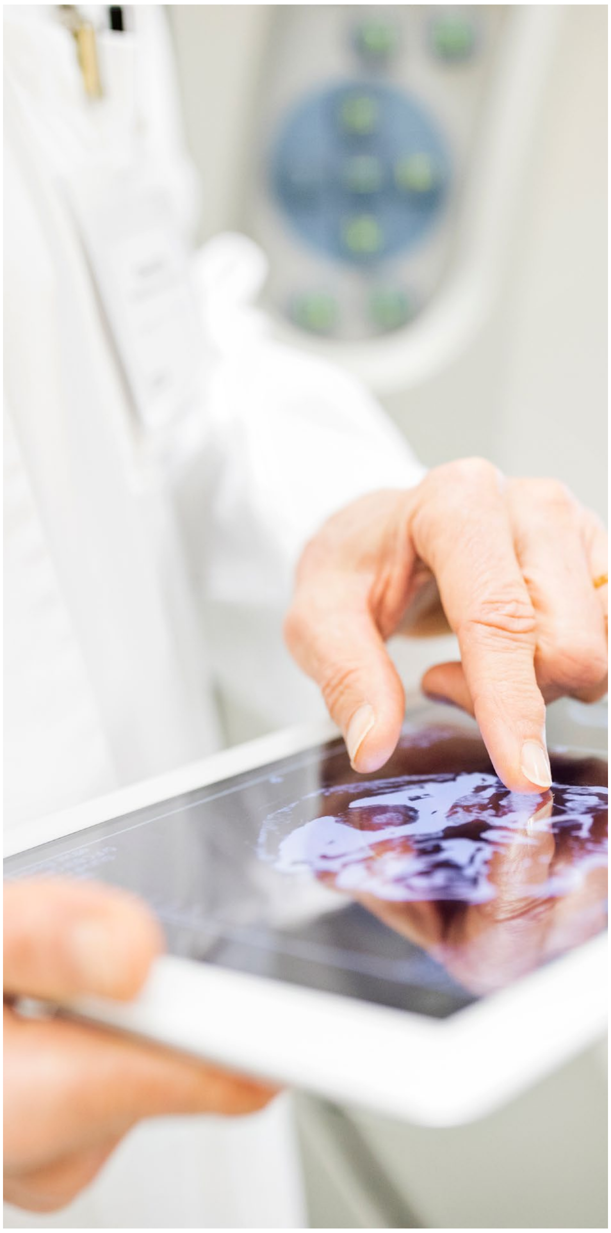
The Unambiguous Exchange and Reuse of Clinical Knowledge

How do we, as humans, convey meaning? Through speech, the written word, body language, other subtle cues or most importantly a combination of these. Around the world we have different ways or words to represent the same meaning.

What do you take when you go to the beach? Flip flops? Jandals? Maybe a pair of thongs? These are just some variations in the English language, not to mention international languages. Ultimately, they are all describing the same thing.

In healthcare, conveying and interpreting clinical meaning is complex, especially when it comes to a clinical information system (CIS). Healthcare professionals already spend a significant amount of their time understanding and interpreting complex intricacies and relationships between clinical data. So the need to use data to link disorders, medications, diagnostic imaging (and much more) along with specific individual patient attributes, is vital in all professional roles.

Does your health service understand the benefits of a CIS that supports clinical terminology systems and services?



Problem

Information within a CIS cannot be easily found, appropriately reused or analyzed. As a result, due to the limitation of timely access to crucial clinical information, clinical workflow and health outcomes such as accurate resource allocation are inhibited.

Research has found that 80%[1] of healthcare data is unstructured and currently unable to be effectively reused or analyzed. This is because free text is often one of the primary sources of data along with classification systems and locally derived codes. This is a barrier to achieving meaningful shared information and knowledge.

Understanding the Clinical System Knowledge Cycle A clinical system knowledge cycle refers to the consumption and production of knowledge across a CIS that informs appropriate clinical, business and population health decisions. This is supported by semantic interoperability through effective data re-use.

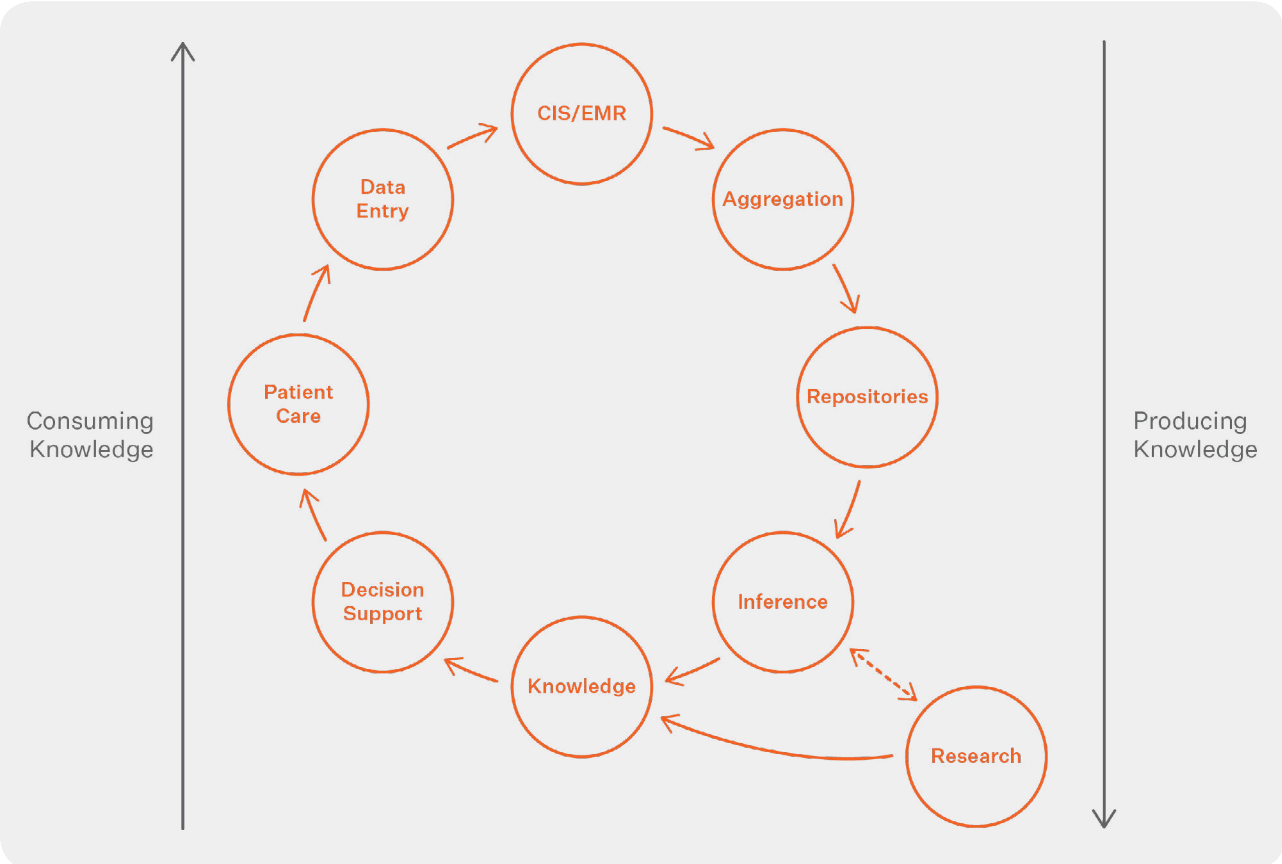


FIGURE 1: Ideal Clinical System Knowledge
Above diagram modified from SNOMED International educational presentation "Fitting SNOMED CT into an EHR system" 2016

[1] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organisations. Technological Forecasting and Social Change, 126, 3-13.

Many CIS do not effectively manage their clinical system knowledge cycle[2]. Currently, it is common for such systems to have a system data cycle which is considered a limited sub-function of a knowledge cycle. This approach is not conducive to developing a strong knowledge cycle with re-use of clinical information for improved workflows and models of care.

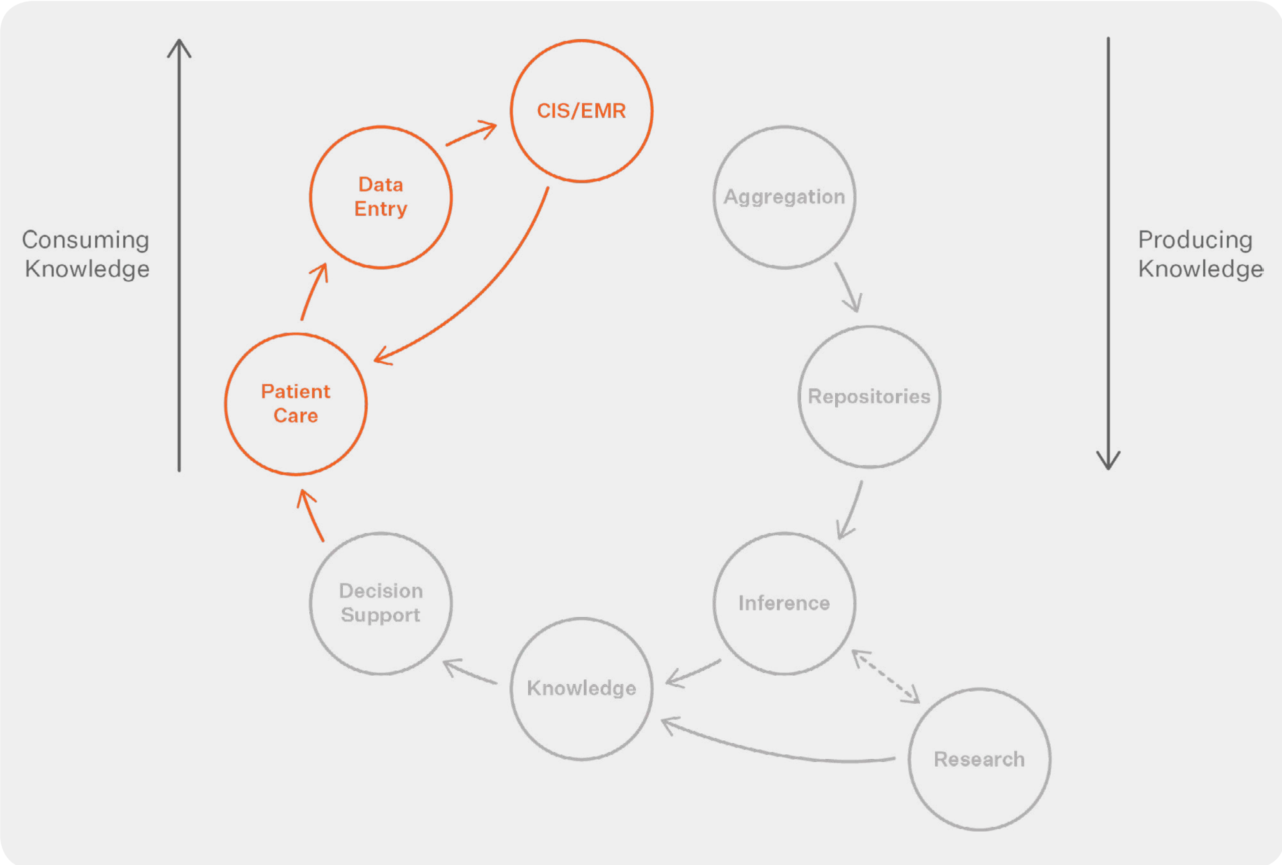


FIGURE 2: Typical Clinical System Data Cycle
Above diagram modified from SNOMED International educational presentation "Fitting SNOMED CT into an EHR system" 2016

[2] Yusof, M. M., Kuljis, J., Papazafeiropoulou, A., & Stergioulas, L. K. (2008). An evaluation framework for Health Information Systems: human, organisation and technology-fit factors (HOT-fit). International journal of medical informatics, 77(6), 386-398.

Solution

Part of the solution to improving clinical knowledge sharing and development for improved decision making is interoperability. This is a vague term so let's break it down.

According to HIMSS, Interoperability can be defined at three levels within healthcare information systems[3].

- **Foundational Interoperability**
    - Basic level of interoperability
    - Exchange of basic, raw data from one system to another
    - Does not require system to interpret data
- **Structural Interoperability**
    - Intermediate level of interoperability
    - Exchange of structural format of data from one system to another
    - Requires receiving system to interpret data field
    - Example: message format standards - HL7 FHIR
- **Semantic Interoperability**
    - Highest level of interoperability
    - Enables exchange, use and re-use of data
    - Requires receiving system to interpret meaning of data
    - Example: SNOMED CT

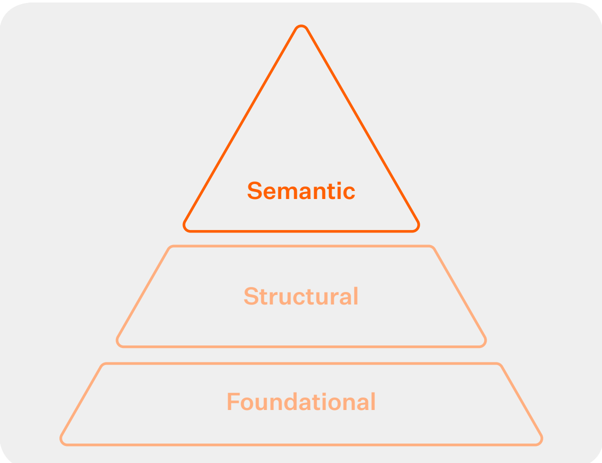


FIGURE 3: Interoperability Levels

Semantic interoperability is the ability to exchange health related data unambiguously and meaningfully between stakeholders[4]. Preserving contextual meaning within a consistent format is needed to support knowledge that is controlled, shareable and reusable.

This meaning can be supported and represented by terminology systems with ontological characteristics. An ontology is a formal naming and definition of things to the level of known truths. These truths can include types, characteristics and relationships in a specific domain. An example of this is: we can confidently say that H1N1 is a type of influenza virus. The statement is always true, regardless of context and is just one of the types of knowledge representation discussed in this paper.

Systematic adoption of controlled, standardised, appropriately flexible terminology systems that support the above are required for any CIS. Effective adoption of these systems can improve primary and secondary data use. This results in improved patient safety and health outcomes which leads to significantly reduced care delivery costs. Such costs can be measured in various ways depending on the business problem needing to be solved.

To support semantic interoperability, foundational and structural interoperability must already exist. Often these exist at the level of technical interoperability including services and messages, transport protocols, data integrity and accessibility. All of which attempt to convey the structure and context of the clinical data being shared/used. Following this solid, more technical interoperability base, semantic interoperability (using standards and terminologies) can support the enriching, exchanging and

[3] http://www.himss.org/sites/himssorg/files/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf

[4] Assessing SNOMED CT for Large Scale eHealth Deployments in the EU, http://assess-ct.eu

reusing of this data as clinically meaningful information. This supports healthcare professionals to represent and reuse clinical meaning across all clinical care settings. This effective reuse of clinical information can further enhance effective healthcare insights and resource allocation, which is especially needed considering our ever increasing population demands.

Quality clinical data is dependent on capturing information in a computer process-able format at the point of care. This aids decision support systems in correlation with clinical guidelines. Evidence based care and service provisioning requires detailed data analysis. At scale, this can support novel technologies including natural language processing with deep learning to accurately predict diagnosis and complete personalised prescriptions along with other decision support mechanisms.

To overcome this issue of inconsistent nomenclature within a CIS or Electronic Health Record (EHR), clinical terminology systems such as Systematised Nomenclature of Medicine – Clinical Terms (SNOMED-CT) and LOINC (Logical Observations Identifiers Names and Codes) can be used as a structured terminology that represents accurate descriptions of clinical information used in clinical practice. This paper will focus on the former clinical terminology system, SNOMED-CT.

Context and purpose of use and re-use

Depending on the intended purpose, audience and outcomes varying levels of information detail and complexity are required. Ultimately, an ideal ontology, in the context of health, would be multipurpose for all users. As an example, a clinician providing clinical care to an individual client does not need to be aware of the service planning for the purposes of funding requirements. Likewise, a Government body or health organisation coordinating funding allocations does not require complex detail of an individual patient's clinical data findings.

What is a Clinical Terminology?

Healthcare professionals can represent complex clinical information in various ways. Clinical terminologies can support a consistent representation of meaning whilst also allow clinicians to have some flexibility with representing information in familiar ways. They enable:

- Consistent naming and identification of clinical knowledge relevant to healthcare
- Unambiguous understanding of relationships between complex clinical concepts and associated attributes to provide richer contextual detail
- Semantic interoperability without loss of detail or change in meaning

Purpose of Clinical Terminology

Primarily, clinical terminologies have been developed to support both a CIS and clinician's providing clinical care. Secondary benefits support the re-use of this information which can support other meaningful representations for other users, organisations and governments. It also supports a mechanism of clinical data governance for improved data collection. To support this, commitment is required from various health organisations to support and educate staff on how to implement effective clinical data governance.

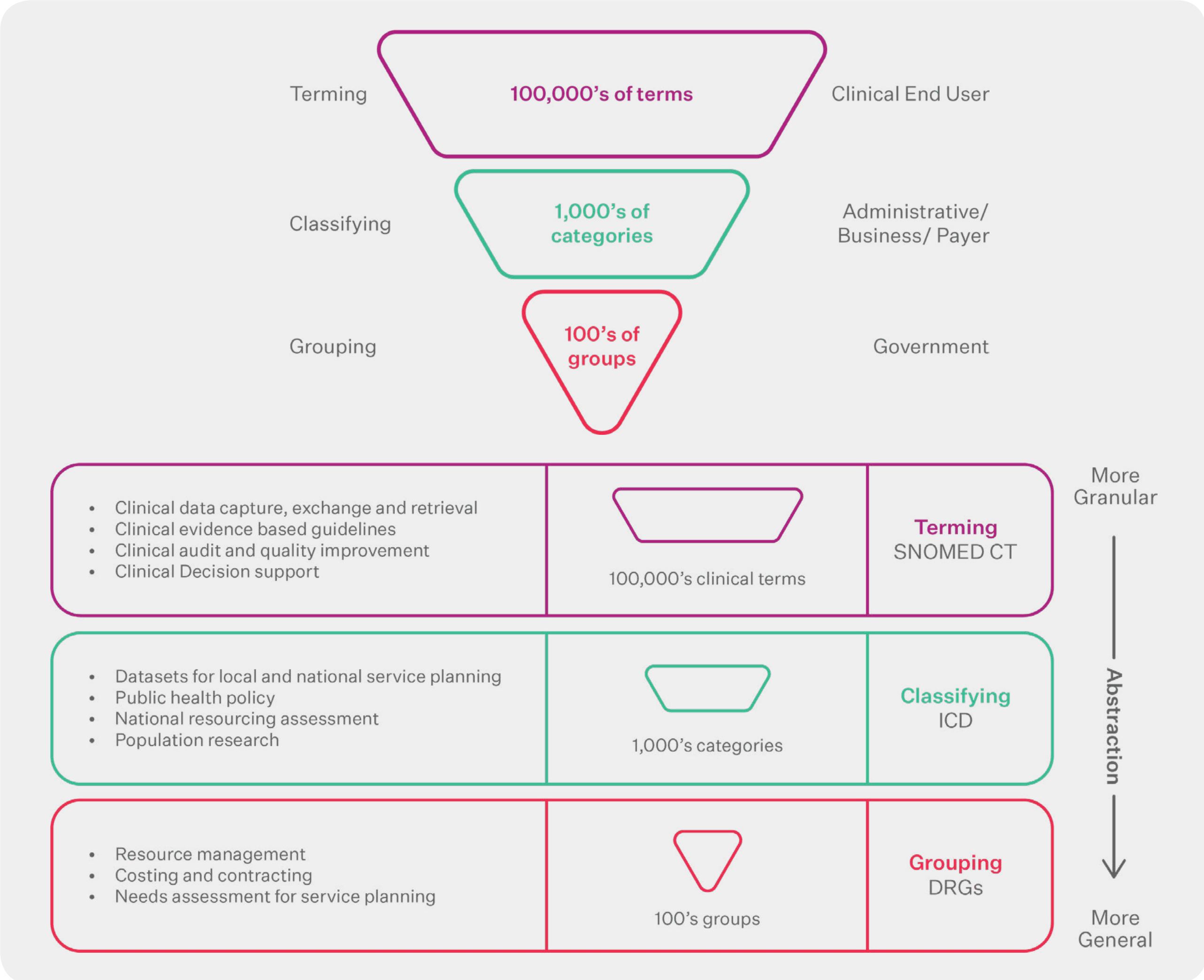


FIGURE 4[5]: Principles of Terminologies, Classification and grouping systems

This diagram shows the level of detail at which clinical information should be collected and reused for secondary purposes and users. As well as some common examples. If we imagine this as a series of sources and filters:

1. **Primary clinical data source**
   a. Detailed data collected and entered by a healthcare professional
2. **Secondary filters**
   b. Detailed data can be 'filtered out' (abstracted) which leaves the more generalised categories that are more useful for statistical and reporting needs

This filter-like mechanism is supported by various standardised health coding systems. Between each system level, there is mapping (linking) which supports the abstraction (generalised representation) of information. This enables information to be displayed in appropriate detail for the intended audience. Both business and user needs should determine which code system and respective mapping/s are most appropriate. There are a range of use cases for mapping between local codes, classification systems and terminologies.

[5] Bowman, S. E. (2005). Coordination of SNOMED-CT and ICD-10: getting the most out of electronic health record systems. Coordination of SNOMED-CT and ICD-10: Getting the Most out of Electronic Health Record Systems/AHIMA, American Health Information Management Association. 37(4-5), 394.

The filtering system appropriately disseminates information to the level of detail required from data source to data re-use based on the use case. This minimises the need for unnecessary manual data entry and increases the potential for more accurate data for re-use.

International Classification of Diseases (ICD), currently in its 10th Revision (ICD-10), is defined by WHO as a classification system designed for "Use for identification of health trends and statistics, and the international standard for reporting diseases and health conditions". It is a mutually exclusive system which means that "lung cancer" is categorised under the category of cancer, and cancer alone. It cannot, however, also fit under the category of a "respiratory system disorder". Use of this system within a CIS results in limitations for advanced clinical analytics and complex decision support.

ICD classifications correspond to Diagnosis-Related Groups (DRGs). These are often used to classify specific inpatient classifications of care provision. These DRG codes relate to needs assessments for service planning, resource management and funding requirements.

For optimal information accuracy, it is best to support clinical users to codify the information at the point of information capture. As for the best user experience, clinical users should not need to know or see that they are coding this information via any user interface. Supporting coding clinical detail, at the point of information capture maintains precision and accuracy of information. This should be done to the level of detail that is known by a user supporting familiar terms, abbreviations, shorthand and spelling variations.

A common CIS user interface limitation is use of vague terms for clinical data entry. This is a limitation because generalised classifications often do not support the level of clinical depth required for accurate knowledge representation for both clinical users and for a CIS. Terms used from classification systems like ICD-10 do not assist with clinical workflow or clinical data quality at the point of care. This is demonstrated in the diagram below.

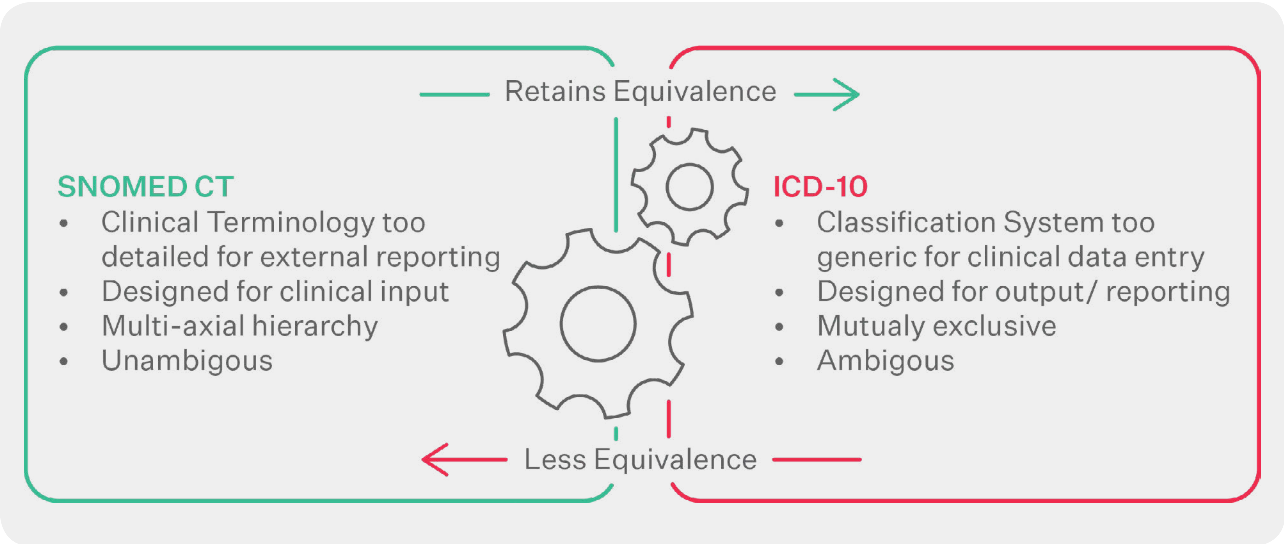


FIGURE 5: Systems and Equivalence

As stated in the well-known “Desiderata for Controlled Medical Vocabularies in the Twenty-First Century”[7] there are some core characteristics of an ideal clinical terminology. Fundamentally, full value of an EHR or CIS will not be realised without effective capture and use of quality, valid data with effective clinical and data governance protocols. There are many clinical terminology systems and classifications in use today but some are more mature in their development and adoption than others.

What is SNOMED-CT?

SNOMED-CT currently is the most comprehensive, multilingual clinical healthcare terminology standard in the world[8]. It contains a large, regularly (six monthly) revised database of clinical concepts which represents clinical knowledge. These concepts are assigned both human readable terms and each of these unambiguous concepts are represented by a unique identifier. From here, clinical terms can be represented in a user interface with the unique identifier being captured and interpreted in the background by an appropriately configured clinical information system or application. There are many other applications of this including clinical knowledge management, decision support mechanisms and population health analytics and prediction.

SNOMED-CT enables a consistent way to index, store, retrieve, and aggregate clinical data across a range of clinical specialities and sites. It also assists in organising medical record content. This reduces inconsistencies with recording/capturing data, which is encoded and used for clinical care of patients and research. It is also a tool that can support clinical decision support and various clinical administration process automation. It has been identified as the preferred national clinical terminology by many countries across the world. This includes; the United States Healthcare Information Technology Standards Panel, the National Health Service in the UK, the Health Information Standards Organisation for its use across the health and disability sector in New Zealand, the ADHA (Australian Digital Health Agency)[9] along with other international government and industry bodies. SNOMED International currently has over 30 member countries with 8 National Extensions.

Many international governments endorse the use of SNOMED-CT as the clinical terminology to be used across health and disability sectors. The reason for this is that the application and use of SNOMED-CT within an EHR, CIS and knowledge management systems support a range of tangible benefits.

Benefits

SNOMED CT has the potential to benefit individuals, populations and evidence based decisions for cost effective care delivery[10,11]. Such benefits include:

Enhancing individual care by:

- Enabling automated clinical decision support systems
- Reducing data collection burdens
- Supporting data with consistent unambiguous semantic information sharing
- Enabling clinical information capture using consistent, commonly used phrases
- Supporting accurate and comprehensive semantic based searches
- Supporting multiple languages

[7] Cimino, J. J. (1998). Desiderata for controlled medical vocabularies in the twenty-first century. Methods of information in medicine, 37(4-5), 394.

[8] SNOMED International SNOMED CT Starter Guide https://confluence.ihtsdotools.org/display/DOCSTAR

[9] https://www.healthterminologies.gov.au/ncts/

[10] http://www.snomed.org/snomed-ct/why-should-i-get-snomed-ct

[11] http://www.snomed.org/snomed-ct/why-should-i-get-snomed-ct/business-case

Enhancing population health by:

- Facilitating population health monitoring
- Reflecting evolutionary growth of changing clinical practices
- Enabling accurate access to relevant information
- Enabling data to support clinical research and improved treatment
- Enhancing audits of care delivery and detailed analysis of clinical records

Enhancing evidence-based healthcare decisions by:

- Supporting links between clinical records and changing clinical guidelines
- Evaluating and enhancing the care delivery experienced by individuals
- Reducing need and cost of inappropriate and duplicate testing and treatment
- Reducing the likelihood, frequency and impact of adverse healthcare events
- Improving cost-effective care delivery and quality of care delivered

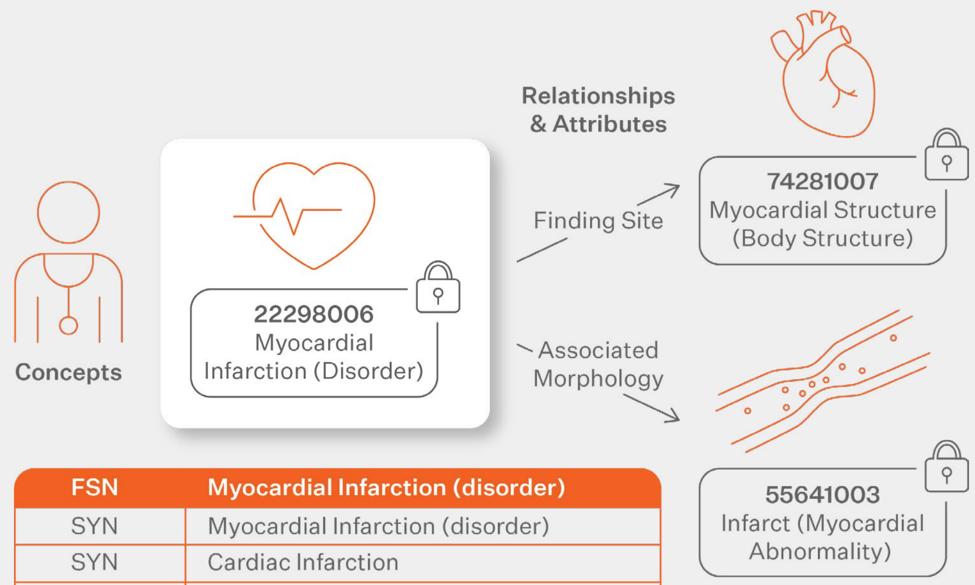
Tangible Cost benefits

Quantifiable cost benefits are difficult to measure due the fact that the economic value of eHealth standards are financially intangible[12]. It is understood though that the cost benefit of implementations of clinical terminologies in Sweden resulted in total rationalisation potential of approximately US$34.7 million per year based on improving nursing documentation and communication alone. The predicted efficiency gains after three months and an overall accumulated potential gain was measured to be US$3.7 million after three years and US$65 million over five years. There are a limited number of documented examples of cost savings due to the limited number of nationally scaled implementations. These examples are limited largely to European Union nations.

[12] Dennis Lee, Nicolette de Keizer, Francis Lau, Ronald Cornet; Literature review of SNOMED CT use, Journal of the American Medical Informatics Association, Volume 21, Issue e1, 1 February 2014, Pages e11–e19, https://doi.org/10.1136/amiajnl-2013-001636

How does SNOMED CT work?

High level example of a SNOMED-CT Concept:



FSN	Myocardial Infarction (disorder)
SYN	Myocardial Infarction (disorder)
SYN	Cardiac Infarction
SYN	Heart Attack
SYN	MI Myocardial Infection
SYN	Myocardial Infarct

FIGURE 6: SNOMED-CT Concept Example

Formal SNOMED-CT Structure of "Myocardial Infarction"

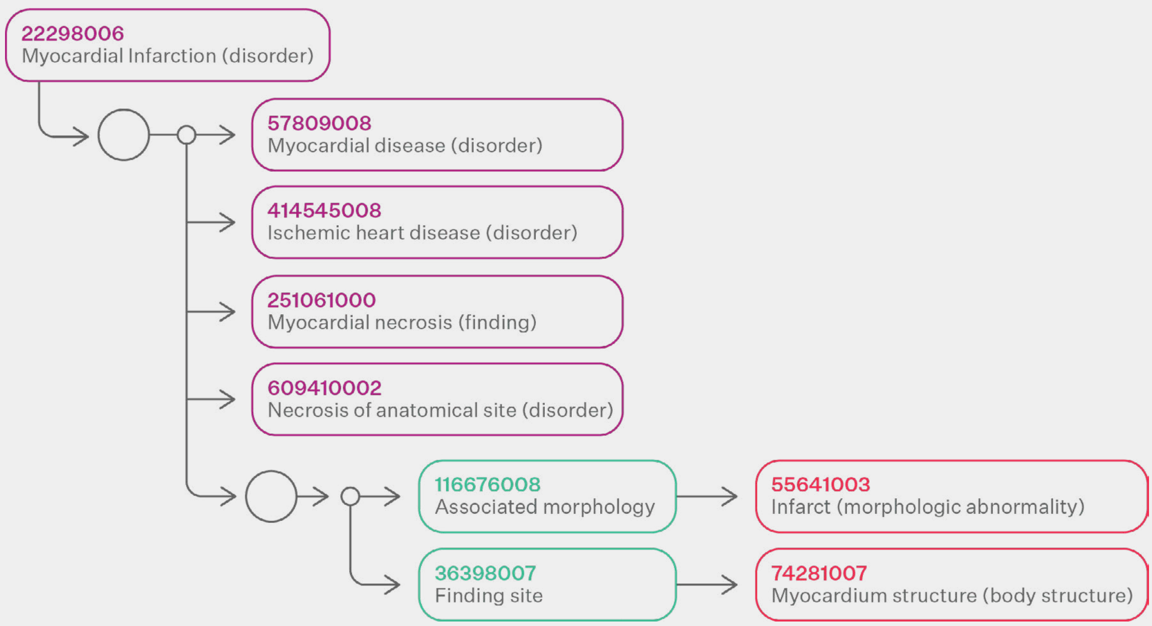


FIGURE 7: SNOMED-CT Structure

SNOMED International partners

There is active collaboration with many other internationals standards organisations and professional clinical bodies to support adoption and use of SNOMED CT with other international standards[14]. This aims to improve and enhance meaningful, consistent data capture and exchange as well as improving safety, functionality and semantic interoperability on a global scale. This includes partnerships with HL7, LOINC, genomic, medical device organisation, GS1, the World Health Organisation among other international organisation[15].

Clinical Terminology Services

Health Level 7 (HL7) Fast Healthcare Interoperability Resource (FHIR)® defines clinical terminology services as services that support healthcare applications use of codes and value sets without implementers or users having to become experts in the explicit details and the included code systems and terminological principles[16]. HL7 FHIR® supports terminology APIs which contain various clinical and medicine based terminology systems (including SNOMED CT). Such services can improve user interfaces, referencing clinical records, implementation efforts, scalability, content management and reduce costs through improved management mechanisms and enhancing many features and benefits mentioned in this paper.

What is Natural Language Processing?

Natural Language Processing (NLP) is a field of Artificial Intelligence which enables computers to analyze and understand the human language. In a health based context, this includes complex clinical language and jargon, because let's face it, it can be treated as another language in itself. The capability of NLP holds significant potential to assist clinicians with understanding, analyzing and interpreting healthcare problems.

How can Natural Language Processing (NLP) applications assist?

Ideally we would want to codify all information capture at the point of care. Defined data fields cannot capture every complex detail that is required by a clinical user. This is where NLP assists in the clinical analysis and reuse of information within free text narratives. This can be further supported by the clinical system knowledge cycle on page 2. There are many clinical service, population health and financial use-cases which can be supported by Machine Learning (ML) and NLP techniques, here are just a few:

1. **Predictive modeling of populations**[17]
   One of the greatest potential applications of NLP is population health analysis through cohort identification for similarity analytics. As systems mature, populations with similar clinical features could be identified through ML. This application of ML can apply a forecasted risk analysis and clinical decision support for individual patients to prevent onset of chronic conditions such as diabetes or rheumatoid arthritis[18]. In the same way, studies have been able to determine if a tumor was stable or showed regression or progression with high sensitivity and specificity.

---

[14] https://www.snomed.org/about/partnerships

[15] https://www.snomed.org/about/partnerships

[16] https://www.hl7.org/fhir/terminology-service.html

[17 & 18] Sheng Yu, Katherine P Liao, Stanley Y Shaw, Vivian S Gainer, Susanne E Churchill, Peter Szolovits, Shawn N Murphy, Isaac S. Kohane, Tianxi Cai; Toward high-throughput phenotyping: unbiased automated feature extraction and selection from knowledge sources. J Am Med Inform Assoc 2015; 22 (5): 993-1000. doi: 10.1093/jamia/ocv034

2. **Automated clinical trial cohort matching**[19,20]
De-identified populations that meet clinical trial criteria can be identified as potential candidates. Upon obtaining appropriate consent, approval and recruitment, this could result in more rapid developments of drug development and treatment advancements. Faster matching of patients to potential treatments which relieve symptoms and prolong life, especially in the case of cancer trials, has potential to assist researchers, clinicians, regulators and patients.

3. **Automated risk/adverse event monitoring/detection**[21,22]
Extraction of data from unstructured clinical documents with support from machine learning algorithms which can assist in monitoring hospital acquired infections (HAI). This can be achieved through event detection within diagnostic reports, discharge summaries and clinic letters which are monitored for temporal details relating to specific infections, diagnoses, physical features and treatments. One example is urinary tract infections (UTI)[23], NLP based monitoring could assist in the identification of patients/cohorts meeting certain risk criteria for UTIs and other conditions to improve patient safety[24].

4. **Automated/semi-automated clinical coding**[25]
Extracting content from unstructured clinical documents (discharge summaries, reports, letters) for semi-automated coding and population health indexing. This would be the basis of a real-world health outcomes solution since prospective problem lists and messaging attributes have well recognised limitations when used in isolation.

5. **Clinical decision support**[26]
Automated report classification and analysis can provide clinicians with guidance of which clinical reports may recommend further analysis or treatment. For example, analysis of clinical documents of individuals identified to have a diagnosis of tuberculosis who are indicated to require further screening or treatment.

6. **Automated de-identification of clinical documents for research**[27] **or to information access processes**[28]
Privacy and sensitivity are significant considerations when analyzing protected health information (PHI). Data used can be de-identified so that personal identifiers will be replaced with non-identifying variable names[29] supported by Named Entity Recognition and co-reference detection[30].

---

[19] Patel, C., Cimino, J., Dolby, J., Fokoue, A., Kalyanpur, A., Kershenbaum, A. & Srinivas, K. (2007). Matching patient records to clinical trials using ontologies. The Semantic Web, 816-829.

[20] Ni, Y., Wright, J., Perentesis, J., Lingren, T., Deleger, L., Kaiser, M., ... & Solti, I. (2015). Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. BMC medical informatics and decision making, 15(1), 28.

[21 & 23] Redder, J.D. et al. Analysing risk factors for urinary tract infection based on automated monitoring of hospital-acquired infection. Journal of Hospital Infection , Volume 92 , Issue 4 , 397 - 400

[22 & 24] Proux D, Hag.ge C, Gicquel Q, Pereira S, Darmoni S, Segond F, Metzger MH. Architecture and Systems for Monitoring Hospital Acquired Infections inside a Hospital Information Workflow -Proceedings of the Workshop on Biomedical Natural Language Processing. Pg 43-48

[25] Stanfill, M. H., Williams, M., Fenton, S. H., Jenders, R. A., & Hersh, W. R. (2010). A systematic literature review of automated clinical coding and classification systems. Journal of the American Medical Informatics Association, 17(6), 646-651

[26] Demner-Fushman D, Chapman WW, Mc- Donald CJ. What can natural language pro- cessing do for clinical decision support? J Biomed Inform 2009;42(5):760–772.

[27] Neamatullah, I., Douglass, M. M., Li-wei, H. L., Reisner, A., Villarroel, M., Long, W. J., ... & Clifford, G. D. (2008). Automated de-identification of free-text medical records. BMC medical informatics and decision making, 8(1), 32.

[28] https://github.com/NICTA/t3as-redact

[29] Stubbs, A., Kotfila, C., & Uzuner, Ö. (2015). Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. Journal of Biomedical Informatics, 58. doi:10.1016/j.jbi.2015.06.007

[30] https://github.com/NICTA/t3as-redact

The benefit of using Deep Learning (DL)

Highly accurate predictions of diagnosis and patient needs through unsupervised deep feature learning[31] has already been modeled with significant results. This has not, however, made full use of clinical terminologies to further support deep semantic description logic for more highly predictive modeling for personalised treatment options, diagnostic support and accurate precision medicine.

Patterns and relationships found through deep learning (DL) across clinical information has the opportunity for uncovering unexpected or previously unknown enhancements and knowledge of clinical best practice guidelines, treatments, diseases and medicine effectiveness. Results from a recent study was able to determine the regression or progression of disease, with a sensitivity of 81% and specificity of 92%[32].

Technology such as DL, ML and NLP support healthcare professionals and clinicians to improve their focus on supporting patients with complex diagnoses, treatments and interventions. This is in combination with the ever-important aspects of interpersonal relationships, empathy, informed decision making, trust and much-needed human attention and care.

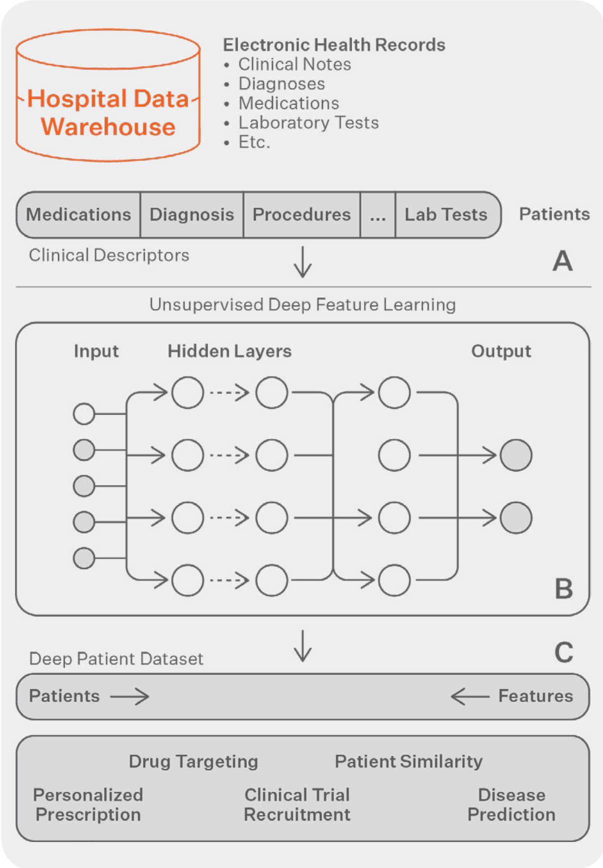


FIGURE 8: Model to Predict the Future of Patients
Above diagram from Deep patient: An unsupervised representation to predict the future of patients from the electronic health records.[33]

[31] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific reports,6.

[32] Pons, E., Braun, L. M., Hunink, M. M., & Kors, J. A. (2016). Natural language processing in radiology: a systematic review. Radiology, 279(2), 329-343.

[33 (FIGURE 8)] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific reports,6.

Conclusion

To provide safe, quality healthcare and achieve semantic interoperability in the future, organisations need to employ consistent terminology systems that match the wider healthcare market. While cost benefits are difficult to measure, they will likely be realised after appropriate assessments of the system are completed[33].

Clinical terminology services simplify the implementation of such systems, reducing information management efforts and enhancing user interfaces, clinical decision support and content management.

More advanced technology is required to better enable clinicians and healthcare professionals to more effectively focus on complex cases and the more human aspects of healthcare.

When considering any implementation of a CIS, it is critical to understand whether these systems can support the needs of all users of clinical and business information, not only through interoperability of data, but interoperability of meaning.

Does your health service understand the benefits of a CIS that supports clinical terminology systems and services?

SNOMED®, SNOMED CT® and IHTSDO® are registered trademarks of International Health Terminology Standards Development Organisation. SNOMED CT® licensing information is available at http://snomed.org/licensing. For more information about SNOMED International and SNOMED International Membership, please refer to http://www.snomed.org or contact info@snomed.org.

[33] Thiel, R., Birov, S., Piesche, K., Højen, A. R., Gøeg, K. R., Dewenter, H., ... & Stroetmann, V. N. (2016, September). The Costs and Benefits of SNOMED CT Implementation: An Economic Assessment Model. In MIE (pp. 441-445).

ORION HEALTH