Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240671

A Missing Link on Semantic Interoperability

Angelo ROSSIMORI [a,1], Sandro GIROLAMI [b] and Gerard FRERIKS [c]
[a] *IRPPS, CNR, Italy*
[b] *Meteda srl, Italy*
[c] *Conexis, The Netherlands*
ORCiD ID: Angelo Rossimori https://orcid.org/0000-0002-6830-0945

Abstract. Syntactic interoperability among health ICT systems is well-established, but achieving semantic interoperability requires more than just exchanging codes. We present a pragmatic, safe, and effective approach towards an ambitious goal: enabling any software to process a critical mass of routine clinical information in a replicable manner across various systems and local contexts. We advocate for the use of reliable, parsimonious coding to handle the most replicable aspects of data processing for routine patient information, while reserving the intricate interpretation of individual patient data nuances for skilled professionals, possibly supported by Artificial Intelligence tools. We suggest coping with routine tasks by focusing on a limited set of a few thousand data elements, named the 'Clinical Documentation Kernel' (CDK). This approach will provide direct benefits to users and assist in the human interpretation of other patient information. Our preliminary study focuses on the 'primitives' and 'qualifiers' that bring the highest value to the health ecosystem in various authoritative scenarios in the field of diabetes.

Keywords. semantic interoperability scenarios, parsimonious coding, perceived data Value, software assistants for repetitive tasks, EHRS across boundaries, EHDS

1. Introduction

The shift from paper to Electronic Health Record Systems (EHRS) signifies a move from static, passive records to dynamic, persistent data from multiple sources. Interoperability is crucial for better care and streamlined operations across multiple authors, facilities, and jurisdictions, and to support research and analytics. Syntactic interoperability (FHIR, IHE) has succeeded, but effective semantic interoperability cannot yet be assured even with advanced terminologies (ICD, SNOMED CT, LOINC) and the dual model approach to EHRS data management (ISO 13606, OpenEHR).

In fact, the desirable goal of semantic interoperability is to enable **any software to process any clinical information in a replicable, reliable manner across any systems and local contexts**. However, this is an almost utopian challenge.

Our compromise is to focus on a limited set of data elements for routine tasks, to produce a **Clinical Documentation Kernel (CDK)**. It will provide direct benefits to users and will aid in the (human) interpretation of the other patient information.

This small subset of structured data elements can meet the needs of professionals and citizens in most repetitive diagnostic and treatment tasks, helping to interpret the richness

---

[1] Corresponding Author: Angelo RossiMori; E-mail: angelo.rossimori@cnr.it.

of the other patient data. We argue that a pervasive use of the CDK allow to harmonize the most relevant data elements across different contexts, i. e., it will allow to effectively and safely handle a critical mass of patient data, fostering more efficient healthcare processes and **yielding a perceived Value** for the users and an important return to the health ecosystems. Professional bodies will have the role of identifying the data elements to be addressed by the CDK for an increasing number of relevant situations, according to evidence-based criteria and consensus.

2. Materials and Methods

To demonstrate our assumption, we focused on diabetes, the chronic disease offering a substantial authoritative clinical knowledge from different perspectives: clinical pathways, outcome measurements, registries, EHRS usage in telemedicine, and national data flows. Our list of selected sources is presented in Table 1.

Table 1. Knowledge sources on diabetes used to extract the lists of data elements.

source	body
NHS RightCare Pathway: Diabetes Diagnosis and initial assessment [1]	NHS England
Care Planning and Annual Review (for both T1 andT2) [1]	NHS England
Outcomes Measurements [2]	ICHOM
Project on Semantic Interoperability for Telemedicine [3]	HDC-CDM NZ
National Patient Flow - Endocrinology Subset [4]	SNOMED CT

In addition, we analysed two examples about the **general usage of codes**: (1) the report published by NHS England providing data on 5.4 billion occurrences coded by 6732 GPs in **SNOMED CT** in the period Aug 2022 - Jul 2023 [5]; (2) **ICPC-3**, a coding scheme created by the World Organization of Family Doctors to enhance the GPs' EPRs [6]. We addressed clinical terms such as findings, diseases and complications, therapeutical activities; instead, we ignored medical devices and drugs ingredients, as already well defined and properly managed by existing interoperability standards in most EHRS.

We isolated **primitive concepts** (e.g. 'retinopathy'), **qualifiers** (or modifiers) either '**generic**' (e.g. laterality, urgency, severity) or '**earmarked**' to primitive concepts (e.g. 'proliferative' for retinopathy) and **semantic links** (e.g. 'complication of'); we also extracted data elements from clinical narratives by cutting free text into terms. We then considered the items overlapping across the different sources, to discuss the reasons behind the different perspectives for each context.

3. Results

Table 2 summarizes the number of diabetes-related items for each source; a detailed report is in [3]. About the NHS collection on SNOMED CT, we measured that **1655 codes** (very often different by one or more qualifiers) are responsible for **90% of 5.4 billion** occurrences; about ICPC-3, we observed that it adopts a **parsimonious approach** to clinical coding, focussing on terms effectively needed in their daily work: **822 main terms** (including **371 'diseases'** and **250 'symptoms'**), with 639 more detailed children.

Table 2. Number of items identified within each source, for the different categories: **Prim**=primitive term, **GQ**=generic qualifier, **EQ**=earmarked qualifier, **SL**=semantic link.

source	Prim	GQ	EQ	SL
Diabetes Diagnosis and initial assessment [1]	139	35	4	3
Care Planning and Annual Review (for both T1 and T2) [1]	68	14	0	1
Outcomes measurement [2]	110	24	21	10
Interoperability across EHR systems for telemedicine [4]	132	41	18	4
National Patient Flow - Endocrinology Subset - [5]	28	5	3	2

4. Discussion

In Table 2, the primitives and qualifiers needed by users to take decisions or effectively support activities are limited to a few dozens, selected according to the peculiar view of each source [3]: e.g. in [1], the initial assessment requires the double of items than the annual review; in [2], 30% of items are on patient feelings; in [3], 33% of items deals with self-care; in [4], the National statistics need a small number of primitives.

Considering the large overlaps among clinical domains and the above notes, we foresee that **a few thousands of items** in the CDK will meet most routine needs to deploy effective, pragmatic software across multiple EHRSs. Each professional can then focus on additional details to fully interpret the other health issues, according to his/her role about the patient.

4.1. Five Levels of Interoperability on Clinical Semantics

Navigating the intricacies of clinical semantics is a daunting yet essential task for the full seamless interoperability of clinical data. We propose **five scenarios** on semantic interoperability that elucidate the varied constraints to which it is subject.

Scenario A defines a **very few coded terms**, at present already processed by most EHRS, plus **a set of headings** to organise relevant categories of terms (e.g. on allergies). This approach satisfies current needs for the International Patient Summary [7] and by the European Health Data Space (EHDS) [8], as it provides a base for consistent data entry and retrieval across health systems in different countries.

Scenario B offers a restricted semantic interoperability in defined contexts, such as **local and/or temporary agreements** between healthcare providers. This is successfully applied in clinical trials protocols, or proprietary software solutions.

Scenario C adopts FHIR resources or openEHR archetypes to use agreed **'internal' information models** inside a clinical statement, representing it with a set of standard components. For example, a laboratory result is a 'property' with 'units', 'normal range', 'alert thresholds', etc. However, "*usually in real-life scenarios, different data models are used to share the same information*" [9]; additionally, achieving a universal standard representation for the admitted values for **non-numeric components** is difficult.

Scenario D will adopt the CDK, with a carefully selected collection of a few thousand primitives and qualifiers, along with their allowed values and codes, approved and maintained by international consensus. It will act as **the missing link** to reduce the variability of terms and data structures, which are often tailor-made to each context. This will enable FHIR and openEHR to effectively manage **a critical mass of routine clinical statements** to allow any software to effectively assist users in their routine tasks.

Scenario E pertains to the free use of natural language and complete coding systems and thus to support most routine tasks, **to preserve any subtle nuances** and detailed descriptors unique to each patient's case. It reflects the complexities of individual health conditions; however, it presents significant challenges in realising a full semantic interoperability, unless the scenario D supports professionals (or Artificial Intelligence software) in interpreting patient information.

4.2. Professional Interpretation, Clinical Judgment and Artificial Intelligence

The uniqueness of each patient cannot conform to a complete "universal" formalization; however, the routine clinical information on each patient managed by the CDK will help professionals better interpret other detailed data items, locally agreed and represented as codes or free text, possibly with future support by Artificial Intelligence.

Data granularity and idiosyncrasies make it nearly impossible to achieve fully integrated EHRS networks with multiple authors in multiple jurisdictions: **human interpretation with direct patient interaction remains mandatory**.

Clinical judgment is mandatory to guide patient-specific decisions, by professionals interpreting EHRS data on the light of the information coming from the patient's reality.

Merging clinical expertise with effective digital data management allows for nuanced patient care, leveraging the strengths of both human judgment and advanced computing.

4.3. The Vision: Bring a Perceived Value to the Users by Processable Clinical Data

A **CDK-based software assistant** can significantly impact on user activities. It can enable precise **generation of highest quality data** for routine tasks, offer customized views on data, and support the management of personalized workflows.

It can also feed **evidence-based algorithms** on clinical decision-making, ensure safe medication prescriptions with **alerts and interaction checks**, and optimize time on repetitive tasks with **scheduling and reminders**. Furthermore, it can promote continuous education and self-evaluation by providing access to authoritative resources and performance analytics.

Unlocking the deployment of these advanced software functionalities in EHRS, validated by authoritative clinical bodies, CDK will catalyse **enhanced clinical outcomes** and **provide a perceived Value** to the users.

In practice, the CDK adopts a shared Information Model that allows the definition of **Shared Information Building Blocks (SIBB)** [10]. These SIBBs, as the **archetypes** in openEHR, allow the documentation of clinical facts, evaluations, plans, actions in a complete as possible context (epistemology). Various Use Cases, based on the most important Clinical Pathways, Protocols, or actual field usage in large jurisdictions, could inspire a large collection of SIBBs and feed the CDK.

Therefore, the SIBB approach will result in the demise of data silos and monopolistic behaviors by ICT-vendors, promoting an important advancement of the market.

4.4. The Role of Professional Bodies

Harmonizing primitives and qualifiers, with explicit references to the algorithms and the use cases which can use them, is critical for maintaining consistency and interoperability among any EHRS, ensuring that data can be seamlessly integrated across any healthcare platform. Professional bodies will play a pivotal role in defining the clinical data that should feed the CDK. Their expertise is crucial for identifying essential data that supports the management of major chronic diseases and other frequent scenarios, e.g. to assess prescription appropriateness on medications and lab tests, minimizing errors and ensuring patient-specific treatment alignment.

The effort of the professional bodies can start either from the long-term management of major chronic diseases or from situations with a high risk of non-appropriateness, also considering the feeding into existing decision support systems. Thus, they can fulfil the most frequent tasks in situations that can quickly bring the most Value for the ‘**Quadruple Aim**’: patient experience, population health, reducing costs, professional satisfaction [11] and then continue their efforts on the other most important scenarios.

This approach ensures that the CDK will gradually meet increasing routine needs in the health ecosystem, breaking down silos between different care settings and local or national jurisdictions. Testing it in real-world deployments will be the ultimate proof of the efficacy of the services embedded in the EHRS.

5. Conclusions

We demonstrated that the amount of clinical data elements required to address most relevant routine situations is manageable. A limited amount of well-organized data can empower software functionalities to streamline clinical processes and support health professionals' decision-making, delivering significant perceived Value to users.

We argue that a few thousand well-structured terms in the CDK can effectively address most repetitive semantic needs in healthcare, streamlining clinical documentation and analysis, especially for long-term integrated care needs. By recognizing the specific challenges presented by five scenarios of semantic interoperability, and the related successful solutions, the health ecosystem can move towards more integrated and patient-centric care, while acknowledging the technological and semantic barriers that must be overcome to realize in future the full potential of interoperable clinical data.

In essence, we rely on the precision of technology to manage foundational data, creating a **secure and efficient backbone** for very large EHRS networks. Concurrently, we entrust the professionals with the critical task of **deciphering and understanding the subtleties and complexities of patient narratives and clinical presentations**.

This dual strategy ensures that while technology facilitates the straightforward aspects of data management, the invaluable human element of clinical judgement and patient interaction is preserved and emphasized. Such a balance aims to maximize the benefits of a digitally enhanced health ecosystem without losing the personalized care and detailed consideration that can only come from a human touch. It recognizes the irreplaceable role of healthcare practitioners in the interpretation of patient data and the delivery of care tailored to the unique aspects of the individual health situation.

Acknowledgements: This activity was promoted in a non-conditional manner by the Association on 'Home and Digital Care' in 'Confindustria Medical Devices'.

References

[1] NHS England. Diabetes Pathway ‘Diagnosis and initial assessment’ and ‘Care Planning and Annual Review (for both T1 andT2)’ https://www.england.nhs.uk/rightcare/toolkits/diabetes-pathway/ https://www.england.nhs.uk/rightcare/wp-content/uploads/sites/40/2018/07/nhs-rightcare-pathway-diabetes.pdf. ‘NHS RightCare Pathway: Diabetes’, nhs-rightcare-pathway-diabetes.pdf

[2] ICHOM. Type 1 and Type 2 Diabetes in Adults Data Collection Reference Guide. https://ichom.org/files/medical-conditions/diabetes-in-adults/dia-reference-guide.pdf

[3] Rossi Mori A, Girolami S, Freriks G. Interim report of the Project on Semantic Interoperability for telemedicine. Heath and Digital Care, Confindustria Dispositivi Medici, March 2024

[4] New Zealand Ministry of Health, National Patient Flow reference sets, SNOMED CT Endocrinology Subset, https://refset.ihtsdotools.org/#/directory?refsetId=11551023

[5] NHS England. '[MI] SNOMED Code Usage in Primary Care, 2021-22'. https://digital.nhs.uk/data-and-information/publications/statistical/mi-snomed-code-usage-in-primary-care/2021-22

[6] WONCA. ICPC-3, https://icpc-3.info/

[7] CEN/TC 251 Health informatics. 'A short and clear introduction to the International Patient Summary (IPS)' https://www.ehealth-standards.eu/2023/09/11/a-short-and-clear-introduction-to-the-international-patient-summary-ips.

[8] European Commission. European Health Data Space (EHDS). https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en

[9] OpenEHR. News and views. Interview with Erik Vermeulen, Global Smart Health Solution leader, EY. https://news.openehr.org/?p=429

[10] opengroup.org. Introduction to Building Blocks https://pubs.opengroup.org/architecture/togaf9-doc/arch/chap33.html

[11] Bodenheimer T and Sinsky C, "From triple to Quadruple Aim: Care of the patient requires care of the provider," Ann Fam Med, 2014, doi: 10.1370/afm.1713.