A Review of Requirements for Information Models in Learning Health Systems

Øivind SKEIDSVOLL SOLVANG[a,c1], Sonja CASSIDY[a,c], Luis MARCO-RUIZ[b],
Ove LINTVEDT[b,c] and Terje SOLVOLL[b,c]

[a] *Department of Strategic ICT, Helse Vest IKT, Bergen, Norway*
[b] *Norwegian Centre for E-health Research, Tromsø, Norway*
[c] *Faculty of Nursing and Health Sciences, Nord University, Bodø, Norway*

Abstract. Learning Health System (LHS) and integrated care are challenged due to a fragmented health data landscape. An information model is agnostic to the underlying data structures and can potentially contribute to mitigating some of the gaps. In a research project, Valkyrie, we are exploring how metadata can be organized and used to promote service coordination and interoperability across levels of care. An information model is viewed as central in this context and as a future integrated LHS support. We examined the literature regarding property requirements for data, information and knowledge models in the context of semantic interoperability and an LHS. The requirements were elicited and synthesized into five guiding principles as a vocabulary to inform the information model design of Valkyrie. Further research on requirements and guiding principles for information model design and evaluation are welcomed.

Keywords. Information model, learning health system, electronic health record

1. Introduction

The reuse of health data to knowledge, such as in a learning health system (LHS), can advance the quality, safety and practice of health care [1]. The LHS process can be described as a three-step iterative cycle (Figure 1): from data to knowledge (D2K), from knowledge to practice (K2P) and from practice to data (P2D) [2]. The P2D step is supported by Electronic Health Record (EHR) systems, which is central to healthcare professionals' work to administer patients' medication and treatment. Consequentially, an EHR is a central source for generating knowledge from data (D2K) [3]. However, EHRs struggle with usability, data quality and semantic interoperability, which have affected the quality, efficiency and cost in health care [3], [4], [5].To reduce the risk of inheriting a legacy of problems from the EHR [3], there is need for a bridge between the often disparate data model structures and the knowledge models. An information model (IM) with its abstract representation of concepts, relationships, and constraints [6] is agnostic to the underlying data structures, which makes it a potential candidate for mitigating current gaps.

In a research project, Valkyrie [7], we are exploring the use and organization of metadata and clinical codes from EHRs through an information mediator which acts as a Virtual Health Record to be accessed by authorized healthcare professionals when

---

[1] Corresponding Author: Øivind S. Solvang, E-mail: oivind.skeidsvoll.solvang@helse-vest-ikt.no.
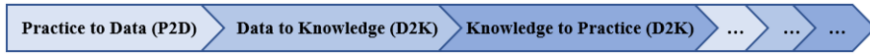
Figure 1. The LHS Process as a three-step iterative cycle (adopted by [2])

needed. We will study how this mediator can contribute to the current interoperability and service coordination problems across care levels. In this context, we see the IM's potential as a translator between the disparate EHRs to promote semantic interoperability. In addition, we see the IM as an extensible prospect for future integrated LHS support. Thus, a literature review on requirements for IM in the context of an LHS was chosen.

The aim of this paper is to examine how property requirements for D2K models in LHS can guide the further information model design of Valkyrie.

2. Methods

To inform the IM model design of Valkyrie, a literature review was conducted to examine property requirements for IMs supporting semantic interoperability and LHS. The databases Medline (through PubMed), Scopus, Web of Science, and IEEE Explore were searched from 2007 in title and abstract. All searches included variations of "Semantic interoperability" and "Learning health system". The Scopus search string was: (TITLE-ABS-KEY (semantic* AND "learning health*") AND PUBYEAR>2007) AND (LIMIT-TO (LANGUAGE,"English")). Search strings in the other databases were in accordance. Additional searches for relevant publications were conducted in ACM Digital Library and Google Scholar. Papers' title and abstract were screened with the inclusion criteria: frameworks, guidelines, principles and requirements for data, information and knowledge models. Duplicates from the included papers were found by sorting on title and controlling that authors' name, abstract and publication date were identical and then removed. Selected papers were read thoroughly. During this process, notes were taken for each publication, and assessed with attention to the relevance of this paper's topic. This excluded publications that focused on methods and techniques related to areas of semantic or learning such as semantic web models, machine learning models and deep learning models, if they had limited or no relation to this paper's topic on model property requirements for interoperability improvements. From the resulting papers, property requirements were elicited and organized according to the D2K step in LHS. With attention to this paper's topic, the property requirements were further synthesized as guiding principles to inform the further IM design of Valkyrie.

3. Results

Seventy-eight publications were identified in the databases during screening of title and abstract using the inclusion criteria. Thirty-five duplicates were identified and removed during sorting on title and controlling the authors' name, abstract and publication date. Thirty-six publications were excluded during the thorough reading of papers due to little or no relevance to the topic of this paper. Three publications were included from ACM Digital Library and one from Google Scholar. D2K model property requirements were elicited from the eleven papers as displayed in Table 1. The guiding principles were synthesized based on the property requirements as displayed in Table 2.

Table 1. Property requirements for D2K models in LHS

Property requirements
Data Accessibility, Timeliness, Authorization, Credibility, Definition, Metadata, Accuracy, Consistency, Integrity, Completeness, Auditability, Fitness, Readability, Structure [8]. Findable, Interoperable, Accessible, Completeness, Reusable common data elements (CDE), Metadata, Integrity, Traceability [9]
Information Software/data model agnostic, Reuse [6]. Provenance, Trust, Reproducibility, Requirements stemming from the context of the LHS: System transparency, Auditability of recommendations, Understandability of data, Validation readiness, Traceability of evidence, Responsibility, Privacy and security, Usability and scalability [10]. Reusability to other domains, Linked data, Flexible, Adaptable (single point of change), Scalable, Maintainable, Sustainable [11]. Provenance, Flexibility, Extensibility, Modularity, Terminology mapping, Consistency [12]. Definition, Metadata, Reuse, Vocabulary mapping [13]. Human readability, Backward compatibility, Vocabulary mapping, Scalability, Extensibility, Sustainability [14]. Provenance, Metadata. Confidentiality, Extensible, Common vocabulary, Reuse [15].
Knowledge Need for metadata to make Computable biomedical knowledge artifacts (CBK) FAIR+T: Findable: Type, Domain, Identification, Purpose Accessible: Location Interoperable: Authorization and rights management, Technical, Integrity, CBK-to-CBK relationship, Reusable, Preservation Trustable: Evidence from use, Evidential basis, Provenance [16]. FAIR, Local adjustments, local preferences and local monitoring, Explanation, Debiasing, Account for generalizability and semantic uncertainty, Elasticity, Extensibility [17].

Table 2. Guiding principles for Valkyrie information models

Principles	Description	Potential achievements
Transparency	An information model contributes to Findable, Accessible, Interoperable and Reusable (FAIR) data [9; 16] by audit trails and provenance [8-10].	Avoids "black boxes" [17] and unexplainable knowledge objects, building trust [10].
Modularity	An information model supports computable biomedical knowledge (CBK) [9; 16] by being agnostic to specific details of data structures [6; 12]	A flexible and scalable architecture [12] reduces maintenance [11].
Openness	An information model supports additional domain models and local preferences [17], and adapts to new knowledge and data structures [11].	Contributes to integrated care [17] and scalable architecture [11].
Readability	An information model supports human and machine readability by supporting additional metadata [13; 15] and international health informatics standards: OpenEHR, HL7 FHIR [6; 11; 14], SNOMED CT and ICD [11; 13; 14].	Enables harmonization [9], completeness [8; 9] and consistency [8], with reduced manual mappings, complex data transformations [11] and semantic uncertainty [9; 17].
Security	An information model supports security, rights management, and access-control [8].	Reduces privacy concerns [10] and builds trust.

4. Discussion

Although the elicited property requirements indicate similarities across D2K such as metadata, provenance and reusability, data models focus on qualitative aspects of data to ensure interoperability and valuable use. Knowledge models focus more on explainable factors to avoid "black boxes" and to build trust. A prerequisite for explainable knowledge models though, are high qualitative data sets. However, when data differs in structures and granularity, it might get difficult to demonstrate explainable knowledge models. The property requirements for IM show a potential to mitigate some of these gaps such as traceability, vocabulary mapping and backward compatibility.

Five guiding principles were synthesized by the authors as an initial vocabulary to inform the design of IM for Valkyrie and for a future integrated LHS candidate. Thus, the principles should be used with caution elsewhere since the meaning of the terms can differ in other contexts. The principles' relative importance of Valkyrie needs further

elaboration, however all five are considered important to plan for early in the design process. Modularity for instance, can ensure a flexible and extensible architecture if considered early, as was done in the TRANSFoRm project with a separation of the information model from the terminology model [12].

The intention of the literature review was not a comprehensive review of all published works in this field and includes only work reported in publications focusing on model property requirements for interoperability and LHS since the authors considered this essential to the Valkyrie project. The publication year was chosen due to the introduction of the LHS concept in 2007. We might have missed some relevant publications due to this timeframe and by limiting the search for publications in English.

More research on requirements and guiding principles for design and evaluation of clinical information models is welcomed, especially in the fields of integrated care and LHS where interoperability improvements are challenged by heterogeneous EHRs.

References

[1] Deans KJ, Sabihi S, Forrest CB. Learning health systems. InSeminars in Pediatric Surgery 2018 Dec 1 (Vol. 27, No. 6, pp. 375-378). WB Saunders.

[2] Menear M, Blanchette MA, Demers-Payette O, Roy D. A framework for value-creating learning health systems. Health research policy and systems. 2019 Dec;17(1):1-3.

[3] McLachlan S, Dube K, Johnson O, Buchanan D, Potts HW, Gallagher T, Fenton N. A framework for analysing learning health systems: Are we removing the most impactful barriers?. Learning health systems. 2019 Oct;3(4):e10189.

[4] Khairat S, Coleman C, Newlin T, Rand V, Ottmar P, Bice T, Carson SS. A mixed-methods evaluation framework for electronic health records usability studies. Journal of biomedical informatics. 2019 Jun 1;94:103175.

[5] Cresswell KM, Worth A, Sheikh A. Integration of a nationally procured electronic health record system into user work practices. BMC medical informatics and decision making. 2012 Dec;12:1-2.

[6] Kuchinke W, Karakoyun T, Ohmann C, Arvanitis TN, Taweel A, Delaney BC, Speedie SM. Extension of the primary care research object model (PCROM) as clinical research information model (CRIM) for the learning healthcare system. BMC medical informatics and decision making. 2014 Dec;14(1):1-6.

[7] Valkyrie - distributed service-oriented architecture for coordinated healthcare services.

[8] Carnicero J, Rojas D. Healthcare Decision-Making Support Based on the Application of Big Data to Electronic Medical Records: A Knowledge Management Cycle. InLeveraging biomedical and healthcare data 2019 Jan 1 (pp. 121-131). Academic Press.

[9] Kush RD, et al. FAIR data sharing: the roles of common data elements and harmonization. Journal of Biomedical Informatics. 2020 Jul 1;107:103421.

[10] Curcin V, Fairweather E, Danger R, Corrigan D. Templates as a method for implementing data provenance in decision support systems. Journal of biomedical informatics. 2017 Jan 1;65:1-21.

[11] Kilintzis V, Chouvarda I, Beredimas N, Natsiavas P, Maglaveras N. Supporting integrated care with a flexible data management framework built upon Linked Data, HL7 FHIR and ontologies. Journal of biomedical informatics. 2019 Jun 1;94:103179.

[12] Ethier JF, Curcin V, Barton A, McGilchrist MM, Bastiaens H, Andreasson A, Rossiter J, Zhao L, Arvanitis TN, Taweel A, Delaney BC. Clinical data integration model. Methods of information in medicine. 2015;54(01):16-23.

[13] Richesson RL, Smerek MM, Cameron CB. A framework to support the sharing and reuse of computable phenotype definitions across health care delivery and clinical research applications. eGEMs. 2016;4(3).

[14] Daniel C, Ouagne D, Sadou E, Paris N, Hussain S, Jaulent MC, Kalra D. Cross border semantic interoperability for learning health systems: the EHR4CR semantic resources and services. 2017 Jan.

[15] Curcin V, Miles S, Danger R, Chen Y, Bache R, Taweel A. Implementing interoperable provenance in biomedical research. Future Generation Computer Systems. 2014 May 1;34:1-6.

[16] Alper BS, Flynn A, Bray BE, Conte ML, Eldredge C, Gold S, Greenes RA, Haug P, Jacoby K, Koru G, McClay J. Categorizing metadata to help mobilize computable biomedical knowledge. 2022 Jan.

[17] Lehmann HP, Downs SM. Desiderata for sharable computable biomedical knowledge for learning health systems. Learning Health Systems. 2018 Oct;2(4):e10065.