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

doi:10.3233/SHTI240664

Development and Implementation of Semantic Information Standards for Oncology

Kees C.W.J. EBBEN[a,b,1], Lonneke R.E. VERMEULEN[a], Astrid C.P. SWINKELS[a], Matthijs SLOEP[a] and Jurrian VAN DER WERF[a]
[a] *Department of Research and Development, Netherlands Comprehensive Cancer Organization (IKNL), Utrecht, The Netherlands*
[b] *Department of Epidemiology, GROW school for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands*
ORCiD ID: Kees C.W.J. Ebben, https://orcid.org/0000-0003-1069-2357

Abstract. Data exchange in oncological healthcare is hindered by insufficient standardization agreements. An Information Standard comprises agreements facilitating accurate communication of care information with the necessary quality and timeliness. We introduce a structured approach to designing, implementing, and maintaining semantic information standards for oncology, supporting information use across medical scenarios. It consists of an element dataset organized into three tiers, ensuring comprehensive documentation and reliable information exchange. These agreements enhance health data interoperability and system functionality, governed by semantic standardization. Together with communication standards, they empower healthcare professionals with extensive medical records and grant patients control over their health data. Consequently, a high-quality semantic information standard supports both providers and patients, and is adequate during development and manageable during maintenance.

Keywords. Healthcare, data elements, concepts, intra-operability, interoperability

1. Introduction

Responsible data management in oncology is crucial for enhancing care and quality. Semantic Information Standards (SIS) play a vital role, granting healthcare professionals access to comprehensive medical records and empowering patients with control over their data [1]. These standards ensure accurate communication between healthcare parties, advancing reliable information exchange.

SISs bridge clinical processes and communication standards while remaining compatible with classification and terminology systems, and other technical and semantic standards. A significant challenge lies in maintaining and operationalizing these standards due to the vast number of concepts involved, exemplified by SNOMED CT's >350,000 unique concepts [2]. Our proposed solution to this challenge can be compared

[1] Corresponding Author: Kees C.W.J. Ebben; E-mail: k.ebben@iknl.nl.

to semantics: elements from a SIS dataset serve as the alphabet's letters, forming an agreed-upon limited set of basic building blocks, anticipating on assembling relevant information units per use case. This approach's greatest advantage lies in its ability to make information exchange components manageable at the lowest level.

This article is the first of a three-part series, delving sequentially into the applied methodology governing a SIS , clinical reporting, and algorithmic decision support. Together, these elucidated methods serve as a robust foundation for effective clinical information management, essential for a learning health system [3].

2. Methods

In the process of digitizing guidelines and standardizing clinical reporting, the National Breast Cancer Platform of the Netherlands (NABON) and the Netherlands Comprehensive Cancer Organization (IKNL) have jointly initiated a project focused on breast cancer. This collaborative effort involves the comprehensive conversion of the entire breast cancer guideline into clinical decision trees, and development of templates for standardized, structured reporting in radiology and multidisciplinary team meetings. The current study conducted analyses specifically targeting the principles for development of a SIS , aligned with the use case of the guideline-based clinical decision tree for the 'primary treatment' process step as published on the Oncoguide platform [4].

A use-case specific SIS consists of a dataset of elements (an atomic unit of data that has precise meaning or precise semantics) obtained from its effectuated use cases with associated concepts (information) and the whole of agreements made for all defined use cases (process, information, and application layers) within the context of a disease or condition. If required, a use case is provided with a communication standard and/or registration standard (application layer) ready for implementation in clinical processes as intended. However, communication standards are out of scope in this document. The development of an SIS is always a collaborative effort involving clinicians, informaticians, and other stakeholders.

2.1. Separation of Clinical Data and Information Model

The distinction between the 'element dataset' and 'use case dataset' is made to achieve a disconnect between clinical data, and the information model. This separation provides flexibility and maintenance advantages in a broad application of (health) data. The 'element dataset' provides a standardized vocabulary that can be applied to any use case, allowing for flexibility in the development of information models. The 'use case dataset' is tailored to specific clinical processes, allowing for maintenance advantages by enabling updates to be made to specific use cases without affecting other processes [Figure 1].

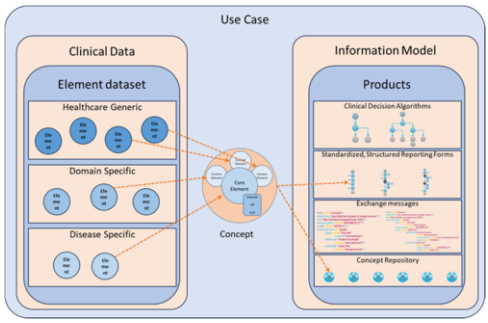

Figure 1. Illustrates a use-case driven semantic information standard where clinical data and information models are separated. Relevant concepts are derived from the element dataset for a specific use case and stored for future reuse.

2.2. Clinical data

For clinical data we use the following modeling principles:

2.2.1. Element dataset

The element dataset lists unique, unambiguous clinical information elements, categorized into Healthcare-generic, Domain-specific, and Disease-specific tiers. Elements are singular representations of clinical information, each distinct and without repetition, such as Domain, Topography, and Diameter.

They are consistently denoted as nouns and accompanied by descriptions, preferred terms, and synonyms. Abbreviations, if used, are limited to classification systems, laboratory values, and acknowledged interventions in clinical practice.

Elements feature properties like discrete, continuous, or textual values. Discrete elements can be a value of a discrete parent element. To prevent arbitrary dichotomization, classification systems like the Karnofsky Performance Status are favored over binary classifications like "fit" and "not fit" for cancer patient abilities.

2.2.2. Tiers

To ensure reusability, we categorize elements into three tiers: Healthcare-Generic, Domain-, and Disease-Specific. Healthcare-Generic includes universally applicable elements, Domain-Specific pertains to specific healthcare domains like Oncology, and Disease-Specific focuses on particular diseases or conditions.

Each element is recorded once in the dataset, in the highest applicable tier, preventing redundancy and ensuring consistency.

2.3. Information model

For information models we use the following modeling principles:

2.3.1. Use case concepts

A use case dataset comprises concepts modeled for specific purposes, assembled from elements in the element dataset. Each concept includes a core element with associated

value set or unit, and optional context elements. For example, the concept of 'Breast Cancer Tumor Diameter', consisting of three elements. Concepts can be reused provided that other use cases define them identically.

All elements, values, and units in the element dataset are assigned a code system and code. Use case concepts must align with the code system definitions. Value lists may be shortened in specific use cases. Elements can serve as core or context elements in concepts, ensuring unambiguous provision of information across diverse use cases.

2.3.2. Level of Context Element Capturing

If context elements of a specific group of concepts contain the same context, for instance, because they have a pathology report as a source, context can also be assigned at the group level to avoid repetitive work. Similarly, context elements at the project level are specified in the metadata of the use case, rather than added directly to concepts, ensuring they are recorded at the highest applicable level.

2.3.3. Clinical Information Models

Clinical data separation from information models is more common. For example openEHR principles include a standardized model for semantic interoperability, archetypes defining clinical content, and a dual-model approach. Moreover, this structure also aligns our methodology with HL7 FHIR, the ISO 11179 metadata registry standard, and application of the FAIR Data Principles, ensuring practical and seamless data exchange.

3. Results

The guideline-based decision tree for 'primary treatment' comprises 62 concepts, with 32 representing patient or disease characteristics, and 30 forming associated recommendations. These concepts consist of 42 elements, where 26 concern to patient or disease characteristics, and 16 concern to recommendations. The distribution of elements across care tiers is 18 for the Healthcare-Generic tier, 20 for the Domain-Specific tier, and 4 for the Disease-Specific tier.

Table 1. Categorization of all elements identified in the 'primary treatment' process step as outlined in the Dutch guideline for breast cancer.

Tier	Elements
Healthcare-generic	Age, Palpability, Surgery type, Drug treatment, Pregnancy, LVEF, Consultation, Hormonal therapy, Topography, Imaging technique, Type of recommendation, Level of Evidence, Body structure, Clinical entity, Inflammatory breast, Diameter, Process step, Anatomical location
Domain-specific	M-status, cN, Positive lymph nodes, cT, Tumor type, Ingrowth, HER2-status, Grade, ER-status, Coloring, Chemotherapy, Targeted therapy, Radiotherapy, PR-status, Tumor behavior, Chemotherapy type, Targeted therapy type, Radiotherapy type, Target area, Fixed lymph nodes
Disease-specific	Invasion risk, Lymph node risk status, Breast cancer tumor type, Breast surgery type

4. Discussion

The breast cancer SIS, developed collaboratively by healthcare professionals and informaticians, employed domain-relevant standards. Despite its limited scope, the element count is lower than concepts, reducing complexity and enabling effective data standardization. Approval from the NABON EHR standardization workgroup validates its relevance.

SISs unify healthcare terminology, they are a means and not a goal in themselves. The application of SISs in clinical practice occurs through implementation in, for example, standardized, structured reporting [5] or (digital representation) of a guideline [6]. Besides, also research gains from standards, crucial for leveraging real-world data (RWD) in cancer therapy research [7]. This alignment supports ongoing learning and aids patient care through uniform practices.

5. Conclusions

This study highlights the crucial role of SISs in promoting seamless interoperability across healthcare domains. Emphasizing their practical implementation in clinical practices and alignment with population-based data, SISs emerge as key tools for enhancing healthcare data reporting and exchange, and fostering continuous learning environments for improved patient outcomes.

References

[1] Nikolski M, et al. Roadmap for a European cancer data management and precision medicine infrastructure. Nat Cancer. 2024 Mar;5(3):367-372, doi: 10.1038/s43018-023-00717-6.
[2] Chang E, Mostafa J. The use of SNOMED CT, 2013-2020: a literature review. J Am Med Inform Assoc. 2021;28(9):2017-26, doi: 10.1093/jamia/ocab084.
[3] Friedman CP, Lomotan EA, Richardson JE, Ridgeway JL. Socio-technical infrastructure for a learning health system. Learn Health Syst. 2024;8(1):e10405, doi: 10.1002/lrh2.10405.
[4] IKNL. NABON Breast Cancer Guideline for Primary Treatment in digital format [Internet]. [Accessed May 18, 2024]. Available from: https://oncoguide.nl/#!/projects/7/tree/10543/101.
[5] Ebben K, et al. Impact on Quality of Documentation and Workload of the Introduction of a National Information Standard for Tumor Board Reporting. Jco Clinical Cancer Informatics. 2020;4:346-56, doi: 10.1200/cci.19.00050.
[6] Hendriks MP, et al. Transformation of the National Breast Cancer Guideline Into Data-Driven Clinical Decision Trees. Journal of Clinical Oncology Clinical Cancer Informatics. 2019;3:1-14.
[7] Nicholson N, Perego A. Interoperability of population-based patient registries. J Biomed Inform X. 2020 Sep;6:100074, doi: 10.1016/j.yjbinx.2020