© 2018 IMIA and Schattauer GmbH

Recent Developments in Clinical Terminologies – SNOMED CT, LOINC, and RxNorm

Oliver Bodenreider[1], Ronald Cornet[2], Daniel J. Vreeman[3]

[1] Senior Scientist, Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD, USA
[2] Associate Professor, Department of Medical Informatics, Academic Medical Center - University of Amsterdam, Amsterdam Public Health research institute, Amsterdam, Netherlands
[3] Director, LOINC and Health Data Standards Regenstrief Center for Biomedical Informatics; Regenstrief-McDonald Scholar in Data Standards Indiana University School of Medicine; Research Scientist Regenstrief Institute, Inc., Indianapolis, USA

Summary

Objectives: To discuss recent developments in clinical terminologies. SNOMED CT (Systematized Nomenclature of Medicine Clinical Terms) is the world's largest clinical terminology, developed by an international consortium. LOINC (Logical Observation Identifiers, Names, and Codes) is an international terminology widely used for clinical and laboratory observations. RxNorm is the standard drug terminology in the U.S.

Methods and results: We present a brief review of the history, current state, and future development of SNOMED CT, LOINC and RxNorm. We also analyze their similarities and differences, and outline areas for greater interoperability among them.

Conclusions: With different starting points, representation formalisms, funding sources, and evolutionary paths, SNOMED CT, LOINC, and RxNorm have evolved over the past few decades into three major clinical terminologies supporting key use cases in clinical practice. Despite their differences, partnerships have been created among their development teams to facilitate interoperability and minimize duplication of effort.

Keywords

Vocabulary, controlled; Logical Observation Identifiers Names and Codes, RxNorm; Systematized Nomenclature of Medicine

Yearb Med Inform 2018:129-39
http://dx.doi.org/10.1055/s-0038-1667077

Introduction

Over the past decades, biomedical ontologies and terminologies have increasingly been recognized as key resources for knowledge management, data integration, and decision support [1]. Among the dozens of ontologies and terminologies available, some have been identified as having high impact on clinical practice and biomedical research [2] and their evolution has been investigated [3].

The recent acceleration in the deployment of electronic health record (EHR) systems has precipitated the emergence of a few terminologies and their wide adoption in the clinical community. Two of them, the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and the Logical Observation Identifiers, Names, and Codes (LOINC®), have become international standards. The last one, RxNorm, is used mostly in the U.S., but similar national drug terminologies exist in other countries (e.g., the NHS Dictionary of medicines and devices (dm+d) [4] in the U.K., the Australian Medicines Terminology (AMT) [5] in Australia) and could have been substituted for RxNorm in this review. In addition to being designated standards mandated for use in U.S. governmental programs, such as the Meaningful Use incentive program [6], these three clinical terminologies have also been selected as the terminological backbone of the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) used for clinical data warehouses internationally by OHDSI, the Observational Health Data Sciences and Informatics collaborative [7].

While SNOMED CT, LOINC, and RxNorm are referenced in many articles investigating aspects of their organization or use, no recent publication has characterized their similarities and differences or analyzed how they can harmoniously contribute to an interoperable health information ecosystem. The objective of this work is to provide a brief review of their history, current state, and future development. This work is also an opportunity to contrast them, and outline areas for greater interoperability among them. In addition to being a review of the literature, this article also reflects the experience of the three authors with the development of these three terminologies, respectively, for at least a decade.

SNOMED CT

Brief History

Since the inception of the Structured Nomenclature of Pathology, SNOP, in 1965, the various versions of SNOMED have developed both in terms of content and underlying representation. Development of content is clearly illustrated by the number of elements in the various systems. SNOP comprised "about

15,000 distinct medical objects, processes, and concepts" [8]. It developed further into SNOMED-2, and later SNOMED International, which contained 150,000 concepts in the mid-1990s [9]. Its successor, SNOMED Reference Terminology (SNOMED RT), contained over 120,000 active concepts [10]. The initial version of SNOMED CT, the merger of SNOMED RT and the UK-based Clinical Terms Version 3, released January 2002, consisted of 278,000 active concepts, a number that has grown to 341,000 in the January 2018 release of SNOMED CT.

The representation has gone through phases of increasing formal rigor. The initial SNOP and SNOMED versions were multi-axial systems that enabled post-coordination. Whereas SNOP started as a 4-axis system, SNOMED International had expanded to using 12 axes: anatomy (topography), morphology (pathologic structure), normal and abnormal functions, symptoms and signs of disease, chemicals, drugs, enzymes and other body proteins, living organisms, physical agents, spatial relationships, occupations, social contexts, diseases/diagnoses and procedures [9]. SNOMED RT abandoned the use of self-standing axes that could be combined into composite codes in favor of a description logic formalism called Ontylog [11], based on the Knowledge Representation System Specification (KRSS) syntax and the K-REP system [9]. Following contemporary naming conventions for description logics, the set of constructors used corresponds to the $\mathcal{EL}^{++}$ language [12]. SNOMED CT has continued to use this description logic as its underlying representation.

The evolution of the representation of "arthritis" through the history of SNOMED provides an illustration of editorial changes over time, with examples from SNOMED-2, SNOMED International, and the current version of SNOMED CT (Table 1).

Table 1 Evolution of the representation of "arthritis" through the history of SNOMED.

Source	Code	Name	Definition
SNOMED-2	D-3060	Inflammatory Athropathy	D-3050 | Disease of Joints | ; M-40000 | Inflammation |
SNOMED International	D1-20050	Arthritis	T-15001 | Joints | ; M-40000 | Inflammation |
SNOMED CT	3723001	Arthritis (disorder)	Formal definition, using concept identifiers: $\equiv$ 64572001 $\sqcap$ $\exists$ RoleGroup.($\exists$ 116676008.23583003 $\sqcap$ $\exists$ 363698007.39352004) Verbose definition, using fully specified names: Equivalent to : Disease (disorder) AND RoleGroup SOME (Associated morphology SOME Inflammation (morphologic abnormality) AND Finding site SOME Joint structure (body structure))

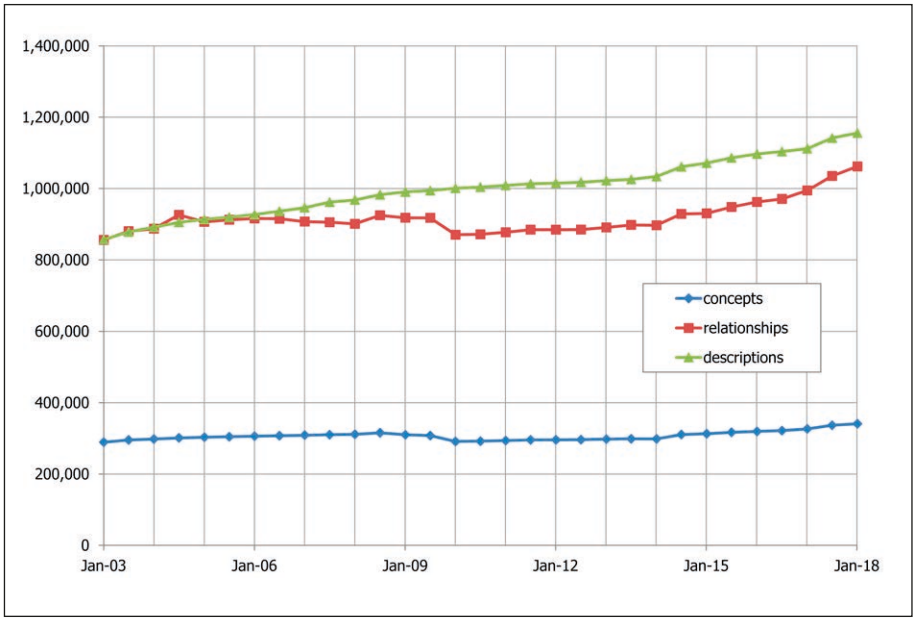

Fig. 1 Evolution of the number of concepts, relationships, and English descriptions in SNOMED CT over time.

Current State

SNOMED CT Content

Since the first release of SNOMED CT in January 2003, updated versions have been released twice a year. Figure 1 shows the evolution of the number of concepts, relationships, and English descriptions over time. It shows that maintenance can lead to a decrease of the number of elements, e.g., in January 2010 when, among others, the veterinary content was removed from the International Release. The January 2018 release contains 341,000 active concepts, 1,062,000 active relationships and 1,156,000 active descriptions. The largest categories of concepts in SNOMED CT are disorders (22%), procedures (17%), body structures (11%), clinical findings other than disorders (10%), and organisms (10%). In SNOMED CT parlance, descriptions are labels that describe the concepts, not textual definitions of the concept, of which SNOMED CT contains only 4,000.

A significant design criterion for SNOMED is to keep concept expressions simple enough to be broadly usable by clinicians, while maintaining faithful representation of concept meaning [11]. Validity of concept expressions is determined by adherence to the concept model underlying SNOMED CT. This concept model specifies

which types of relationship are allowed for which concepts, and what the allowed values are. For example, the concept model specifies that "Method" is an allowed attribute relationship for "Procedure" concepts, for which the value should be a type of "Action". The January 2018 release uses 82 relationship types, almost twice the number of the initial 42 relationship types. Since July 2017, the concept model is available in machine-readable form, and distributed as separate tables which are part of the semi-annual releases. Other tables provide mappings to several versions of the International Classification of Diseases (ICD-9-CM, ICD-10 and ICD-O). The current release format, called Release Format 2 (RF2), supports versioning, providing access to any previous release of SNOMED CT.

SNOMED CT Adoption

When ownership of SNOMED CT was transferred to the newly formed International Health Terminology Standards Development Organization (IHTSDO, recently renamed SNOMED International) in 2007, this organization consisted of nine member countries, with a joint population approaching 500 million. In the majority of the inaugural member countries (USA, UK, Canada, Australia, and New Zealand), the official language was English, and a minority (Netherlands, Sweden, Denmark, Lithuania) had other official languages.

Over time, the organization has expanded, now covering 32 countries as of May 2018 [13], with a total population of over 2 billion, and including a broad range of languages. Each year, the SNOMED CT Expo [14] provides a forum for EHR vendors, health terminology specialists, and the community of practice to exchange best practices and measure progress towards the implementation of SNOMED CT across the world. The primary role of SNOMED CT as a reference terminology does not require localization for the official language. Still, to be understood by developers and clinicians, a complete or partial translation is an asset. Alternatively, interface terminologies in the everyday language can be developed and mapped to SNOMED CT [15]. Partial or full translations of SNOMED CT have been developed in Danish, Dutch, French, Spanish, and Swedish.

Membership is a prerequisite for adoption, but it is not sufficient. While translations and interface terminologies may facilitate adoption, regulations have had a strong impact on adoption at a national level. Such regulations are in place in the USA, where EHR systems are required to use SNOMED CT for documenting problem lists, procedures, and some clinical findings, such as smoking status [16], and in the UK, where many health information systems "must use SNOMED CT as the clinical terminology standard within all electronic patient level recording and communications before 1 April 2020" [17]. Adoption of SNOMED CT was described in [18] and is monitored by SNOMED International [19].

SNOMED CT Collaborations

SNOMED CT is not developing in isolation, but increasingly collaborating and harmonizing with other relevant standards in the area of structured and standardized storage and exchange of biomedical data. As described in [3], this includes mapping, as well as binding of information model and terminology.

Mappings are maintained between SNOMED CT and a number of terminology systems. These include the World Health Organization (WHO) classifications (e.g., versions ICD-10 and ICD-O of the International Classification of Diseases), as well as the International Classification of Primary Care (ICPC-2), the International Classification for Nursing Practice (ICNP), and LOINC. The latter is especially important in the context of the specifications of the U.S. Meaningful Use incentive program, in which LOINC is the primary choice for specifying attributes, and SNOMED CT the system of use for the relevant attribute values. In other words, LOINC is used to specify the question (e.g., 29308-4: "what is the diagnosis?"), and SNOMED CT to specify the answer (e.g., 3723001: "Arthritis"). Adoption of this principle by both the Regenstrief Institute and SNOMED International has formed the basis for a cooperation agreement in 2013[1].

[1] https://loinc.org/download/cooperation-agreement-between-ihtsdo-and-regenstrief-institute-2013-07/

SNOMED CT also has collaborations for specific domains. In the context of rare diseases, collaboration with Orphanet leads to harmonization of content between SNOMED CT and ORDO, the Orphanet ontology of rare diseases [20]. Medical device terminology is addressed in collaboration with the Global Medical Device Nomenclature Agency (GMDNA). Finally, Kaiser Permanente's Convergent Medical Terminology (CMT) provides concepts and descriptions to be considered for inclusion.

Moreover, SNOMED CT forms the backbone for the development of national extensions by member countries of SNOMED International. National extensions typically contain concepts that are important in a given country, but not in scope for the international release of SNOMED CT. Other extensions, such as the veterinary extension, contain content specific to a given community of practice.

Future Directions

The adherence to $\mathcal{EL}^{++}$ and, since 2009, the provision of a script to convert SNOMED CT into OWL (Web Ontology Language) representation have enabled the use of SNOMED CT in generic tools based on Semantic Web technologies, such as Protégé [21] and description logic reasoners, such as SnoRocket, Pellet, and FaCT++ [22]. This has resulted in a broad range of interests and developments. On the one hand, there are organizations that rely on the terminology as provided by SNOMED CT, mainly using the hierarchical relationships. On the other hand, there is an interest in the use of reasoners, which are essential for processing fully defined concepts in an extension, and in expanding the language underlying SNOMED CT, for example to include so-called concrete domains (e.g., dose strength of medication) or to support negation, e.g., to explicitly express "non-viral disorders" [23]. The challenge is to provide a balance between these seemingly conflicting requirements, for which SNOMED International has launched a "proposal to enhance SNOMED CT's logic capabilities" late 2017 [24]. This proposal will lead to the introduction of new tables, in which the OWL representation of SNOMED

CT content will be provided. In the longer run, this OWL representation may allow for the use of a more expressive language, but the resulting relationships table will still reflect the current expressivity. This accommodates concomitant use of simpler tools for processing the SNOMED CT hierarchy and use of a more expressive language that can be processed using default description logic reasoners.

LOINC

Brief History

Logical Observation Identifiers, Names, and Codes (LOINC®) is a clinical terminology for identifying health measurements, observations, and documents. LOINC was initiated in 1994 by the Regenstrief Institute, a non-profit medical research organization associated with Indiana University. By 1994, many electronic systems were sending clinical information as discrete results using messaging standards such as Health Level Seven (HL7) or ASTM 1238 (American Society for Testing and Materials). Inside these messages, laboratories and clinical systems used local, idiosyncratic names and codes to identify which test was being reported. This was problematic for data exchange and aggregation because of the large resources it takes to map codes between every participating system.

To solve this problem, Regenstrief organized the LOINC Committee to develop a common terminology for laboratory and clinical observations [15, 16]. Existing terminologies were not granular enough, focused on coding for billing rather than clinical results delivery, or did not fit with the messaging models being used. Because such a standard did not exist, the LOINC Committee embarked on creating a terminology with the appropriate level of granularity for defining the names of observations used in laboratory and clinical information systems. Since its creation, LOINC has continued to be developed and published by the Regenstrief Institute as a freely available global standard with a rich set of implementation tools. Today, LOINC is used by a diverse global community who propel its continuous development.

LOINC's primary role is to provide identifiers and names for observations [25]. Here, we use observation as a generic term for health data represented in a particular way. In different domains, these might be called tests, variables, or data elements. Within and among health IT systems, observations are communicated with a structure that has two key structural elements. The first element identifies what the observation is, e.g., diastolic blood pressure, hematocrit, tobacco smoking status. The second element carries the result value of the observation, e.g., 80 (mmHg), 40 (%), or "current every day smoker". When used together, these two elements carry the instance of a specific test result for a given patient. A common pairing is to use LOINC as the standard code for the observation, and SNOMED CT as the standard code for the observation value when needed. This approach is endorsed by the developers of both terminologies and fits their design purpose.

Some mistakenly believe that LOINC is only for laboratory tests. It is true that the first release of LOINC in May 1995 contained only terms for laboratory testing, but by December of 1996, LOINC had already added about 1,500 clinical measurement terms (vital signs, ECG measures, etc). Now, more than 20 years and 60 releases later, LOINC has grown significantly in other domains, including radiology [26], standardized survey instruments and patient-reported outcomes measures [27], clinical documents [28], nursing management data [29], and nursing assessments [22]. The LOINC Committee itself is now composed of three major composite committees: Laboratory, Clinical, and Radiology.

Current State

LOINC Content

Regenstrief continues to update LOINC and publish twice-yearly releases. New concepts are added to LOINC based on submissions from end users. The current version (December 2017, version 2.63) contains more than 86,000 terms covering the full scope of laboratory testing (chemistry, microbiology, molecular pathology, …etc.) and a broad range of clinical measurements (e.g., vital signs, ECG, patient-reported outcomes, …etc.). In addition to distributing the terminology, Regenstrief makes available at no cost a variety of supporting tools and resources, including the Regenstrief LOINC® Mapping Assistant (RELMA®) and online search application [30].

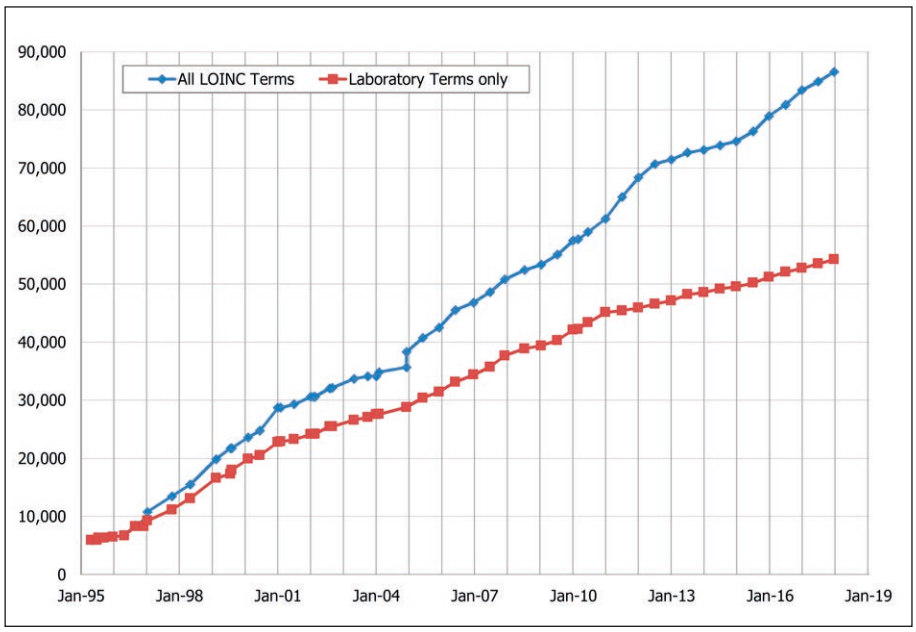

Fig. 2 Evolution of the number of LOINC terms (all terms, and laboratory terms only) over time.

LOINC uses a semantic data model containing six major and up to four minor attributes to create fully-specified names for concepts.[16] The major attributes of the LOINC name are:

1. component (e.g., what is measured, evaluated, or observed)
2. kind of property (e.g., mass, substance, catalytic activity)
3. time aspect (e.g., 24-hour collection)
4. system type (e.g., context or specimen type within which the observation was made)
5. type of scale (e.g., ordinal, nominal, narrative)
6. type of method (e.g., procedure used to make the measurement or observation).

The atomic elements that make up each LOINC term name are called "Parts" and are also assigned identifiers. The combination of attribute values produce term names that are detailed enough to distinguish among similar observations. Of the six attributes, only the method is optional and used only when necessary to distinguish among clinical important differences.

For example, the molar concentration of sodium measured in the plasma (or serum) with quantitative result is represented in LOINC as shown in Table 2, along with the correspondence between LOINC Parts and their equivalent concepts in SNOMED CT.

Over time, LOINC has not only grown in size (Figure 2), but also developed additional data structures and content around its main codes for individual observations. The LOINC release contains a basic hierarchy that organizes LOINC codes into a set of is-a relationships. LOINC now has a detailed model for representing enumerated collections of observations, such as laboratory panels (complete blood count), assessment instruments (e.g., PHQ-9), data sets (National Trauma Data Standard), and forms (e.g., U.S. Standard Birth Certificate). This content is published in a special release artifact called the LOINC Panels and Forms File, with the current version (December 2017) containing more than 3,000 panel terms. LOINC also has a detailed model for connecting observation terms to structured answer lists. These answer lists can be defined by extension or intension, and linked to observation terms with different types (e.g., example, preferred, normative). This content is published in the LOINC Answer File, with the current version (December 2017) containing links between 15,966 unique LOINC terms and 3,239 unique answer lists composed of coded LOINC Answers, and including mappings to other terminologies such as SNOMED CT where they exist. LOINC now also publishes the atomic elements (called Parts) that make up each LOINC term name. The LOINC Part File includes the Part identifiers and names, links between Parts and LOINC terms, and mappings from LOINC Parts to other terminologies such as SNOMED CT and RxNorm where they exist.

Table 2 Representation of molar concentration of sodium measured in the plasma (or serum) with its quantitative result in LOINC (2951-2) and correspondence between LOINC Parts concepts and SNOMED CT.

Attribute	Value (Part)	LOINC code	SNOMED CT code
Component	Sodium	LP15099-2	304050002
Property	SCnc – Substance Concentration (per volume)	LP6860-3	118556004
Timing	Pt – Point in time (Random)	LP6960-1	123029007
System	Ser/Plas – Serum or Plasma	LP7576-4	122592007
Scale	Qn – Quantitative	LP7753-9	30766002
Method	--		

In the last decade, Regenstrief has put special emphasis on crafting narrative summaries for content within LOINC. These annotated summaries focus primarily on explaining precisely what is being measured, how the observation is performed, what it is used for, and its clinical relevance. Having such definitions within LOINC greatly enhances the ability of LOINC users to make accurate mapping choices. Currently, there are about 10,200 definitions for 10,000 unique parts as well as term-level definitions for about 10,000 LOINC terms.

LOINC Adoption

LOINC has become widely adopted as the standard for laboratory and clinical observations in the USA and internationally. Today, there are more than 60,000 registered users from 170 countries and it has been translated into 18 variants of 12 languages [31]. More than 30 countries have adopted LOINC as a national standard. There are many different kinds of LOINC users, including reference labs, radiology centers, health-related federal agencies, care organizations, professional societies, health information exchange networks, insurance companies, health IT vendors, in vitro diagnostic (IVD) testing vendors, health app developers, and more.

Adoption of LOINC has enabled many kinds of large scale informatics applications. Here we mention just a few examples. National health record systems in Austria (called "ELGA") [32] and Estonia (called "ENHIS") use LOINC for standard coding of laboratory tests. National health programs in Australia, Malaysia, the Netherlands, New Zealand, Qatar, Rwanda, Saudi Arabia, Turkey, and many other countries use LOINC to enable interoperability. These programs support initiatives in maternal health, disease surveillance, international cross-border patient care, and more. In France, the public health code for practice of laboratory medicine adopted the French specification of IHE XD-LAB profile for laboratory reporting, which includes LOINC for lab test identification. Many countries, such as Brazil, Italy, Spain, and Canada set national policies that enable interoperability with regional health information exchanges. In Canada, for example, those exchanges now include nearly all lab results across all provinces.

Within the USA, the Meaningful Use incentive program requires LOINC in messages reporting laboratory test results, exchanging medical summaries, and sending data to cancer registries and public health agencies. About 20 U.S. federal agencies have adopted LOINC in various programs [33]. The U.S. Food and Drug Administration (FDA) will be requiring LOINC for lab test

data in regulated studies starting after March 2020, and they have articulated a broad vision of using real-world evidence in post market surveillance that depends on standardized data. The 2017 Interoperability Standards Advisory of the Office of the National Coordinator for Health Information Technologies (ONC) lists LOINC for many interoperability needs, including functional status, laboratory tests, imaging diagnostics, nursing observations, vital signs, and social determinants of health. The Centers for Medicaid and Medicare Services adopted LOINC for the patient assessment instruments required in post-acute care settings. Large-scale research networks, such as PCORnet, the National Patient-Centered Clinical Research Network [34], OHDSI, the Observational Health Data Sciences and Informatics research group [28], and the FDA's Mini-Sentinel [29], all use LOINC in their common data models.

LOINC Collaborations

Regenstrief is committed to working with developers of health data standards that are complementary to LOINC, including syntax standards for data exchange and other terminology standards. Regenstrief and HL7 have a long-standing collaboration; a few joint work highlights include clinical genomics guides, claims attachments specifications, and approaches for representing vocabulary standards in Fast Healthcare Interoperability Resources (FHIR) terminology services (Regenstrief Institute plans to make core LOINC content available via a FHIR API as part of its normal release process beginning Summer 2018). Regenstrief and the IEEE Standards Association, developer of the 11073™ standards, are collaborating to enhance the interoperability of traditional medical devices and personal health devices. Regenstrief is an active member of the Health Standards Collaborative (HSC) which provides an executive forum for senior leadership of the U.S. healthcare standards development community to improve interoperability. Recently, Regenstrief worked with the in vitro diagnostic (IVD) Industry Connectivity Consortium (IICC) on a new specification for publishing vendor IVD tests associated with a set of LOINC codes that identify the distinct observations produced by the test [35].

In 2013, Regenstrief and SNOMED International formed a landmark long-term collaborative relationship to link the rich clinical semantics of SNOMED CT to LOINC, which provides extensive coverage of laboratory tests and clinical measurements. Regenstrief and the Radiological Society of North America (RSNA) have unified the RadLex™ Playbook and LOINC radiology terms to produce a single, comprehensive standard for radiology procedures with a shared governance.

Future Directions

The growth in user adoption and continued innovation in diagnostic testing continue to fuel requests for new LOINC content. In particular, global initiatives in precision health are expanding the interest in representing genomic data. LOINC has been actively involved in the efforts to develop new models of reporting clinical genomic results, and includes codes for cytogenetic or mutation analysis tests, specific chromosomal alteration or mutation testing, and fully structured discrete genetic test reporting [31]. In addition, efforts such as a collaboration with the Clinical Pharmacogenetics Implementation Consortium (CPIC) are facilitating the application of pharmacogenetics to clinical practice by developing guidelines for clinicians [32]. Precision health initiatives are also driving interest in social, behavioral, and environmental determinants of health, and LOINC is expected to keep adding content for representing assessment instruments and community-level variables.

Significantly less effort is needed to achieve interoperability when data is standardized upstream at the producing systems. In this regard, we are particularly excited about efforts by IVD vendors to adopt and publish mappings from their internal codes to LOINC codes. The LIVD standard [35] will greatly improve the efficiency and consistency with which laboratories can deploy LOINC. We also anticipate that measurement devices and data collection apps will increasingly incorporate standard terminologies into their emitted data, which will ease their incorporation into downstream systems.

RxNorm

Brief History

At the dawn of the twenty-first century, there was no standard terminology for drugs in the U.S. [36]. While many companies provide information about drugs for use in clinical information systems (e.g., to support clinical decision), each drug knowledge base defines its own codes and names for medications, making it difficult to exchange information across clinical information systems and to retrieve information from different systems. For example, the same transdermal patch delivering 0.583 milligrams of nicotine per hour for 24 hours (e.g., to help with smoking cessation) is referred to with the following codes and names in three of the major drug knowledge bases:

- 2707 nicotine 14 mg/24 hr transdermal film, extended release
- 102712 Nicotine 14 MG/24 HR Transdermal Patch, Extended Release
- 016426 NICOTINE 14 mg/24 hour TRANSDERM PATCH, TRANSDERMAL 24 HOURS

In addition to capitalization differences, there is a lack of standardization in naming dose forms (transdermal film vs. transdermal patch) and units (24 hr vs. 24 hour), making it difficult to parse names from multiple systems.

RxNorm was created to address the lack of standardization in drug names, and to make drug terminologies interoperable by integrating them into a reference system [34]. Since the 1990s, the National Library of Medicine (NLM) has released the Unified Medical Language System (UMLS), a terminology integration system in which names and codes from all major biomedical terminologies are integrated, and equivalent terms across vocabularies are identified. RxNorm can be thought of as a specialized version of the UMLS. While both UMLS and RxNorm are built upon existing vocabularies, one major difference between the two is that UMLS generally does not create names for biomedical entities. In contrast, RxNorm creates a "normal form" for every drug entity it integrates. In RxNorm parlance, normal forms are standardized terms for drug entities. For

example, the normal form for the nicotine patch discussed above is "24 HR Nicotine 0.583 MG/HR Transdermal System", to which RxNorm permanently assigns the concept unique identifier 198029. Unlike the UMLS, RxNorm also defines a rich network of named relationships among the various types of drug entities it integrates (e.g., ingredient, brand name, generic drug product, branded drug product).

The main use cases RxNorm was designed to support include electronic prescribing, drug information exchange, and mapping across drug vocabularies (e.g., for medication reconciliation purposes). Standard names and codes for drugs were also expected to facilitate the development of standard clinical decision support rules involving medications.

RxNorm started in 2002. It was first released through the UMLS [37] and has been published as an independent terminology with monthly releases since November 2004, and weekly updates since 2008 to reflect drugs recently marketed in the U.S. market. The number of sources integrated in RxNorm has grown from 5 to 13.

Current State

RxNorm Content

Sources. RxNorm currently integrates terminology information from most drug knowledge base vendors (e.g., First DataBank, Multum, Micromedex, Gold Standard), as well as the drug component of standard terminologies (e.g., SNOMED CT, MeSH). RxNorm also integrates sources from several U.S. federal agencies, including the Food and Drug Administration (FDA) Structured Product Labels, the Veterans Health Administration (VHA) National Drug File, the Centers for Medicare & Medicaid Services (CMS) Formulary Reference File, as well as the list of vaccines administered (CVX) maintained by the National Center of Immunization and Respiratory Diseases at the Centers for Disease Control and Prevention (CDC). More recently, RxNorm has also integrated international drug resources, such as the Anatomical Therapeutic Chemical (ATC) Classification System and DrugBank, a drug resource used in many research projects [36].

Organization. The RxNorm drug model distinguishes between generic and branded drug entities and identifies three major definitional elements for drug products, namely ingredient, strength, and dose form, along with two additional elements, quantity factor and qualitative distinction [38]. The major types of drug entities include ingredient (e.g., Azithromycin), brand name (e.g., Zithromax), clinical drug (e.g., Azithromycin 250 MG Oral Tablet), and branded drug (e.g., Zithromax 250 MG Oral Tablet). Generic and branded packs are collections of clinical and branded drugs, respectively (e.g., Z-PAK, a branded pack of 6 tablets of 250 milligrams of azithromycin). While drugs are sold mostly pre-packaged in some countries (e.g., individual tablets in blister packs), packs are available for a minority of drugs in the U.S. In addition to the six major drug entities, RxNorm provides entities for navigational purposes. Clinical and branded drug components associate ingredient (or brand name) and strength information, and clinical and branded dose forms associate ingredient (or brand name) and dose form information. As shown in Figure 3, the various types of drug entities in RxNorm are organized into a graph that can be easily traversed, enabling users to navigate among types of entities (e.g., to find the branded drugs associated with a given ingredient).

While its main organization principles have remained centered on the ingredient-strength-dose form triad, RxNorm has added definitional features to accommodate distinctions, such as duration for extended release forms and transdermal systems (e.g., the quantity factor "24 HR" in "24 HR Nicotine 0.583 MG/HR Transdermal System") and size of unit of presentation for injectable medications (e.g., the quantity factor "40 ML" in "40 ML Ciprofloxacin 10 MG/ML Injection"), as well as qualitative elements for specific drugs (e.g., the qualitative distinction "Sugar-Free" in "Sugar-Free Cholestyramine Resin 4000 MG Powder for Oral Suspension"). A drug product in RxNorm is fully defined by its set of ingredient, strength, dose form, quantity factor, and qualitative distinction values. Types of ingredients include multi-ingredients (e.g., Sulfamethoxazole / Trimethoprim) and "precise ingredients" (e.g., Atorvastatin calcium, Morphine Sulfate), generally denoting, salts, esters, and complexes of base substances.

As illustrated in the examples above, RxNorm normal forms reflect the definitional features of drug entities. Additionally, RxNorm explicitly links drug entities to these features, which supports efficient processing. For example, the generic nicotine patch "24 HR Nicotine 0.583 MG/HR Transdermal System" is linked to its ingredient (Nicotine), strength (0.583 MG/HR), dose form (Transdermal System), and quantity factor ("24 HR").

Coverage. The scope of RxNorm is different from that of drug knowledge bases. RxNorm focuses on drug names and codes. In other words, clinical information (e.g., indications, drug classes, and drug-drug interactions) and administrative information (e.g., drug pricing) are out of scope for RxNorm. Although it integrates international sources (e.g., ATC, DrugBank), RxNorm focuses on drug products marketed in the U.S. Finally, non-therapeutic radiopharmaceuticals, bulk powders, contrast media, food, dietary supplements, and medical devices, such as bandages and crutches, are all out of scope for RxNorm [38].

In addition to the codes from drug terminologies and knowledge bases, RxNorm integrates codes from the FDA's National Drug Code (NDC) Directory, which serve as product identifiers for drugs in billing transactions, and contains links to Structured Product Labels (i.e., package inserts) submitted by drug manufacturers to the FDA.

The February 2018 edition of RxNorm includes 11,697 (base) ingredients, 6,053 brand names, 18,486 clinical drugs, 10,425 branded drugs, 380 generic packs, and 456 branded backs. Figure 4 shows the evolution of the number of active RxNorm concepts over time[2].

[2] The dramatic variation in number of concepts simply reflects changes in editorial policies (e.g., creation of dose form groups for each clinical drug in October 2011 and massive removal of obsolete brand names in 2013).

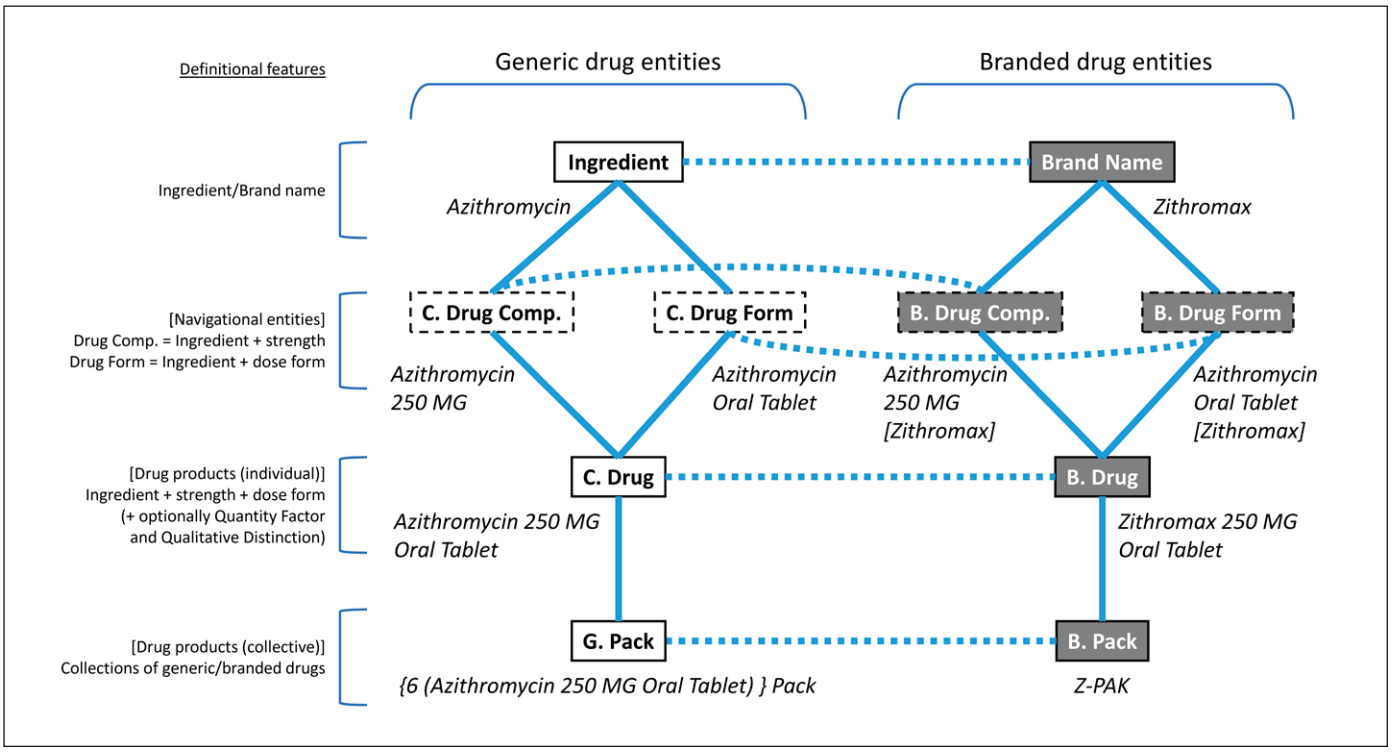

Fig. 3 Organizational structure of RxNorm, with its different types of drug entities (ingredient, brand name, clinical drug component, branded drug component, clinical drug, branded drug, generic pack, branded pack), using Azithromycin products as an example. Generic entities are on the left-hand side and branded entities are on the right hand-side. The definitional elements for each type of drug entity are indicated on the left. The lines between types of drug entities represent named relationships in RxNorm. (Relationship names are omitted for simplicity).

Permanent identifiers. With each monthly update, the RxNorm content is kept current and in sync with drugs available on the U.S. market, i.e., new drug products are added and drug products no longer available are retired. RxNorm identifiers are never reused and can be safely used as permanent identifiers for drugs in clinical data warehouses and prescription datasets. However, any given release of RxNorm only contains detailed information about active drug products in that release.

RxNorm Adoption

Publication and usage. RxNorm is published on a fixed schedule, with monthly releases on the first Monday of each month. RxNorm is published as a set of relational files, with a schema similar to that of the UMLS Metathesaurus. As with the UMLS, users need to complete a license agreement to access to the RxNorm files. There is no

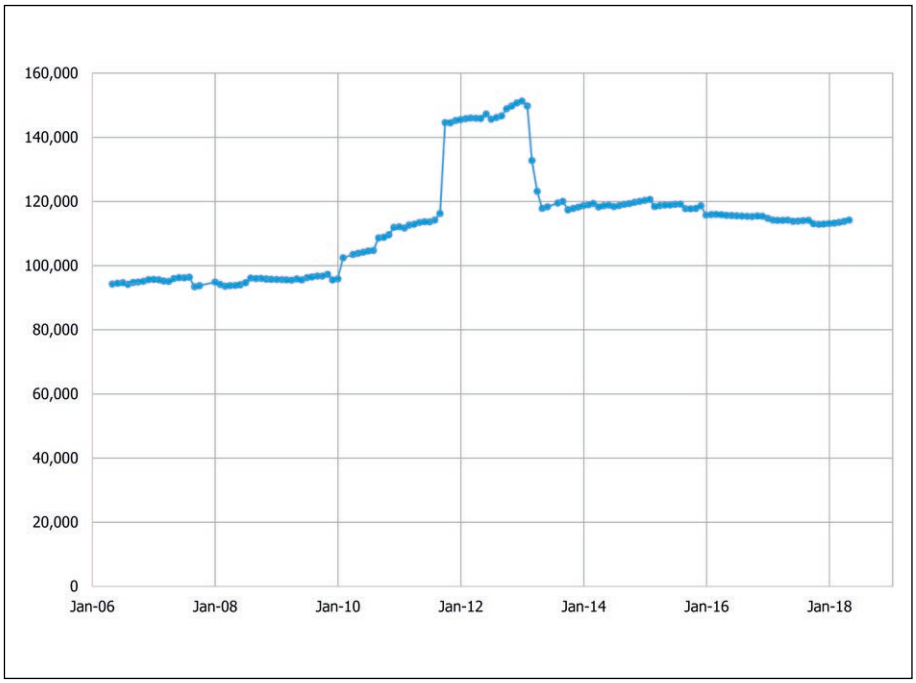

Fig. 4 Evolution of the number of active RxNorm concepts over time.

cost for accessing RxNorm, but some of the RxNorm content is subject to intellectual property restrictions, namely the names and codes from proprietary sources. The "prescribable subset" of RxNorm is a subset restricted to sources from Federal Agencies. Free of proprietary information, these files are publicly available [38].

Also publicly available are graphic and programming interfaces developed to expose some of the content of RxNorm. The RxNav browser [39] allows users to explore RxNorm from a variety of names and codes (including proprietary names and codes), but only returns the information that is publicly available in RxNorm. Similarly, the RxNorm application programming interface (API) [40] allows users to integrate RxNorm in their applications. RxNav and the RxNorm API are also closely integrated with companion resources, facilitating access to additional information, such as drug classes and drug-drug interaction information.

The RxNorm files are downloaded about 1,000 times each month. RxNav has over 2,000 unique users and serves some 500,000 queries annually. The RxNorm API has over 20,000 unique users and serves some 800 million queries annually. Usage of RxNorm has increased markedly after the 2014 EHR certification criteria designated RxNorm as the vocabulary for medications as part of the Meaningful Use incentive program.

Use cases. The main use cases for RxNorm are presented below:

- *Electronic prescribing*. The National Council for Prescription Drug Programs (NCPDP) is a standards development organization. Its SCRIPT standard for e-prescribing [41] requires RxNorm as its standardized medication nomenclature.
- *Information exchange*. RxNorm is often used as the drug vocabulary for exchanging drug medication information across clinical or administrative systems. For example, the U.S. Department of Defense (DoD) and Department of Veterans Affairs (VA) have relied on RxNorm to mediate drug information across their respective electronic medical record systems [40].
- *Formulary development*. The Centers for Medicare & Medicaid Services (CMS) use RxNorm in their Formulary Reference File, as part of the guideline for Medicare drug benefits [42].
- *Reference value sets*. The drug value sets used in electronic clinical quality measures for the Meaningful Use incentive program are defined in reference to RxNorm [43].
- *Analytics*. Increasingly, RxNorm is used as the standard for drugs in clinical data warehouses. For example, OHDSI, the Observational Health Data Sciences and Informatics research group, uses RxNorm for representing drugs as part of its Observational Medical Outcomes Partnership (OMOP) common data model (CDM) [44]. PCORnet, the National Patient-Centered Clinical Research Network makes similar use of RxNorm in its common data model [45].

Of note, analytics was not among the use cases RxNorm was initially designed to support. One specific issue here is that many drug identifiers recorded in clinical data warehouses may no longer be valid in the current release of RxNorm and detailed information about the corresponding drugs may be missing. To address this issue, the RxNorm API has developed functions to support a history mechanism for RxNorm entities and codes from the National Drug Code (NDC) Directory [46]. Moreover, RxNorm is purposely biased towards drugs marketed in the U.S. To support their international analytics efforts, the OHDSI research group has extended RxNorm to drug terminologies used in other countries [47].

RxNorm Collaborations

In the development of RxNorm content, NLM has worked in close collaboration with the vendors of drug knowledge bases, with federal partners, and with representatives of the pharmacy services industry represented by NCPDP. Similarly, the development of RxNav and the RxNorm API have greatly benefited from the feedback provided by their user community. For the past five years, NLM has held an annual DailyMed/RxNorm Jamboree Workshop to bring together the RxNorm stakeholders.

To extend the usefulness of RxNorm despite its limited scope, NLM has also initiated partnerships with providers of clinical information that can be linked to RxNorm. For example, NLM has developed companion APIs to link RxNorm drugs to various drug classification systems and to publicly available sources of drug-drug interaction information [48]. Drug classes for RxNorm drugs can also be explored through the RxClass application [49].

Future Directions

The new Medication Reference Terminology (MED-RT) was recently announced as the evolutionary successor to the Veterans Health Administration's (VHA) National Drug File-Reference Terminology (NDF-RT). While both NDF-RT and MED-RT have the purpose of linking drugs to clinical information (e.g., indications, mechanism of action, physiologic effect), MED-RT differs from NDF-RT in that it uses RxNorm identifiers for the drugs, instead of the proprietary identifiers NDF-RT was using. The use of RxNorm identifiers in MED-RT will greatly facilitate the interoperability between MED-RT and clinical data warehouses that use RxNorm identifiers for drugs.

RxNorm will continue to improve compliance with international standards, such as standards developed by the International Organization for Standardization (ISO) for the identification of medicinal products (IDMP) [50]. As part of this effort, RxNorm will explicitly represent information, such as the basis of strength substance (BoSS), and will refine its representation of strength to further normalize units and facilitate computation of medication doses. These changes will enhance interoperability between RxNorm and the international drug model in SNOMED CT, and facilitate the use of RxNorm as a national extension of SNOMED CT for medications.

RxNorm is also working in close collaboration with the developers of Fast Healthcare Interoperability Resources (FHIR), in particular to expose the RxNorm content as medication resources [51]. The availability of RxNorm medications in FHIR represents

an important step, because FHIR medication resources play a key role in other resources defined for adverse events, medication dispensation, and medication orders.

Discussion

Similarities and Differences

SNOMED CT, LOINC, and RxNorm are all integrated in the UMLS Metathesaurus [37], which identifies equivalences among the concepts they share and provides a common interface to access them. Additionally, drugs from SNOMED CT are integrated in RxNorm, drugs from LOINC are mapped to RxNorm, and "parts" concepts from LOINC are mapped to SNOMED CT (Table 2). These mappings create tight, curated integration among them, independent of the UMLS. SNOMED CT, LOINC, and RxNorm have all been selected to support interoperability not only from a regulatory perspective in the Meaningful Use incentive program, but also from a practical perspective in the OMOP common data model developed by the OHDSI research group. All three are also actively supporting FHIR.

Despite recent efforts to standardize terminology services, SNOMED CT, LOINC, and RxNorm each have a specific terminology model, and rely on different formalisms and specific tooling for their development. While the benefits of converting LOINC to a description logic formalism have been demonstrated [52], it remains difficult to adapt legacy tooling and change development practices. Although built natively with description logics, SNOMED CT can only move slowly to a more expressive dialect.

Uniform Terminology Services

With growing use of clinical terminologies in health IT applications, so has the need for enterprise terminology services that can make available code systems, value sets, and mappings across different terminologies through well-defined interfaces. Early work to define such terminology services [52, 53] matured into specifications such as the joint HL7 and OMG standard Common Terminology Services 2 [53]. These efforts informed development of terminology services capabilities in the rapidly spreading Fast Healthcare Interoperability Resources (FHIR) standard [54]. The core FHIR standard and reference servers now have native support for SNOMED CT, LOINC, and RxNorm, which is opening many possibilities for dynamically accessing these terminologies in developer - and implementer - friendly manner via common web technologies and an open application programming interface (API). Key resources such as the Value Set Authority Center developed by the NLM [55] and enterprise terminology servers (e.g., Ontoserver, Distributed Terminology System) already support FHIR-based access to these terminologies.

Integrating SNOMED CT, LOINC and RxNorm

Over the past few years, collaboration among the developers of SNOMED CT, LOINC, and RxNorm has increased. The Memorandum of Understanding signed between SNOMED CT and LOINC developers has paved the way for leveraging SNOMED CT for the representation of the building blocks of LOINC (e.g., substances, organisms) and for a more consistent representation of clinical and laboratory observations in SNOMED CT. Similarly, the publication of the new international drug model in SNOMED CT facilitates the development of compatible national extensions, such as RxNorm, and better support for IDMP in RxNorm also helps to make it more consistent with SNOMED CT.

However, while this evolution leads to greater compatibility and interoperability, integration of SNOMED CT, LOINC, and RxNorm still requires mappings among the three terminologies. Moreover, these three terminologies use different formalisms and tools for their representation, have their own release cycles and versioning mechanisms, which makes their seamless integration non trivial, if at all possible. A further step towards effective integration is being pursued under the auspices of the Veterans Health Administration (VHA) and the Healthcare Services Platform Consortium (HSPC), under the codename SOLOR [56]. The proposal calls for using description logics for the representation of the three terminologies, while only SNOMED CT currently uses this formalism.

Conclusions

With different starting points, representation formalisms, funding sources, and evolutionary paths, SNOMED CT, LOINC, and RxNorm have evolved over the past few decades into three major clinical terminologies supporting key use cases in clinical practice. Despite their differences, partnerships have been created among their development teams to facilitate interoperability and minimize duplication of effort. Further integration has been proposed, but will require additional resources to bring these terminologies closer together. The benefits of the integrated terminologies in terms of homogenous semantics and inherent interoperability should, however, outweigh the complexity added to the system. Meanwhile, the availability of these three clinical terminologies through the unified terminology services of FHIR will already facilitate their usage in support of interoperability in healthcare.

Acknowledgments
This work was supported in part by the Intramural Research Program of the NIH, National Library of Medicine and by contract HHSN276201400008C from the National Library of Medicine.

References

1. Bodenreider, O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. Yearb Med Inform 2008:67-79.
2. Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. Yearb Med Inform 2006:124-35.
3. Cornet R, Chute CG. Health concept and knowledge management: Twenty-five years of evolution. Yearb Med Inform 2016:S32-41.
4. NHS Business Service Authority. Dictionary of medicines and devices (dm+d). 2018.
5. Australian Digital Health Agency. Australian Medicines Terminology; 2018.
6. Marcotte L, Seidman J, Trudel K, Berwick DM,

Blumenthal D, Mostashari F, et al. Achieving meaningful use of health information technology: a guide for physicians to the EHR incentive programs. Arch Intern Med 2012;172(9):731-6.
7. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform 2015;216:574-8.
8. Dunham GS, Pacak MG, Pratt AW. Automatic indexing of pathology data. J Am Soc Inf Sci 1978;29(2):81.
9. Spackman KA, Campbell KE, Cote RA. SNOMED RT: a reference terminology for health care. Proc AMIA Annu Fall Symp 1997:640-4.
10. Wang AY, Barrett JW, Bentley T, Markwell D, Price C, Spackman KA, et al. Mapping between SNOMED RT and Clinical terms version 3: a key component of the SNOMED CT development process. Proc AMIA Symp 2001:741-5.
11. Spackman KA, Dionne R, Mays E, Weis J. Role grouping as an extension to the description logic of Ontylog, motivated by concept modeling in SNOMED. Proc AMIA Symp 2002:712-6.
12. World Wide Web Consortium (W3C). OWL 2 Web Ontology Language Profiles (Second Edition); 2018.
13. SNOMED International. Members 2018 [Available from: https://www.snomed.org/members/].
14. SNOMED International. SNOMED CT Expo; 2018.
15. Schulz S, Rodrigues JM, Rector A, Chute CG. Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. Stud Health Technol Inform 2017;245:940-4].
16. Health IT Buzz Blog. EHR Certification Criteria for SNOMED CT will help doctors transition to ICD-10 2013 [Available from: https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/ehr-certification-criteria-snomed-ct-doctors-transition-icd10/].
17. NHS Business Service Authority. SNOMED CT 2018 [Available from: https://www.nhsbsa.nhs.uk/clinical-services/dental-snomed-ct].
18. Lee D, de Keizer N, Lau F, Cornet R. Literature review of SNOMED CT use. J Am Med Inform Assoc 2014;21(e1):e11-9.
19. SNOMED International. SNOMED in action; 2018.
20. Orphanet. The portal for rare diseases and orphan drugs 2018 [Available from: https://www.orpha.net/].
21. Stanford Center for Biomedical Informatics Research. Protégé 2018 [Available from: https://protege.stanford.edu/].
22. Dentler K, Cornet R, ten Teije A, de Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web 2011;2(2):71-87.
23. Piro R, Nenov Y, Motik B, Horrocks I, Hendler P, Kimberly S, et al., editors. Semantic Technologies for Data Analysis in Health Care2016; Cham: Springer International Publishing; 2016.
24. SNOMED International. Consultation for SNOMED CT Description Logic Enhancements 2018 [Available from: https://www.nhsbsa.nhs.uk/clinical-services/dental-snomed-ct].
25. McDonald CJ, Huff SM, Suico JG, Hill G, Leavelle D, Aller R, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clin Chem 2003;49(4):624-33.
26. Vreeman DJ, McDonald CJ. Automated mapping of local radiology terms to LOINC. AMIA Annu Symp Proc 2005:769-73.
27. Vreeman DJ, McDonald CJ, Huff SM. LOINC(R) - A Universal Catalog of Individual Clinical Observations and Uniform Representation of Enumerated Collections. Int J Funct Inform Personal Med 2010;3(4):273-91.
28. Frazier P, Rossi-Mori A, Dolin RH, Alschuler L, Huff SM. The creation of an ontology of clinical document names. Stud Health Technol Inform 2001;84(Pt 1):94-8.
29. Subramanian A, Westra B, Matney S, Wilson PS, Delaney CW, Huff S, et al. Integrating the nursing management minimum data set into the logical observation identifier names and codes system. AMIA Annu Symp Proc 2008:1148.
30. Regenstrief Institute. Regenstrief LOINC® Mapping Assistant (RELMA®) and online search application 2018 [Available from: http://search.loinc.org].
31. Vreeman DJ, Chiaravalloti MT, Hook J, McDonald CJ. Enabling international adoption of LOINC through translation. J Biomed Inform 2012;45(4):667-73.
32. Seerainer C, Sabutsch SW. eHealth Terminology Management in Austria. Stud Health Technol Inform 2016;228:426-30.
33. Vreeman DJ. LOINC - An overview and background 2018 [Available from: https://danielvreeman.com/loinc-overview-background/].
34. Collins FS, Hudson KL, Briggs JP, Lauer MS. PCORnet: turning a dream into reality. J Am Med Inform Assoc 2014;21(4):576-7.
35. IVD Industry Connectivity Consortium. LIVD – Digital format for publication of LOINC to vendor IVD test results 2018 [Available from: http://ivdconnectivity.org/livd/].
36. Sperzel WD, Broverman CA, Kapusnik-Uner JE, Schlesinger JM. The need for a concept-based medication vocabulary as an enabling infrastructure in health informatics. Proc AMIA Symp 1998:865-9.
37. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(Database issue):D267-70.
38. National Library of Medicine. RxNorm technical documentation 2018 [Available from: https://www.nlm.nih.gov/research/umls/rxnorm/docs/index.html].
39. National Library of Medicine. RxNav 2018 [Available from: https://rxnav.nlm.nih.gov/.
40. National Library of Medicine. RxNorm application programming interface (API) 2018 [Available from: https://rxnav.nlm.nih.gov/].
41. National Council for Prescription Drug Programs. SCRIPT standard for electronic prescribing 2018 [Available from: https://www.ncpdp.org/NCPDP/media/pdf/EprescribingFactSheet.pdf].
42. Centers for Medicare & Medicaid Services. Formulary Guidance 2018 [Available from: https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/RxContracting_FormularyGuidance.html].
43. National Library of Medicine. Value Set Authority Center 2018 [Available from: https://vsac.nlm.nih.gov/].
44. Knowledge Base workgroup of the Observational Health Data Sciences and Informatics collaborative. Large-scale adverse effects related to treatment evidence standardization (LAERTES): an open scalable system for linking pharmacovigilance evidence sources with clinical data. J Biomed Semantics 2017;8(1):11.
45. Patient-Centered Clinical Research Network. PCORnet, the National Patient-Centered Clinical Research Network 2018 [Available from: http://www.pcornet.org/].
46. U.S. Food & Drug Administration. National Drug Code Directory 2018 [Available from: https://www.accessdata.fda.gov/scripts/cder/ndc/].
47. Observational Health Data Sciences and Informatics. RxNorm Extension - an OHDSI resource to represent international drugs 2018 [Available from: http://www.ohdsi.org/web/wiki/doku.php?id=documentation:international_drugs].
48. National Library of Medicine. Drug Interaction API 2018 [Available from: https://rxnav.nlm.nih.gov/InteractionAPIs.html#].
49. National Library of Medicine. RxClass 2018 [Available from: https://mor.nlm.nih.gov/RxClass/.
50. International Organization for Standardization]. Revised IDMP standards to improve description of medicinal products worldwide 2018 [Available from: https://www.iso.org/news/ref2234.html].
51. Fast Healthcare Interoperability Resources. Resource Medication - Content 2018 [Available from: https://www.hl7.org/fhir/medication.html].
52. Adamusiak T, Bodenreider O. Quality assurance in LOINC using Description Logic. AMIA Annu Symp Proc 2012;2012:1099-108.
53. Object Management Group. About the Common Terminology Services 2 Specification version 1.2 2018 [Available from: https://www.omg.org/spec/CTS2/].
54. Health Level 7. FHIR - Terminology module 2018 [Available from: https://www.hl7.org/fhir/terminology-module.html.]
55. National Library of Medicine. FHIR® Terminology Service for VSAC Resources 2018 [Available from: https://cts.nlm.nih.gov/fhir/].
56. SOLOR. SOLOR 2018 [Available from: http://solor.io].

Correspondence to:
Dr. Olivier Bodenreider
National Library of Medicine
8600 Rockville Pike - MS 3826 (Bldg 38A, Rm 9S904)
Bethesda, MD 20894 - USA
Tel: +1 301 827 4982
E-mail: olivier.bodenreider@nih.gov